Studien aus dem Max-Planck-Institut
für Sozialrecht und Sozialpolitik

81

Christian Günther

Artificial Intelligence, Patient Autonomy and Informed Consent

Nomos

Studien aus dem Max-Planck-Institut
für Sozialrecht und Sozialpolitik

Volume 81

Christian Günther

Artificial Intelligence, Patient Autonomy and Informed Consent

Nomos

Open Access funding provided by Max Planck Society.

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de

a.t.: München, LMU, Diss., 2024

ISBN 978-3-7560-2239-7 (Print)
978-3-7489-4891-9 (ePDF)

1st Edition 2024

Published by
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Production of the printed version:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN 978-3-7560-2239-7 (Print)
ISBN 978-3-7489-4891-9 (ePDF)

DOI https://doi.org/10.5771/9783748948919

Online Version
Nomos eLibrary

Preface

This book represents the outcome of my research as a doctoral student at the Max Planck Institute for Social Law and Social Policy. Its manuscript was submitted as a doctoral thesis to the Law Faculty of the Ludwig Maximilian University of Munich on 13th June 2023. Barring minor amendments, the substance of the text has remained unchanged. Aside from considerations of formality and practicality, this aligns with the fundamental hypothesis that a forward-looking response to a rapidly evolving technology could be derived from the existing legal framework(s). No doubt the subject of this work, the uses of artificial intelligence in medicine, as well as the laws of California and the United Kingdom, have developed in the intervening period and will continue to develop. Nonetheless, the arguments advanced here have retained their relevance and it is hoped that they can make a valuable contribution to the broader discussions concerning: the relationship between law and technology, the appropriate uses of artificial intelligence in healthcare systems and the legal protections afforded to patient autonomy.

My profound gratitude goes to my supervisor, Prof. Dr. Ulrich Becker, who supported me throughout this process, continually provided me with constructive feedback and criticism and pushed me to strive for excellence. Likewise, I must thank Prof. Dr. Jens Kersten for writing the second opinion to my thesis and Prof. Dr. Andreas Spickhoff, for authoring a position statement. Their generous assessments, as well as critical remarks, have given me much to consider as I continue researching this topic.

As I have benefited greatly from my time at the institute, especially its warm collegial atmosphere, I would like to thank the following colleagues for our exchanges, discussions and joint projects: Anika, Franciska, Hung-Sheng, Irene, Irene, Kristine, Lauren, Linxin, Madeleine, Rick, Simone, Teodora, Tim and Yifei. A special mention is due to the institute's library team and its head, Henning Frankenberger, who shied no efforts to provide us researchers with the necessary resources. I would further like to express my gratitude to Prof. Dr. Timo Minssen and the researchers at the Centre for Advanced Studies in Bioscience Innovation Law at the University of Copenhagen for the opportunity to undertake a research stay there and to present my work.

Last, but certainly not least, I must thank my family. I have been lucky enough to have Ina and Michael as my parents. Over the years they have fostered my intellectual curiosity and offered me their unwavering support. I owe them the greatest debt. I am grateful to my brother Sebastian for sharing his unique perspectives on all things innovation- and technology-related. Finally, I want to express my heartfelt gratitude to my wife, Samantha, who stayed up countless nights with me to see this thesis through and never stopped making me the happiest man in the world.

Content Overview

Table of Contents

List of Tables

Zusammenfassung

Technologische Entwicklungen schreiten in modernen, postindustriellen Gesellschaften bekanntermaßen rapide voran. Hierbei stellt sich immer wieder aufs Neue die Frage nach der Zweckmäßigkeit des bestehenden Rechts und den rechtsdogmatischen Möglichkeiten den Herausforderungen, welche diese Entwicklungen unweigerlich belgeiten, zu begegnen. Die folgende Arbeit analysiert am Beispiel von medizinischen Anwendungen der künstlichen Intelligenz (KI) und den Voraussetzungen, die das britischen und kalifornischen Common Law an die informierte Einwilligung des Patienten stellen, die Interaktion zwischen Innovation und Recht. Dabei hinterfragt sie drei, in diesem Zusammenhand regelmäßig diskutierte, Annahmen: (1) Das Recht ist vor allem ein Instrument, um außerrechtliche, innovationsverwandte Maßstäbe zu verwirklichen. (2) Rechtliche Lösungen, als auch juristisches Denken, bleiben weitestgehend hinter technologischen Entwicklungen zurück. (3) Außerrechtliche Regulierungsmöglichkeiten, wie die Anpassung von Systemarchitekturen, bieten vergleichbar attraktive – vor allem schnelle und unmittelbar effektive – Lösungen.

Beginnend mit der zugrundeliegenden Technologie, ist festzuhalten, dass zurzeit vielfältige klinische KI entwickelt, geprüft und für die Nutzung in den Gesundheitssystemen von vor allem wohlhabenderen Staaten zugelassen werden. KI wird als Technologie definiert, die in der Lage ist, die Art von Aufgaben zu bewältigen, die menschliche Experten bisher durch ihr Wissen, ihre Fähigkeiten und ihre Intuition gelöst haben. Die Entwicklung klinischer KI mit solchen Fähigkeiten wird insbesondere durch den Ansatz des maschinellen Lernens (ML) ermöglicht. Fortgeschrittene Medizinprodukte, welche diesen Ansatz verfolgen, bilden den Gegenstand dieser Untersuchung.

Für solche intelligente Programme werden zunächst bestimmte Charakteristika identifiziert und Anwendungsfälle klassifiziert. Wo eine KI in die Behandlung des individuellen Patienten integriert ist, müssen vor allem die unterschiedliche Qualität dieser Interaktionen, einschließlich unterschiedlicher Automatisierungsgrade, kategorisiert werden. Viele intelligente Medizinprodukte werden menschliches Wissen nur komplementieren. Sie werden nicht unmittelbar die Expertise schmälern, die medizinische Ex-

perten zur Diagnose, Prognose oder Behandlung des Patienten beitragen. Andere Produkte werden diese Expertise teilweise ersetzen. Anstelle einer Begutachtung oder Beratung durch zwei menschliche Experten, wird der Patient zukünftig zum Beispiel von einem Arzt und einer KI evaluiert. Alternativ werden weniger qualifizierte menschliche Fachkräfte Aufgaben ausführen können, die bisher Expertenwissen erforderten. In einer sehr begrenzten Anzahl von Fällen wird die KI Teile des klinischen Entscheidungsprozesses sogar determinieren. Sie wird zum Beispiel eine überraschende Erkenntnis beitragen, die dann Berücksichtigung finden muss, oder sie wird bestimmte Befunde triagieren, sodass sie steuert, welcher Patient am dringendsten menschliche Aufmerksamkeit erfordert.

Neben diesen Automatisierungsgraden werden weitere Eigenheiten mit durch ML getriebenen Medizinprodukten assoziiert. Diese umfassen: (1) Mängel bei der Leistungsbewertung klinischer KI, insbesondere wenn sie für unterschiedliche Gruppen und/oder Umgebungen eingesetzt werden; (2) ein unvollständiges Verständnis der Funktionsweise von KI im Einzelfall, welches, trotz sich abzeichnender technologischer Lösungen, fortbestehen wird; (3) eine daraus folgende Angewiesenheit des menschlichen Entscheidungsträgers auf allgemeines Wissen über die Technologie, um sie in eine kooperative Entscheidungsfindung integrieren zu können; und (4) ein Einwirkung von ML-Geräten auf die Entscheidungsfindung durch die Ausnutzung von unterbewussten Neigungen der menschlichen Partei.

Hieraus geht hervor, dass die variablen medizinischen Einsätze von ML unzweifelhaft viele Vorteile mit sich bringen, sie jedoch auch Gefahren für das Individuum und die Gesellschaft bergen. Eine besondere Herausforderung ergibt sich aus der Annahme, dass medizinische Entscheidungen neben naturwissenschaftlichen Erkenntnissen auch eine normative Evaluierung erfordern und dass diese Evaluierung richtigerweise in das Ermessen des individuellen Patienten fällt. Die Auslagerung solcher Entscheidungen an eine KI, welche die obengenannten Besonderheiten aufweist, droht die Fähigkeit von Patienten zu untergraben, individuelle, den persönlichen Umständen und Überzeugungen angemessene Wertungen zu treffen.

Die Signifikanz dieser Herausforderung folgt daraus, dass der vorherrschende bioethische Diskurs und das Recht liberaler Staaten den Grundsatz der Patientenautonomie anerkennen und dem Patienten daraufhin das Recht zugestehen, entsprechende Entscheidungen zu treffen. Dieser Grundsatz stellt sowohl ein etabliertes, als auch ein sich selbst fortentwickelndes Phänomen dar. Er ist etabliert, denn das Recht des

Individuums medizinische Eingriffe unter bestimmten Voraussetzungen abzulehnen, ist seit längerer Zeit in vielen Gesellschaften anerkannt. Der Autonomie-Grundsatz befindet sich zugleich in einer Phase der Evolution, da dem Patienten erst vor relativ kurzer Zeit ein umfassendes autonomiebasiertes Interesse an der Partizipation an der medizinischen Behandlung zugemessen wurde und dem Arzt entsprechende informationelle Pflichten, die der Realisierung dieses Interesses dienen, auferlegt wurden. Was diese Pflichten konkret, im Lichte sich ändernder sozialer Umstände und Empfindungen, erfordern, muss anhand ethischer und rechtlicher Maßstäbe fortwährend aufs Neue evaluiert werden und diese untergehen dabei selbst bedeutsamen Entwicklungen.

Was diese Maßstäbe für medizinische Entscheidungen durch KI bedeuten, welche zum ersten Mal nicht-menschlichen Akteuren die Anwendung von Fachwissen auf den Einzelfall erlauben, ist eine drängende Frage. Zugleich kann, aufgrund der Neuheit dieser Anwendungen, noch nicht erwartet werden, dass betroffene Rechtssysteme – also solche, wo eine Einführung von KI in das Gesundheitssystem momentan voranschreitet – dies bereits konkret erörtert haben.

Diese Arbeit untersucht rechtsvergleichend wie die normativen Maßstäbe des Vereinigten Königreich und der Vereinigten Staaten (spezifisch durch eine Fallstudie des kalifornischen Staates) auf dieses Problem reagieren. Dabei entwickelt sie eine Methode, um die Eigendynamik dieser Maßstäbe zu erfassen und die Anpassungsfähigkeit des Common Laws im Lichte technologischer Entwicklungen zu evaluieren.

Unter den Vorgaben der funktionellen rechtvergleichenden Methode, umschreibt die Arbeit zuerst das vorrechtliche Problem, das *tertium comparationis*. Dies bezieht sich auf die Verletzung der Patientenautonomie. Um zu ermitteln welche KI-Eigenschaften hierzu beitragen, wird ein prozeduraler Autonomiebegriff angewendet, welcher auf den Prozess abstellt, durch den das Individuum Entscheidungen trifft. Insbesondere assoziiert er autonome Entscheidungen mit den Werten und Überzeugungen des Einzelnen und mit der Ausübung einer gewissen Rationalität.

Dieser Begriff ermöglicht die Identifizierung von vier KI-Herausforderung für die Patientenautonomie: (1) Eine Ungewissheit bei der Nutzung von ML-unterstützten Medizinprodukten, welche Analogien zu innovativen Behandlungsformen mit generischen risikobezogenen Merkmalen rechtfertigt. (2) Das zielgerichtete Handeln der KI, welches vergleichsweise unabhängig von menschlicher Einflussnahme ist, ermöglicht es der Technologie, einige normativ wichtige Entscheidungen ohne bedeut-

same Einbeziehung des Patienten zu treffen. (3) Dieses Handeln beeinflusst und beeinträchtigt zu einem geringeren Maße auch den kooperativen medizinischen Entscheidungsprozess und die Möglichkeiten des Patienten, rationale Abwägungen vorzunehmen. (4) Die nicht offensichtliche Ersetzung von menschlicher Expertise durch KI-gestützte Informationen, stellt ein erkenntnistheoretisches Problem für den Patienten dar, da er sich normalerweise zu Recht auf die Aussagen menschlicher Experten verlassen und ihnen Vertrauen schenken kann. Dieses Vertrauen lässt sich nicht ohne Weiteres auf maschinell erzeugte Erkenntnisse übertragen.

In einem nächsten Schritt wird die Relevanz dieses Autonomiebegriffs und der daraus folgenden Probleme in den beiden Rechtssystemen erörtert. Es ist festzuhalten, dass das prozedurale Verständnis der Autonomie eine vertretbare Konkretisierung der Patientenautonomie im Medizinrecht beider Länder darstellt. Weiter kann dieses Konzept als ein rechtliches Prinzip konstruiert werden. Dies erklärt unter anderem, wie autonomiebasierte Argumentation rechtsintern mit anderen Normen interagiert, den dogmatischen Anforderungen dieser Normen gerecht wird, und nichtsdestotrotz eine Flexibilität und Anpassungsfähigkeit schafft, welche dem Recht selbst eine dynamische Leitfunktion verleiht.

Anhand des konkretisierten Prinzips wird dann geprüft, inwieweit die Einzelnormen, welche die Einwilligung des Patienten unter dem Aspekt der Patientenautonomie verlangen, den Herausforderungen von medizinischen KI begegnen können. Spezifisch sind dies im Vereinigten Königreich und Kalifornien zwei Common-Law-typische, deliktsrechtliche Ansprüche: die Haftung für vorsätzliches Handeln (*battery*) und die Haftung für fahrlässiges Handeln (*negligence*). Dabei werden im Lichte der KI-Herausforderungen Entwicklungsmöglichkeiten unter dem Autonomieprinzip sowie die Grenzen solcher Möglichkeiten aufgezeigt. Dies entwickelt ein vorausblickendes, nuanciertes Bild davon, wie das Recht mit den neuen technologischen Problemen von medizinischen ML-Anwendungen umgehen kann.

In beiden Ländern ist abzusehen, dass das Autonomieprinzip nur bis zu einem gewissen Grad in der Lage sein wird, notwendige Änderungen in der Gestaltung sowie in der Interpretation spezifischer Normen herbeizuführen. Das heißt, dass die Anpassungsfähigkeit des Common Law ab einem Punkt an seine Grenzen stößt. Daraufhin wird erörtert, wie eine komplementäre, gesetzliche Regelung auf den Einsichten der vorangegangenen Analyse aufbauen könnte, sodass sie dem Grundsatz der Patientenautonomie, als auch den Eigenheiten der jeweiligen Systeme, gerecht wird.

In einem finalen Abschnitt kehrt die Arbeit zu den drei anfangs erörterten Grundannahmen zum Verhältnis zwischen Recht und Technologie zurück. Die entwickelte und angewandte Methodik verdeutlicht die proaktive Dynamik, welche dem Recht innewohnt, mit gesellschaftlichen Entwicklungen einherschreitet und sie auf vielfältige Weise auch anleitet. Es ist ein Missverständnis, das Recht nur als ein System strikter Regelungen zu konzipieren, die der Innovation im Wege steht (oder sie alternativ begünstigt) und periodisch der neuen „Realität" angepasst werden muss. Darüber hinaus wurden die angepriesenen Vorteile von außerrechtlichen Lösungen im Fall der klinischen KI nicht identifiziert: Ein unmittelbarer Schutz der Patientenautonomie oder eine inhärente, schnellere Anpassungsfähigkeit wurden bei keiner der analysierten technologischen Lösungen festgestellt.

Table of abbreviations

AB	Automation Bias
AI	Artificial Intelligence
AUC	Area Under the Curve
CACI	Judicial Council of California Civil Jury Instructions
CRT	Chemoradiotherapy
DPA	Data Protection Act
DNN	Deep Neural Network
EU	European Union
EU GDPR	European General Data Protection Regulation
ECHR	European Convention on Human Rights
FDA	Food and Drug Administration
IVF	In Vitro Fertilisation
MICRA	Medical Injury Compensation Reform Act
MCA	Mental Capacity Act
ML	Machine Learning
NCCT	Non-Contrast Computerised Tomography
NHS	National Health Service
UK	United Kingdom
UK GDPR	United Kingdom General Data Protection Regulation
U.S.	United States

Chapter 1: Introduction

> 'The life of the law has not been logic: it has been experience. The felt necessities of the time, the prevalent moral and political theories, intuitions of public policy, avowed or unconscious, even the prejudices which judges share with their fellow-men, have had a good deal more to do than the syllogism in determining the rules by which men should be governed. The law embodies the story of a nation's development through many centuries, and it cannot be dealt with as if it contained only the axioms and corollaries of a book of mathematics. In order to know what it is, we must know what it has been, and what it tends to become. We must alternately consult history and existing theories of legislation. But the most difficult labor will be to understand the combination of the two into new products at every stage.' - Oliver Wendell Holmes Jr.[1]

I. Problem statement

The law must adapt, and more accurately must be adapted by actors and institutions, to address the evolving needs of society. As the introductory quote eloquently states, this is no easy feat. At the heart of this work lies one specific problem of adaptation that many societies are only just beginning to confront: the medical uses of artificial intelligence (AI).

A preliminary definition of AI can be given as a technology that is capable of performing the kinds of tasks that humans solve by drawing on their intuition, knowledge and skill.[2] It is a technology that is rapidly gaining acceptance in the clinical sphere. Devices incorporating machine learning (ML), the currently dominant form of AI, are being developed, rigorously examined and approved for these purposes.[3]

---

1 Holmes Jr. *The Common Law* (1881) 1-2.

2 The justification for this definition will be provided in Chapter 1.

3 This will be explored in Chapter 2, but one can already note that in October 2022 the U.S. Food and Drug Administration added 178 entries to their list of AI/ML enabled medical devices: U.S. Food & Drug Administration, 'Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices' (5.10.2022) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-le

The novelty distinguishing this technology relates to its capabilities. It does not only provide information to which medical professionals can apply their own expertise, their medical judgment. It provides such expertise and judgment itself. Indeed, in some circumstances it has demonstrated the ability to displace human specialists in the performance of demanding cognitive tasks and even to determine aspects of clinical decision making.

The societal challenges that this generates are related to the assumption that all such decision making has a normative component and that this type of choice is properly the prerogative of the patient.[4] Such a prerogative is both well-established and relatively new. It is well-established because the right of persons to take medical decisions has been widely recognised in many societies for some time – at least in so far as they have a right to reject interferences from other actors, including clinical professionals.

It is relatively new because this negative and defensive framing did not actually suffice to place the patient in a position where they could determine the course of their care. To make a meaningful decision the patient needed the information that the professional, usually the physician, possessed. Without it the control in the clinical encounter continued to rest with the professional and the patient's ability to choose remained subject to their discretionary judgment. It was within the physician's power to decide whether to share their expertise regarding a recommendation and, if so, how far.[5] In short, there was a real danger that the patient would have normative positions thrust upon them under the guise of clinical expertise.[6]

The primary legal instrument that emerged to rebalance the professional-patient relationship is the doctrine of 'informed consent'.[7] Hereunder the medical profession is obligated to disclose and discuss certain classes of information with the patient, to facilitate their decision making. This

---

arning-aiml-enabled-medical-devices> accessed 19.3.2023. For an analysis of these figures over time see: Selanikio, 'A Closer Look at FDA's AI Medical Device Approvals' (12.10.2022) <https://www.futurehealth.live/blog/2022/10/10/closer-look-at-fda-ai-approvals> accessed 19.3.2023.

4 Kennedy in Byrne, *Rights and Wrongs in Medicine* (1986) 8.

5 ibid 13-15.

6 Veatch, 'Doctor Does Not Know Best: Why in the New Century Physicians Must Stop Trying to Benefit Patients' (2000) 25(6) The Journal of Medicine and Philosophy p. 701.

7 The term itself was coined in the Californian case of *Salgo v. Leland Stanford Jr. University Bd. of Trustees* (1957) 154 Cal.App.2d 560, 578. In its specification this mechanism can itself embody the described tension and development: Jackson in McLean, *First Do No Harm* (2016) 279.

doctrine has been widely recognised today,[8] and it is said to have led to widespread rejection of paternalistic practices in medicine.[9]

In spite of these achievements, informed consent remains a highly dynamic and open-textured doctrine, subject to contestation and evolution. Partly this must be related to the fact that, although it has been realised through legal norms, the doctrine has developed – and remains closely associated with – the bioethical discourse on patient autonomy.[10] It is this value that has generally been stated to provide the primary justification for the creation and operation of the relevant norms.[11]

Returning to the law's ability to adapt to shifts in circumstances, expert decision making by an AI/ML device poses the danger of reinvigorating the mismatch between expert knowledge and individual choice. It could overturn the ethical and legal position that has come out in favour of protecting the patient's access to the information. With this background in mind one can now formulate the challenges that emerge from clinical AI use more precisely. Given that AI is providing the expertise that was previously the preserve of human professionals, what specific information must be granted to a patient regarding AI/ML to maintain an acceptable balance between specialist skill and their personal decision-making? How must AI expertise be framed and conveyed to them? Does the technology fit within established paradigms? Could it necessitate the creation of novel ones?

To answer these questions our analysis must also engage with a much more fundamental question that goes towards the relationship between law and innovation. One must consider, particularly, the method that should be deployed in this type of analysis. It is asked: how is the law's interaction with technological change to be assessed?

One influential view of the relationship between law and innovation envisages a 'legal lag' between developments within (one aspect) of society

8 Vansweevelt and Glover-Thomas, *Informed Consent and Health: A Global Analysis* (2020).

9 Montgomery, 'Law and the Demoralisation of Medicine' (2006) 26(2) Legal Studies p. 185, 187; Sharpe and Faden, *Medical Harm: Historical, Conceptual, and Ethical Dimensions of Iatrogenic Illness* (2001) 67.

10 Brazier and Cave, *Medicine, Patients and the Law* (Sixth Edition 2016) 67.

11 Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009) 149-150.

and the adaptation of legal norms.[12] Specifically in the case of novel technologies, the paradigmatic position leads one to believe that an innovation emerges, causes frictions and reveals flaws in rules that were previously rigidly established.[13] According to these external demands, the rules are then ideally adapted to fit the new state of affairs.[14] As further developments emerge and the law is forced to respond, this cycle repeats.

Hallmarks of this view are the belief that the adaptation to technology is an external response to the limitations of legal rules and that this should be guided by rationales that stand behind or beyond the law, i.e. are extra-legal themselves.[15] For some contexts it has even be argued that the nature of regulation must itself become technological rather than legal, to maintain an effective oversight over innovations.[16]

One can already begin to see that the above framing of medical AI's challenges does not fall neatly into this paradigm. It does not purport to react to a development that has already taken place, but to anticipate, to guide, the response to a process that is only just beginning to take hold. In addition, the demands of the informed consent doctrine are themselves relatively novel, in a state of development and they are closely connected to a more abstract norm of patient autonomy that, alongside its legal manifestation, has fundamental ethical significance. This combines a factual dynamism with a normative dynamism that is intertwined with the law's operation. The paradigmatic view does not engage with this perspective, but the present work must account for it.

In consequence, the stated research problem calls for a legal methodology that is forward-looking and takes seriously the distinct contribution of

12 Dror, 'Law and Social Change 1958-1959' (1959) 33(4) Tulane Law Review p. 787; Rustad and Koenig, 'Cybertorts and Legal Lag: An Empirical Analysis' (2003) 13(1) Southern California Interdisciplinary Law Journal p. 77, 77-80.

13 Mandel in Brownsword, Scotford and Yeung, *The Oxford Handbook of Law, Regulation and Technology* (2016) 228.

14 Rustad and Koenig, 'Cybertorts and Legal Lag' (2003) 13(1) Southern California Interdisciplinary Law Journal p. 77, 77-80, 118-122; Brownsword and Somsen, 'Law, Innovation and Technology: Fast Forward to 2021' (2021) 13(1) Law, Innovation and Technology p. 1, 4-6. See also the concrete illustrations provided by: Kirby in Brownsword and Yeung, *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes* (2008) 368-373.

15 Mandel in Brownsword, Scotford and Yeung, *The Oxford Handbook of Law, Regulation and Technology* (2016) 230-231.

16 Lessig, 'The Constitution of Code: Limitations on Choice-Based Critiques of Cyberspace Regulation' (1997) 5(2) CommLaw Conspectus: Journal of Communications Law and Policy p. 181.

the legal process to the resolution of non-legal, specifically technological, problems. The development of such an approach is itself no easy feat and will be considered in the next section.

II. Methodology and research question

To offer a response to the unique challenges associated with medical AI and patient autonomy, as well as the more general questions hanging over them, a number of analytical steps ought to be further elaborated.

A. Comparative method

A first significant choice is made to evaluate the developing problem through a legal comparison and, in particular, through the use of the functional method. As the problem-statement implies, the object of the present investigation is a real-life situation – AI/ML's impact on patient decision making – with which legal systems are anticipated to interact. The envisaged application of legal doctrine to this situation, conceptualising the law as thereby fulfilling a role within society, sits comfortably within the functionalist tradition.[17]

Admittedly, the consideration of some pervasive informed consent doctrines may seem anomalous within this framework, since they possess an undeniable legal dimension. However, this is justified both by the need to account for the existing clinical reality, which is partially shaped by the law, and by the already mentioned, intertwined nature of legal and wider bioethical analyses. Positing that it is normatively significant to obtain a patient's informed consent, and why this is so, is central to understanding the common problem with which legal systems are faced.[18]

Understood in these terms, the benefits to be gained from a legal comparison recommend themselves. Assessing the responses to a novel problem can improve the understanding of this phenomenon and its relation to the law, potentially providing insights into different approaches and how these can inform one another. Relatedly, it can indicate what the appropri-

17 Michaels in Reimann and Zimmermann, *The Oxford Handbook of Comparative Law* (Second Edition 2019) 347-348.

18 ibid 374.

ate societal response to the problem could be.[19] All in all the comparison envisaged here has both descriptive and evaluative components, asking: how will the existing legal structures be applied to AI/ML in medicine? What does this tell one about the law's interaction with a novel technological phenomenon? Does it point the way towards an appropriate response, given the deeper normative significance of the problem?

To employ the comparative method, it is of course necessary to make a selection of countries to compare. Three factors are considered determinative in this respect. A first limitation flows from the necessity for potential comparators to experience a sufficiently widespread adoption of AI/ML in their healthcare systems. This renders the factual problem applicable and relevant – something worthy of a legal response. Such adoption takes not only a certain societal prosperity as a prerequisite, but also a preparedness to embrace technological change at scale, specifically in the healthcare sector.

Second, the outlined problem has the aforementioned ethical, normative dimension. To enable the identification of a common problem under this head one should select legal systems that are hypothesised to have sufficiently similar conceptions of patient autonomy. This is what will render the conceptualisation of AI's challenges to patient decision making comparable.

Third, the question hanging over the law's creative, pro-active (rather than reactive) role in adapting to technological change suggests the fruitfulness of examining legal systems that employ methods of legal reasoning that themselves seek to achieve a certain dynamism, capable of being directly responsive to wider societal values and shifts. This arguably points towards the examination of a common law system. As Green and Sales have pointed out 'the common law (and particularly the law of obligations under the common law) is an entity that has been defined by its capacity for evolution, incremental development and flexibility'.[20]

Not only does this family of legal systems attribute a generative role to the judge, but they demarcate methods of legal reasoning that are designed to guide:

19 Zweigert and Kötz, *Einführung in die Rechtsvergleichung: Auf dem Gebiete des Privatrechts* (Third Edition 1996) 14.

20 Green and Sales, 'Law, Technology and the Common Law Method in the United Kingdom' [2023](5) Europäische Zeitschrift für Wirtschaftsrecht p. 205, 205.

> complex judgments at every step–from the identification and interpretation of foundational principles, over the determination and weighing of applicable social propositions, to the assessment of the social congruence and systemic consistency of doctrinal propositions, and the adjustment between these requirements as well as the demands of doctrinal stability, which is effected through the choice between different modes of legal reasoning and overturning[21]

Of course one should not overstate the differences between common law and civil law systems.[22] However, the creative role envisaged for the judge and the development of methodologies for requisite forms of responsive legal reasoning are arguably two aspects in which significant differences persist.[23] In so far as one wishes to assess how the open-textured doctrine of informed consent, shaped as it is by a wider societal shift, can interact with the deployment of medical AI, it is hypothesized that such a system offers the most intriguing case studies.

It is on these bases that this thesis examines the United Kingdom (UK) and the United States (U.S.). Regarding the first element, these are states that are leading seats of AI innovation,[24] and they undoubtedly possess the resources for implementing AI within their healthcare systems.[25] This is already evidenced by the fact that there are hundreds of medical devices

21 Hohmann, 'The Nature of the Common Law and the Comparative Study of Legal Reasoning' (1990) 38(1) The American Journal of Comparative Law p. 143, 158; See similarly: Green and Sales, 'Law, Technology and the Common Law Method in the United Kingdom' [2023](5) Europäische Zeitschrift für Wirtschaftsrecht p. 205, 211-212.

22 See: Lawson, 'The Family Affinities of Common-Law and Civil-Law Legal Systems' (1982) 6(1) Hastings International and Comparative Law Review p. 85. Further one should be mindful of the differences in the possibilities for adaptation within distinct types of civil law systems: Beck, Demirgüç-Kunt and Levine, 'Law and Finance: Why Does Legal Origin Matter?' (2003) 31(4) Journal of Comparative Economics p. 653.

23 See also Calbresi's seminal analysis, relating the traditional common law reasoning of the courts to the task of balancing adaptation and 'the burden of inertia': Calabresi, *A Common Law for the Age of Statutes* (1985) 117-119.

24 Comparing these jurisdictions to other European nations, see: Franke, 'Artificial Intelligence Diplomacy: Artificial Intelligence Governance as a New European Union External Policy Tool' (2021) <https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2021)662926> accessed 26.3.2023 10-11.

25 Note in this respect also the relevance of the digital divide between developed and developing nations noted in: World Economic Forum, 'The 'AI divide' between the Global North and Global South' (16.1.2023) <https://www.weforum.org/agenda/2023/01/davos23-ai-divide-global-north-global-south/> accessed 26.3.2023.

employing AI that have applied for and have been granted market access in these jurisdictions.[26]

Things are admittedly more complex when one strives to account for the reimbursement of the clinical uses of the technology and the commissioning of it by providers.[27] As is to be expected from our preliminary analysis of the technology, integration of AI/ML into existing systems is still an emerging phenomenon in these respects.[28] Nevertheless, the two countries do boast significant strategic frameworks for the purposes of promoting all dimensions of AI adoption as the technology's development proceeds.[29] In sum therefore, there are good reasons for supposing that AI is beginning to assume a role in clinical decision making in these states.

A further reason for selecting these jurisdictions is the relation between their understanding of patient autonomy that will inform the identification of relevant AI/ML challenges. Without reaching a definitive conclusion,

26 This will be explored in detail in Chapter 2. See also the guidance issued on how such technologies fit within existing regulatory frameworks: U.S. Food & Drug Administration, 'Artificial Intelligence and Machine Learning in Software as a Medical Device' (2021) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> accessed 6.3.2022.

27 Davenport and Glaser, 'Factors Governing the Adoption of Artificial Intelligence in Healthcare Providers' (2022) 1(1) Discover Health Systems; Parikh and Helmchen, 'Paying For Artificial Intelligence in Medicine' [2022](5) NPJ Digital Medicine.

28 Few devices are being reimbursed in the U.S. and UK healthcare systems: Parikh and Helmchen, 'Paying For Artificial Intelligence in Medicine' [2022](5) NPJ Digital Medicine; Davenport and Glaser, 'Factors Governing the Adoption of Artificial Intelligence in Healthcare Providers' (2022) 1(1) Discover Health Systems; Chaudhury, 'AI in Health in the United Kingdom: An Overview for SME's and Research Institutes on the Trends, Challenges and Opportunities for AI Applications in the British Healthcare Sector' (2021) <https://www.rvo.nl/sites/default/files/2021/06/AI-in-Health-UK-market-report_0.pdf> accessed 26.3.2023.

29 In the U.S. there is a slurry of such strategies – covering also the healthcare sphere – by federal and state policymakers, non-profit organisations and private companies: Zhang and others, 'Artificial Intelligence Index Report 2022' (2022) <https://aiindex.stanford.edu/report/> accessed 26.3.2023. Specifically regarding reimbursement and adoption one can point to the Centers for Medicare and Medicaid Services' adoption of concrete measures, including reimbursement pathways, for AI: AI Healthcare Coalition, 'AI Healthcare Coalition Appreciates CMS' Efforts to Support Access to Innovative AI Services' (2021) <https://ai-coalition.org/news/ai-healthcare-coalition-appreciates-cms-efforts-to-support-access-to-innovative-ai-services> accessed 26.3.2023. In the United Kingdom the central role for AI strategies in healthcare has been granted to the NHS AI lab, which is responsible for accelerating the adoption of AI in healthcare by providing initiatives and guidance. See NHS Transformation Directorate, 'The NHS AI Lab' <https://transform.england.nhs.uk/ai-lab/> accessed 26.3.2023.

which must necessarily await the comprehensive examination in Part II., one can hypothesise on this connection.

For these purposes it is notable that the bioethical, and related legal, literature has been at pains to delineate an Anglo-American approach to autonomy. Often this is deployed to distinguish other, especially continental European, systems that place more emphasis on other values, such as solidarity, which shape their understanding of the concept.[30] It is also manifested in the pronouncements of American judges and law academics who draw on the close cultural and legal affinities with the UK to emphasise the historical continuity of their adopted concept.[31] In turn British judges have borrowed from their transatlantic counterparts in shaping relevant legal standards in light of changing perceptions of society and individual interests.[32] These interests and shifts were implicitly deemed sufficiently similar to warrant such inferences.

With this one can transition to the common law nature of the selected legal systems. Building on the historical heritage already referred to, both the UK and U.S. engage norms that often evolved from the same legal sources and which remain interconnected by similar modes of legal reasoning.[33] It has been argued that these norms are particularly well-suited to

30 Gaille and Horn, 'Solidarity and Autonomy: Two Conflicting Values in English and French Health Care and Bioethics Debates?' (2016) 37(6) Theoretical Medicine and Bioethics p. 441; Prainsack and Buyx, 'Thinking Ethical and Regulatory Frameworks in Medicine From the Perspective of Solidarity on Both Sides of the Atlantic' (2016) 37(6) Theoretical Medicine and Bioethics p. 489. One must consider the burgeoning, differentiated analyses from other regions too: Orfali, 'A Journey Through Global Bioethics' (2019) 16(3) Journal of Bioethical Inquiry p. 305.

31 *Natanson v. Kline* (1960) 186 Kan. 393, 406-407; Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 220. See also *In re Gardner* where the court cited John Stuart Mill in their identification of this common concept: *In re Gardner* (Me. 1987) 534 A.2d 947, 950.

32 *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 886, 899; *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [53]. See also: Maclean, 'The Doctrine of Informed Consent: Does It Exist and Has It Crossed the Atlantic?' (2004) 24(3) Legal Studies p. 386.

33 For one attempt to ascertain more nuanced differences between American and English legal cultures and reasoning, even concerning analytically identical formulations, see Posner, *Law and Legal Theory in England and America* (2003) 39-45. See also Duxbury who highlights the interconnectedness of such reasoning: 'If a national court has to (…) apply a contentious common law principle, and it is brought to the court's attention that the court (…) of another common law jurisdiction, has already considered at length and in the same language how that (…) principle might best be understood, it would be somewhat odd (though not improper) if that national court

accommodating legal rules to the demands of technological change.[34] The judge has the ability to 'extend a rule to a new set of facts provided the differences are irrelevant in light of the principles underlying the rule'.[35]

For the present context it is concretely envisioned that the judge-led adaptation of common law can serve to accommodate the shifts in social value caused by technological innovation.[36] This has led some commentators to state a preference for the common law over legislative solutions within a common law system: 'Often it is better to wait and allow the common law to develop its response before rushing in with new statutes. Because common law rules are inherently more flexible, statutory law reform risks reducing the law's ability to respond to future change'.[37]

Although not framed comparatively, these insights reinforce the above findings concerning the nature of these systems' legal reasoning. The common law judge ideally keeps abreast of societal developments and is prepared to shape and apply norms to functionally similar scenarios (i.e. scenarios posing sufficiently similar normative problems) even where the precise factual matrix has been drastically altered by novel innovations.[38]

---

were to refuse so much as to take a glance at the foreign court's ruling': Duxbury, 'The Law of the Land' (2015) 78(1) The Modern Law Review p. 26, 30-31.

34 See Moses, 'The Legal Landscape Following Technological Change: Paths to Adaptation' (2007) 27(5) Bulletin of Science, Technology & Society p. 408; Moses, 'Adapting the Law to Technological Change: A Comparison of Common Law and Legislation Courts and Parliament' (2003) 26(2) UNSW Law Journal p. 394; Sales and Green critique the generalised nature of this argument: Green and Sales, 'Law, Technology and the Common Law Method in the United Kingdom' [2023](5) Europäische Zeitschrift für Wirtschaftsrecht p. 205. The advantages in adaptation to social change, including technological progress, also informed Calabresi's analysis of the common law's role in an age of statutes: Calabresi, *A Common Law for the Age of Statutes* (1985) 69-80.

35 Moses, 'The Legal Landscape Following Technological Change' (2007) 27(5) Bulletin of Science, Technology & Society p. 408, 411.

36 'Technological developments in society may also be a relevant type of social value, adding momentum to the need for changes in the law, until the tipping point of incremental adjustment in the existing law is met': Green and Sales, 'Law, Technology and the Common Law Method in the United Kingdom' [2023](5) Europäische Zeitschrift für Wirtschaftsrecht p. 205, 210. The reference to the tipping point is representative of these authors' sound argument that the common law possesses only a limited flexibility and can only adjust after a sufficient amount of pressure has built up. Nevertheless, the fundamental point regarding flexibility and adaptability stands.

37 Moses, 'The Legal Landscape Following Technological Change' (2007) 27(5) Bulletin of Science, Technology & Society p. 408, 416.

38 The statement of Lord Nicholls of Birkenhead in *Re Spectrum Plus Ltd* is instructive: 'judges themselves have a legitimate law-making function. It is a function they have long exercised. In common law countries much of the basic law is still the common

Selecting the UK and U.S. allows one to examine how well-suited two common law jurisdictions are to adapt to AI innovation in the informed consent context. Crucially also, it allows one to understand and contrast how such adaptation proceeds under a specified legal doctrine.

B. Scope of inquiry

With the legal comparison to be undertaken established, it is now necessary to restrict its object. One limitation emerges from the outlined nature of the comparison. If one aim is to gauge a legally dynamic adaptation to an emerging technological problem, and this directs us towards the examination of common law legal systems, then it stands to reason that the legal mechanisms to be examined should not primarily be statutory in nature. Principally we are directed towards judge-made law.

This excludes certain regimes generally relevant to patient consent in the clinical environment. For example, both jurisdictions have certain statutory regimes for the disclosure of information on the basis of strict product liability and on the basis of data protection considerations. Neither of these will be examined in any depth in this work, since they lack the breadth and flexibility to adapt to the fundamental normative challenges associated with novel impairments of meaningful patient decision making. They are not amenable to the kinds of generative development associated with the common law. Furthermore, although both England and the U.S. possess statutory regimes relevant to the consent of research subjects, these will not be considered. Indeed, here one can adduce the additional rationale that this work's principal concern is the therapeutic deployment of AI.

Given this exclusion, the focus of this work will be made up of the consent obligations that have been developed under the established common law mechanisms of negligence and battery. In both jurisdictions these are the primary instruments determining what information a patient must be provided with under therapeutic circumstances. Battery requires that an individual's person not be interfered with unless there is a relevant legal

law. The common law is judge-made law. For centuries judges have been charged with the responsibility of keeping this law abreast of current social conditions and expectations. That is still the position. Continuing but limited development of the common law in this fashion is an integral part of the constitutional function of the judiciary': *Re Spectrum Plus Ltd* [2005] UKHL 41, [2005] 2 AC 680 [32].

justification, including an individual's valid consent. By comparison negligence requires that healthcare professionals exhibit due skill and care in the performance of their tasks, including in the disclosure of information to the patient. The term 'informed consent' may be applied to refer to the cumulative demands that the courts have imposed on healthcare professionals under the head of these requirements.[39]

This terminology closely reflects the connection between the developed legal doctrine and the underlying bioethical discourse surrounding patient autonomy. This value arguably provides the principle underlying the common law's fashioning and refashioning of battery's and negligence's more specific rules in light of patient consent.

Lastly for this section, it must be noted that the focus on common law mechanisms must be accompanied by a restriction of jurisdictional scope. In the UK the implications of this delimitation are less severe given that the highest court, the UK Supreme Court, often judges on common law principles, including issues of tort law and informed consent.[40] Although its decisions are then strictly binding only in the relevant jurisdiction in which the case arose – Scotland, Northern Ireland, or England and Wales – they command considerable respect and persuasive authority in all constituent nations and have generally been followed across the UK.[41] For ease of exposition the author sometimes refers to the law of a specific nation (especially England where a majority of cases arise), but this should not be taken to represent a hard and fast limitation on the scope of the research.

Things are very different in the United States. The federal nature of the system precludes the U.S. Supreme Court from judging on most matters of common law, which fall in the preserve of the individual states. Commenting on one leading exposition of tort law in the U.S. Gardner has critically remarked 'They give the impression that they are writing about the tort law of the United States of America. But it is doubtful whether there is such a thing. There are at least 51 legal systems in the United States with tort

39 Maclean, 'The Doctrine of Informed Consent' (2004) 24(3) Legal Studies p. 386, 392-393. Although the author notes a distinction that can be drawn between 'real consent' required for battery and the 'informed consent' under negligence.

40 The leading informed consent decision of *Montgomery v Lanarkshire Health Board* is itself a Scottish case, but it has shaped the common law across the UK: *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430.

41 Bankowski, MacCormick and Marshall in MacCormick and Summers, *Interpreting Precedents: A Comparative Study* (2016) 315.

jurisdictions'.[42] There is no doubt some uniformity of principle amongst these jurisdictions, but there are also important differences. Any detailed examination of what specific mechanisms require, and how legal reasoning engages with these requirements, must therefore be conducted at the level of the individual system, the individual state. Although cases from other states may serve as persuasive authority for the determination of specific questions, as in the common law more widely, these are not automatically or routinely followed.

In consequence, the present analysis will not examine the approach of all U.S. jurisdictions. Rather, California is selected as a targeted case studied, enabling a nuanced evaluation of this system's requirements. One reason for selecting this state is the finding that its Supreme Court is consistently one of the most cited courts in the country.[43] This suggests that it is a state whose pronouncements of the law are more influential and can be expected to provide a better approximation to the state of affairs across the U.S. Nevertheless, it is by no means representative and significant divergent approaches will be touched upon where relevant in the course of the analysis.

A second reason is the particular role that California has played in the development of the informed consent doctrine. By operating in a manner responsive to wider socio-cultural changes, it has reinterpreted and reshaped the common law. Not only is it a decision of this jurisdiction – *Salgo v. Leland* – that provided the terminology of 'informed consent', but the seminal Californian Supreme Court case *Cobbs v. Grant* outlined applicable principles that have been widely cited and applied in California's sister jurisdictions.[44] Many more examples of such influential generative activity will be provided in Chapter 7.[45] Suffice it to say here that it is a

42 Gardner, *Torts and Other Wrongs* (2019) 1-2, referring primarily to Goldberg and Zipursky, 'Torts as Wrongs' (2010) 88(5) Texas Law Review p. 917.

43 Hinkle and Nelson, 'The Transmission of Legal Precedent among State Supreme Courts in the Twenty-First Century' (2016) 16(4) State Politics & Policy Quarterly p. 391, 399.

44 *Salgo v. Leland Stanford Jr. University Bd. of Trustees* (1957) 154 Cal.App.2d 560; *Cobbs v. Grant* (1972) 8 Cal.3d 229. For a high profile case outside of California citing these principles see, for example: *Mink v. University of Chicago* (N.D.Ill. 1978) 460 F.Supp. 713, 716-718.

45 Touching on cases like: *Truman v. Thomas* (1980) 27 Cal.3d 285 (imposing informed consent requires where there was a refusal of treatment) and *Moore v. Regents of University of California* (1990) 51 Cal.3d 120 (extending informed consent requirements beyond the realm of treatment risks and alternatives to include professional conflicts of interest).

state that is hypothesised to be particularly well suited to the adaptation of informed consent requirements to the dynamic challenges posed by medical AI.

C. Legal reasoning and technological innovation

Within this scope it is a fundamental aim of this work to further an understanding of the law's dynamic and variable interaction with the phenomenon of social change and specifically technological innovation.[46] The method developed here seeks to anticipate the possibilities for applying and adapting common law mechanisms to meet the challenges posed to patient autonomy by one such innovation: medical AI/ML.

As the quote at the beginning of this chapter illustrates, the law's ability to adapt to changing social, extra-legal circumstances has been a topic of long-standing interest. It has informed foundational schools of thought such as legal realism[47] and the economic analysis of the law.[48] In the selection of our comparator countries it has also been discussed that the adaptability of the law has informed examinations of different legal families (common law vs. civil law systems) and of different sources of law (common law vs. legislation).

Further, it is important to note that the responsiveness to external factors cannot be regarded as uniform across different legal subject areas. While the focus of this work, private law, has traditionally been regarded as less responsive to social and political influences,[49] this is not at all true for other

46 That AI is one instantiation of such technological innovation is the key point for our purposes: Liu and others, 'Artificial Intelligence and Legal Disruption: A New Model for Analysis' (2020) 12(2) Law, Innovation and Technology p. 205. For a more general definition of technology see: Crootof and Ard, 'Structuring Techlaw' (2021) 34(2) Harvard Journal of Law & Technology p. 347, 348-349.

47 Cohen, 'Transcendental Nonsense and the Functional Approach' (1935) 35(6) Columbia Law Review p. 809; Gilmore, 'Legal Realism: Its Cause and Cure' (1961) 70(7) The Yale Law Journal p. 1037.

48 Landes and Posner, *The Economic Structure of Tort Law* (1987). On the knock-on impact of this school of thought on Lessig's prominent theory of the regulation of technology (specifically cyberspace) see: Mayer-Schonberger, 'Demystifying Lessig' [2008](4) Wisconsin Law Review p. 713, 723-724.

49 Robertson in Robertson and Tang, *The Goals of Private Law* (2009) 269-279. This characterisation holds at least in the sense of the formally recognised influences on legal actors and commentators in this area: Landes and Posner, *The Economic Structure of Tort Law* (1987) 5-6.

fields, such as social law[50] and constitutional law.[51] Even generalisations within a given field must be questioned and this is one aspect that our methodology will uncover and emphasise. Structured adaptation under conditions of uncertainty and flux is a ubiquitous aspect of legal orders and many normative analyses thereof, but it is recognised to be highly variable. It exists in a state of tension between the law's stabilising function and its need to provide guidance that is appropriate to the subjects' circumstances.[52]

For technological innovation there has been a recurring problematization of this tension and its implications for the capacity of law to regulate the underlying phenomenon.[53] Although the literature on this subject remains highly fragmented and multi-faceted,[54] associated challenges are

50 'Sozialrecht zeichnet sich aber durch seine besondere Funktionalität aus. Es beruht nur in wenigen Grundzügen auf normativen Vorgaben, seine Existenz wie seine Ausgestaltung sind stärker von sozialpolitischen als von verfassungsrechtlichen Erwägungen geprägt. Und selbst diese Erwägungen sind oft hinter den einmal geschaffenen und dann ein eigenes Beharrungsvermögen entwickelnden Einrichtungen kaum mehr erkennbar': Becker, 'Sozialrecht und Sozialrechtswissenschaft' (2010) 65(4) Zeitschrift für öffentliches Recht p. 607, 619; Zacher in Zacher, Maydell and Eichenhofer, *Abhandlungen zum Sozialrecht* (1993) 17. The need for a legal framework responsive to social policy, so as to accommodate specifically the demands of the welfare state, has also been highlighted in common law systems: Harris in Harris, *Social Security Law* (2000) 3-5; Calabresi, *A Common Law for the Age of Statutes* (1985) 74.

51 It is impossible to provide a comprehensive listing of the literature that engages with the potential for the adaptation of constitutions in accordance with extra-legal pressures. Examples include: Johnson and Yi Zhu, *Sceptical Perspectives on the Changing Constitution of the United Kingdom* (2023); Harrison and Boyd, *The Changing Constitution* (2006); Ackerman, 'The Storrs Lectures: Discovering the Constitution' (1984) 93(6) The Yale Law Journal p. 1013. Interesting approaches falling under this rubric can, naturally, also be found in different legal families: Hornung, *Grundrechtsinnovationen* (2015).

52 Becker, 'Sozialrecht und Sozialrechtswissenschaft' (2010) 65(4) Zeitschrift für öffentliches Recht p. 607, 618.

53 The distinctness of the technological nature of change from other processes of social change and the particular legal problems posed by this are highlighted in: Moses, 'Why Have a Theory of Law and Technological Change?' (2007) 8(2) Minnesota Journal of Law, Science & Technology p. 589. See also: Crootof and Ard, 'Structuring Techlaw' (2021) 34(2) Harvard Journal of Law & Technology p. 347, 349-350; Price, 'The Newness of New Technology' (2001) 22(5-6) Cardozo Law Review p. 1885, 1888; Calabresi, *A Common Law for the Age of Statutes* (1985) 74-77.

54 Guihot, 'Coherence in Technology Law' (2019) 11(2) Law, Innovation and Technology p. 311; Brownsword, Scotford and Yeung, *The Oxford Handbook of Law, Regulation and Technology* (2016) 7-8: 'Any attempt to identify an overarching purpose or com-

widely and consistently discussed. Above all these discussions tend to exhibit three prevalent assumptions that frame the law as a problematic regulator of innovation: (1) the pre-eminence of an instrumentalist perspective that judges 'good outcomes', attained through the law or otherwise, primarily according to policy-orientated yardsticks (2) the identification of a substantial, unavoidable difference in the speed of adaptation or change between technology and the law (3) a preference for non-legal regulatory modalities that are said to be better placed to realise policy-objectives in relation to technological innovation.

In the remainder of this section the nature of these assumptions and the support they enjoy in the literature is outlined. Simultaneously, by indicating the manner in which the present methodology challenges, and to some extent departs from, them it is anticipated how the ensuing analysis of AI/ML devices and informed consent will contribute to the fundamental conversation concerning the law's relationship to innovation.

1. Instrumentality of law

The first theme to be found in the literature relates to the focus on policy-orientated reasoning and the assumption that legal modes of regulating technology are and ought to be directed by these ends.[55] Given the open-ended nature of policy-orientated reasoning, it is difficult to pin down one definitive account in this respect; the relevant ends are highly, potentially infinitely, variable. However, it is clear that prominent examples include the need: to support and promote innovation, to minimise harm and/or to suitably allocate economic resources and incentives.[56] These objectives are taken to be the *raison d'être* for legal and extra-legal technology regulation. To the extent that the legal system falls short in accomplishing the tasks set

mon identity in the multiple lines of inquiry in this field may well fail to recognize the richness and variety of the individual contributions and the depth of their insights'.

55 Brownsword, 'Law Disrupted, Law Re-Imagined, Law Re-Invented' (2019) 1 Technology and Regulation p. 10 24-27. This is also evident in Guihot's exposition of technology law: Guihot, 'Coherence in Technology Law' (2019) 11(2) Law, Innovation and Technology p. 311, 322-325.

56 Cockfield and Pridmore, 'A Synthetic Theory of Law and Technology' (2007) 8(2) Minnesota Journal of Law, Science & Technology p. 475, 503-504; Brownsword, *Law 3.0: Rules, Regulation, and Technology* (2021) 21-22; Hoffmann-Riem, *Innovation und Recht - Recht und Innovation: Recht im Ensemble seiner Kontexte* (2016) 28-35.

according to these malleable yardsticks, the law is subject to censure and ought to be reformed, supplemented or abrogated.[57]

In responding to this prevalent assumption, this work builds on critical accounts that object to a conception of the law as purely an instrument for the realisation of external values.[58] Instrumentality is partly objectionable from a purely descriptive perspective because it posits an inaccurate unidirectional relationship between law, society and technology. But more than this, it is objectionable on normative grounds because it ignores the way in which values are transformed through their interaction with the legal process and it does not account for the wider relevance of the outputs of this process.[59] Assuming an instrumentalist perspective consequently closes off or distorts this contribution of the law to society's adjustments to technological change.

57 Brownsword, *Law 3.0: Rules, Regulation, and Technology* (2021) 17-20.

58 For critical overviews see: Tranter, 'The Law and Technology Enterprise: Uncovering the Template to Legal Scholarship on Technology' (2011) 3(1) Law, Innovation and Technology p. 31; Tribe, 'Technology Assessment and the Fourth Discontinuity: The Limits of Instrumental Rationality' (1973) 46(3) Southern California Law Review p. 617, 631: 'By focusing all but exclusively on how to optimize some externally defined end state, policy-analytic methods distort thought, and sometimes action, to whatever extent process makes–or ought to make–an independent difference'. For recent examples of such an approach see: Marchant in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (2011); Brownsword, 'Law Disrupted, Law Re-Imagined, Law Re-Invented' (2019) 1 Technology and Regulation p. 10, 15; Brownsword, *Rights, Regulation, and the Technological Revolution* (2008) 241.

59 Gutwirth and others have expressed a deep-seated dissatisfaction with the state of discourse in this area with which the present author sympathises: 'what we wanted to express was a disappointment towards a position that now dominates the legal discussions around new technologies—while at the same time rendering it impossible to go further. We do not consider that regulation is a terminus. On the contrary, we rather see it as a point where to start in order to build a more interesting legal appreciation of the emergence of new technologies. At the end of the present paper, it is not a mystery that we would see this legal appreciation formulated in the terms of the legal practice itself, rather than in the terms of what, for lack of better words, we are forced to qualify as political science. We trust that to ask the lawyers themselves how they deal with new technologies would always be more interesting and more enlightening than to define some very sophisticated program, however balanced and nuanced it might be, in order to avoid their escape': Gutwirth, Hert and Sutter in Brownsword and Yeung, *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes* (2008) 216. See also: Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (2015) 163-165.

Beginning with the descriptive component, critical commentators have picked up on the fact that instrumental accounts tend to imply a certain relationship between technology, society and the legal system. Specifically, it is claimed that they exhibit 'technological determinism' – positing a relatively simple, unidirectional influence of technological progress on the law.[60] Law serves extraneous ends *because* innovation will inevitably affect its functioning, whereas accelerating and self-reinforcing technological progress could not be stopped or, at any rate, only at an unrealistic cost.[61] In this manner, law and society are bound to conform to the new reality shaped by innovation and legal actors – scholars, judges, legislators – must adjust their thinking as best they can.[62]

However, such a determinism is objectionable because it is inaccurate. This is evidenced by the fact that it is widely eschewed in other areas of social science enquiry.[63] To the extent that accounts in the law and technology field reproduce this narrative, they are ignoring the way in which technology is socially constructed and, partially also, constructed by the surrounding legal framework.[64]

---

60 Jones, 'Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw' [2018](2) University of Illinois Journal of Law, Technology & Policy p. 249, 253-260; Mayer-Schonberger, 'Demystifying Lessig' [2008](4) Wisconsin Law Review p. 713, 737-738.

61 This is all the more true as there are now synergies between fundamental fields of technological development that speed up the progress of the others, including: 'nanotechnology, biotechnology, robotics, information and communication technology (ICT), and applied cognitive science': Allenby in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (2011) 7-11.

62 Jones, 'Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw' [2018](2) University of Illinois Journal of Law, Technology & Policy p. 249, 256; Tranter, 'The Law and Technology Enterprise' (2011) 3(1) Law, Innovation and Technology p. 31, 76: 'The message is that it is law's function to implement policy, not to debate the merits of policy'.

63 Jones, 'Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw' [2018](2) University of Illinois Journal of Law, Technology & Policy p. 249, 253-259; Mayer-Schonberger, 'Demystifying Lessig' [2008](4) Wisconsin Law Review p. 713, 739-740. For a good overview of the intellectual development of the approach in the social sciences see: Winner, 'Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology' (1993) 18(3) Science, Technology, & Human Values p. 362; Grattet, 'Sociological Perspectives on Legal Change: The Role of the Legal Field in the Transformation of the Common Law of Industrial Accidents' (1997) 21(3) Social Science History p. 359, 363-365.

64 '[T]echnologies must be understood as integrated cultural, economic, institutional, and built phenomena. To consider technology to be merely artifacts, while appropri-

The described arguments indicate that the descriptive criticism maintains some relevance. However, as many legal scholars do now address the interrelationship between law, technology and society,[65] and seemingly to a growing extent,[66] it does not constitute the focus of this analysis. In particular, critics of law as a regulatory tool submit that the legal framework plays a role in shaping technological development and society's reaction to that development.[67] This role is often encapsulated precisely in the arguments of those who fear that it threatens the instrumental goal of furthering, promoting (rather than stifling) innovation.[68]

Where the existing accounts still fall short is arguably in respect of the aforementioned normative dimension. They maintain a limited outlook on the nature of legal norms and fail to account for the wider relevance of values that are constructed or transformed through the legal process. Even if the law is seen to have an influence on innovation, avoiding technological determinism, it is still framed in terms of quite limited standards.

Specifically, in so far as the primary emphasis of commentators remains (in a relatively narrow positivistic fashion)[69] on rules, this allows them to frame the legal system as a discretionary tool that can realise any ex-

---

ate in many cases, would lead to gross over-simplification and dysfunctional analysis and policy formulation if applied to technology systems': Allenby in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011) 7.

65 Kaminski charts how one prominent early scholar of cyberlaw addressed this dimension, but occupied a minority position: Kaminski, 'Technological "Disruption" of the Law's Imagined Scene: Some Lessons from Lex Informatica' (2021) 36(3) Berkeley Technology Law Journal p. 883, 888. See also: Moses, 'Why Have a Theory of Law and Technological Change?' (2007) 8(2) Minnesota Journal of Law, Science & Technology p. 589, 599-600.

66 For a recent nuanced engagement with this point see: Crootof and Ard, 'Structuring Techlaw' (2021) 34(2) Harvard Journal of Law & Technology p. 347, 355-356.

67 Lessig, *Code: Version 2.0* (2006) 61-80; Brownsword, *Law 3.0: Rules, Regulation, and Technology* (2021) 51-53; Cockfield and Pridmore, 'A Synthetic Theory of Law and Technology' (2007) 8(2) Minnesota Journal of Law, Science & Technology p. 475, 497-500.

68 'In some cases, lethargic development of new legislation or adaptation of existing legislation in response to scientific discovery or development can impede research and innovation, resulting in blocking of new technology': Marchant in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011) 25.

69 Cf. Tranter, 'The Law and Technology Enterprise' (2011) 3(1) Law, Innovation and Technology p. 31, 69.

ternal policy objective.[70] Rules themselves could apparently be interpreted, amended, abrogated or replaced simply in accordance with the desired ends of a relevant authority.[71] When examined in isolation they appear easy to refashion in a way that is directly transparent to desired values. With a background assumption that technological progress amounts to social progress, it can only be expected that the innovation context heightens the perceived malleability of the law even further.[72]

Of course, there are also voices that recognise a broader role for the law in guiding societal adjustment to innovation. This perspective is promoted mostly by those who assert the importance of basic human rights standards.[73] The recognition of the relevance of these fundamental normative goalposts provides an invaluable contribution to the law and technology literature, moving away from an instrumentalist perspective and recognising the law's unique potential in this area.[74] However, it arguably does not

70 It is notable how often the framing is focused on the specific, rule-based framework. See: Mandel in Brownsword, Scotford and Yeung, *The Oxford Handbook of Law, Regulation and Technology* (2016); Rustad and Koenig, 'Cybertorts and Legal Lag' (2003) 13(1) Southern California Interdisciplinary Law Journal p. 77; Kaminski, 'Technological "Disruption" of the Law's Imagined Scene' (2021) 36(3) Berkeley Technology Law Journal p. 883, 892. Epstein presents a version of this position in response to economic analyses of the law: Epstein, 'The Static Conception of the Common Law: Legal and Economic Perspectives' (1980) 9(2) The Journal of Legal Studies p. 253, 269-273.

71 The apparent unfettered discretion that a focus on rules granted judges was one of Dworkin's criticisms of the positivist conception of law: Dworkin, *Taking Rights Seriously* (1987) 37-39.

72 Jones, 'Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw' [2018](2) University of Illinois Journal of Law, Technology & Policy p. 249, 256-257.

73 Brownsword, 'Law Disrupted, Law Re-Imagined, Law Re-Invented' (2019) 1 Technology and Regulation p. 10, 27; Murphy, *New Technologies and Human Rights* (2009). Cf. Askland who hints at a broader role: 'If the changes are unavoidable, but their form is unspecified, then law ought to be able to negotiate about that form. This negotiation role surely should not be limited to law, but as law reflects important ethical, social and economic values, it ought to be included among the negotiators': Askland in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (2011) *xvii*.

74 Sator makes an instructive statement in this regard 'Human rights are indeed important since they provide us with a framework for articulating some basic normative structures for the governance of the information society, in the awareness of the human values at stake. It is true, authoritative formulations, doctrinal developments and social understanding of human rights cannot provide us with a complete regulatory framework: economical and technological consideration must be taken into account,

go far enough in recognising the pervasive presence of relatively abstract legal norms (i.e. norms that are not fundamental rights and which do not originate in public or international law) that co-exist, conflict and possess relevance to society's co-evolution with innovation.

The method of this work will bring out this dimension by emphasising the role of principles in legal argumentation. Such principles may be motivated by external normative influences, but they have manifestly been translated into unique terms that allow them to perform a structured role in the resolution of conflicts.[75] This prevents an arbitrary repurposing of legal mechanisms and, critically, it demonstrates how the law offers instructive guidance going beyond its immediate area of application.

2. Technological dynamism and legal inertia

The second assumption that is widespread in the literature focuses on the inertia of legal regulation. It is stated that legal regulation lags, and must lag, behind technological developments.[76] Technological innovation embodies

---

while legal traditions and political choices play a decisive role in many regards (even with regard to the very understanding of human rights and their balance). However, the human-rights discourse still play an important role: it identifies some basic fundamental needs and entitlements, it links our understanding of such needs and entitlements to successes and failures of human history, it enables us to provide a context for our analysis of the new issues emerging in the information society, linking such analysis to a rich background including legal cases as well as social, political and legal debates': Sartor in Azevedo Cunha and others, *New Technologies and Human Rights: Challenges to Regulation* (2013) 19. Another commentator who highlights the benefits of human rights, especially their ability to be flexibly adapted to local circumstances is: Somsen in Murphy, *New Technologies and Human Rights* (2009) 115.

75 Compare Brownsword who has undertaken a normative analysis of 'a newly recognised interest in human dignity', selecting a contested understanding of that interest and hypothetically integrating this interest into the existing tort system, but not deeming it necessary to conceive of the value itself in a way that transforms it into a legal norm: Brownsword, 'An Interest in Human Dignity as the Basis for Genomic Torts' (2003) 42(3) Washburn Law Journal p. 413, 416-419. In Part II. we will see similar trends in the literature on autonomy, treating it purely as a value that is utilised in legal reasoning.

76 Friedman and Ladinsky, 'Social Change and the Law of Industrial Accidents' (1967) 67(1) Columbia Law Review p. 50, 73; Rustad and Koenig, 'Cybertorts and Legal Lag' (2003) 13(1) Southern California Interdisciplinary Law Journal p. 77; Marchant in

the idea of rapid, even revolutionary, change.[77] This change bypasses or disrupts existing legal paradigms.[78] These paradigms must be adjusted or abandoned, in line with the instrumentalist approach, to serve external goals. This represents an orthodox understanding of the 'legal lag',[79] or the 'pacing problem'.[80]

Although connected, this narrative of the law's inflexibility in the face of change is more commonly made explicit in the literature on the interaction between law and technology than the instrumentalist thesis.[81] Still, one would be remiss if one did not properly account for the spectrum of positions that have emerged from the basic assumption that the law struggles to adapt to the speed of innovative developments.

At one end of this spectrum there is a refined, but unfaltering account of the pacing problem. It is submitted that the law does possess capabilities for adaptation, but that technological change challenges these in a manner, or at a scale, that other social developments do not.[82] In consequence, the gap in relation to the law's ever-shifting regulatory target remains a problem of considerable magnitude – at the very least given the speed of change that modern societies are experiencing.[83] At the other end of the spectrum, there are a miscellany of approaches that view the law as more adaptable than it

---

Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011).

77 Price, 'The Newness of New Technology' (2001) 22(5-6) Cardozo Law Review p. 1885, 1886, 1912-1913.

78 Brownsword, *Law 3.0* (2021) 17-20.

79 Rustad and Koenig, 'Cybertorts and Legal Lag' (2003) 13(1) Southern California Interdisciplinary Law Journal p. 77, 77-78.

80 Marchant in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011) 23.

81 Friedman and Ladinsky, 'Social Change and the Law of Industrial Accidents' (1967) 67(1) Columbia Law Review p. 50, 73.

82 Askland in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011) *xviii-xix*.

83 'On the face of it, coherentism belongs to relatively static and stable communities, not to the dynamic and turbulent technological times of the Twenty-First Century not as a response to unauthorised drones at airports, or to dangerous or distressing online content, or to accidents involving robot carers (…) to assume that traditional legal frameworks enable regulators to ask the right questions and answer them in a rational way seems over-optimistic': Brownsword, 'Law Disrupted, Law Re-Imagined, Law Re-Invented' (2019) 1 Technology and Regulation p. 10, 17; Marchant in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011) 25-27; Post, 'Against "Against Cyberanarchy"' (2002) 17(4) Berkeley Technology Law Journal p. 1365, 1374-1384.

is generally given credit for and innovation is not generally understood to represent an exceptional challenge in this respect.[84] Naturally these latter approaches still submit, however, that law can only change to a limited extent. Stability remains an important reference point.[85]

The interesting question that emerges from this line of thinking for the present work, is not that this spectrum exists, which has been the subject of considerable discussion,[86] but the manner in which the law's adaptability is properly conceived from a methodological standpoint. To the extent that some adaptability is associated with the law, one must ask how this is properly represented and accounted for in the doctrinal analysis of a fast-evolving, allegedly disruptive innovation, such as AI. Only once the law's potential for dynamism has been scrutinised within a given context, should one make a specific judgment on the existence of an allegedly irreducible legal inertia.[87]

One starting point is certainly to attempt to anticipate the nature of an emerging innovation and its wider implications.[88] This work addresses this

84 Tranter offers an overview of early legal analyses of technology that were optimistic in this regard: Tranter, 'The Law and Technology Enterprise' (2011) 3(1) Law, Innovation and Technology p. 31, 38; Morgan's specific analysis of tort law goes so far as to state an 'adaptability hypothesis' according to which tort law can be developed to respond to the harms of new technologies: Morgan in Brownsword, Scotford and Yeung, *The Oxford Handbook of Law, Regulation and Technology* (2016) 524-526. In a recent contribution Ard has also considered that, at the level of doctrine, changes demanded by technological innovation are often less problematic than widely perceived: Ard, 'Making Sense of Legal Disruption' (2022) Forward Wisconsin Law Review p. 42, 46-48.

85 Predictability and stability are valuable contributions of the law's functioning and they usually constitute an asset rather than an insurmountable problem: Kaminski, 'Technological "Disruption" of the Law's Imagined Scene' (2021) 36(3) Berkeley Technology Law Journal p. 883, 888.

86 Berman, *Law and Society Approaches to Cyberspace* (2007) *xiii-xiv*.

87 Friedman and Ladinsky, 'Social Change and the Law of Industrial Accidents' (1967) 67(1) Columbia Law Review p. 50, 73-75.

88 Tranter, 'The Law and Technology Enterprise' (2011) 3(1) Law, Innovation and Technology p. 31, 38; Cockfield, 'Towards a Law and Technology Theory' (2003) 30(3) Manitoba Law Journal p. 383, 384-388; Askland in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (2011) *xv-xvii*. In contrast Moses highlights the difficulty of anticipating such developments: Moses, 'Why Have a Theory of Law and Technological Change?' (2007) 8(2) Minnesota Journal of Law, Science & Technology p. 589, 599-600.

aspect in Part I. But the question remains how a pre-existing legal doctrine then adjusts itself in the light of the identified shifting circumstances.[89]

A common method is to situate legal responses on a continuum of more liberal approaches and more restrictive, conservative one.[90] Naturally, this represents a simplification. But more concerningly, it often does not represent a considered attempt to track legal reasoning at all. Instead, it is based on an instrumental perspective. For example, the choice of a court to assume one stance or the other is framed by reference to consequentialist concerns or by reference to the goal of making other regulatory interventions more likely.[91] Given this, such approaches fail to engage with the law's internal capacity for adaptation. Again, the posited vision is that of external pressures refashioning a legal system that has already failed to 'get with the programme'.

More detailed considerations emerge in works that examine how legal adaptation proceeds according to uniquely legal demands.[92] Here the pro-

89 'Altered flows of information, resulting from new technologies, change the balances that previously existed in a legal framework. But it is hard to know when those changes undo the preexisting formulaic approaches to a task': Price, 'The Newness of New Technology' (2001) 22(5-6) Cardozo Law Review p. 1885, 1913. One common pitfalls is to think too narrowly, to focus only on narrow analogical arguments in this regard. This has provided a convincing basis for critique: Post, 'Against "Against Cyberanarchy"' (2002) 17(4) Berkeley Technology Law Journal p. 1365, 1373-1376.

90 Cockfield, 'Towards a Law and Technology Theory' (2003) 30(3) Manitoba Law Journal p. 383, 407-408; Brownsword, *Rights, Regulation, and the Technological Revolution* (2008) 167; Crootof and Ard, 'Structuring Techlaw' (2021) 34(2) Harvard Journal of Law & Technology p. 347, 379-386.

91 Respectively: Cockfield, 'Towards a Law and Technology Theory' (2003) 30(3) Manitoba Law Journal p. 383, 407-408; Brownsword, *Rights, Regulation, and the Technological Revolution* (2008) 182-184.

92 Gutwirth and others have recognised these procedural aspects by forcefully asserting that external, technological questions must be transformed into 'legal matter' to become 'the object of the legal operation': Gutwirth, Hert and Sutter in Brownsword and Yeung, *Regulating Technologies* (2008) 204-205. In the context of tort law, Bell and Ibbetson have analysed and affirmed the distinct role that legal doctrine plays in the process of adaptation to social change. Such adaptation is not free but is directed by existing forms of thought and the need to ensure coherence within a system of norms: Bell and Ibbetson, *European Legal Development: The Case of Tort* (2012), 162-163. This is true especially within judge-made law, but legislative activity too is subject to the implicit constraints of the legal operation, interacting with established modes of thought and often proceeding incrementally: Bell and Ibbetson, *European Legal Development: The Case of Tort* (2012) 171-172. It is a separate question in how far legislative activity is bound by its own rules and in how far this is detrimental when confronted with novel or extraordinary circumstances. I have commented on

cess of legal reasoning is conceived of as providing the potential for a creative activity, allowing a dynamic and contestable crystallisation of legal commitments.[93] When interacting with novel social circumstances, judicial actors must be mindful of the limitations of their enterprise, they are performing legal operations, but this does not preclude creativity – there can be '"obligations" of creativity'.[94]

Under such an account it is evident that the law's response to technological uncertainty inherently makes a generative contribution that is distinctly related to the legal process.[95] It is not exhausted by an analysis of limiting rules that are subject to adaptation according to vicissitudes of external necessity and external evaluations: 'One must take into account the law's own dynamics, the own *devenir* of the law'.[96] Overall this is hypothesised to yield a more nuanced and informative account of the law's co-evolution with technology and with society more generally.

The present work, through its common-law lens, engages precisely with this generative dimension of the law and seeks to offer one concretisation of it. Again, by focusing its analysis on the role of principles – which impact and alter existing, potentially inapposite, norms in a structured manner – a richer account of the law's potential for adaptation is provided. The underpinnings for this approach are developed in Part II. and its implications are seen in Part III., where it is examined how well the law is placed to adapt to the autonomy-based challenges posed by medical AI.

---

this elsewhere: Günther, 'Legal vs. Extra-Legal Responses to Public Health Emergencies' (2022) 29(1) European Journal of Health Law p. 131.

93 An excellent exposition of this point is provided in the German context by: Hoffmann-Riem, *Innovation und Recht - Recht und Innovation: Recht im Ensemble seiner Kontexte* (2016) 80-84. See also Kaminski, 'Technological "Disruption" of the Law's Imagined Scene' (2021) 36(3) Berkeley Technology Law Journal p. 883, 892-895.

94 Gutwirth, Hert and Sutter in Brownsword and Yeung, *Regulating Technologies* (2008) 205-208. This general point going towards the creative potential of legal restrictions also echoes elements of Lon Fuller's work, especially his conceptualisation of a 'liberating limitation': Fuller in Winston, *The Principles of Social Order: Selected Essays of Lon L. Fuller* (Revised Edition 2001) 66.

95 Gutwirth, Hert and Sutter in Brownsword and Yeung, *Regulating Technologies* (2008) 207-209, 214-215. The authors here also refer to Dworkin and MacCormick's theories, which will be utilised in our understanding of legal reasoning. However, while identifying their relevance to the underlying issue, they mainly refer to them in order to distinguish them from Latour's description of legal practice (the theory that they themselves rely on).

96 ibid 216.

3. Desirability of non-legal regulation

If one pairs the instrumental assessment of law with the view that there is a substantial lag in the legal system's adjustment to technological change, then one can already see why the legal regulation of technology can be judged ineffective. A third strand in the law and technology literature adds to these views the conviction that, not only does technology evolve quicker in terms of generating societal challenges, but that it is simultaneously better placed to offer solutions to these challenges through relevant designs and architectures.[97] To be clear, this does not mean that law cannot have a role in framing technological solutions, but only that it should be supplemented to a sufficient degree by these and that it should perform a secondary function.[98]

Having outlined the nature of the first two assumptions, we can deal with this third aspect more succinctly. For, it builds upon the *relative* slowness of the law and it is based upon an instrumental mentality that prizes the *relative* effectiveness of technological solutions. Specifically, the potential benefits of technological over legal regulation relate to the direct, unmediated effect of the former and the rapidity with which any necessary adjustments can be made.[99]

For example, a change in the design of cyberspace – effectively *ex ante* regulation through code – has been described as the application of a rule 'through a kind of physics'.[100] It has often been repeated in the context of cyberlaw that the architecture, the design of a technical environment can leave users with no choice other than to obey. This ensures compliance

97 Famously Lessig referred to such solutions as West Coast Code, as opposed to the legal East Coast Code: Lessig, *Code* (2006) 71-74. For a recent critical examination of the relationship between code and law as regulatory modalities and the influences between the two see: Kähler in Kuhli and Rostalski, *Normentheorie im digitalen Zeitalter* (2023).

98 Reidenberg, 'Governing Networks and Rule-Making in Cyberspace' (1996) 45(3) Emory Law Journal p. 911, 927-930; Lessig, *Code: Version 2.0* (2006) 114-119; Brownsword, *Law 3.0: Rules, Regulation, and Technology* (2021) 28-30.

99 Reidenberg, 'Lex Informatica: The Formulation of Information Policy Rules through Technology' (1997) 76(3) Texas Law Review p. 553, 577-581. Here Reidenberg refers also to 'jurisdictional advantages', which are excluded in the following as they are less relevant to the regulation of physical technologies such as AI medical devices.

100 Lessig, *Code* (2006) 81-82.

much more effectively than the legal system's behaviour-guiding norms could ever hope to achieve.[101]

Similarly, it has been remarked that customisations in the technological sphere can be achieved with lower costs and increased flexibility (in the sense of a personal tailoring of measures) than would be possible for legal regulation.[102] Such increased adaptability offers one apparent response to the pacing problem. Whereas the law is cumbersome, time-consuming and expensive to change, technology is not. This has also been used as an argument for the overhaul of legal thinking in favour of a more instrumentalist mindset: since non-technological solutions score comparatively worse under the selected parameters they are said to be fundamentally disrupted and it only makes sense to exhibit a general tendency towards technological solutions.[103]

This thesis will touch upon the nature of extra-legal solutions in the clinical AI context. The Part I. analysis of ML devices seeks to anticipate whether changes to the underlying technology can immediately and effectively make contributions to the resolution of relevant normative problems. Rather than framing this as an assessment that arises from a disruption of legal thinking, however, this step is conceived of as an orthodox element in the functional, comparative evaluation of the law. The touted benefits of technological modalities will then be questioned in the final evaluation of this work, drawing lessons from the concrete example of clinical AI.

4. Summation

Throughout the following evaluation of the selected legal systems' response to the autonomy challenges posed by clinical AI, the significance of these underlying methodological commitments will emerge in manifold ways. It should not be forgotten that one aim of the specific analysis is to present a dynamic and nuanced perspective of the way in which the law adapts to emerging technologies. This provides a further thread connecting its three parts and, in the final analysis, the fundamental insights to be gleaned from this case study will be elaborated upon.

---

101 Berman, *Law and Society Approaches to Cyberspace* (2007) *xvi-xvii*; Brownsword, *Law 3.0: Rules, Regulation, and Technology* (2021) 23-25.

102 Reidenberg, 'Lex Informatica' (1997) 76(3) Texas Law Review p. 553, 579-581; Lessig, *Code: Version 2.0* (2006) 126-127.

103 Brownsword, *Law 3.0: Rules, Regulation, and Technology* (2021) 22-25.

D. Research question

Having established the method to be applied in this work, the research question can now be formulated. Drawing on the common law's ability to realise principles as well as laying down specific norm, it is asked:

> Can the common law doctrine of informed consent ensure an adequate protection of patient autonomy as artificial intelligence is introduced into medicine?

It is with respect to this question that the state of the existing research must be ascertained and, subsequently, the requisite analytical structure must be provided.

III. State of the art

Regarding the informed consent requirements that are to be applied to AI, there is a valuable repository of academic literature dealing with the ethical dimension of this doctrine. Above all such commentators identify types of information that may be necessary to meet these challenges.[104] In doing so they complement the concerns raised in professional guidance and policy documents,[105] but they fail to construct or deploy an underlying normative ideal by reference to which legal instruments could be applied or assessed.

In comparison, some specific debates go further – seeking to understand what particular bioethical conceptions of autonomy entail for interactions with the patient and the information that must be disclosed to them to secure the protection of this conception.[106] While offering insightful systematisations of challenges and emphasising the connection between general

104 See for example: Gerke, Minssen and Cohen in Bohr and Memarzadeh, *Artificial Intelligence in Healthcare* (2020). Although insightful, the authors touch only relatively briefly on informed consent issues, focussing their analysis on other matters. See also: Di Nucci, Jensen and Tupasela, 'Ethics of Medical AI: The Case of Watson for Oncology' (5.12.2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3432317> accessed 5.4.2020.

105 Nuffield Council on Bioethics, 'Bioethics Briefing Note: Artificial Intelligence (AI) in Healthcare and Research' (2018) <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> accessed 17.6.2022.

106 Miguel Beriain, 'Should We Have a Right to Refuse Diagnostics and Treatment Planning by Artificial Intelligence?' (2020) 23(2) Medicine, Health Care, and Philosophy p. 247; See also: Ploug and Holm, 'The Four Dimensions of Contestable AI Diagnostics - A Patient-Centric Approach to Explainable AI' (2020) 107 Artificial

standards and specific norms, these debates remain of limited relevance to the present analysis. If at all, they do not engage with the relevant legal material and reasoning in the requisite depth.

Within the legal literature it is possible to find some preliminary analyses of the issue of informed consent. Many of these are regrettably cursory in nature,[107] while others focus on quite specific aspects of the doctrine.[108] In the UK context, one of the most insightful articles attempting to seriously grapple with the informed consent obligations of AI has been framed almost exclusively by reference to risk-disclosure.[109] Although this author does appeal to a certain understanding of patient autonomy, he does not offer a detailed analysis of its nature.[110] Nor is there any extensive scrutiny of relevant legal norms.[111]

By comparison Glenn Cohen is one author who has engaged with the issue of informed consent to AI use much more comprehensively, and through an analysis of U.S. common law no less. However, his analysis is fundamentally different from the present. For one, he makes a blanket assessment, spanning the different common law jurisdictions of the U.S. The extent to which this yields a persuasive picture of the operation of specific legal mechanisms has been questioned above.

Intelligence in Medicine; Ploug and Holm, 'The Right to Refuse Diagnostics and Treatment Planning by Artificial Intelligence' (2020) 23(1) Medicine, Health Care, and Philosophy p. 107; Bjerring and Busch, 'Artificial Intelligence and Patient-Centered Decision-Making' (2021) 34(2) Philosophy & Technology p. 349; Rubel, Castro and Pham, *Algorithms and Autonomy* (2021); Debrabander and Mertes, 'Watson, Autonomy and Value Flexibility: Revisiting the Debate' [2021] Journal of Medical Ethics p. 1043.

107 Dismissing the possibility of novel informed consent obligations *tout court* see: Schönberger, 'Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications' (2019) 27(2) International Journal of Law and Information Technology p. 171, 188: 'even AI applications in riskier areas would not add anything novel. An explanation of the inner workings of the respective algorithms would not empower patients to make an informed choice about a given treatment'. For a brief, practical approach see also: Keating and Wright in Hervey and Lavy, *The Law of Artificial Intelligence* (2021).

108 Nolan, 'Artificial Intelligence in Medicine - Is Too Much Transparency a Good Thing?' [2023] The Medico-Legal Journal Onlinefirst.

109 Kiener, 'Artificial Intelligence in Medicine and the Disclosure of Risks' (2020) 36(3) AI & Society p. 705.

110 ibid 706.

111 Although one does find scattered references to prominent case law: ibid 706-708.

Cohen also focuses on drawing incremental analogies to relatively limited classes of existing case law.[112] He terms this a doctrinal approach and distinguishes it expressly from a normative approach that could appeal to the principle of patient autonomy.[113] Although his treatment of this dimension is relatively cursory,[114] he does not purport to offer a close analysis. Rather, he states explicitly: 'The goal is to begin a conversation, not definitively answer it'.[115]

One manner in which to interpret the present research, is to respond to this call and to formulate one such answer. Yet it also goes further. The normative approach accounts for the outlined relationship between legal reasoning and technological innovation and anticipates the legal developments that may be generated in response. Without it one risks ignoring both the law's generative potential and its guiding function. In short, the positive, proactive role that it plays in shaping responses to innovation.

Overall, then, there is a real gap in the existing literature regarding the implications of AI/ML technologies for the legal doctrine of informed consent. What is missing is precisely a normative analysis of the underlying legal mechanisms to anticipate the common law's response to this novel societal challenge.

IV. Outlook and structure

To answer the research question in a manner that takes seriously the generative and dynamic nature of common law reasoning, the following analysis is split into four parts.

Part I. begins with the underlying factual phenomenon. It seeks to describe and categorises the AI currently found in the healthcare systems of the two outlined jurisdictions. Alongside this, it conveys the foundational understanding of the technology and its operation in clinical environments that is necessary for any serious legal assessment thereof. On the basis of these insights, it determines the normative challenges posed by AI/ML devices from the perspective of a procedural, rationalist account of patient autonomy.

112 Cohen, 'Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?' (2020) 108(6) The Georgetown Law Journal p. 1425, 1444-14449.
113 ibid 1449, 1557.
114 ibid 1557.
115 ibid 1449.

Part II. argues that this normative conception of autonomy offers one defensible interpretation of the jural concept found in the UK and U.S. jurisdictions. It further asserts that this concept is properly understood to operate as a common law principle in the domain of medical law. With this understanding one can frame a general standard, that is able to: justify existing norms, institute norm change, aid interpretation, generate new norms, create exceptions to rules and, potentially, to ground actions directly. In short, it lays the foundation for an analysis that goes beyond narrow deductive or incremental analogical reasoning. At the same time, it recognises that the doctrinal pressures exerted by the legal framework and other applicable norms cannot be ignored. Creative reasoning must be conducted within limitations.

Part III. turns to a comprehensive analysis of the specific legal mechanisms of negligence and battery in the UK and California. These offer a response to AI's autonomy challenges by requiring forms of consent and information disclosure. Where possible – i.e. where realistically permissible within the recognised constraints – argumentation is framed by reference to the autonomy principle. The varying strengths of such forms of argumentation, in light of countervailing normative considerations, is also accounted for. In the final analysis, a detailed picture emerges of the kinds of situations in which a patient may be able to assert a legal right to certain classes of information that can safeguard their autonomous decision making.

Chapter 8 represents an overall assessment of the relevant findings and demonstrates the wider significance of the pursued approach. It is analysed in how far the anticipated adaptations of the specific tort law instruments in the compared jurisdictions suffice to meet the novel autonomy challenges posed by medical AI/ML devices. It will be seen that there are several grounds for concluding that the common law's generative process provides only imperfect solutions. Nevertheless, its outputs play a guiding role. The law's operation transforms a technological problem for society and helps to discern normatively desirable resolutions. This is further demonstrated through an analysis of existing legislative schemes in England and California that supplement the common law and which are based upon a closely connected rationale. Ultimately, the three prevalent assumptions concerning the relationship between law and innovation are re-examined and critiqued. In light of the preceding arguments, it is not possible to claim that, as a matter of course, extra-legal solutions adapt more rapidly and provide effective, normatively defensible, resolutions to technological problems. Nor is it true that the law can be, or should be, straightforwardly

instrumentalised to serve innovation-related ends. Finally, the law's multifaceted potential for adaptation is differentiated and categorised. This refutes the claim that there is a problematic and inevitable disjuncture between legal thinking and fast-paced innovation.

Part I:
Practical and theoretical foundations

Chapter 2: Artificial intelligence's use in medicine

An accurate description of artificial intelligence's use in medicine must precede any normative or legal assessment thereof. The following chapter provides this in four stages. Section I. provides a definition of artificial intelligence (AI), a description of the structure and development of relevant techniques and the factors affecting their performance. Sections II. to IV. utilise this understanding to anticipate those aspects of AI functioning that will be problematic for the value of patient autonomy. Specifically Section II. offers case studies on the degrees of clinical AI automation, Section III. outlines the literature on the interpretability of AI and Section IV. explores their relation to established human clinical expertise.

I. Artificial intelligence

A. Definition

AI has a long history as a field of multidisciplinary research. It is related to the fields of computer science,[116] statistics/mathematics,[117] engineering and neuroscience.[118] The aims of this research have shifted over time, partly corresponding to the demands of each particular discipline, and they remain contested. It is unclear whether the goal is to emulate a general form of intelligence,[119] or to solve practical problems that had previously

116 Alpaydin, *Machine Learning* (Revised Edition 2021) 19.
117 Morik in Bauer and others, *Applications in Statistical Computing* (2019).
118 Vieira, Pinaya and Mechelli in Mechelli and Vieira, *Machine Learning: Methods and Applications to Brain Disorders* (2019) 1.
119 Pennachin and Goertzel in Goertzel and Pennachin, *Artificial General Intelligence* (2007) 1.

been understood to require a certain degree of human intelligence – i.e. to build 'smart tools'.[120] One may also ask what it even means to speak of intelligence in systems. Is it a matter of structure, of behaviour, of rationality, of particular cognitive functions or of a capacity to solve specific problems?[121] All these are questions thrown up by the goal of developing and/or understanding intelligent machines. It is therefore unsurprising that it has been difficult to arrive at a consensus in the definition of AI. The outlined approaches and the aims underlying them may be related in many ways, but they are not compatible.[122]

Any research touching on AI must grapple with this issue, so that it may at least offer a working of definition of the subject.[123] A definition must be found that is tailored to the goal of the investigation. Presently this is to analyse a certain set of challenges (for autonomy) that arise from technologies – which both exist and are in a continuous, rapid state of development – and that aim to solve relatively narrow challenges in a certain context (health care). The focus is squarely on the *application* of such technologies. This is a different investigation than one that abstractly or theoretically seeks to outline the nature of AI.[124] Therefore, rather than getting hung up on 'abstract notions of intelligence', the goal is to facilitate the analysis of 'useful artifacts'.[125]

Oriented towards these priorities, our concern is with intelligence as 'the ability to solve hard problems'.[126] This approach has been summed up as a 'Capability-AI' definition, which serves as a referent '[f]or people whose

120 Nilsson, *The Quest for Artificial Intelligence* (2009) 508-518.

121 These contrasting approaches are explored in: Wang in Wang, Goertzel and Franklin, *Artificial General Intelligence 2008: Proceedings of the First AGI Conference* (2008).

122 Wang, 'On Defining Artificial Intelligence' (2019) 10(2) Journal of Artificial General Intelligence p. 1, 13-14.

123 ibid 2-6.

124 For similar approaches see Matheny and others, 'Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril' (2019) <https://nam.edu/artificial-intelligence-special-publication/> accessed 5.4.2020 13-14: 'This publication does not address the hypothetical (…) and focuses instead on the current and near-future uses and applications of AI'; Turner, *Robot Rules: Regulating Artificial Intelligence* (2019) 15: 'this book does not seek to lay down a universal, all-purpose definition of AI which can be applied in any context. Its aim is much less ambitious: to arrive at a definition which is suited to the legal regulation of AI'.

125 Russell, 'Rationality and Intelligence' (1997) 94(1-2) Artificial Intelligence p. 57, 57.

126 Minsky, *The Society of Mind* (First Edition 1988) 71.

interest in AI mainly comes from its potential applications'.[127] For these 'the intelligence of a system should be indicated by its problem-solving capability'.[128] Therefore, to offer a useful definition of AI we must relate it to a certain problem-solving context, concurrently restricting the research agenda. Here this context is circumscribed by the healthcare field.

It has been a longstanding objective of the computer science community to develop programmes that can master the kinds of tasks that human medical experts solve by drawing on their intuition, knowledge and skill. In the past there were already some limited advances in this field.[129] But it is only over the past decade or so that this objective has increasingly been realised. Computers are now able to perform complex cognitive functions associated with clinical decision making and especially with diagnostic, prognostic and therapeutic tasks.[130] To a lesser degree they are also assisting with motoric actions and interpersonal aspects of healthcare, which form demanding components of human medical expertise in their own right.[131] It is the technologies demonstrating these capabilities in the healthcare field that constitute 'AI' for the purposes of this work.

B. Machine learning: the underlying technology

With this definition in mind our research can primarily be fixed on one type of AI technology: machine learning (ML). ML is distinct from previously dominant AI methods, referred to as good old fashioned or symbolic AI, and it has played an indispensable role in recent advances. It is devices with ML components that have begun to exhibit the outlined clinical expertise, most especially the cognitive capabilities involved in diagnoses, prognoses and the proffering of therapeutic advice.[132]

127 Wang, 'On Defining Artificial Intelligence' (2019) 10(2) Journal of Artificial General Intelligence p. 1, 10-11.

128 ibid 10.

129 See: Szolovits, *Artificial Intelligence in Medicine* (1982); Yu, Beam and Kohane, 'Artificial Intelligence in Healthcare' (2018) 2(10) Nature Biomedical Engineering p. 719, 719-722.

130 Braude in Schramme and Edwards, *Handbook of the Philosophy of Medicine* (2017) 702: 'Cognition refers to all mental processes related to knowledge, including but not limited to memory, attention, perception, representational schemas, consciousness, and language'.

131 ibid 706-712.

132 See Section II. below.

ML itself is a complex concept, serving as an umbrella term for many more specific techniques that accomplish these tasks.[133] As Burrell notes:

> popular machine learning models include neural networks, decision trees, Naïve Bayes, and logistic regression. The choice of model depends upon the domain (i.e. loan default prediction vs. image recognition), its demonstrated accuracy in classification, and available computational resources, among other concerns. Models may also be combined into 'model ensembles,'[134]

These techniques share the ability to automatically learn from data and to improve with experience: 'developers in ML program computers to find solutions on their own'.[135] They are able to autonomously adjust various aspects of their structure to find an efficient way to accomplish certain tasks, such as making classifications and making predictions about as-yet unseen data.[136] In addition, in contrast with traditional AI types these structures are also said to operate at a sub-symbolic level[137] – rather than consisting of units that are easily interpretable, with a clear conceptual meaning,[138] ML techniques tend to consist of 'fine grained dynamical features that are below the conceptual level'.[139] Relatedly, rather than attempting to model tasks on the basis of their logical description, making clear logical inferences,[140] the newer ML technologies function largely on the basis of probability theory – enabling these machines to deal with the ambiguities and uncertainties of life.[141] In making decisions ML algorithms are able to settle 'into an equilib-

133 Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective* (Second Edition 2020) 2-4.
134 Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) Big Data & Society p. 1, 5.
135 Zednik, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34(2) Philosophy & Technology p. 265, 267.
136 Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Interpretable* (2019) 13-14.
137 Berkeley, 'The Curious Case of Connectionism' (2019) 2(1) Open Philosophy p. 190.
138 Sun in Frankish and Ramsey, *The Cambridge Handbook of Artificial Intelligence* (2014)114.
139 Berkeley, 'The Curious Case of Connectionism' (2019) 2(1) Open Philosophy p. 190, 200.
140 Morik in Bauer and others, *Applications in Statistical Computing* (2019) 130.
141 Alpaydin, *Machine Learning* (Revised Edition 2021) 32-35.

rium state in which a majority of (potentially contradictory) constraints are simultaneously satisfied'.[142]

Four overarching classes of ML are often outlined according to the purpose of the ML and the nature of its training. These are: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning.[143] Supervised learning refers to the situation where the aim is to draw an inference from an input (e.g. an X-ray) to an output decision (e.g. the identification of a pathology). Here there is the assumption that human experts possess the requisite knowledge to connect the two variables, with them labelling the output that should be reached.[144] Supervised learning is a means for the computer to teach itself this human knowledge by developing its own way of connecting the input to the output that the human supervisor says is correct. The focus of this type of learning is classification: 'the aim is to predict the class of each observation' – i.e. to predict the appropriate label.[145]

Unsupervised learning by contrast does not rely on labelling by humans. It seeks to discover ways of grouping data according to its own criteria, discovering 'patterns, classes or distinctive features that cannot be readily interpreted by a human observer or necessarily judged against established gold standards or ground truth.[146] One common application for unsupervised learning is clustering.[147] For example, the algorithm may cluster groups of patients together according to traits which it determines to be similar and arriving at new categories (labels) for them.[148] Other important tasks include detecting novelties or outliers in the data or reducing its

142 Boden in Frankish and Ramsey, *The Cambridge Handbook of Artificial Intelligence* (2014) 95.

143 E.g. Chang in Riaño, Wilk and Teije, *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings* (2019).

144 Deo, 'Machine Learning in Medicine' (2015) 132(20) Circulation: Cardiovascular Quality and Outcomes p. 1920, 1920.

145 Scarpazza and others in Mechelli and Vieira, *Machine Learning: Methods and Applications to Brain Disorders* (2019) 46.

146 Kellmeyer in Mechelli and Vieira, *Machine Learning: Methods and Applications to Brain Disorders* (2019) 337.

147 Igual and Seguí in Igual and Seguí, *Introduction to Data Science* (2017).

148 Li and others, 'Unsupervised Analysis of Transcriptomic Profiles Reveals Six Glioma Subtypes' (2009) 69(5) Cancer research p. 2091.

complexity ('dimensionality').[149] This may be done by grouping correlated features together – leaving only a smaller group of uncorrelated principal components for analysis.[150] Very often these forms of unsupervised learning are not applied in their own right, but are important elements in supervised learning approaches.[151] So-called autoencoders are one modality that may be used to generate labels for use in supervised learning in clinical research.[152] As such, unsupervised learning offers possibilities to automate dimensions of ML development, including aspects of feature selection and data processing.

This illustrates that, while it is important to understand the different purposes underlying the classes – and their contribution to AI capabilities and a more independent functioning – in evaluating practical applications it is often difficult to draw sharp distinctions. They must be considered as likely elements of a complex whole. The same thing is also intimated by semi-supervised learning. Under this head, ML techniques are utilised to learn from both labelled and unlabelled data, enabling the model to account for gaps in the human expertise that is provided. These algorithms are capable of recognising and incorporating patterns from the unlabelled data.[153] A concurrent benefit is that this mixture can limit the impact of biases that could be introduced into the programme by human labelling.[154]

The final class is reinforcement learning. Here the machine is target-oriented.[155] It reacts to signals of rewards/penalties, (usually in the form of a numerical value) rather exact labels for desired outcomes.[156] Obtaining such labels that sufficiently correspond to the relevant situation would be

149 Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective* (Second Edition 2020) 12.
150 Jolliffe, *Principal Component Analysis* (Second Edition 2002) 1.
151 Schmidhuber, 'Deep Learning in Neural Networks: An Overview' (2015) 61 Neural Networks p. 85, 89.
152 Stevens and others, 'Recommendations for Reporting Machine Learning Analyses in Clinical Research' (2020) 13(10) Circulation: Cardiovascular Quality and Outcomes 782-793, 783.
153 Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective* (Second Edition 2020) 12.
154 Chang in Wulfovich and Meyers, *Digital Health Entrepreneurship* (2020) 75.
155 Naeem, Rizvi and Coronato, 'A Gentle Introduction to Reinforcement Learning and its Application in Different Fields' (2020) 8 IEEE Access p. 209320, 209322.
156 Sun in Frankish and Ramsey, *The Cambridge Handbook of Artificial Intelligence* (2014) .111-112.

impractical in relation to many interactive, uncertain problems.[157] Thus positive/negative signals are automatically generated *via* a trial and error approach in the course of the relevant interactions with the environment.[158] The ML agent must both exploit ways in which it has effectively solved the relevant problem in the past (signalled by the relevant reward) and explore new ways in which the problem may be solved in order to progressively improve – these issues do not arise in either supervised or unsupervised learning.[159] Reinforcement learning too can be paired with other types of ML techniques and may fulfil functions that are closely analogous to those usually fulfilled by supervised learning in clinical research and practice.[160]

Each of the 'classes' of AI offer different capabilities. Their selection and combination can serve reduce the need for human input and oversight. Supervised learning is often associated with a greater degree of human involvement, as developers establish criteria for the machine's functioning, and with a more consistent operation, as functioning is optimised on a stable dataset.[161] This is to be contrasted with unsupervised learning, which was seen to do away with different kinds of human engineering, and with reinforcement learning. The latter's ability to learn from changing, uncertain environments means that it can pursue dynamic, interactive responses.[162]

Therefore, selecting unsupervised or reinforcement learning approaches – or pairing them with supervised learning – is indicative of the higher degree of automation (independence from human input and action) that makes AI technology less predictable and controllable.[163] A potentially un-

157 Sutton and Barto, *Reinforcement Learning: An Introduction* (Second Edition 2018) 2.

158 Vieira, Pinaya and Mechelli in Mechelli and Vieira, *Machine Learning* (2019) 13.

159 Sutton and Barto, *Reinforcement Learning: An Introduction* (Second Edition 2018) 3.

160 Stember and Shalu, 'Deep Reinforcement Learning With Automated Label Extraction From Clinical Reports Accurately Classifies 3D MRI Brain Volumes' (17.6.2021) <https://arxiv.org/pdf/2106.09812> accessed 6.3.2022: here reinforcement learning is used to classify 2D and 3D brain images.

161 Strauß, 'Deep Automation Bias: How to Tackle a Wicked Problem of AI?' (2021) 5(2) Big Data and Cognitive Computing p. 1, 5.

162 Shortreed and others, 'Informing Sequential Clinical Decision-Making Through Reinforcement Learning: An Empirical Study' (2011) 84(1-2) Machine Learning p. 109, 111.

163 Strauß, 'Deep Automation Bias: How to Tackle a Wicked Problem of AI?' (2021) 5(2) Big Data and Cognitive Computing p. 1, 5-9.

precedented degree of automation *via* ML techniques allows the technology to operate in ways that are less closely and less obviously aligned with the goals of human users in practice.[164]

C. Specific features of ML models: the example of deep neural networks

To understand further, specific characteristics of ML it is worth illustrating a model through which these types of learning and their respective objectives are implemented. Hereby it is notable that one kind of algorithm is not necessarily restricted to one type of training outlined above. For instance, deep learning is one prominent approach that can be deployed to pursue all classes of AI learning: supervised, unsupervised, semi-supervised and reinforcement.[165] One deep learning algorithm may even be trained in sequence – first in an unsupervised and then in a supervised manner.[166] Again, one can see that a close connection between the different training methods often exists in practice. To provide some more concrete explanations of how ML functions, this section focuses on a type of deep learning, a deep neural network (DNN), which is trained in a supervised fashion.[167]

I. Sub-symbolic functioning

DNNs are species of artificial neural networks. Neural networks are made up of connected nodes that are often called neurons after their biological inspirations.[168] Neurons are essentially separate processing units where

164 ibid 5-6.

165 López-Rubio, 'Computational Functionalism for the Deep Learning Era' (2018) 28(4) Minds & Machines p. 667, 670.

166 ibid 671.

167 This particular modality features prominently in many viable clinical AI and illustrates many of ML's distinct attributes: 'Almost every type of clinician, ranging from specialty doctor to paramedic, will be using AI technology, and in particular deep learning, in the future. This largely involved pattern recognition using deep neural networks (DNNs) (…) that can help interpret medical scans, pathology slides, skin lesions, retinal images, electrocardiograms, endoscopy, faces, and vital signs': Topol, 'High-Performance Medicine: The Convergence of Human and Artificial Intelligence' (2019) 25(1) Nature Medicine p. 44, 44.

168 Buckner, 'Deep learning: A Philosophical Introduction' (2019) 14(10) Philosophy Compass p. 1, 2.

computations occur, taking values from the signals sent along the connections between them.[169] In a neural network these neurons are arranged in layers and function in parallel (i.e. many units can carry out their computations at the same time).[170] In the simplest form the network receives input signals in an input layer, processes this data in a hidden layer and provides a human-interpretable 'decision' *via* an output layer.[171] More exactly, the first layer receives inputs in the form of values from an input signal (e.g. corresponding to pixels in an image) and the inputs in the next layer come from the output signals generated by the activation of this initial layer of neurons.[172]

The activation of neurons is determined by three types of computation occurring in the input-neuron interaction, termed: weight, bias and activation function. Weight is the multiplication applied to each single input before it is computed in the neuron where it is received.[173] Within the relevant neuron two computations are then performed simultaneously.[174] Each neuron adds a constant term, *a bias*, to the sum of received weighted inputs. This pushes the sum in a direction that keeps the output in a desired range.[175] The resulting value is passed through a non-linear function that ultimately determines the degree of the neuron's activation and thus the output it provides to serve as an input for the next layer.[176] Hence this is the aforementioned activation function.

Overall one should note the sub-symbolic nature of this process. Each neuron is carrying out the simple task of utilising various values without clear semantic meaning and computing these *via* the outlined processes, which may be identifiable but are likewise without well-defined content. Depending on the size of the network, thousands or millions of these processes can occur simultaneously and in sequence.

---

169 Rumelhart, Hinton and McClelland in Rumelhart, James L. McClelland and PDP Research Group, *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations* (1999) 47.

170 ibid 47.

171 Buckner, 'Deep learning' (2019) 14(10) Philosophy Compass p. 1, 2.

172 Michelucci, *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks* (2018) 84.

173 ibid 32-34.

174 ibid 34.

175 Erb, 'Introduction to Backpropagation Neural Network Computation' (1993) 10(2) Pharmaceutical Research p. 165, 167.

176 Vieira and others in Mechelli and Vieira, *Machine Learning: Methods and Applications to Brain Disorders* (2019) 159.

Particularly DNNs are liable to contain complex combinations of such computations. Instead of having just one hidden layer, they contain many hidden layers between the input and output. These can possess different activation functions, fulfilling different roles and they can combine the sub-symbolic operations of neurons to achieve increasingly abstract and sophisticated representations of the data. The ability to automatically learn hierarchical representations distinguishes deep learning from shallow networks and renders it so successful as an ML tool.[177] Yet these representations are still distributed – that is, spread across neurons that each represent small, feature-like entities rather than concepts.[178] It is difficult to isolate a concept by focussing on a pattern of neurons and to understand or anticipate how changing the functioning of some elements, will impact the representation as a whole.[179]

Consequently, it will not be possible to straightforwardly associate these representations in the hidden layers with human-understood concepts and looking into the hidden layers does little to indicate the interpretable criteria that are being used to reach the outcome.[180] This mode of functioning exemplifies many of the challenges that will be discussed later in terms of interpretable AI and the black box problem in Section IV.

2. The training process

With this outline of DNN structures we should turn to the training process. Exploring this further promotes our understanding of AI-related automation and interpretability. All ML algorithms will have some features that

177 ibid 158.

178 Rumelhart, Hinton and McClelland in Rumelhart, James L. McClelland and PDP Research Group, *Parallel Distributed Processing, Volume 1* (1999) 47.

179 Zednik, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34(2) Philosophy & Technology p. 265, 280: 'there is often no way of knowing in advance whether an intervention on a single parameter will change the relevant system's behavior entirely or else affect it in a way that is mostly or entirely imperceptible'. See also Rudin and others, 'Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges' (2022) 16 Statistics Surveys p. 1, 32: referring to DNN's they state: 'concepts that are completely unrelated could be activated on the same axis' and that 'vectors in the latent space are "impure" in that they do not naturally represent single concepts'.

180 There are attempts to extract certain kinds of criteria, I will come onto this below when I discuss explainable AI.

will be altered with experience; these are known as learnable parameters. In a DNN these are, for example, the weights of connections between neurons, which will be changed automatically in response to a learning rule.[181]

To illustrate this one can turn to a common type of supervised learning for DNN's known as backpropagation.[182] Here an input is provided to a model with randomly initialised learnable parameters. Signals are propagated forward through the network in the manner described above, activating a pattern of neurons and providing a given output.[183] This output is automatically compared with a labelled set of correct decisions (hence this is supervised learning) *via* an error function.[184] This 'represents how far off the network is from making accurate predictions based on the input'.[185] Initially it is to be expected that the DNN will be quite mistaken, given the random values it is created with. To optimise performance, the errors are propagated back up the layers to the front of the network, adjusting the weights differently according to a pre-determined rule. Hereby a single forward- and backpropagation can cause the machine to update its parameters across a complex, multilayer network.[186]

By incrementally adjusting these weights in response to training samples, which may be presented one at a time or in small batches, a DNN 'can converge on the solutions to a wide range of classification and decision problems'.[187] Such problems span the areas of natural language processing, computer vision, speech recognition and robotics.[188] One can therefore see how these capabilities are acquired without close human involvement and direction. This allows for a greater degree of automation, even in a super-

181 Rumelhart, Hinton and McClelland in Rumelhart, James L. McClelland and PDP Research Group, *Parallel Distributed Processing, Volume 1* (1999) 46.

182 Hosseini and others in Pedrycz and Chen, *Deep Learning: Concepts and Architectures* (2020) 2.

183 Erb, 'Introduction to Backpropagation Neural Network Computation' (1993) 10(2) Pharmaceutical Research p. 165, 167.

184 Buckner, 'Deep learning' (2019) 14(10) Philosophy Compass p. 1, 2. This can also be called loss function: Vieira and others in Mechelli and Vieira, *Machine Learning* (2019) 161; or cost function: Hosseini and others in Pedrycz and Chen, *Deep Learning* (2020) 9.

185 Hosseini and others in Pedrycz and Chen, *Deep Learning* (2020) 9.

186 ibid 10.

187 Buckner, 'Deep learning' (2019) 14(10) Philosophy Compass p. 1, 2.

188 Dube, *An Intuitive Exploration of Artificial Intelligence: Theory and Applications of Deep Learning* (2021) Part II.

vised learning context, and affects interpretability, as the system's design is not made dependent on human understanding.[189]

3. Data and performance evaluation

DNNs are only one example of ML models that learn to achieve a distinctive kind of functioning through a relatively automated training process. At the same time, the way in which human engineers select and structure the use of data in ML development remains a crucial prerequisite for success and for an assessment of the algorithm's performance. Stepping back from our analysis of DNNs, we can frame this aspect more widely, as an issue for ML models in general.

A basic distinction to be made here is between offline and online ML algorithms. This marks the difference between those that do not draw on data from the application environment to improve performance and those that do. The values of offline models are determined by their interactions with the dataset used for training and are then locked in. After the ML system is understood to be sufficiently capable of performing the human defined task on the training data, the learning algorithm ceases to be applied so that the model ceases to update its learnable parameters with the presentation of new data. When ML models are online they will initially still be trained on a training dataset, but they will continue to learn from incoming information during application. This entails a greater independence and unpredictability of performance, limiting the value of *ex ante* assessments.

These features of online AI do not recommend themselves to the sensitive healthcare field and for the moment it appears that the ML applications that have received regulatory approval are offline.[190] For this reason, the following discussion will be couched in terms of offline ML models. The characteristics of online learning will be mentioned where relevant.

189 Burrell, 'How the Machine 'Thinks'' (2016) 3(1) Big Data & Society p. 1, 6-7.

190 Minssen and others, 'Regulatory Responses to Medical Machine Learning' (2020) 7(1) Journal of Law and the Biosciences p. 1, 5. At the same time the FDA has, for instance, already made an action plan for the regulation of online algorithms: U.S. Food & Drug Administration, 'Artificial Intelligence and Machine Learning in Software as a Medical Device' (2021) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> accessed 6.3.2022.

Even in the case of offline learning, the relative independence of the learning process, the nature and quality of data and the way in which it is used to train and validate the machine's functioning are central to the performance it can learn to achieve. Additionally, given the limited knowledge of ML models' decision-making criteria, as exemplified by DNNs, close attention to these data-related processes will also provide critical information on the AI's performance.[191] It is during training where it is easy to follow which inputs are matched correctly to their outputs. For example, whether a patient's condition is correctly identified from their electronic health record.

Developers will then be able to derive evaluative metrics, assessing how accurately the algorithm is performing its function across the entire training data set. For example, it may arrive at the correct diagnosis in X% of cases. This has practical implications. If an algorithm performs poorly in categorising outputs in training data sets it may be subject to 'underfitting'.[192] That is, the model may be too simple to capture complex relations in the data, leading to a low accuracy.[193] In consequence, non-learnable parameters (also known as hyperparameters)[194] may have to be tuned, or a different model may have to be selected to accomplish the relevant task. By contrast, if the algorithm has a high accuracy, one may say it performs well in the training environment, although it will still be expected to fall short in some cases.

This indicator is still inadequate for a realistic judgment of AI utility. To truly understand whether there is a basis for applying an AI in a practical setting one further needs to know something about its generalisability. Even if a model is assessed as sufficiently accurate in a training environment, i.e. it has not underfit, this may be the result of a phenomenon known as 'overfitting'. Here machine learning is so flexible that it simply memorises the quirks of the training data.[195] A good example from the medical field

191 Krishnan, 'Against Interpretability: A Critical Examination of the Interpretability Problem in Machine Learning' (2020) 33(3) Philosophy & Technology p. 487, 495-496.

192 Vieira, Pinaya and Mechelli in Mechelli and Vieira, *Machine Learning: Methods and Applications to Brain Disorders* (2019) 34-35.

193 Tayo, 'Simplicity vs Complexity in Machine Learning — Finding the Right Balance' (11.11.2019) <https://towardsdatascience.com/simplicity-vs-complexity-in-machine-learning-finding-the-right-balance-c9000d1726fb> accessed 6.3.2022.

194 Vieira, Pinaya and Mechelli in Mechelli and Vieira, *Machine Learning* (2019) 32.

195 Sejnowski, *The Deep Learning Revolution* (2018) 43.

is provided by Narla and others who designed an algorithm for diagnosing the malignancy of a skin lesion. They found that:

> the algorithm appeared more likely to interpret images with rulers as malignant. Why? In our dataset, images with rulers were more likely to be malignant; thus the algorithm inadvertently 'learned' that rulers are malignant[196]

This association between a ruler and malignancy is dependent on the precise nature of the training data. It incidentally holds for a particular set, but for obvious reasons it is unlikely to transfer to new examples. Thus, even if accuracy is high for one data set (the algorithm is not biased) it may fall dramatically when used for unseen examples (the algorithm has overfit). Given the complexity of DNNs they are particularly prone to exhibit this defect – even if various mitigation techniques are being developed.[197]

To have a basic indicator of generalisability, i.e. an absence of overfitting, AI development almost ubiquitously includes a testing stage. Here data that is not used during the training phase is presented to the trained algorithm in a separate testing phase and performance is assessed. If accuracy remains high, one can have some confidence that the model has not overfit during training and that, in consequence, it may be applicable to new data.

Going even further for medical AI, it is important to demonstrate that the ML can generalise to the specific clinical environment in which it is to be used. This is necessary for several reasons. Even if an algorithm is neither inappropriately underfit nor overfit, there is no guarantee that the data used for testing or training is representative of the groups on which it is to be deployed. A factor that must be assessed against the background of a well-documented propensity of ML, including medical ML, to reach decisions that disadvantage already vulnerable groups. This state of affairs can result from the selection of training and test data, from biased engineering of features and from a failure to consider the biases that may arise in human-machine interactions.[198]

196 Narla and others, 'Automated Classification of Skin Lesions: From Pixels to Practice' (2018) 138(10) The Journal of Investigative Dermatology p. 2108, 2108. See similarly: Afnan and others, 'Interpretable, Not Black-Box, Artificial Intelligence Should Be Used for Embryo Selection' [2021](4) Human Reproduction Open p. 1, 3.

197 Vieira and others in Mechelli and Vieira, *Machine Learning* (2019) 164.

198 Thomasian, Eickhoff and Adashi, 'Advancing Health Equity with Artificial Intelligence' (2021) 42(4) Journal of Public Health Policy p. 602. See also Alon-Barkat and Busuioc, 'Human-AI Interactions in Public Sector Decision-Making: 'Automation

Fundamentally one must remember that, unlike humans, AI are not able to 'basically recognize and adapt to changed situations or contexts'[199] and that performance may drop if there are relevant distinguishing characteristics between training or validation and deployment:

> Proper assessment of real-world clinical performance and generalisation requires appropriately designed external validation involving testing of an AI system using adequately sized datasets collected from institutions other than those that provided the data for model training. This will ensure that all relevant variations in patient demographics and disease states of target patients in real-world clinical settings are adequately represented in the system where it will be applied[200]

Training and validating ML performance on the right type of data is particularly important in healthcare where decisions have significant consequences and where there are numerous variables that can lead to different divisions between population groups. For instance, in a study investigating the IDx-DR device (which will be examined further below) it was found that its accuracy was robust for sex, race and ethnicity, but that there was a greater specificity for those aged over 65.[201]

Such variability will also often be co-determined by the highly variable clinical contexts. Price provides an excellent example of this when he distinguishes between AI application in high-resource settings (where they are often trained and tested) and low-resource settings (where they may be most needed and deployed):

> The most significant problem with applying algorithms developed in High-Resource Hospitals in lower-resource settings is that those algorithms are likely to make diagnoses and treatment recommendations that are systematically suboptimal in those lower-resource settings. These can arise in at least two different ways: differences in diagnoses and treatment recommendations based on systematically different patient

Bias' and 'Selective Adherence' to Algorithmic Advice' (2023) 33(1) Journal of Public Administration Research and Theory p. 153.

199 Strauß, 'Deep Automation Bias: How to Tackle a Wicked Problem of AI?' (2021) 5(2) Big Data and Cognitive Computing p. 1, 3.

200 Kelly and others, 'Key Challenges for Delivering Clinical Impact with Artificial Intelligence' (2019) 17(1) BMC Medicine p. 1, 4.

201 Abràmoff and others, 'Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices' (2018) 1 NPJ Digital Medicine p. 1.

> populations, and differences in recommended treatments based on treatment rankings whose order shifts with available medical resources[202]

Similarly Cohen and others note:

> a model to identify patients with sepsis that was derived from data at ten community hospitals may need to be changed for use in a tertiary care center that serves a large transplant population or in hospitals that do not have an ICU[203]

This illustrates the many nuances at play in the healthcare context that point towards the need for clarity about the type of data that is used in AI training and testing and about the nature of its development.

Ideally it demands some form of external validation, in the sense of validation in the application-environment, definable in terms of specific regions or even hospitals.[204] There is also a need for such external analyses to be continually updated to account for any relevant shifting in patient populations that may occur over time.[205]

More generally it is necessary to not only assess the technical features of the algorithm's functioning, but also its interaction with the relevant medical environment. Knowing about generalisability in the abstract may not translate to improved outcomes in practice if the AI is not used properly, if it is not integrated into the clinical workflow or its recommendations are not accepted by users.[206]

Indeed, if the AI is making a recommendation without having accounted for the wider operation of the healthcare environment and the causal factors at play there, then it may be outright dangerous. This is exemplified by an algorithm that indicated that patients with a history of asthma have a

---

202 Price II, 'Medical AI and Contextual Bias' (2019) 33(1) Harvard Journal of Law and Technology p. 65, 91.

203 Cohen and others, 'The Legal and Ethical Concerns That Arise From Using Complex Predictive Analytics in Health Care' (2014) 33(7) Health Affairs (Project Hope) p. 1139, 1143.

204 NHSX, 'NCCID case study: Setting standards for testing Artificial Intelligence' (21.2.2022) <https://www.nhsx.nhs.uk/ai-lab/explore-all-resources/develop-ai/nccid-case-study-setting-standards-for-testing-artificial-intelligence/> accessed 6.3.2022.

205 Kelly and others, 'Key Challenges for Delivering Clinical Impact with Artificial Intelligence' (2019) 17(1) BMC Medicine p. 1, 3.

206 Garg and others, 'Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review' (2005) 293(10) The Journal of the American Medical Association p. 1223, 1235-1236.

lower risk of dying from pneumonia.[207] This association only held because patients with such a history were prioritised for intensive treatment, which was the case in the training data. If the AI decision does not account for the specific proactive dimension, then it could disincentive intensive treatment, putting asthmatics at a greater risk of death. Ultimately, having some evidence from prospective, external validation of clinical ML models is a prerequisite for adequately gauging whether it alters clinical practice or improves clinical outcomes.[208]

It is therefore of some concern that a lack of evidence from such evaluations has been a noted defect in medical AI development.[209] Nor is such validation mandated to achieve market access. There is evidence that both the U.S. Food and Drug Administration (FDA) and Notified Bodies in the European Union (EU) have approved devices without such evidence.[210] Further it appears that some AI will, even if used in a clinical or hospital setting, not have passed through these regulatory procedures given that

207 Caruana and others, 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission' (Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney NSW Australia, 10.8.2015-13.8.2015).

208 Brajer and others, 'Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission' (2020) 3(2) JAMA Network Open 1-14, 2.

209 Kelly and others, 'Key Challenges for Delivering Clinical Impact with Artificial Intelligence' (2019) 17(1) BMC Medicine p. 1, 4; Topol, 'High-Performance Medicine' (2019) 25(1) Nature Medicine p. 44, 45; Freeman and others, 'Use of Artificial Intelligence for Image Analysis in Breast Cancer Screening Programmes: Systematic Review of Test Accuracy' (2021) 374 BMJ (Clinical Research Edition) 1-15. Four common defects of AI studies in one field of medicine have also been outlined as: (1) No generalisability assessment (2) unbalanced data (3) small sample size and (4) a limited reporting of performance metrics: Curchoe and others, 'Evaluating Predictive Models in Reproductive Medicine' (2020) 114(5) Fertility and Sterility p. 921, 923.

210 Angus, 'Randomized Clinical Trials of Artificial Intelligence' [2020](11) The Journal of the American Medical Association p. 1043: 'the US Food and Drug Administration recently approved AI-enabled decision support tools (also called software as medical devices or SaMDs) for diagnosis of diabetic retinopathy on digital fundoscopy and early warning of stroke on computed tomography scans. In neither instance was approval based on any RCT evidence that the information provided by the SaMD improved care'. The former (IDx-DR) device has also received a CE-Mark: IDx LLC, 'Fully Automated Diagnostic Device Receives CE Certification; IDx LLC Planning For Rollout Across Europe' (6.5.2013) <https://www.prnewswire.com/news-releases/fully-automated-diagnostic-device-receives-ce-certification-idx-llc-planning-for-rollout-across-europe-206263101.html> accessed 7.3.2022.

they may not be classified as medical devices.[211] One can also see that in the UK the design of guidance for the testing of medical AI for implementation is very much still at the proof of concept stage.[212] Currently there is a situation where medical AI implementation is progressing rapidly, with a risk that the outlined performance metrics are not collected or not made available.

Lastly, it has also been thrown into doubt in how far traditional clinical trials or indicators, as well as enforced centralised oversight, can be achieved for AI in healthcare.[213] For instance, area under the curve (AUC) measures are 'a core professional method of evaluating diagnostic tools', allowing professional radiologists to gauge whether any such tool improves accuracy.[214] While developers of diagnostic AI also regularly cite this aggregate value as a primary indicator of performance – including in regulatory applications and research publications – its utility for this purpose has proven limited.[215] It has been hypothesised that this is due to the implicit, subjective and variable nature of the professional know-how that ML models are seeking to capture and the contestable assumptions that must therefore be made during training and testing.[216]

---

211 E.g. Currently it appears that Watson for Oncology (a device utilising ML techniques) has been implemented in hospitals in the U.S. without being subject to regulatory oversight. See: Ross and Swetlitz, 'IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close' (5.9.2017) <https://www.statnews.com/2017/09/05/watson-ibm-cancer/> accessed 28.3.2023. For a more general overview of the U.S. situation see: Price II, Sachs and Eisenberg, 'New Innovation Models in Medical AI' (2022) 99(4) Washington University Law Review p. 1121, 1125-1126, 1150-1151; Price II, 'Distributed Governance of Medical AI' (2022) 25(1) SMU Science & Technology Law Review p. 3: highlighting both the limits of central regulatory involvement and the limits on effective central oversight in the case of involvement.

212 NHSX, 'NCCID case study: Setting standards for testing Artificial Intelligence' (21.2.2022) <https://www.nhsx.nhs.uk/ai-lab/explore-all-resources/develop-ai/nccid-case-study-setting-standards-for-testing-artificial-intelligence/> accessed 6.3.2022.

213 Price II, 'Artificial Intelligence in Health Care: Applications and Legal Implications.' (2017) 14(1) The SciTech Lawyer p. 10, 11 and see: Price II, 'Distributed Governance of Medical AI' (2022) 25(1) SMU Science & Technology Law Review p. 3.

214 Lebovitz, Levina and Lifshitz-Assaf, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly p. 1501, 1507.

215 ibid 1510.

216 ibid 1510: for instance it was found that 'only a narrow subset of the relevant diagnosis inputs was captured in the datasets underlying the ML model'.

4. Summary

On the basis of the described attributes of ML structure and development we should bear in mind the following aspects as we go forward. We are focussing on ML algorithms, which take multiple forms and can pursue different types of cognitive clinical problem-solving tasks. It was shown how the innate structure of one ML technique, a DNN, makes it difficult to comprehend the criteria by which it arrives at a decision. We also know that an ML model's performance is importantly shaped by the data on which it is trained as well as tested, giving rise to risks for certain groups and in certain implementation environments, and that a meaningful evaluation metric for this performance presupposes further validation procedures – especially in the protean medical context. That such an evaluation has taken place is not currently guaranteed by the regulatory environment and it is a shortcoming of many development procedures. As such, a degree of uncertainty around these metrics ought to be taken as a general, albeit not universal feature, of current medical AI.

With this description of AI one can now anticipate the sources of autonomy-related problems that will constitute the focus of this book's discussion. The first are the capabilities that were formerly the preserve of human professionals, but which are now, to some degree, exercised autonomously by the AI. Specifically, it must be elaborated how AI are solving the outlined cognitive problems of clinical practice and how they are interacting with human decision-makers. Secondly, one needs to have a deeper understanding of the ways in which ML may be understood to have or to lack the quality of interpretability in the medical context. This includes an elaboration of how existing AI are interpretable or not and the ways in which interpretability is or is not being enhanced by scientific or technological means.

The remainder of this chapter will focus on a description of these aspects. Section II. uses case studies to demonstrate the capabilities of medical AI and their relationship with established medical practice and human practitioners. Section III. elaborates on this relationship by exploring general limitations on human oversight over machine assistance in decision-making. This indicates why machine-generated problems persist in expert-mediated contexts. Section IV. details the state of the art on ML interpretability.

II. Capabilities of clinical AI: case studies

At the time of writing there are already dozens of medical ML applications that possess the features outlined above and which have reached the implementation stage in the United States and the United Kingdom.[217] This is well-documented in survey studies cataloguing the regulatory approvals that grant market access to AI/ML-based medical devices.[218] Such approved devices may not include all instances of mature AI, but it can be supposed that they capture a significant proportion.[219]

Indeed, the nature of many of these applications has been verified through searches of medical device databases and through a review of the public resources offered by developers. In this respect it is important to note that medical devices from the United States may be overrepresented in this book's analysis, owing to the easily accessible federal database and the detailed information that is offered on devices' nature, benefits, risks and intended types of use. The EU and UK systems do not yet offer a functional equivalent, although the European EUDAMED database, which is currently under development, may remedy some of these omissions. In any case, by focussing on devices with market access, as well as those whose practical feasibility is supported by robust external evidence,[220] we are adhering to the outlined definition of AI. It is their capabilities as useful problem-solving tools that drive their approval and implementation.

---

217 One must recall in this regard also the caution in Chapter 1, that this is not to be equated with reimbursement and widespread adoption in the healthcare system. There are additional hurdles to this, beyond the exhibition of the discussed capabilities.

218 Muehlematter, Daniore and Vokinger, 'Approval of Artificial Intelligence and Machine Learning-Based Medical Devices in the USA and Europe (2015–20): A Comparative Analysis' (2021) 3(3) The Lancet Digital Health p. 195; Benjamens, Dhunnoo and Meskó, 'The State of Artificial Intelligence-Based FDA-Approved Medical Devices and Algorithms: An Online Database' (2020) 3 NPJ Digital Medicine. It is to be noted that existing U.S. and EU approvals are considered for these purposes, with the latter still conferring UK market access until 30 June 2023: GOV.UK, 'Regulating medical devices in the UK: What you need to do to place a medical device on the Great Britain, Northern Ireland and European Union (EU) markets' (1.1.2022) <https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk> accessed 7.3.2022.

219 See for example the reference to Watson for Oncology *supra*.

220 This may take the form of evidence that the relevant ML is in being trialled for use or that it has demonstrated utility in clinical practice.

This section presents several case studies from this subset of devices. The purpose of these case studies is manifold. They illustrate concrete cognitive problems that ML is capable of solving. At the same time, they will also touch upon some of the problems associated with the opacity of these devices. At bottom however, they are intended to emphasise the non-uniform manner in which AI are interacting with human clinical expertise and are assuming new roles for machines in medical decision-making.[221] They offer a context-specific differentiation of the kinds of automation that are taking place and highlight that 'automation is not a straightforward perspective, but a choice'.[222] Understanding that choices are being made, and the kinds of trade-offs involved, provides concrete reference points for the subsequent legal analysis.

The following section is accordingly structured by reference to the different types of interaction that are envisaged between human and ML agents: (A.) the complementation of human expertise, the form of which essentially remains unchanged (B.) the partial replacement of human cognitive capabilities (C.) direct control over dimensions of clinical decision-making.

A. Devices complementing human expertise

There are medical ML devices that do not lessen the human cognitive capabilities that are brought to bear on a particular patient's treatment. Rather, they are intended to provide an additional resource that guides and supports such human decision-making.

A good example is provided by the AI-Pathway Companion Prostate Cancer that is on the EU market.[223] It is an AI that provides treatment advice. Specifically:

> Natural Language Processing is used to extract and compile data relevant to the decision-making process from the radiology, pathology,

221 See Strauß, 'Deep Automation Bias: How to Tackle a Wicked Problem of AI?' (2021) 5(2) Big Data and Cognitive Computing p. 1 for a general comment on the different types of automation underlying AI implementation.

222 Tsoukias in Papathanasiou, Zaraté and Freire de Sousa, *EURO Working Group on DSS: A Tour of the DSS Developments Over the Last 30 Years* (2021) 156.

223 Siemens Healthineers, 'AI-Pathway Companion Prostate Cancer from Siemens Healthineers approved for use in Europe as medical device' (3.3.2020) <https://www.siemens-healthineers.com/fr-be/press-room/press-releases/pr-aipathwaycomp-ce.html> accessed 7.3.2022.

> genetics, and lab results (…) Algorithms search through the prostate cancer guidelines for recommendations that suit the patient's individual disease status based on his or her current available data. (…) Based on this data, AI-Pathway Companion Prostate Cancer displays the patient's current clinical situation and offers guideline-based recommendations for further steps to provide treatment in accordance with the medical evidence[224]

A common problem is encountered here, in that it is not clarified what types of techniques are leveraged to achieve the relevant functioning.[225] However, the explicit role of natural language processing in extracting data and offering guideline-conform recommendations is almost certain to utilise ML and, more than likely, deep learning.[226] This helps to accomplish the demanding cognitive task of providing bespoke judgments in individual cases. The device offers therapeutic advice that is tailored to the specific patient, to the system within which they are being treated and to the medical evidence. Its capabilities are brought to bear to visualise the current situation and the possible treatment options. The device thereby frames the exercise of human expertise and seeks to optimise treatment decisions. There is no intention to replace elements of human decision-making; the AI is a sophisticated support tool that assists multidisciplinary teams of human experts.[227]

---

224 ibid.

225 Muehlematter and others note this shortcoming even for the more transparent database: Muehlematter, Daniore and Vokinger, 'Approval of Artificial Intelligence and Machine Learning-Based Medical Devices in the USA and Europe (2015–20)' (2021) 3(3) The Lancet Digital Health p. 195, 201.

226 'Various NLP architectures—including rule-based, machine learning-based, and hybrid models—have been developed and studied to enhance the accuracy of clinical concept extraction. With the emergence of deep learning models, research on clinical concept extraction has shifted from traditional machine learning models that rely heavily on semantic and lexical features manually crafted by domain experts to deep learning models that can automatically learn feature representations (eg, word embeddings) from large volumes of unlabeled clinical text': Yang and others, 'Clinical Concept Extraction Using Transformers' (2020) 27(12) Journal of the American Medical Informatics Association p. 1935, 1935-1936.

227 Siemens Healthineers, 'AI-Pathway Companion Prostate Cancer from Siemens Healthineers approved for use in Europe as medical device' (3.3.2020) <https://www.siemens-healthineers.com/fr-be/press-room/press-releases/pr-aipathwaycomp-ce.html> accessed 7.3.2022.

Another example is the Acumen Hypotension Prediction Index Software, which utilises machine learning to carry out prognoses. Specifically it indicates the 'patient's likelihood of future hypotensive events' during surgery – by offering a risk score for a specific range of time.[228] It is a CE-marked and FDA approved device.[229] In addition it is to be noted that the implementation of this device was supported by a randomised control trial that demonstrated an 'ability to influence physician actions and change proximate patient outcomes'.[230] As with the AI-Pathway Companion Prostate Cancer, this software does not seek to replace physician expertise. It serves to provide information to a clinician who is responsible for the patient's condition and 'no therapeutic decisions should be made based solely on the Hypotension Prediction Index (HPI) parameter'.[231] Nevertheless, one can see that this device is introducing a unique form of judgment and influencing physician behaviour. Rather than relying on the established, overt clinical signs of hypotension, which occur relatively late, the data-driven model extracts and analyses subtle information from arterial waveforms to arrive at a prediction score in a way that was previously impossible.[232] This machine-determined prediction score and the early action it allows are the primary benefits that are offered to clinical decision-makers.[233]

228 U.S. Food & Drug Administration, 'De Novo Classification Request for Acumen Hypotension Prediction Index Feature Software' (16.3.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN160044> accessed 7.3.2022; Angus, 'Randomized Clinical Trials of Artificial Intelligence' [2020](11) The Journal of the American Medical Association p. 1043.

229 Edwards Lifesciences, 'Edwards' Acumen Hypotension Prediction Index Launches In The U.S.' (18.3.2022) <https://www.edwards.com/ns20180319> accessed 7.3.2022; the latter's use of the de novo classification pathway is in itself something remarkable as it (1) illustrates the novelty of this type of device (2) imposes higher regulatory burdens than the more commonly used 501(k)procedure.

230 Angus, 'Randomized Clinical Trials of Artificial Intelligence' [2020](11) The Journal of the American Medical Association p. 1043.

231 U.S. Food & Drug Administration, 'De Novo Classification Request for Acumen Hypotension Prediction Index Feature Software' (16.3.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN160044> accessed 7.3.2022.

232 Hatib and others, 'Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis' (2018) 129(4) Anesthesiology p. 663, 664.

233 U.S. Food & Drug Administration, 'De Novo Classification Request for Acumen Hypotension Prediction Index Feature Software' (16.3.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN160044> accessed 7.3.2022.

This category represents the least independent mode of AI functioning, being closely aligned with established uses of computers as decision support systems. The AI essentially collects, curates and presents knowledge to a fully-qualified human specialist to instrumentalise as they see fit.[234] Nevertheless, there is empirical data that documents the, significant influence of automated tools in clinical decision making. For instance, non-AI evidence-retrieval tools have been shown to have the capacity to improve the accuracy of decisions, boost confidence in conclusions but also, negatively, introduce new kinds of errors.[235] As will be discussed in Section III. below, there is an increased potential for ML-based devices, purporting to contribute a degree of specialist judgment, to have such impacts.

B. Devices (partially) replacing pre-existing cognitive capabilities

The models in the previous section introduced new types of judgment into the medical decision-making process, but they left intact existing human cognitive resources which are to be exercised in a patient's treatment. In this section we will see that the use of ML devices may form part of a choice to diminish such resources, even if they do not eliminate them entirely. Two examples serve to make this point.

The first is the IDx-DR device, which has received FDA approval and the CE-mark.[236] This uses multiple algorithms to look for different types of lesions that indicate diabetic retinopathy in recorded images of patients' eyes.[237] This is a condition that can lead to blindness in people with

234 See Kazzazi, 'The Automation of Doctors and Machines: A Classification for AI in Medicine (ADAM framework)' (2021) 8(2) Future Healthcare Journal p. 257, 260.

235 Westbrook, Coiera and Gosling, 'Do Online Information Retrieval Systems Help Experienced Clinicians Answer Clinical Questions?' (2005) 12(3) Journal of the American Medical Informatics Association p. 315.

236 U.S. Food & Drug Administration, 'De Novo Classification Request for IDx-DR' (11.4.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?ID=DEN180001> accessed 7.3.2022; IDx LLC, 'Fully Automated Diagnostic Device Receives CE Certification; IDx LLC Planning For Rollout Across Europe' (6.5.2013) <https://www.prnewswire.com/news-releases/fully-automated-diagnostic-device-receives-ce-certification-idx-llc-planning-for-rollout-across-europe-206263101.html> accessed 7.3.2022.

237 Grzybowski and Brona, 'Analysis and Comparison of Two Artificial Intelligence Diabetic Retinopathy Screening Algorithms in a Pilot Study: IDx-DR and Retinalyze' (2021) 10(11) Journal of Clinical Medicine p. 1, 4-5.

diabetes if not identified and treated. Without the device human-based screening would have to be carried out. This requires specially trained graders or ophthalmologists.[238] With the device, an eye care institution can rely on non-physician operators to detect this condition, without a need for specialists to over-read the results.[239] On the basis of the device's output a decision can be made on whether to refer the patient for further examination. The AI is said to function autonomously in this respect, although it remains embedded in a clinical institution.[240] It is envisaged that this institution will provide relevant information to patients - for instance that 'IDx-DR does not treat retinopathy'[241] – and will provide for their wider eye-examination needs. Essentially IDx-DR therefore replaces the patient's access to one relatively narrow type of proficiency: that necessary for screening for diabetic retinopathy. The level of human expertise that is brought to bear on this decision is reduced, supplanted by the machine. The recourse to equivalent human capabilities may become *de facto* barred but the AI does not abrogate human involvement altogether. Again, it bears emphasising that this situation is different to the previous two examples where the patient's access to human resources in their care remained undiminished.

The second, comparable example is Mia (Mammography Intelligent Assessment). This is a deep learning software that, inter alia, has been trialled in several NHS Trusts for use as a second reader in mammograms, diagnosing breast cancer and determining whether women should be recalled for further examination.[242] The expertise that is being supplied here is that of a second human expert. One human professional and the AI will initially be involved. If the human and AI agree, then the relevant decision to recall or not to recall is made. If they disagree then a further human arbitrator

238 ibid 1.

239 'No need for specialist overread or telemedicine call backs': Digital Diagnostics, 'IDx-DR' <https://www.digitaldiagnostics.com/products/eye-disease/idx-dr/> accessed 7.3.2022.

240 'IDx-DR is for medical professionals who want to provide patients with rigorously validated and ethically designed diagnostic results at the point-of-care': ibid.

241 U.S. Food & Drug Administration, 'De Novo Classification Request for IDx-DR' (11.4.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN160044> accessed 7.3.2022.

242 United Lincolnshire Hospitals NHS Trust, 'ULHT trialling artificial intelligence software to support breast cancer screening' (16.8.2019) <https://www.ulh.nhs.uk/news/ulht-trialling-artificial-intelligence-software-to-support-breast-cancer-screening/> accessed 7.3.2022.

is engaged to make the final decision.[243] Evidently therefore the AI is altering the established pattern of human expert involvement. Whereas previously two human specialists would have been routinely involved in recall decisions, this is now reduced to one. But unlike with the IDx-DR example, the human knowledge brought to bear on every decision should remain – qualitatively if not quantitatively – at the same level: that of a human specialist.

Of course one could go further than these examples; the strongest sense in which AI are often portrayed as replacing the expertise of professionals is by granting the patient direct access to the device's abilities, without the involvement and/or oversight of other humans.

This appears to be occurring for certain very narrow diagnostic tasks, such as the detection of atrial fibrillation. For example, the FibriCheck device is 'indicated for self-testing by patients who have been diagnosed with, or are susceptible to developing, atrial fibrillation and who would like to monitor and record their heart rhythms on an intermittent basis'.[244] However, it is ambiguous quite how far these decisions are really removed from human mediation. It is noted on the FibriCheck website that 'state-of-the-art machine learning algorithms (…) automatically interpret these results and create an output towards the healthcare professional, who can visually confirm these findings'.[245] A degree of physician involvement is therefore very much envisaged.[246]

It is also to be emphasised that the capabilities of such devices appear to be at the lower end of the scale – fulfilling a highly targeted diagnostic task, which will usually have to be located in a broader care context. Topol makes

243 See Sharma and others, 'Large-Scale Evaluation of an AI System as an Independent Reader for Double Reading in Breast Cancer Screening' (2021) Pre-Print p. 1.

244 U.S. Food & Drug Administration, 'FibriCheck 510(k) Summary' (28.9.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K173872> accessed 7.3.2022.

245 Fibricheck, 'What is FibriCheck and how does it work?' <https://www.fibricheck.com/what-is-fibricheck-and-how-does-it-work/> accessed 7.3.2022.

246 This can also be found in comparable devices. For instance one study evaluating a similar device emphasised the importance of transmitting relevant information to the physician: 'The [Kardia Band] is the first smartwatch accessory cleared by the FDA and available to the general public without a prescription that claims to instantaneously detect AF and transmit this information to a patient's treating physician': Bumgarner and others, 'Smartwatch Algorithm for Automated Detection of Atrial Fibrillation' (2018) 71(21) Journal of the American College of Cardiology p. 2381.

this point by arguing that medical AI will not lead to a lack of clinician oversight 'across all conditions, across all time' and that forms of partial and conditional autonomy (where there is more general human oversight in the background) are likely the highest degrees of automation that ML will introduce into healthcare.[247] Overall human decision-makers are seen as 'critical checks, in their roles as decisional mediators'[248] and they are likely to appropriately fulfil this purpose in relation to targeted, verifiable tasks. Much more of this will be made in Section IV. below.

Consequently the focus of this book's analysis is not on fully automated clinical ML models. There may be important ways in which machines replace human cognitive capabilities, but human professional involvement is not altogether precluded. Again, it will be seen in Section III. how this constellation nevertheless enables AI to influence medical processes. Moreover, it will be seen in Part III. that this has significant implications for the types of legal analyses that are conducted.

C. Devices determining dimensions of clinical decision-making

Among devices that augment or partially replace human expertise a separate category is due to those that effectively determine a significant aspect of the clinical decision-making process, even in the case of physician involvement. Such a distinction must necessarily be a matter of degree, as the framing of various medical decisions, or the weakening of professional oversight, can prove influential in various ways (see Section III. below and Chapter 3). In this sense, this category of AI poses merely the most direct challenge to the notion of human mediation by altogether precluding the possibility for effective oversight of certain decisions.

Accipio Ix provides an illustration of a medical device making such interferences. It has received a CE-mark and FDA approval.[249] It utilises a form

247 Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (2019) 87. Kazzazi also notes the majority of current AI applications 'demonstrate a clear and narrow function that is both accurate and safe (…) but require human instrumentation in order to prove effective' Kazzazi, 'The Automation of Doctors and Machines' (2021) 8(2) Future Healthcare Journal p. 257, 260.

248 Alon-Barkat and Busuioc, 'Human-AI Interactions in Public Sector Decision-Making' (2023) 33(1) Journal of Public Administration Research and Theory p. 153, 156.

249 U.S. Food & Drug Administration, 'Accipiolx 510(k) Summary' (26.20.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182177>

of DNN known as a convolutional neural network to identify instances of acute intercranial haemorrhages in non-contrast computerised tomography (NCCT) head scans.[250] It thereby provides a diagnostic indication that is intended to assist, not replace, the assessment of a clinical professional. The professional can now draw on the device's assessment, but: '[i]ts results are not intended to be used on a stand-alone basis for clinical decision-making (…) or otherwise preclude clinical assessment of CT cases'.[251] To this extent, the device aims to complement human expertise, which remains fully applicable, much like the AI-Pathway Companion Prostate Cancer or the Acumen Hypotension Prediction Index Software.

What sets Accipio Ix apart is the fact that it triages the analysed NCCT scans for the clinician, who will often be engaged in the time-intensive task of analysing a large number of such scans. The machine is able to leverage its capabilities to read the scan almost instantaneously and this allows it to prioritise the review of patients without the possibility for human oversight. This benefits some and disadvantages others. Even if the human makes the ultimate decision on diagnosis and treatment, the machine has essentially made a triaging choice that represents a judgment on how to balance the needs of different patients.[252] Without effective human mediation this choice is susceptible to the shortcomings associated with AI decision-making. As Voter and others note in this context: 'a poorly performing [decision support system] can hinder a clinician by highlighting false-positive studies and promoting premature closure in falsely negative studies'.[253]

---

accessed 7.3.2022; MaxQ AI, Ltd, 'MaxQ-AI Receives CE Mark Approval for Accipio™Ix Intracranial Hemorrhage Artificial Intelligence Software Platform' (22.5.2018) <https://www.prnewswire.com/news-releases/maxq-ai-receives-ce-mark-approval-for-accipioix-intracranial-hemorrhage-artificial-intelligence-software-platform-300652488.html> accessed 7.3.2022.

250 MaxQ Artificial Intelligence, 'ACCIPIO®—Solution Architecture and Design: A White Paper' <https://www.maxq.ai/resources> accessed 7.3.2022.

251 U.S. Food & Drug Administration, 'Accipiolx 510(k) Summary' (26.20.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182177> accessed 7.3.2022.

252 One should note attempts to generalise these kinds of triage capabilities: For instance in the UK's NHS there has been a trial of an AI symptom checker seeking to '"triage out" avoidable attendances that present at emergency departments, to combat growing demand': Mahase, 'Birmingham Trust and Babylon Health Discuss Pre-A&E Triage App' (2019) 365(l2354) BMJ (Clinical Research Edition).

253 Voter and others, 'Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Intracranial Hemorrhage' (2021) 18(8) Journal of the American College of Radiology p. 1143, 1144.

III. Interpretability of AI

A task that is necessarily prior to understanding how decision-makers interact with these forms of ML functioning, is to determine what they will be in a position to know about them. As has been referred to, the technology underlying many cutting-edge medical AI is subject to a pervasive problem of 'opacity', sometimes referred to as the 'black box problem'.[254] Hereby human users of ML algorithms are said to systematically lack epistemic access to a relevant element;[255] an element that is often referred to as a specific kind of knowledge or understanding.[256] This section seeks to expand upon the nature of this element for a clinical specialist and a patient that are using AI. This involves stipulating what factors are (1) relevant and (2) systematically obscured by the outlined uses of ML.

The relevance of certain facts on AI functioning is often discussed in the literature under the head of interpretability, understandability, transparency or accountability.[257] The precise meaning of these concepts, representing opacity's counterpart, is hotly contested and it is left intentionally ambiguous here. The aim is to provide an overview of the kinds of factors that fall under this head, the reason for their being obscured and whether there are trends that are likely to resolve these issues as AI-deployment progresses in the near-future. As interpretability and opacity are relative concepts[258] it is worth noting again that the perspectives of the physician and patient are deemed central to this discussion and that the latter's understanding is often mediated through the former.

254 Zednik, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34(2) Philosophy & Technology p. 265, developing the concept of Humphreys, 'The Philosophical Novelty of Computer Simulation Methods' (2009) 169(3) Synthese p. 615, 618.

255 Zednik, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34(2) Philosophy & Technology p. 265, 268-269.

256 See e.g. Rudin and others, 'Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges' (2022) 16 Statistics Surveys p. 1, 2-11.

257 Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487, 488; Binns, 'Algorithmic Accountability and Public Reason' (2018) 31(4) Philosophy & Technology p. 543, 544.

258 Humphreys, 'The Philosophical Novelty of Computer Simulation Methods' (2009) 169(3) Synthese p. 615, 618.

A. Prevalent types of opacity

An objection that is often levelled at ML algorithms is that their 'inner workings' are not accessible to users.[259] One way in which this criticism has been framed is as an inability to examine the code of the model, 'the variables – learnable parameters and/or abstract representational structures – that mediate the transformation of inputs to outputs'[260] and how these behave or have behaved when interacting with data. The missing piece of information is that users are not aware of the internals of the black box model.[261] A possible source for this inability is technical; it stems from the characteristic complexity of ML models, paired with time- and capability-limitations on users.[262] An adequate inspection of the internal operations becomes unfeasible in practice.

More commonly, such a lack of knowledge of the makeup of an ML tool, will also stem from a non-technical source, from corporate secrecy.[263] This was seen to generate uncertainty even around the basic fact of whether devices employ ML techniques. While such secrecy is neither an inevitable fact of AI deployment nor exclusive to it,[264] the fostering of opacity as an 'intentional form of self-protection by corporations intent on maintaining their trade secrets and competitive advantage'[265] with respect to devices using AI is well-documented in medicine and beyond.[266] Specifically with regard to ML, the maintenance of corporate secrecy around model func-

259 Lakkaraju and Bastani, '"How Do I Fool You?": Manipulating User Trust via Misleading Black Box Explanations' (AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 07.02.2020-09.02.2020); Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487, 495.

260 Zednik, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34(2) Philosophy & Technology p. 265, 271.

261 Lakkaraju and Bastani, '"How Do I Fool You?": Manipulating User Trust via Misleading Black Box Explanations' (AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 07.02.2020-09.02.2020).

262 Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (2019) 51(5) ACM Computing Surveys p. 1, 6.

263 Burrell, 'How the Machine 'Thinks'' (2016) 3(1) Big Data & Society p. 1, 3-4.

264 Rudin cites the proprietary COMPAS risk prediction tool as an example of a non-ML model that would, but for corporate secrecy, be interpretable: Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206, 209.

265 Burrell, 'How the Machine 'Thinks'' (2016) 3(1) Big Data & Society p. 1, 3.

266 Rudin and Ustun, 'Optimized Scoring Systems: Toward Trust in Machine Learning for Healthcare and Criminal Justice' (2018) 48(5) Interfaces p. 449.

tioning can be seen as an intentional strategy that companies adopt to avoid responsibility for the quality of individual predictions, while at the same time being able to capitalise from the purported capabilities of providing accurate predictions at the granular level.[267]

Ultimately, whether it is caused by corporate secrecy or technical complexity, an inability of users to understand the form, content and functioning of the underlying computational mechanisms is one prevalent form of opacity in ML.

The second form of opacity concerns the way in which the functioning of ML algorithms is related to external factors and forms of justification. Hereby there are difficulties in understanding the 'environmental patterns and regularities' that are being tracked, which 'features of the environment' are represented by variables in the relevant model and how the generation of outputs *via* causal mechanisms in the model relates to justificatory reasons for it.[268] The technical features of common ML techniques, like DNNs, contribute substantially to a disjuncture between (1) a technical knowledge of the computations being performed by the model (in so far as the user or another actor is able to access these) and (2) a set of general criteria that can serve to indicate to (any) human what information is flowing through it and how and why it is being manipulated and classified in a relevant way.[269] It is neither possible to understand all of the model at once, nor to intuitively explain the role of specific elements.[270] In consequence, even when access to the internal workings of an algorithm is granted, obtaining knowledge about its 'reasoning process',[271] 'the underlying rationale of …

267 Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206, 210.

268 Zednik, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34(2) Philosophy & Technology p. 265, 279; Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487, 493-494.

269 Funer, 'The Deception of Certainty: How Non-Interpretable Machine Learning Outcomes Challenge the Epistemic Authority of Physicians' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 172.

270 Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487, 490.

271 Afnan and others, 'Interpretable, Not Black-Box, Artificial Intelligence Should Be Used for Embryo Selection' [2021](4) Human Reproduction Open p. 1, 2.

machine-learning components?'[272] and the link of this to some form of ground truth is not forthcoming in many ML devices.

For example, if an AI is being used to select embryos for in vitro fertilisation (IVF) treatment, then one may be left asking what categories have been learned and applied. One may wonder whether cleavage rate, symmetry, etc. are relevant and, if so, how are they combined with other factors.[273] It may not be deducible whether a particular aspect of the patient's personal life has already been taken into account,[274] or what benefits, risks and limitations can be associated with a utilisation of the device.[275] It may further not be possible to check for errors, including the identification of forms of reasoning that are obviously wrong, in real-time.[276]

Beyond this, there are two important relations between mathematical processes and general criteria, which are particularly relevant to the physician-patient interaction, but which are obscured by this disjuncture. On the one hand there is the 'difficult task of translating information about causal processes into considerations relevant to the justification of a categorization'.[277] For users of AI and for decision-subjects this makes it unclear what kind of justification-strategy is being pursued for the output, since there are many different kinds for any given decision.[278] As a result, a physician or patient will struggle to incorporate the recommendation into their wider decision-making processes.[279] It is hidden from examination whether the selection and implementation of a justification is of a kind that would be deemed adequate by the patient.

---

272 Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (2019) 51(5) ACM Computing Surveys p. 1, 2.

273 Afnan and others, 'Interpretable, Not Black-Box, Artificial Intelligence Should Be Used for Embryo Selection' [2021](4) Human Reproduction Open p. 1, 4.

274 Funer, 'The Deception of Certainty' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 175.

275 Afnan and others, 'Interpretable, Not Black-Box, Artificial Intelligence Should Be Used for Embryo Selection' [2021](4) Human Reproduction Open p. 1, 4.

276 Afnan and others, 'Interpretable, Not Black-Box, Artificial Intelligence Should Be Used for Embryo Selection' [2021](4) Human Reproduction Open p. 1, 3; Parikh and others, 'Why Interpretable Causal Inference is Important for High-Stakes Decision Making for Critically Ill Patients and How To Do It' (2022) Preprint.

277 Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487, 494.

278 Binns, 'Algorithmic Accountability and Public Reason' (2018) 31(4) Philosophy & Technology p. 543, 544.

279 Funer, 'The Deception of Certainty' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 173.

On the other hand, there are inevitable normative and epistemological assumptions that are embedded in the finished device and in the way that it has been developed and tested. Specifically in medicine, there will have to be a 'selection of parameters deemed relevant, the weighting of each possible treatment goal, the choice of means to achieve the selected goals with their respective consequences for the patient's life'.[280] In addition, there may be constraints embedded in an ML algorithm that represent an attempt to prevent algorithmic discrimination.[281] Whether intentional or inadvertent, these are choices with normative implications that are contestable[282] and they may be obscured by the second type of opacity, especially if they are made against a backdrop of corporate opacity.

The same may be said of epistemic assumptions that can be found in the code itself, as well as in the ways in which it is optimised and tested. For instance, an AI may be trained and assessed only with one kind of uncertainty in mind, ignoring another.[283] Or, ML results may be categorised into 'good' and 'poor' quality, giving the impression of high capabilities, when a distinction should in fact be made between results of a similar 'good' quality.[284]

In light of the omnipresence of such assumptions – and without knowing how the mathematical processes of the algorithm are incorporating them into their reasoning – the provision of external epistemic justifications for the AI's use by the developer gains added significance. External evaluation metrics – such as accuracy and reliability – will have to be drawn upon to serve as arguments for the deployment of clinical ML devices.[285] Users will have to assess the rigour and the underlying assumptions incorporated into

280 ibid 173.

281 Binns, 'Algorithmic Accountability and Public Reason' (2018) 31(4) Philosophy & Technology p. 543, 547.

282 One useful example of such choices in the domain of fairness can be found in: Grgic-Hlaca and others, 'Human Perceptions of Fairness in Algorithmic Decision Making' (WWW '18: Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23.04.2018-27.4.2018).

283 Bhatt and others, 'Explainable Machine Learning in Deployment' (Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27.01.2020-30.01.2020).

284 Afnan and others, 'Interpretable, Not Black-Box, Artificial Intelligence Should Be Used for Embryo Selection' [2021](4) Human Reproduction Open p. 1, 2.

285 Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487, 495-496.

the training data (so-called 'pre-model techniques'),[286] as well as those that are implicit in the performance metrics gathered afterwards.

A good example is provided by Lebovitz and others.[287] In their study, U.S. hospital managers were considering whether to adopt five ML models and they took into account, amongst other things, the expertise of those who created 'ground truth' labels for the supervised training of ML models, and the compatibility of such labels with professional standards and the professional 'know-how' of local experts.[288] They also performed their own local studies of the devices.[289]

For the adoption and clinical use of otherwise opaque ML devices, external evaluative metrics gain an added significance. Yet it is contested whether they can truly stand in for an analysis of the machine's reasoning process.[290] Corporate opacity continues to pose issues, even for this alternative rationale,[291] and given the problems highlighted in Section I.C.3. there are still pitfalls that remain common, perhaps intractable,[292] in their collection.[293]

286 Okay, Yildirim and Ozdemir, 'Interpretable Machine Learning: A Case Study of Healthcare' (2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 10.31.2021-11.2.2021).

287 Lebovitz, Levina and Lifshitz-Assaf, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly p. 1501.

288 ibid 1512-1517.

289 ibid 1510.

290 Funer, 'The Deception of Certainty' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 172-173.

291 Lebovitz and others highlight this when they document the frustration of managers who were not able to assess the source of the ground truth labels for one ML tool: Lebovitz, Levina and Lifshitz-Assaf, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly p. 1501, 1509.

292 Funer, 'The Deception of Certainty' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 176; Bjerring and Busch, 'Artificial Intelligence and Patient-Centered Decision-Making' (2021) 34(2) Philosophy & Technology p. 349, 368.

293 Curchoe and others, 'Evaluating Predictive Models in Reproductive Medicine' (2020) 114(5) Fertility and Sterility p. 921, 923.

B. Solutions to interpretability problems

Given that this is a forward-looking analysis, it is important to ask whether problems around interpretability will persist as AI-based devices are deployed more widely in medicine, or whether solutions can be anticipated. The focus in this regard is not so much on how the veil of corporate secrecy may be pierced, but on how the underlying technical difficulties may be alleviated. The former is dependent on contingent regulatory interventions that do not appear to be existent or forthcoming at this stage. Some extent of secrecy is therefore taken as a pervasive, albeit not inevitable, fact of clinical AI. By contrast, there is widespread agreement that, before regulatory interventions can be effective in dealing with the black box problem, there would have to be adjustments to the currently unavoidable difficulties associated with the underlying technology.[294] Two avenues can be pursued in this regard that are often distinguished by the labels explainability and interpretability.[295]

1. Explainability

Explainability is used where ML devices are deployed without alterations to their opaque functioning but with an additional model bootstrapped to them.[296] This latter model will be designed to be more understandable and to approximate the device's functioning, reconstructing an *explanation* for the way in which the model is working, without actually tracking the causal processes within it.[297] Such techniques may seek to indicate the logic behind a model's overall performance,[298] to measure one specific property,

294 Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (2019) 51(5) ACM Computing Surveys p. 1, 2; Rudin and Ustun, 'Optimized Scoring Systems: Toward Trust in Machine Learning for Healthcare and Criminal Justice' (2018) 48(5) Interfaces p. 449, 450.

295 Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206.

296 Laugel and others, 'The Dangers of Post-Hoc Interpretability: Unjustified Counterfactual Explanations' (Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 8.10.2019-8.16.2019) 2802.

297 Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (2019) 51(5) ACM Computing Surveys p. 1, 10-11.

298 ibid 11. This is sometimes referred to as global explainability.

such as 'sensitivity to attribute changes',[299] or to establish reasons for specific results.[300]

For instance, saliency maps are employed to provide grounds for the output of an image analysis.[301] Such maps are intuitive because they visualise the parts of an image that an ML device is concentrating on. In medicine they may be used to indicate whether the algorithm is picking up on an area or pathology that a human expert would deem relevant or irrelevant.[302]

Counterfactual reasoning techniques are another prominent type of approach that explains ML decisions. They aim to identify a factor in the input data that, if changed, would alter the prediction.[303] Such models are easily interpretable by humans as they 'provide a minimal amount of information capable of altering a decision, and they do not require the data subject to understand any of the internal logic of a model in order to make use of it'.[304]

2. Interpretability

Interpretability is used to refer to devices that are developed with the aim of being more transparent and understandable to humans. By contrast with explainable approaches, the model itself and the actual causal reason for a given output are made more accessible to users.[305]

In effect this means that different kinds of constraints are imposed on the device to align its computational functioning with human capacities

299 ibid 14.

300 ibid 13-14. This is often termed local explanation to contrast it with the aforementioned global type.

301 Saporta and others, 'Benchmarking Saliency Methods for Chest X-Ray Interpretation' [2022](4) Nature Machine Intelligence p. 867.

302 Saporta and others, 'Benchmarking Saliency Methods for Chest X-Ray Interpretation' [2022](4) Nature Machine Intelligence p. 867, 867-868.

303 Laugel and others, 'The Dangers of Post-Hoc Interpretability: Unjustified Counterfactual Explanations' (Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 8.10.2019-8.16.2019) 2802.

304 Wachter and others, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2017) 31(1) Harvard Journal of Law & Technology p. 841, 851.

305 Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206, 206.

for reasoning.[306] For instance, a constraint of sparsity (a low number of features)[307] on some forms of data may align the model with humans' limited capacity for handling 'at most 7 ± 2 cognitive entities at once'.[308]

Such constraints cannot, however, be determined in the abstract. Sparsity would make little sense in different contexts: the analysis of an image would not be easier to comprehend if it uses only a limited number of pixels.[309] Here a useful interpretability constraint would be for the model to pursue a process of case-based reasoning. This emulates a well-established human problem-solving technique: by leveraging known solutions to past situations, it solves a new problem.[310] For example, by highlighting what feature of an image has been extracted, what features in past images it compares this to and how it has been combined with other information to reach the new result, the computer would engage in a human-interpretable form of reasoning.[311] In a healthcare context a physician could then understand that a given output has been provided for a current patient because of the way(s) they compare to past patients.[312]

For the outlined DNN's more interpretable models may be designed by disentangling the flow of information through the neurons. Whereas it has been described above how neurons are typically associated with several concepts. so that it is difficult to determine how altering individual variables will impact overall model performance, there are attempts to associate neurons in a certain layer with individual concepts.[313] For instance, when classifying a bedroom, different neurons in one layer may be associated with '"lamp," "bed," "nightstand," "curtain."' and '[a]ll information about the concept up to that point in the network travels through that concept's

306 ibid 206.

307 Carvalho, Pereira and Cardoso, 'Machine Learning Interpretability: A Survey on Methods and Metrics' (2019) 8(8) Electronics p. 832.

308 Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206, 206.

309 Rudin and others, 'Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges' (2022) 16 Statistics Surveys p. 1, 5.

310 ibid 25.

311 ibid 26-28.

312 Parikh and others, 'Why Interpretable Causal Inference is Important for High-Stakes Decision Making for Critically Ill Patients and How To Do It' (2022) Preprint.

313 Rudin and others, 'Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges' (2022) 16 Statistics Surveys p. 1, 31-33.

corresponding neuron'.[314] By seeing what concepts are being identified, the flow of information becomes more transparent to human users.

3. Evaluation

Both types of approaches to solving the opacity problem of AI have benefits and disadvantages. Yet the reason why this section refers to the interpretability of AI, rather than their explainability, is that only the former offers satisfactory solutions in the sensitive healthcare context. While the explainable AI movement purports to offer a solution, and thereby to increase public trust, protect individual rights and mediate the risks posed by black box tools,[315] it has the potential to worsen many of the issues associated with them.

The most glaring issue has been pointed out by Rudin: by definition the models approximating to the black box are only that, approximations, and they may generate 'explanations that are not faithful to what the original model computes'.[316] In medicine there is a resulting risk that for a percentage of cases, or a specific subset of them, the provided reasons do not match with the predictions that the medical device is making. This is evidenced by the experiments of Saporta and others, where the utilised heatmaps were more likely to be erroneous for certain kinds of pathologies.[317] This would limit trust in the explanation,[318] re-invigorating or compounding doubts about the original ML device.

Moreover, even where explanations do match predictions, the two may be appealing to 'completely different features'.[319] Laugel and others frame this problem in the context of the aforementioned counterfactual explana-

314 ibid 28.

315 Such arguments are deployed in Wachter and others, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2017) 31(1) Harvard Journal of Law & Technology p. 841.

316 Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206, 207.

317 Saporta and others, 'Benchmarking Saliency Methods for Chest X-Ray Interpretation' [2022](4) Nature Machine Intelligence p. 867, 873-874.

318 Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206, 207.

319 ibid 207.

tions. They point out that explanations may be generated by unjustified counterfactuals which are disconnected from the ground truth.[320] Similarly, Lakkaraju and Bastani have highlighted how explainable models may offer predictions that are highly faithful to the original black box models and yet hide the fact that they rely on morally problematic factors (such as race and gender) by offering alternative reasons for the decisions.[321]

Lastly, where predictions and common features are utilised by the two models, the explanations may not provide enough details to make sense of what the black box is doing with them.[322] A heatmap indicating that a black box is concentrating on parts of the image does not indicate how that image is used.[323]

Explanations will be liable to mislead users in some cases, either regarding the predictions *per se* or the factors used in generating them, and the knowledge they provide about the operation of the original model is inherently limited. Rather than opening the black box, they merely deliver 'summary statistics' and 'trends in how predictions are related to the features'.[324] For these reasons this work takes the stance that explainable AI techniques do not offer a satisfactory solution to the types of opacity outlined above.

Things are different with interpretable AI techniques. By definition they provide insights into what the relevant device is actually doing. Therefore, they do provide a solution to the opacity problem, although, even here it remains contested how the techniques used for these purposes relate to an epistemological concept of interpretability.[325] Moreover, there is no universal or perfect technique for interpretability. Within a given context there are outstanding questions, such as: how should one select between

320 Laugel and others, 'The Dangers of Post-Hoc Interpretability: Unjustified Counterfactual Explanations' (Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 8.10.2019-8.16.2019) 2.

321 Lakkaraju and Bastani, '"How Do I Fool You?": Manipulating User Trust via Misleading Black Box Explanations' (AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 07.02.2020-09.02.2020).

322 Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence p. 206, 208.

323 ibid 208.

324 ibid 208.

325 Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487, 492-493.

them in a given context? When is one aiming for sparsity, monotonicity, decomposability or other factors?[326]

What's more, there are considerable challenges in the adequate technological realisation of interpretable models. Even their proponents acknowledge that it is often easier to construct accurate black box models, than to develop comparable interpretable ones.[327] This means that, although perhaps not inevitable, there are still technological limits on the knowledge that can be generated about high-performing ML models. For instance, the aforementioned association of neurons in DNNs with individual concepts is not comprehensive: residual neurons are needed within the relevant layer to deal with uncategorised information, insights into other layers will be limited and opacity around the way in which the concepts are combined persists.[328] More generally, there are serious challenges to the satisfaction of user-defined interpretability constraints.[329]

Perhaps most relevant for the present work is the fact that such interpretable techniques are still not required, or sufficiently ubiquitous, in medical devices.[330] Regulators should bear in mind that the right pressure on developers may lead to accurate, interpretable solutions. However, for the foreseeable future, users of many ML devices will be dealing with technologies that, although desirable on other grounds, and although purportedly offering explanations, remain opaque in the outlined manner.

IV. Human-AI collaboration in the healthcare environment

Equipped with an understanding of ML technology, having seen the capabilities ML devices possess, and also anticipating the difficulties surrounding AI interpretability, we are now in a position to describe how humans and

326 Rudin and Ustun, 'Optimized Scoring Systems: Toward Trust in Machine Learning for Healthcare and Criminal Justice' (2018) 48(5) Interfaces p. 449, 450. These authors ask: 'What are the desired characteristics of an interpretable model, if one exists?'.

327 ibid 450.

328 Rudin and others, 'Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges' (2022) 16 Statistics Surveys p. 1, 33-35.

329 Rudin and Ustun, 'Optimized Scoring Systems: Toward Trust in Machine Learning for Healthcare and Criminal Justice' (2018) 48(5) Interfaces p. 449, 451.

330 ibid 449. The authors hypothesise that this is partly due to a misaligned incentive, whereby the black box nature of models shields developers from accountability.

AI are likely to work together when the technology is implemented in the healthcare environment.

The point of departure for this section, and the assumption underlying much of the previous analysis, is the need for AI to collaborate with human experts. Although ML techniques are demonstrating that they can execute demanding cognitive tasks at scale, quickly and effectively, they lack the requisite contextual understanding and empathy with the individual circumstances of the patient that are so central to medical decision making.[331] Moreover, as will be discussed in the next chapter, there is an important difference between the trust that is placed in a human professional and the reliance placed upon a technical tool. Amongst other things, patients can and do assume that a physician is representing and acting in their interests. No such assurances can be made regarding AI.

For these reasons it is said that human actors and ML devices have different strengths that may complement each other to achieve optimal outcomes.[332] Three features that frame this collaborative effort are picked out here: the nature of the choices made by healthcare actors when they use AI, the knowledge that such actors must have to sensibly utilise ML technology and the potential for the device to influence the human element of the interaction.

A. Choices in the use of ML devices

Clinicians and hospitals may choose whether to use AI and which AI to use for a given kind of task,[333] although both choices will be influenced by the institutional setting, such as a system's reimbursement framework. Once this choice has been made, a further contextual decision will often be necessary in relation to a particular patient. Namely, whether to use

331 Holley and Becker, *AI-First Healthcare: AI Applications in the Business and Clinical Management of Health* (2021), 58-61.

332 The idea between man-computer symbiosis is well-established and its implications are now being worked out in the healthcare sphere: see ibid 49-71.

333 Providing case studies from the U.S. context see: Lebovitz, Levina and Lifshitz-Assaf, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly p. 1501. It is also the assumption running through: Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

the AI in that specific clinical encounter and how to use it: for what purposes, in preference over which other techniques, in combination with what other sources of information and with what degree of confidence its results should be accepted and acted upon.[334]

From the perspective of professional-AI cooperation it is particular this latter dimension that is significant. Where an AI complements human expertise there must still be a contextual decision as to how to treat the AI's results. For example, even where an AI does not necessarily offer an accurate diagnosis for a given decision, it may provide an efficient prompt for the kind of analysis that is likely to lead to a better diagnosis (if properly integrated into the clinical workflow and considered as one piece of evidence alongside others).[335]

Things may be slightly different where devices partially replace cognitive expertise or determine dimensions of the clinical decision-making process. The very choice to rely on the AI for a task – e.g. as a tool for triaging or as a second reader in radiology – may constitute a standardised solution that is applied without individual consideration of patients. Nonetheless, there will still be an operational decision to introduce the patient to this system and a localised assessment of how to act after the patient has been processed by the AI device. For instance, if an emergency department triages chest X-rays as a matter of course and the patient's chest X-ray is classified as non-urgent, and the AI is known to be relatively reliable in this regard, then this may shape the assessment of the patient's condition by the doctor that makes the post-triage decisions.

The key point is that it is a pervasive choice whether and how to divide one's specialist labour with an AI tool. In the final instance this will involve a human decision that is granular, having to be specified for the circumstances of a particular patient, conducted against the backdrop of operational decisions that were made at a more abstract level.

334 'During clinical decision making, clinicians should determine appropriate confidence in AI-derived information and balance this with other sources of clinical information': ibid.

335 Lebovitz and others provide the example of Bone Age and Brain Tumor Segmentation tools, which were explored as a way of improving diagnostic processes by professionals and managers in spite of known flaws in their outputs: Lebovitz, Levina and Lifshitz-Assaf, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly p. 1501, 1512.

B. User knowledge of ML devices

In Section III.A. above it has been argued that, relative to the physician and patient, current ML models will be opaque or uninterpretable. Specifically in the sense that the reasoning process of the AI, the range of factors it considers and the goals that it pursues will be relatively inaccessible. By contrast, given the collaborative context assumed here and the role that medical professionals play in imparting information to patients, the focus of this section is the knowledge that a healthcare provider will *possess*, since it is a prerequisite for a realistic use of the technology.

There can be little doubt that some such knowledge is necessary to make the choices outlined in the previous section. A recent policy document published by the UK's NHS AI Lab and Health Education England highlights that: 'clinicians will need to understand when AI-derived information should and should not be relied upon, and how to modify their decision making process to accommodate and best utilise this information' and they will need to be in a position 'to confidently evaluate, adopt and use AI'.[336] In the U.S. context Lebovitz and others present a nuanced evaluation of the different kinds of information that the managers and professionals of a hospital had to possess (and partially generate) in order to use AI effectively in their practice.[337] Moreover, forceful cases are being made for the integration of AI literacy courses into the curriculum of medical students.[338] These assessments support the position advocated earlier: understanding AI performance is a complex task, requiring knowledge and skills that are not readily transferable from those exercised in the evaluation of existing technologies by healthcare providers.

To decide how much confidence to place in an ML device, a professional user must be assumed to have an awareness of the broad tasks that an AI is

336 Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

337 Lebovitz, Levina and Lifshitz-Assaf, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly p. 1501.

338 See: Wood, Ange and Miller, 'Are We Ready to Integrate Artificial Intelligence Literacy into Medical School Curriculum: Students and Faculty Survey' (2021) 8 Journal of Medical Education and Curricular Development 1-5; Ngo, Nguyen and van Sonnenberg, 'The Cases for and against Artificial Intelligence in the Medical School Curriculum' (2022) 4(5) Radiology: Artificial intelligence p. 1.

designed to accomplish and have access to some type of performance evaluation that measures its suitability for this task, as highlighted in Section I.C.3.

Precisely what further information is needed will depend on a range of contextual factors. A professional may be able to place partial reliance on research studies and official guidance as to the quality and applicability of the device to a given context.[339] As was seen however, AI may also capture aspects of medical know-how that are contested and subjective and they may be useful only for a relatively specific clinical environment. This places an additional burden on health professionals using the AI: they must be aware of these limitations and they ought to critically evaluate how they can employ these tools in the context of their own experience, the applicable professional standards and other clinical evidence.

Towards this end, a basic understanding of AI model design, data collection and processing, as well as validation will be essential. For example, understanding that a supervised learning model functions on the basis of human designed labels will direct attention towards the individuals who have labelled the data and the assumptions that they have made.[340] This is the foundation for a realistic assessment of the purposes, anticipated benefits and risks that are posed by the tool,[341] as well as allowing the doctor to relativize some of these influences through their own input.

Engaging with the limitations of performance evaluations should further point practitioners towards more general features of AI, such as their relative independence and the ambiguity of the goals that they pursue. The selection of 'good' embryos for implantation, or the ranking of 'best' treatment options requires a great deal of careful analysis. The nature of these labels and how they can be combined with human preconceptions must (almost inevitably) be considered in AI deployment.

339 'Several standards and tools have or are being developed for medical devices and clinical research to guide approaches to the evaluation of AI products, including the National Institute for Health and Care Excellence (NICE) evidence standards framework.': Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

340 Lebovitz, Levina and Lifshitz-Assaf, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly p. 1501, 1509-1510.

341 ibid 1515.

Moreover, the fact that a relatively unique engagement with external studies is required is itself an outgrowth of the AI's novelty and opacity. The physician must have an awareness of the direct inaccessibility of conceptual features – as well as the reasoning processes – that the machine has relied upon and they must see that AI use gives rise to pervasive forms of uncertainty. Once again this can be expected to frame a purposive collaboration with the tool and the assessment of its risks and benefits – requiring reference to independent human expertise (whether one's own or that of a colleague) or to objectively ascertainable clinical indicators.

C. ML influence

Leading on from the previous discussion, the professional ought to be aware of the capability of ML technology to shape the human element of the collaborative decision through cognitive biases. Such biases are a pervasive element in human action and it is well-documented that healthcare professionals are susceptible to heuristic replacements in response to automated decision aids.[342]

Several different kinds of biases have been deemed relevant to the deployment of medical AI,[343] but a consideration of automation bias (AB) serves to highlight the evidence that ML will shape the judgement of clinical professionals even under a collaborative framework.

AB can be defined as the tendency of humans to overly rely on automation.[344] It can lead to 'errors resulting from the use of automated cues as

342 In this sense it contrasts with other areas where AI may be deployed: Alon-Barkat and Busuioc, 'Human-AI Interactions in Public Sector Decision-Making' (2023) 33(1) Journal of Public Administration Research and Theory p. 153, 165.

343 A non-exhaustive list includes: automation bias, aversion bias, alert fatigue, confirmation bias and rejection bias: Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

344 Although the term is often used alongside related concepts like 'automation-induced complacency', it has emerged as a central focus for discussions of automation misuse in healthcare. See: Parasuraman and Manzey, 'Complacency and Bias in Human Use of Automation: An Attentional Integration' (2010) 52(3) Human Factors p. 381, 394-395; Goddard, Roudsari and Wyatt, 'Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators' (2012) 19(1) Journal of the American Medical Informatics Association p. 121; Lyell and Coiera, 'Automation Bias and

a heuristic replacement for vigilant information seeking and processing'.[345] Although the phenomenon remains underexplored,[346] Lyell and Coiera state the general problem in this way:

> When it performs well, automation can reduce errors and improve decision performance. It also, however, has the potential to introduce *new types of errors*. One particularly significant risk is that users may become overreliant on automation, especially when a [clinical decision support system] tool is less than perfectly accurate or reliable, leading to decision errors[347]

Taking such mental shortcuts is especially ill-suited to the task of arriving at accurate decisions with the novel and singular ML technology.[348] It was seen to be exceptionally difficult to calibrate one's confidence in the outputs of ML models, even in light of an in-depth critical analysis. How, then, would such calibration be achieved in an abbreviated form? Moreover, AI decision making introduces controversial value judgments and relies on an independent human element to provide context and to tailor a decision to the specific needs of a situation. A biased engagement with an ML device would not only lead to an inaccurate decision, but to one that is missing the broader human component.

Verification Complexity: A Systematic Review' (2017) 24(2) Journal of the American Medical Informatics Association p. 423.

345 See: Mosier and others, 'Automation Bias: Decision Making and Performance in High-Tech Cockpits' (1998) 8(1) The International Journal of Aviation Psychology p. 47; Goddard, Roudsari and Wyatt, 'Automation Bias' (2012) 19(1) Journal of the American Medical Informatics Association p. 121, 121; Parasuraman and Manzey, 'Complacency and Bias in Human Use of Automation' (2010) 52(3) Human Factors p. 381, 391: Lyell and Coiera, 'Automation Bias and Verification Complexity' (2017) 24(2) Journal of the American Medical Informatics Association p. 423, 423.

346 Goddard, Roudsari and Wyatt, 'Automation Bias' (2012) 19(1) Journal of the American Medical Informatics Association p. 121; Lyell and Coiera, 'Automation Bias and Verification Complexity' (2017) 24(2) Journal of the American Medical Informatics Association p. 423; Schemmer and others, 'On the Influence of Explainable AI on Automation Bias: Research in Progress' (19.4.2022) <https://arxiv.org/pdf/2204.08859> accessed 6.6.2022.

347 Lyell and Coiera, 'Automation Bias and Verification Complexity' (2017) 24(2) Journal of the American Medical Informatics Association p. 423, 423 (emphasis added).

348 It is notable how seriously this issue was considered in Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

The significance of the problem should therefore be clear. However, on the basis of existing studies it is difficult to assess its precise scope. On the one hand, errors associated with AB appear more prone to occur in the medical sphere. This has been associated with the complexity of decision-making and the large amounts information, ambiguity and detail that characterise it.[349] As a result of these circumstances, biases arise even in single-task environments (i.e. where the human user only performs one task concurrently).[350] For example, in the assisted detection of abnormalities in mammography scans it has been found that in the absence of a prompt, physicians were less likely to classify abnormal cases correctly.[351]

On the other hand, there are many case-specific factors that will co-determine the influence of automated tools, including: the intensity of the workload,[352] the skill of the user, and the difficulty of the cases.[353] The generally higher skills of medical professionals, and their ability to synthesise many different forms of knowledge,[354] may suggest a more limited role for AB.

More specifically for AI, the technology's nature may also be argued to increase the likelihood that there is an undue dependence on automation. One reason for this is the task complexity for which the devices are designed. A distinguishing feature of ML algorithms is the capacity to accomplish cognitive tasks that require sophisticated capabilities. It is likely therefore that they will function in environments that will induce AB, even if they focus on a single task – as many of the examples outlined in Section II. do.

The second, related, reason why AB is hypothesised to occur more frequently is that some uses of ML were seen to lower the levels of human

349 Lyell and Coiera, 'Automation Bias and Verification Complexity' (2017) 24(2) Journal of the American Medical Informatics Association p. 423, 429.

350 ibid 426.

351 Alberdi and others, 'Effects of Incorrect Computer-Aided Detection (Cad) Output on Human Decision-Making in Mammography' (2004) 11(8) Academic Radiology p. 909.

352 Goddard, Roudsari and Wyatt, 'Automation Bias' (2012) 19(1) Journal of the American Medical Informatics Association p. 121, 124-125.

353 Povyakalo and others, 'How to Discriminate Between Computer-Aided and Computer-Hindered Decisions: A Case Study in Mammography' (2013) 33(1) Medical Decision Making p. 98, 106.

354 Funer, 'The Deception of Certainty' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 169.

expertise brought to bear on a case. This has been found to increase the vulnerability to AB.[355]

Thirdly, a greater reliance is likely to be placed on ML systems. The very fact that humans may attribute more sophisticated capabilities to them than previous automatic systems, and that the performance of many will be advertised as having ultra-high accuracy and/or reliability, may be a cause for AB.[356] This is because it has been found that a greater accuracy may engender reliance or trust and trusting users are less likely to detect failures of machines.[357]

Explainable AI techniques may also feed into this misplaced reliance. It has been argued that 'the sole existence of an explanation could increase the reliance of the human on the AI which increases AB'.[358] Furthermore, it has been seen that explanations may be targeted toward inducing certain behaviours in human users, such as placing confidence in the AI, rather than truly offering an interpretation of considered factors and applicable forms of reasoning.

Similar issues will no doubt arise from the design of user interfaces for ML devices. Framing treatment options to indicate the levels of confidence that an AI has in different predictions, or offering an AI output as a default option, will aim to bring about human responses to the technology that are deemed desirable. Presenting such information, and the form of that presentation, will itself contribute to ML devices' potential for inducing cognitive biases and influencing judgments, going even beyond AB.

With the significance and probability of ML's potential to induce cognitive biases thus stated, it can be understood as an important piece of information about AI functioning that medical professionals must consider in reaching collaborative judgments.[359] A healthcare provider should be

355 Povyakalo and others, 'How to Discriminate Between Computer-Aided and Computer-Hindered Decisions' (2013) 33(1) Medical Decision Making p. 98, 106.

356 Sujan and others, 'Human Factors Challenges for the Safe Use of Artificial Intelligence in Patient Care' (2019) 26(1) BMJ Health & Care Informatics p. 1, 3.

357 Lyell and Coiera, 'Automation Bias and Verification Complexity' (2017) 24(2) Journal of the American Medical Informatics Association p. 423, 424.

358 Schemmer and others, 'On the Influence of Explainable AI on Automation Bias', (19.4.2022) <https://arxiv.org/pdf/2204.08859> accessed 6.6.2022.

359 'The propensity towards these biases may be affected by choices made about the point of integration of AI information into the decision making workflow, or the way such information is presented. Interviewees for this research highlighted that enabling clinicians to recognise their inherent biases, and understand how these affect their use of AI-derived information should be a key focus of related training and

aware of AI-induced biases and they must attempt to assess and mitigate their impact on decision making.[360]

V. Conclusion

The aim of this chapter was to provide the empirical, technical characteristics that are necessary for an assessment of AI's autonomy-related challenges in the following work. Stated succinctly the dimensions that should be emphasised going forward for these purposes are: (1) the choices represented by the introduction of new human-machine interactions that provide alternatives to the *status quo*: human specialists, (2) the variable quality of these interactions, including different degrees of automation, (3) the deficiencies in the performance evaluation of clinical AI, especially where deployed for variable groups and/or environments, (4) a lack of understanding about AI functioning, which will remain in spite of emerging technological solutions, (5) the fact that AI use will involve choices for healthcare providers in individual clinical encounters and that these must normally be able to regulate their cooperative decision making with AI on the basis of their broader knowledge of the technology and how it functions, and (6) the ML devices' influence on this cooperative interaction.

education. Failure to do so may lead to unnecessary clinical risk or the diminished patient benefit from AI technologies in healthcare.': Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

360 How far this and other strategies can serve to combat AB is admittedly contested: Lyell and Coiera, 'Automation Bias and Verification Complexity' (2017) 24(2) Journal of the American Medical Informatics Association p. 423, 429-430.

Chapter 3: Bioethical autonomy and artificial intelligence

This chapter investigates how the application of artificial intelligence (AI) may undermine the adherence to the value of autonomy in medicine. By combining an understanding of the factual background with a bioethical analysis, the aim is to identify those challenges that represent comparable problems across the UK and the U.S. jurisdictions. As outlined in Chapter 1, for the purposes of the legal comparison underlying this work, autonomy challenges stemming from AI constitute the *tertium comparationis*: a comparable pre-legal problem between legal systems to which each may offer a different response.

An appropriate definition of autonomy is central to the establishment of the comparison. If the definition is drawn too narrowly, then there is a risk that diverging conceptions of that value, which can be found reflected in the law, could be excluded *ex ante*. If the definition is not detailed enough however, then one risks missing or underconceptualising the core challenges posed by medical machine learning (ML).

The way to strike this balance is to recognise an important facet of the common law's private law reasoning that is the subject of both of our jurisdictional analyses: within limitations this reasoning is policy-based and aspirational.[361] Although national specificities must be respected, there is often an (implicit) assumption that judges are involved in a collective process of reasoning in line with recognised principles.[362] The nature of these principles will be unspecified to a certain degree and contested – allowing for comparable argumentation regarding their nature.[363] Much more of this will be made in Part II.

For present purposes this means that a conception of autonomy will be developed that is initially external to the law, lends itself to an identification of AI problems and yet can be inserted into its argumentative structure. Section I. develops this understanding of autonomy, it's broad nature. Section II. draws on elements of this theory to frame the autonomy challenges posed by AI as going towards: certain key beliefs about the goals being

361 Robertson in Robertson and Tang, *The Goals of Private Law* (2009).
362 Duxbury, 'The Law of the Land' (2015) 78(1) The Modern Law Review p. 26, 47-48.
363 Sunstein, *Legal Reasoning and Political Conflict* (1998) 35-61.

pursued in the patient's care, the changing expertise of human users, the risks related to AI and the technology's potential for manipulation.

I. The procedural conception of autonomy

It is a trite observation that autonomy derives from the terms of self (*autos*) and law (*nomos*). It is also widely recognised that what this entails is controversial. Broad distinctions have been drawn between variously defined theories of autonomy, such as 'libertarian', 'liberal' and 'communitarian'.[364]

In the practice-oriented field of bioethics there have been attempts to reconcile these different approaches to arrive at a workable middle path. Most notably this is purportedly done by the widely operationalised theory of principlism developed by Beauchamp and Childress.[365] These authors outline a principle according to which autonomy is 'self-rule that is free from both controlling interference by others and from limitations, such as inadequate understanding, that prevent meaningful choice'.[366] The principle is thereby elaborated at a level of abstraction that allows diverging specifications of a core normative content and it is often described as a middle-level guide to action that can yield more specific rules in specific scenarios.[367]

While this theory therefore points towards two potential sources of autonomy violations, control of the patient and their limited understanding, it does not offer very much by way of specification. One is not provided with the kind of general grounds that determine which controlling interferences are permissible and which are not, or which deficiencies of understanding render it inadequate.[368] Even if one can agree that violations are related to interferences and limitations this does not take one that much further, since it is generally recognised that humans are always

364 Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009) 11-17.

365 Beauchamp and Childress, *Principles of Biomedical Ethics* (Fifth Edition 2001).

366 ibid 58-59.

367 Wolf, 'Shifting Paradigms in Bioethics and Health Law: The Rise of a New Pragmatism' (1994) 20(4) American Journal of Law & Medicine p. 395, 400; that Principlism's distinction between rules and principles does not follow a Dworkinian model, but is merely a matter of abstractness has also been noted by Paulo, *The Confluence of Philosophy and Law in Applied Ethics* (2016) 119.

368 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 11.

subjected to some forms of entirely legitimate limitation and interference.[369] This account does not enable us to deal with the central issue at hand: to identify *which* AI-influences or AI-related limitations are challenging autonomy and *why*.

To offer general grounds that help with this identification, this chapter will utilise a procedural and, specifically, rationalist account of bioethical autonomy that has been developed by Johnathan Pugh.[370] Hereby the autonomy of an individual depends on the way in which they came to make their decision.[371] It presupposes that beliefs and actions flow from the values that individuals hold *and* from a certain responsiveness to reasons.[372]

The strength of this account stems especially from the fruitful way in which it allows one to conceptualise the challenges raised by AI in medicine and from the fact that it can be integrated into the legal doctrines of the selected jurisdictions, as described in Part II. Below it will be seen that the use of AI in medicine concerns the insertion of a device into the deliberative process, changing the way in which the doctor and ultimately the patient, arrive at their decision. What is problematic about such an insertion, is the sense that it upsets the deliberative-facilitative process of doctor-patient decision making and that it undermines the ability to act on the basis of reasons.

Pugh begins his work with the declared aim of elucidating 'the nature and forms of influence that can subvert autonomy'.[373] By distinguishing

369 This is recognised by relational accounts of autonomy, which take the social-embeddedness of individuals and their autonomy as a central premise: Stoljar, 'Informed Consent and Relational Conceptions of Autonomy' (2011) 36(4) The Journal of Medicine and Philosophy p. 375, 376.

370 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020). Pugh is building on an existing strand of thought that has developed in bioethics, see: Savulescu in Rhodes, Francis and Silvers, *The Blackwell Guide to Medical Ethics* (2008); Ploug and Holm, 'Doctors, Patients, and Nudging in the Clinical Context--Four Views on Nudging and Informed Consent' (2015) 15(10) The American Journal of Bioethics p. 28.

371 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 5. Fischer and Ravizza develop a similar approach for moral responsibility and highlight that their account focuses on 'the characteristics of the *actual sequence that leads to the action*': Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (2000) 37.

372 Again there are established theories that call for a similar responsiveness outside of the bioethical context: Wolf, *Freedom Within Reason* (1993); Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (2000).

373 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 3.

between a decisional dimension of autonomy and a practical dimension of autonomy, he is able to provide interlocking, general reasons for distinguishing interferences that are legitimate from those that constitute autonomy violations.

A. Decisional autonomy

Decisional autonomy relates to the ability to make one's own decisions. It incorporates both a cognitive and a reflective element. Under the cognitive element the agent ought to follow the norms of theoretical rationality, which means arriving at beliefs in ways that are based on appropriate evidence, inductive reasoning, are not inconsistent, etc.[374] These beliefs must then also be placed in the broader context of the agent's other convictions 'about both descriptive and evaluative features of the world'.[375]

The reflective element of decisional autonomy responds to the intuition that autonomy requires one's reasons for action to be one's own.[376] That is, regardless of what is generally accepted or treasured, an individual can shape their own unique system of beliefs and desires. Determining exactly how to identify the relevant, internalised states is not straightforward. We each have many beliefs and desires that occur to us without being endorsed, we may consider them fleetingly or non-seriously. These may have some importance too, but they are not taken to define who we are or what we want. To make decisions truly one's own, there is a philosophical tradition that holds it necessary for an agent to form evaluative judgments about what they have reasons to do.[377] To act autonomously, individuals must

374 Pugh draws on well-established conceptions of practical reasoning. Similar accounts can be found in: Baron, *Rationality and Intelligence* (2005) 90; Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (2000) 71. The latter maintain that individuals must be receptive to what reasons there are, which involves an understandable pattern of (actual and hypothetical) reasons-receptivity.

375 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 38.

376 Hyun, 'Authentic Values and Individual Autonomy' (2001) 35(2) The Journal of Value Inquiry p. 195, 196. See also: Christman, 'Autonomy and Personal History' (1991) 21(1) Canadian Journal of Philosophy p. 1. A version of this intuition is also expressed by: Frankfurt, 'Freedom of the Will and the Concept of a Person' (1971) 68(1) The Journal of Philosophy p. 5; Dworkin, *The Theory and Practice of Autonomy* (2012).

377 Christman, 'Autonomy and Personal History' (1991) 21(1) Canadian Journal of Philosophy p. 1, 4-6.

reflect upon their 'desires and beliefs, forming attitudes towards them'.[378] This evaluative *reflection* should occur at some point, although it need not be at the point of action[379] and although it may be unconscious.[380]

In essence this means that self-governance requires a hierarchical structuring of beliefs and desires, with some being higher, in the sense that they have lower beliefs or desires as their object.[381] It also means that there is an internal condition of evaluation. Individuals exercise a degree of control in making beliefs and desires their own: they apply personal standards in deciding whether these are true and whether they are good to act upon.[382] Given his rationalist emphasis, Pugh holds that evaluative attitudes, whether about beliefs or about one's understanding of the good, must be theoretically rational and not just aimed at one's own subjective understanding of the truth or the good.[383]

These more targeted beliefs and desires are respectively termed acceptances and preferences.[384] Pugh follows Ekstrom in going further and distinguishing a subset of these artefacts that constitute one's self. Ekstrom develops a coherentist model whereby one authorises certain preferences and acceptances 'when they cohere with one's other preferences and acceptances', they 'hold together firmly, displaying consistency and mutual support'.[385] There must be some judgement as to the value of the belief and action in the context of one's wider character system in order for them to relate to the agent's true self.[386]

---

378 Ekstrom, 'A Coherence Theory of Autonomy' (1993) 53(3) Philosophy and Phenomenological Research p. 599, 599.

379 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 49, citing Savulescu, 'Rational Desires and the Limitation of Life-Sustaining Treatment' (1994) 8(3) Bioethics p. 191, 199-200.

380 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 49; Ekstrom, 'A Coherence Theory of Autonomy' (1993) 53(3) Philosophy and Phenomenological Research p. 599, 603.

381 This is inspired by Frankfurt's seminal theory: Frankfurt, 'Freedom of the Will and the Concept of a Person' (1971) 68(1) The Journal of Philosophy p. 5.

382 Ekstrom, 'A Coherence Theory of Autonomy' (1993) 53(3) Philosophy and Phenomenological Research p. 599, 606-607.

383 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 51-52.

384 Ekstrom, 'A Coherence Theory of Autonomy' (1993) 53(3) Philosophy and Phenomenological Research p. 599, 600.

385 ibid 608.

386 ibid 610-612. Pugh elaborates on the details of this and amends some part of the theory: Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 50-54. His amendment ensures that autonomous action is possible in light of conflicting,

Cohering acceptances and preferences can be identified by reference to their being: long-lasting, defensible (being well-supported with reasons and thus resilient to challenge) and comfortably owned by the individual (they are not conflicted in acting on them).[387] In this manner, one has a rational justification of a given element that is in line with one's wider character. The outcomes are authorised preferences or acceptances.[388] This will be used as a shorthand for desires and beliefs that reflect the agent's self, since they are an outcome of the reflective process.

In addition, Pugh argues that certain beliefs pertaining to a decision need not only be arrived at rationally and upon reflection, but must also be true.[389] These are decisionally necessary true beliefs. If one is mistaken about them, then one cannot arrive at an autonomous decision.[390] The key thought behind this approach is that the content of these beliefs is so central to the relevant choice – so significant for the connection of one's values to one's actions – that, without them, one does not really control one's decision at all.[391] Given this connection to action, the nature of these beliefs will be defined below, with a view to the practical dimension of autonomy.

Overall, these conditions may make it seem like autonomy is difficult to exercise for a normal patient and thus an interest that seldom requires protection. Beauchamp has claimed that 'the conditions of (…) reflective control are so demanding in this theory that either many human actors will be excluded as persons or their actions will be judged nonautonom-

---

difficult choices where one may act in a less than fully rational manner, but not irrationally. Sub-optimal choices can still cohere with one's character sufficiently.

387 Ekstrom, 'A Coherence Theory of Autonomy' (1993) 53(3) Philosophy and Phenomenological Research p. 599, 608-609.

388 'One's preferences, I suggest, are authorised-or sanctioned as one's own-when they cohere with one's other preferences and acceptances': ibid 608.

389 Pugh makes clear that he draws on a long-established intellectual tradition by relating this to the 'Aristotelian claim that actions performed from reasons of ignorance are non-voluntary': Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 131.

390 ibid 35.

391 This is encapsulated in Wolf's claim that: 'an agent cannot have the kind of freedom and control necessary for responsibility unless, when making choices about values and actions, she can understand the significant features of her situation and of the alternatives among which her choice is to be made. That is, an agent cannot be free and responsible unless she can sufficiently see and appreciate the world for what it is': Wolf, *Freedom Within Reason* (1993) 117.

ous'.[392] This may be termed an anti-paternalist concern and, although it is impossible to mount a full defence of Pugh's theory here, it is important to counter this argument because, as will be seen in subsequent chapters, this is a kind of objection that has been raised within the law itself and within academic analyses of the law.[393]

As Beauchamp rightly points out, the outlined theory requires individuals to be able to engage in certain processes of rational evaluation. It would be misguided to deny that such a bar is set. Yet its demandingness is arguably grossly overstated. In particular, the requisite capabilities are not of a kind that is unachievable for those that we would want to describe as autonomous agents.[394] Nor is particular emphasis placed on the intellectual calibre of these agents – the requisite processes may even be engaged entirely unconsciously.[395] If one takes seriously the criterion that autonomy requires self-government, as per the definition with which we started this chapter, then setting conditions that enable this control will be unavoidable. Pugh's theory seems well-placed to strike the balance.

B. Practical autonomy

Practical autonomy is used to refer to conceptions of positive and negative freedom that constitute distinct prerequisites for effective action.[396] Freedom in this sense is used to refer to some restraint – this can be positive or negative – whereas autonomy is broader and encapsulates the above, deliberative elements.[397] Negative freedom requires that there be no debilitating factor or force preventing an agent from achieving an end that they are motivated to achieve and positive freedom holds that no factor enabling that end is absent (this is taken to refer to capacities, but also to the possession of information, understanding etc.)[398]

392 Beauchamp, 'The Failure of Theories of Personhood' (1999) 9(4) Kennedy Institute of Ethics journal p. 309, 313.
393 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 199.
394 ibid 201.
395 ibid 200-201.
396 Pugh is drawing on an established tradition of discussions of freedom/liberty: Berlin and Harris, *Liberty* (Second Edition 2017); MacCallum, 'Negative and Positive Freedom' (1967) 76(3) The Philosophical Review p. 312.
397 Cf. Christman, 'Autonomy and Personal History' (1991) 21(1) Canadian Journal of Philosophy p. 1, 2-4.
398 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 123-125.

Negative freedom appears relatively self-explanatory and need not be specified further for our purposes. Positive freedom, by contrast, ought to be considered because it points one towards a broader dimension of autonomy and lies beneath the problems to be elaborated in the clinical sphere. Namely, respecting another's autonomy sometimes entails assisting another, enabling them to achieve their ends. This may involve education, dialogue and decision aids, as well as an expansion of options and/or information about those options.[399]

One may also note the interconnectedness with theoretical rationality here; adhering to or enhancing this dimension also makes the realisation of practical autonomy more likely, being conducive of acting in ways that are likely to further the pursuit of one's ends, even if it does not guarantee success.[400] Significantly, one can see that the facilitation of understanding depends on putting the patient in a position where they are able to grasp the relation of inputs to the realisation of their own values.[401]

Furthermore, returning to the notion of decisionally necessary true beliefs, positive freedom requires that, for certain discrete categories of beliefs, the agent's understanding must map onto the true nature of the world.[402] In Parfit's terminology, utilised by Pugh, certain apparent reasons must be real reasons.[403] These categories of reasons are vital for the realisation of one's ends, so that they must not only be arrived at rationally, but they must actually be true.

To flesh out what these categories are Pugh draws on a modal test. Hereby he asks whether the agent could hold a relevant false belief and still achieve their desire's objective in a relevantly similar situation (a 'nearby possible world').[404] If they can, then the belief is not decisionally necessary, if they cannot, then it is decisionally necessary. An example that is given is 'the fact that an intervention will be painful or invasive'.[405] It is implied that without being aware that their treatment involves such factors the patient

399 ibid 144-145.
400 ibid 22.
401 ibid 134-135, 157.
402 For a similar, but by no means identical account (given that Pugh merely refers to true belief rather than knowledge) see: Mueller, 'The Knowledge Norm of Apt Practical Reasoning' (2021) 199(1-2) Synthese p. 5395. This provides an overview on the debate surrounding a knowledge criterion.
403 Parfit, *On What Matters: Volume One* (2011) 111; Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 25-26.
404 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 128-136.
405 ibid 173.

will not be able to achieve their goals in their care. Anticipating a discussion that will be elaborated upon in Section II., one could also cite the example of information that concerns the fundamental purpose of a procedure, such as the conditions that are to be ascertained *via* a genetic test. Without such information the patient cannot align the goals that are pursued by the procedure with their own desires and beliefs.

The precise nature of the modal test, its validity and implications, need not be elaborated upon for our purposes. Rather, it is taken to illustrate the differing levels of significance that can be attached to information and that a particularly significant interference occurs where a lack of understanding altogether precludes a patient from aligning their care with their objectives. Contrast this with a situation where a patient decides to pursue a certain intervention, but does not know the future state of the world (whether a risk will eventuate or not).[406] This does not preclude effective action. Rather, making such judgment calls may be seen as a condition of practical agency.[407]

All in all, the notion that the absence of important beliefs may sever the connection between the agent's situation and the pursuit of their desired ends, is a fruitful one for identifying particularly severe autonomy violations, which may demand a unique institutional response. It does not, however, require a patient to conform their decision making to an overarching objective framework. It must always be borne in mind that the kinds of necessary beliefs are limited both in number and scope. Under the reflective dimension the pride of place in the balancing of reasons and interests must be left to the individual patient.

C. Summation

In sum, AI's challenges will be conceptualised as affecting one or more of these dimensions. In the course of this I will draw on the kinds of considerations adduced under Pugh's approach. These include a cognitive dimension of autonomy – which posits a decisional process where a minimal degree of theoretical rationality and an agent's reflective capabilities are brought to bear – as well as a practical dimension, which demands that

406 ibid 132.

407 Rid and Wendler, 'Risk-Benefit Assessment in Medical Research – Critical Review and Open Questions' (2010) 9(3-4) Law, Probability and Risk p. 151, 154.

the patient's decision-making process is not interfered with and, above all, that this process is also adequately facilitated by a suitable informational environment.

II. The challenges posed by clinical AI to procedural autonomy

On the above account patient autonomy demands that the patient be able to adequately reason about their clinical choices, so that these can be aligned with their own desires and goals. There are several ways in which the embedding of AI into clinical decision making can cause this endeavour to go wrong.

AI will sometimes constitute an external (non-agential)[408] influence that undermines the cognitive aspect of the patient's decisional autonomy. Most clearly this is the case where its use causes a patient to fail to hold a decisionally necessary true belief, but it may also happen where that individual is led to sustain theoretically irrational beliefs. Both of these may be referred to as forms of informational manipulation.[409]

ML technology may also affect the practical dimension of autonomy. One must distinguish here between an interference with negative and positive freedom. The former is limited where one is restricted from pursuing the end that one has decided to pursue. Ordinarily the functioning of AI cannot be expected to undermine this dimension: it is a supporting element in a decision-making process that does not have the ability to restrain autonomous action.[410] Where a patient's negative freedom may become relevant, is where AI use is made non-optional by a medical professional. Yet this purported autonomy interference must be located in the wider context

408 Pugh is clear that his account differs from Principlism in that non-agential forces can violate autonomy: Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 61. This is a further reason why the rationalist conception of autonomy is well-suited to identify the challenges posed by narrow AI.

409 ibid 60. I do not address another form of manipulation brought up by Pugh, psychological manipulation, as this occurs where the patient's desires are changed without appealing to their cognition. AI are less likely to bring about such changes, as they engage only indirectly with patients' ends, especially in the context of physician mediation. For example, automation bias does not change the desires of the patient.

410 I hold the same to be true in relation to coercion: AI technology is not designed to issue coercive threats and the possibility of AI-mediated use having the features of coercion does not warrant discussion in the abstract.

of medical practice, which is premised upon offering individuals a highly restricted choice-set that is shaped by the importance of other values.[411] Without more, offering medical treatment that is conditional on AI use is unlikely to constitute a significant autonomy violation and will not be considered further here.

Things are different with the dimension of positive freedom. Recall that such freedom is limited where one is not provided with the enabling factors to act in pursuit of one's goals. It focuses our attention on creating those conditions of decisional autonomy in AI use that allow for effective individual action. For example, while one may restrict a patient's negative freedom by denying them access to a non-AI alternative, one may still have an obligation to promote their awareness of the existence of non-AI alternatives. Indeed, an individual's autonomy may be substantially undermined where there are relevant differences between these alternatives and an individual is not put in the position to make a meaningful choice amongst them.

With these broad categories of interreference in mind, three challenges will be focussed on: causing a patient to fail to hold decisionally necessary true beliefs about goal-directed AI action, failing to assist the patient's in forming beliefs about the nature of AI-human cooperation and failing to facilitate the patient's understanding of AI devices, including their general risk characteristics and their influence on decision making.

A. The need to form true beliefs about AI's goal-directed action

Drawing on the insights of the last chapter we know that, when AI is inserted into medical decision making, it influences that process. Sometimes this influence has implications for the patient's pursuit of their plans and policies. Specifically, one way in which AI threatens to undermine autonomy stems from its ability to pursue goals relatively independently; if not independently from the designer, then at any rate independently from the user and subject.

This has been discussed in the bioethical literature in relation to Watson for Oncology, an AI similar to our case study of the AI-Pathway Companion Prostate Cancer. Both devices draw on ML techniques to complement

411 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 140.

human expertise, providing an additional resource for decision-making. Moreover, Watson analyses the patient's condition and proposes tailored treatment options, ranking them in different colours to indicate their desirability, which appear to be related to the chances of achieving 'disease-free survival'.[412]

One commentator, McDougall, has argued that because this end is not tailored to the individual patient, since it drives the treatment decision and since there is no encouragement to reflect on the value-laden nature of decisions, AI like Watson for Oncology violate patient autonomy.[413] This violation is said to be incurable by physician mediation because '[r]espect for patient autonomy means that patients' values should drive the ranking process. The patient's own values should be overtly shaping treatment decision making as a primary parameter, not a secondary consideration'.[414]

This account arguably overstates the significance of AI autonomy violations, missing the key situations in which the technology's goal-directed action becomes challenging. The rationalist approach to autonomy provides a more nuanced explanation of relevant violations, one that fits better with our understanding of the technology and with the rationale behind the legal protections of patient autonomy.[415]

Before critiquing her position, we can note that McDougall is surely correct in the claim that many medical devices based on ML do not tailor their recommendations to patient values. As such, they potentially pursue ends that diverge from those that some patients may wish to realise in their care.[416]

One cannot deny that this technology will, almost by necessity, have to have recourse to financial considerations. At a minimum, to be useful, a recommendation must accord with the treatments that the particular

412 Di Nucci, Jensen and Tupasela, 'Ethics of Medical AI' (5.12.2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3432317> accessed 5.4.2020.

413 McDougall, 'Computer Knows Best?: The Need for Value-Flexibility in Medical AI' [2019](45) Journal of Medical Ethics p. 156, 157-158.

414 McDougall, 'Computer Knows Best?' [2019](45) Journal of Medical Ethics p. 156, 158.

415 One relatively brief attempt has already been made to utilise Pugh's theory in this respect: Debrabander and Mertes, 'Watson, Autonomy and Value Flexibility' [2021] Journal of Medical Ethics p. 1043.

416 Bjerring and Busch highlight the conceptual issues of sensibly encoding preferences and values into artificial systems: Bjerring and Busch, 'Artificial Intelligence and Patient-Centered Decision-Making' (2021) 34(2) Philosophy & Technology p. 349, 360.

health system reimburses.[417] For many ML devices an additional goal may be to increase resource conservation and efficiency. At its most extreme, the designers of a particular AI may covertly seek to shape its reasoning to further their own financial interests.[418] Moreover, ML models may also weigh considerations that, while not financial, are nevertheless extrinsic to the patient's medical need. For instance, where an AI is online (i.e. continuously learning) a need to improve performance may influence how, and indeed what, decisions are made. The aim of furthering the 'research' that advances the AI would then become an operative goal in the treatment of the patient.[419]

Even here, however, the claim must be qualified. For some tasks assigned to AI the goals will be so narrowly defined that a greater capacity for independent action does not become relevant.[420] The IDx-DR case study illustrates how AI can be used to pursue a narrowly defined, singular goal: to diagnose diabetic retinopathy on the basis of an image-analysis of the patient's retina. If the patient knows the broad purpose of the procedure (say to check for one detrimental eye condition associated with diabetes), then it is unlikely that there will be a divergence of goals between human and machine. This is so even where there is no further attempt to incorporate patient values.

Admittedly, things are different for AI like Watson or the AI-Pathway Companion Prostate Cancer. They perform a broad array of complex tasks,

417 '[IBM] said that the system can be customized to reflect variations in treatment practices, differences in drug availability and financial considerations': Ross and Swetlitz, 'IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close' (5.9.2017) <https://www.statnews.com/2017/09/05/watson-ibm-cancer/> accessed 28.3.2023.

418 Although not involving AI, a situation like this has already arisen with drug-prescription software: Mann, 'Health Care Software Firm Fined $145M In Opioid Scheme With Drug Companies' (1.2.2020) <https://www.npr.org/2020/02/01/801832788/healthcare-software-firm-fined-145m-in-opioid-scheme-with-drug-companies?t=1615393792393> accessed 10.3.2021.

419 Cf. Grubb and Pearl's distinction between therapeutic and non-therapeutic touching, noting that '[t]he law does not assume that a patient who consents to treatment also consents to research': Grubb and Pearl, *Blood Testing, AIDS, and DNA Profiling: Law and Policy* (1990) 12-13.

420 Funer makes a similar distinction by reference to AI that do and do not depend 'on statistically based, normatively uncontroversial information': Funer, 'Accuracy and Interpretability: Struggling with the Epistemic Foundations of Machine Learning-Generated Medical Information and Their Practical Implications for the Doctor-Patient Relationship' (2022) 35(1) Philosophy & Technology, Article no 5.

involving the analysis of different data sources to visualise the patient's situation and possible treatment options. This visualisation and the presentation of treatment options will inevitably contain ideas about the ends that are to be pursued. As already discussed, this could involve a preference for the pursuit of longevity over quality of life. Here there is real potential for misalignment between the agent's desired ends and the one's that the AI is programmed, or has programmed itself, to pursue.

Whether this potentiality is then actualised depends on further factors that restrict the subset of problematic AI. Most importantly, as de Nucci has pointed out[421] and as we already saw in Chapter 2, it is still unclear whether an assisting AI/ML device can be said to drive a medical decision. No doubt it exerts some influence and capitalises upon some biases,[422] but the intention behind human involvement, especially human expert involvement, is to serve as a check on these influences and it is at least unclear when, or how far, these checks are ineffective.[423]

McDougall's assertion that AI drives medical decisions is a strong generalisation, especially if one uses Watson for Oncology, an AI merely complementing human expertise, as an example. If this generalisation is questioned, then it is plausible that any divergent patient goals can be meaningfully introduced into the subsequent deliberations between healthcare professional and patient. McDougall simply stipulates that autonomy demands that patients' values should drive the ranking process and that they should overtly shape treatment decision making as a primary parameter.

Pugh's autonomy account is intended to move beyond stipulations. Hereunder, AI could only drive clinical decision making in such a strong manner, irreversibly subverting patient values, if it hid its objectives and

421 Di Nucci, 'Should We Be Afraid of Medical AI?' (2019) 45(8) Journal of Medical Ethics p. 556.

422 For instance, Debrabander and Mertes note that Watson taps into the 'order effect': Debrabander and Mertes, 'Watson, Autonomy and Value Flexibility' [2021] Journal of Medical Ethics p. 1043.

423 McDougall appears to acknowledge that the influence of AI is an open question, calling for further empirical assessments: McDougall, 'No We Shouldn't Be Afraid of Medical AI; It Involves Risks and Opportunities' (2019) 45(8) Journal of Medical Ethics p. 559. Similarly Kudina and de Boer acknowledge that 'the line between supporting medical decisions and determining them may be thin if not carefully reflected upon': Kudina and Boer, 'Co-Designing Diagnosis: Towards a Responsible Integration of Machine Learning Decision-Support Systems in Medical Diagnostics' (2021) 27(3) Journal of Evaluation in Clinical Practice p. 529, 533.

thereby induced a failure to hold a decisionally necessary true belief.[424] This would sever the connection between the patient's actions and the goals that they wish to pursue. Under Pugh's modal test, this requires there to be no relevantly similar scenario in which the patient could fail to hold the relevant belief about the AI's objectives and yet achieve their end.

Such a state of affairs seems highly unlikely where a human expert meaningfully mediates AI use for the patient. Where an AI merely complements a human professional, this professional can use their specialised body of knowledge to assess the goals that AI outputs can achieve, as well as the implications of this and what other values and evidence are in play. On this basis they can discursively engage with the patient about their plans, desires and goals. This is arguably suited to generate the kind of *post facto* critical reflection that Pugh alludes to elsewhere.[425] Subsequently it is hard to identify elements of deliberative or practical autonomy that are inescapably undermined by such an AI output. Contrary to McDougall's claim, it is not necessary for a patient to know the AI's goals in order to align their action with their preferences in most circumstances.

Watson's case itself exemplifies the outlined distinction. Even if users do not know that the machine is optimising for life expectancy, the professional should still be able to incorporate the machine's outputs into their wider body of knowledge and identify the options that they judge to promote life expectancy, promote quality of life, etc. and to discuss this with the patient. The patient must then only be able to reflect upon those commitments and to incorporate them into their practical reasoning, choosing to endorse or reject a course of action accordingly.

This is not to say that an agent's autonomy remains entirely unimpaired however. Deliberating reflectively about the utilisation of AI outputs would admittedly be more straightforward if one knew the goals that an AI is striving to accomplish – or if one could exert a degree of control over these goals.[426] The problems associated with weaker forms of AI influence will be explored separately below.

For now it is found that, if one can deliberate in a roundabout way and reach a decision independently from the AI, then neither dimensions of autonomy is fatally undermined. Such general reflection, accompanied

424 In Section II.C.2. below we will discuss the weaker influence that AI can have on decision making.

425 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 138.

426 This is McDougall's proposed technological solution: McDougall, 'Computer Knows Best?' [2019](45) Journal of Medical Ethics p. 156, 158-159.

by an assisted decision process for the patient, is arguably maintained in many instances where AI complement or partially replace expert cognitive capabilities.

Where an AI's pursuit of goals irredeemably violates autonomy, is where it is not only used to accomplish a task that has no single, well-defined aim, but where this also partially determines an aspect of the medical procedure (i.e. where a device falls under the third class of case studies discussed in Chapter 2). One may refer back to the triaging example of Accipio Ix in this respect. The influence of this ML device's output irreversibly manifests itself as soon as it brings certain cases to the physician's attention above others. A patient or physician cannot counteract this influence by incorporating it into their reasoning because there is no opportunity to reflect on the ends underlying it.[427] Let us exemplify the resulting autonomy violations through two hypothetical examples of ML that pursue a general, or not very well-defined, aim and partially determine a decision.

The first example leans on Accipio Ix. While this AI arguably has a relatively narrow purpose, the identification and prioritisation of acute intracranial haemorrhages, there is significant potential for AI to be used in the triaging of patients more widely, with one NHS Foundation Trust for example already seeking to implement AI to support a general triaging service.[428] The very broad purpose of such devices will be to prioritise different clinical options in light of patients' conditions. Many different goals could be pursued under this head – focusing on the most acute conditions, ensuring an efficient allocation of resources, counteracting potential

427 Of course, in emergency settings autonomy may not even come into play in relation to triaging, as the patient may lack capacity, but it is not clear that this will always be so. Moreover, the point applies to the use of AI in the prioritisation of AI more generally, as the next example shows.

428 E.g. Mahase, 'Birmingham Trust and Babylon Health Discuss Pre-A&E Triage App' (2019) 365(l2354) BMJ (Clinical Research Edition).The Babylon Symptom checker app discussed here appears to rely on some forms of AI for some aspects of its service provisions, but they do not appear as sophisticated as the AI techniques discussed by Marchiori and others, 'Artificial Intelligence Decision Support for Medical Triage' [2020] AMIA Annual Symposium Proceedings p. 793. In the U.S. context see, for example: Johns Hopkins Technology Ventures, 'Digital Health Start-up That Assists Emergency Department Decision Making Acquired' (2022) accessed 17.3.2023.

discrimination, ensuring some targeted access to human contact and care, etc.[429]

Moreover, the rationale for giving AI this role is that they alleviate the intense pressure on a limited workforce.[430] As such, it stands to reason that AI will play a preliminary, largely unsupervised role: engaging the patients in a dialogue, undertaking a personalised assessment of their symptoms, giving them healthcare recommendations and in ensuring a targeted use of telehealth and physical services.[431] This could not only considerably speed up the availability of triaging services in-house, but even make it generally available to patients *via* web- and mobile-applications. Even if there is a backstop of human decision making therefore, the AI initially determines an aspect of the decision for which it will be hard to incorporate patient preferences.

The second example draws on the recognition that making and disclosing certain diagnoses has irreversible impacts on patients' lives (their plans, policies, etc). This is well-illustrated by discussions surrounding autonomy and predictive genetic testing, including the 'right not to know' certain classes of information.[432] Concerns arise particularly from the uncertain scope of many such tests and from the possibility of incidental diagnoses being made.[433] In some circumstances the concern is not only that the patient is unaware of certain risks or dangers (a false positive or negative),

429 '[W]hen we leverage AI to detect people with high potential for chronic disease, we must understand the AI goal. Is the purpose of the AI to make people with that disease process healthier, or is it to reduce costs? In the former case, the AI would prioritize the highest-risk patients to receive healthcare that would intervene in or even prevent development of the disease. In the latter case, the AI would appear discriminatory if it prioritized on the low end those who were at highest risk for poor outcomes, with the result that they received the least amount of healthcare': Holley and Becker, *AI-First Healthcare: AI Applications in the Business and Clinical Management of Health* (2021) 59-60.

430 Kim and others, 'A Data-Driven Artificial Intelligence Model for Remote Triage in the Prehospital Environment' (2018) 13(10) PloS One.

431 Marchiori and others, 'Artificial Intelligence Decision Support for Medical Triage' [2020] AMIA Annual Symposium Proceedings p. 793.

432 E.g. Andorno, 'The Right Not to Know: An Autonomy Based Approach' (2004) 30(5) Journal of Medical Ethics 435-9.

433 Bunnik and others, 'The New Genetics and Informed Consent: Differentiating Choice to Preserve Autonomy' (2013) 27(6) Bioethics p. 348, 350-351.

but that the patient has not consented to categories of tests with certain aims at all.[434]

AI that function by reference to widely defined goals will involve similar concerns. They may be initiated in a procedure for one broad, unspecified purpose and then generate knowledge in surprising ways, without the possibility for direct oversight. For example, an analysis of an ECG may, when supplemented by ML techniques, provide a non-invasive method for determining potassium levels – a feat that was previously impossible.[435] Once this knowledge is generated, a doctor may not be able to discuss its disclosure without revealing its existence.[436]

Drawing on the AI-Pathway Companion Prostate Cancer and case studies like it, one can imagine an AI tool that is used to conduct a multi-purpose analysis of existing patient data and which finds a rare form of cancer in an unanticipated way.[437] This finding cannot be brought up abstractly to the patient after the fact in order to engage their reasoning and judgment. It is important that the patient is made aware *ex ante* that a diagnosis with this purpose and the intendent aims was a possibility.

Lastly, it is also important to realise that these issues do not arise in a binary fashion, but on a sliding scale. Some AI that partially replace human expert decision-making will be suited to drive medical decision-making, although they do not determine an aspect of the clinical process. The purposes of other AI may appear relatively narrow, as with Accipio Ix, but this may obscure contentious judgments. To provide the clearest illustration of this category of autonomy challenges posed by AI, the two examples given here are admittedly futuristic. Yet they draw on existing clinical studies and

434 ibid 352; It is telling that Hostuic also frames the relevant issues in terms of the autonomy to initiate actions: Hostiuc in Hostiuc, *Clinical Ethics at the Crossroads of Genetic and Reproductive Technologies* (2018) 233. See also: Herring and Foster, '"Please Don't Tell Me": The Right Not to Know' (2012) 21(1) Cambridge Quarterly of Healthcare Ethics p. 20.

435 Dillon and others, 'Noninvasive Potassium Determination Using a Mathematically Processed ECG: Proof of Concept for a Novel "Blood-Less, Blood Test"' (2015) 48(1) Journal of Electrocardiology p. 12; Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (2019) 61-68.

436 Andorno, 'The Right Not to Know' (2004) 30(5) Journal of Medical Ethics 435-9, 436. This offers another analogy to genetic testing.

437 Although Watson for Oncology is used in a relatively restricted context, it already managed to surprise doctors by identifying a rare form of the disease: Rohaidi, 'IBM's Watson Detected Rare Leukemia In Just 10 Minutes' (16.8.2016) <https://www.asianscientist.com/2016/08/topnews/ibm-watson-rare-leukemia-university-tokyo-artificial-intelligence/> accessed 4.9.2022.

the features of approved devices and, more importantly, they point towards problems that exist more widely and less visibly.

It is my contention that such scenarios – those where ML devices are used to pursue broadly defined goals and in a way that is partially determinative of the clinical decision – result in a situation where a patient fails to hold a decisionally necessary true belief. If a patient wishes to realise certain authorised preferences through a clinical intervention and it involves this kind of AI use, then they will need to accurately gauge the goals that the AI is intended to pursue. Without such beliefs they simply have no opportunity to engage their practical reasoning and cannot align their decisions and actions with their desired ends. If a patient wishes to be triaged in order to receive the most effective care, while the operating AI in fact maximises for efficiency in resource use, then the patient must be aware of this discrepancy before their case is triaged. If an open-ended AI analysis is being conducted of the patient's data, then they must know the aims of this analysis before it is conducted and that this may yield significant categories of diagnoses.

Again, we may note that things are different were there to be an effective physician-mediator. Even if the AI pursues a covert goal and, in spite of expert involvement, this is allowed to determine the actual process of decision-making, there would at least be some similar scenarios where the expert intervenes to correct the misalignment. With meaningful physician mediation there is, at most, a risk of divergence. Without it, there will be a systematic discrepancy, the extent of which depends on how well the particular patient's ends align with the AI's assumed framework.

In summation, the argument in this section has been that there is a subset of AI where there is a systemic mismatch between the general, non-personalised goals that the technology pursues and the ones that the agent may wish to structure their clinical reasoning and decision making by. Moreover, this mismatch cannot be remedied by physician mediation *ex post*, since there will be no practical reasoning left to be done on the relevant point. In consequence, there is a very strong mandate that information be provided about the goal-directed nature of AI, including some general information about the aims of its use.

B. Theoretical rationality and changes in human-AI expertise

Recall that another injunction of our autonomy concept was that the patient must not be placed in a position where they sustain theoretically irrational beliefs. Drawing on this notion, the starting point for this section is the hypothesis that the reasoning of patients depends on certain justifiable assumptions about the expertise of medical professionals, which includes a degree of deference to such professionals' authority. The latter's privileged epistemic status appears widely accepted. Accordingly, it will be assumed that medical expertise confers some sort of domain-relative epistemic authority on medical professionals (they are 'experts').[438] In consequence, a patient will justifiably believe in the truthfulness of a medical opinion because it is provided by a medical professional and they have reason to believe, or conversely have no reason to doubt, the expert status of this professional.[439] To form a justified, theoretically rational belief they do not need to comprehensively evaluate the validity of every claim for themselves.[440]

Against this background, it is concerning that the reliance on certain AI tools is likely to give rise to situations where the patient has no grounds for changing their assumptions regarding the human professional, even though the basis for their expert status has fallen away. In particular, in our second class of case studies one can assume that there will be circumstances where the patient believes their medical professional to be an expert, or to be exercising their expertise, when in fact their relevant specialist knowledge is being derived primarily, if not solely, from an AI device. For example, in cases like IDx-DR, given the nature and complexity of the diagnosis, a patient may expect that there will be a person evaluating them with expert knowledge, when in fact they are only an augmented assistant.[441]

438 Funer, 'The Deception of Certainty' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 169.

439 Wagemans states this succinctly: 'the claim that the opinion involved is an ''expert opinion'' functions as an argument for the truth or acceptability of that opinion': Wagemans, 'The Assessment of Argumentation from Expert Opinion' (2011) 25(3) Argumentation p. 329, 331.

440 Goodwin, 'Accounting for the Appeal to the Authority of Experts' (2011) 25(3) Argumentation p. 285.

441 This is of course subject to the caveat that the patient may be informed of the assistant's status. But this is itself a remedy for a potential violation and as such will be returned to below.

Such examples constitute grounds for asking whether the patient's justified reliance on expert knowledge cannot be translated into a justified reliance on a human non-expert whose technical capabilities are augmented by AI. Why would the reasoning of a patient be justified when the relevant information comes from an expert, but not justified when it comes from a human-AI combination?

Two grounds will be adduced here. First, the theoretical appropriateness of relying on expert knowledge is premised in part on the ability of the advisee to critically interrogate it, which will not be possible where the only or primary source of a piece of knowledge is an opaque AI. Second, and relatedly, an individual's practical reasoning with expert knowledge is often premised on trust. The kind of justification that AI relies on may provide grounds for believing in its reliability, but it is not capable of instilling trust in the same way.

Regarding, the first ground, Walton has highlighted 'the critical nature of dialogue between the expert and the user of expert advice in argumentation' as a basis for the rational acceptance of an argument from expert authority.[442] Such a dialogue may not always be easy, even with human experts,[443] but 'interpretation, questioning and clarification' is possible.[444]

Concretely this means that, where information is conveyed in a practical decision-making setting such as medicine,[445] by someone claiming to possess expert status, the presumptive force of this information can be tested by the advisee. Even if they lack expert knowledge they can assess, for example, the credibility of the expert status of an individual, the consistency of the claim made by them with the claims of other experts, and the evidence that is relied upon to back up their claims.[446] Likewise, if someone puts forward another expert's views, then a secondary dialogue can be presupposed, in relation to which the same kinds of critical questions can be asked.[447] If these questions cannot be adequately answered, the appeal to expert opinion loses much of its force.[448]

AI's challenge to theoretical rationality emerges partially from the fact that it precludes these kinds of questions from being answered. Funer

442 Walton, *Appeal to Expert Opinion: Arguments from Authority* (1997) 112.
443 ibid 113-114.
444 ibid 114.
445 ibid 118-119.
446 ibid 258.
447 ibid 121-122.
448 ibid 258-259.

is one commentator who has picked up on this dimension, arguing that relevant AI techniques evade 'such deliberative scrutiny'.[449] But one can also see that such a conclusion follows directly from our discussion of AI opacity in the previous chapter. Assessing AI credibility is much more complex and not easily demonstrated by established methods. The precise evidence that the AI relies upon will not be apparent – even in cases of explainable AI. For it will not be possible to provide an intuitive, faithful explanation of individual elements being processed through the ML model and the kinds of human-interpretable justifications being applied to them. Moreover, comparing an ML device's output with the opinions of other expert will be more difficult in scenarios where a non-interrogable device is providing a non-specialist professional with the requisite expertise. It is entirely plausible that they themselves will lack the context to aid the patient's practical decision making in this respect.

Stated simply, a situation involving a replacement of overall human expertise with AI expertise – i.e. where the ML device is the only agent possessing a requisite level of knowledge and skill with regard to an aspect of the patient's care – a patient will at least have to engage with this expertise differently to make a rational decision. As our previous discussion of explainable AI demonstrates, this shortcoming cannot be remedied by technological means alone; by attempting to make a secondary ML model that is responsive to critical inquiry. The responses given by an explainable model only approximate to the actual decision-making process and they will sometimes provide the patient with a misleading impression. Therefore, even if a patient knows an AI is used and they attempt to exercise their theoretical rationality *vis-à-vis* expertise normally, by critically interrogating ML functioning, they ought not to attribute the same kind of weight to the responses they receive.

Such a concern leads on to our second ground. It could be rational to adhere to the recommendations of actors, even where other avenues of critical questioning are limited, if they are sufficiently reliable or trusted.[450] Indeed, this will be the natural recourse of an individual who is uncertain whether a certain bias has influenced a human expert, or whether they are lying – possibilities that are arguably comparable to the shortcomings

449 Funer, 'The Deception of Certainty' (2022) 25(2) Medicine, Health Care and Philosophy p. 167, 174.

450 Krishnan, 'Against Interpretability' (2020) 33(3) Philosophy & Technology p. 487.

of explainable ML models.[451] Walton submits that a patient must make a subjective assessment in these circumstances of reliability or trustworthiness.[452]

For our purposes it is important to distinguish between these two concepts. There is a fundamental differentiation between trust in human agents and reliance on technical tools. Ryan has elaborated this distinction in depth, stating that trust exhibits a 'concern about the trustee's motivation for action',[453] whereas reliability lacks this concern and is preoccupied with the evaluation of past behaviour and its extrapolation in order to make predictions.[454] Reliability indicates that 'the network tends to track the truth'.[455] By contrast, the motivational aspect, which is highly pertinent to the clinical sphere, appeals to the anticipation of the trustor that the trustee will act with good will towards them, exercising their skill in their interest.[456] This posits a subjective, affective attitude towards the trustee that allows for the trustor's rational evaluation of their actions and recommendations.[457]

With regard to AI, it has been convincingly argued that such an anticipation is simply not possible.[458] Nothing in our technical analysis alters this viewpoint. Quite the reverse. The connection between the values an ML device tracks, and the values of patients was argued to be tenuous, especially where the AI could exercise a wide discretion. Furthermore, nothing in the technology's design suggested an affective attitude towards an individual or a specific case. It was found to be an important role of human professionals to exercise their practical judgment to introduce relevant qualifications in the context of individual decisions.

In conclusion, if an AI replaces an element of human expertise in the clinical decision-making process, then a patient must be given the oppor-

451 Walton, *Appeal to Expert Opinion: Arguments from Authority* (1997) 115.
452 Walton himself recognises this 'trustworthiness question': ibid 115, 258. Although he does not distinguish it from reliability, as will be done below.
453 Ryan, 'In AI We Trust: Ethics, Artificial Intelligence, and Reliability' (2020) 26(5) Science and Engineering Ethics p. 2749, 2752.
454 ibid 2759.
455 Smart and others, 'Why Reliabilism Is Not Enough: Epistemic and Moral Justification in Machine Learning' (AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 07.02.2020-09.02.2020), 374.
456 Ryan, 'In AI We Trust' (2020) 26(5) Science and Engineering Ethics p. 2749, 2752-2753.
457 ibid 2760-2761.
458 ibid 2761-2762.

tunity to adjust their assessment of these recommendations accordingly. In such situations it will not be possible, or at least more difficult, for them to adequately bring their theoretical rationality to bear on the relevant recommendation. Nor will they be in a position to trust the technology in a manner that they are accustomed to trust a human expert. In consequence, they must be put in a position where they can adjust their confidence in the relevant output or recommendation.

Given our analysis in Chapter 2, this challenge is most likely to arise where an AI device leads to a diminishment of the overall level of expertise brought to bear on a decision: where a less qualified individual performs a role that was previously reserved for a more qualified one. By comparison, it will not occur in the case of, for example, Mia (Mammography Intelligent Assessment) where the technology merely offers a second opinion without influencing the initial, separate human decision, which remains trusted and interrogable.

The distinction between these categories must necessarily remain fluid. Not least due to the dimension of informational manipulation that will be evaluated below. This may diminish the extent to which a human actor can independently bring their expertise to bear on a decision, even if they do possess a requisite degree of skill.

C. Positive freedom and the task of ensuring an adequate understanding of AI

The autonomy of a patient may also be challenged on the basis that they do not possess an adequate understanding of the AI device itself. Two categories of information will be adduced here. First, understanding the way in which the ML technology underlying AI is generically related to the risk profile surrounding its use allows a patient to act effectively in pursuit of their ends. It enables them to surpass the threshold level of positive freedom that is required for them to deliberate and act effectively. Second, given that AI devices will increase the potentially manipulative influences that a patient is exposed to, understanding AI's relation to such influences is necessary to assess the risk that this poses to the patient's autonomy.

1. General risk characteristics of AI

Risks are a function of the probability that a harmful event occurs in the future and of the extent of that event's harmful impact.[459] Such risks are often related to, and balanced against, the future, expected beneficial impact that is associated with the same event.[460]

There can be no question that specific clinical AI are intended to provide such clinical benefits and that they simultaneously pose risks in terms of the physical well-being of the patient, which constitutes one commonly accepted way to understand harm in the bioethical context. A diagnostic AI such as the Acumen Hypotension Prediction Index Software, described in Chapter 2, exemplifies how the technology shifts the risk-profile of an intervention. By expanding 'the diagnostic and monitoring abilities currently available in operating rooms, which fail to predict hypotension at an early stage',[461] it ameliorates an existent risk (provides a relative benefit). Yet it also creates new dangers, such as the prospect of incorrect, unnecessary interventions in response to false positive readings. This would create the chance of a different kind of physical harm materialising, whose exact likelihood is the result of the design and use of the AI.[462] Like most clinical interventions, we may therefore suppose that AI devices give rise to the possibility of a harmful event occurring in the future and that they co-determine its likelihood.

Holding beliefs about such risks is highly relevant to patient autonomy, although their absence does not disconnect a patient from the pursuit of their ends. Rather, it is a piece of information that increases a patient's understanding so that they can exercise their practical rationality more effectively, for example by placing them in a better position to assess and

459 Rid and Wendler, 'Risk-Benefit Assessment in Medical Research – Critical Review and Open Questions' (2010) 9(3-4) Law, Probability and Risk p. 151, 152. See also: Perry in Lewens, *Risk: Philosophical Perspectives* (2007) 190.

460 Rid and Wendler, 'Risk-Benefit Assessment in Medical Research – Critical Review and Open Questions' (2010) 9(3-4) Law, Probability and Risk p. 151, 151-152.

461 Davies and others, 'Ability of an Arterial Waveform Analysis-Derived Hypotension Prediction Index to Predict Future Hypotensive Events in Surgical Patients' (2020) 130(2) Anesthesia and Analgesia p. 352, 352.

462 U.S. Food & Drug Administration, 'De Novo Classification Request for Acumen Hypotension Prediction Index Feature Software' (16.3.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN160044> accessed 7.3.2022.

reason about their different options.[463] It is imaginable that such information will play a significant part in a patient's decision-making process, allowing them to gauge whether they should rely on a procedure or choose an alternative course of action.

Concretely, thinking back to Chapter 2 one could imagine a patient who is made aware that their X-Ray is being prioritised according to an automated procedure with a certain risk profile or even with a propensity for certain mistakes. They may then be in a position to consider whether their case is being subjected to an obvious error and respond accordingly – seeking out additional attention from their care team. Risk disclosure thereby becomes an enabling factor that is necessary to secure an appropriate degree of positive freedom for the patient.

This relation between AI and particular risks does not secure the patient very much information about the AI being used in their care, however. Indeed, as Schönberger has noted, 'even AI applications in riskier areas would not add anything novel' to the information that the patient requires.[464]

Arguably, this relatively mundane disclosure provides insufficient information about AI-related risk. In particular, a patient may assume that the risk assessment around AI is comparable to that carried out for other procedures involved in their medical care. Yet there are general characteristics of ML technology that complicate the picture and which a patient must be placed in a position to assess for themselves.

An analogy can be drawn to the characterisation of innovative treatments. The innovative nature of a procedure is not merely an aspect of the specific type or magnitude of risks that the patient is facing in their care. Rather, the classification as 'innovative' itself conveys generic information to the patient that it is important for them to accommodate in their reasoning process if they are deliberating about their options. Especially when this involves a choice between an innovative and non-innovative alternative.

The first piece of generic information is that there is normally a lack of scientific testing of novel procedures. These do not require one to proceed according to the ordinary standards of validated scientific knowledge: 'in offering innovative treatment, the physician is working on a hunch or

463 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 166.

464 Schönberger, 'Artificial Intelligence in Healthcare' (2019) 27(2) International Journal of Law and Information Technology p. 171, 188.

scientific theory that has not been adequately investigated or researched'.[465] Second, and relatedly, there is a general assumption that there is a greater degree of uncertainty in such situations about the nature or degree of physical harms to which the patient is exposed.[466] Similar arguments in favour of a category-driven approach have been advanced in relation to the unlicensed or off-label use of medical products and devices.[467]

While it cannot be assumed that the classification of procedures as novel or off-label exactly tracks their risk profile – and that this profile differs categorically from 'normal' medical interventions[468] – nevertheless a relevant categorisation is seen as an important risk-related factor for the patient to be aware of, to consider and to weigh in their decision-making.[469]

One explanation, under Pugh's account, stems from the fact that the provision of some information about risks can aid reasoning, but that an overly complex description of relevant factors does not further, and will in fact hinder, it.[470] Given that a detailed explanation of the (lack of) scientific evidence or approval, or the mechanisms associated with this, and the different reasons for uncertainty could negatively impact a patient's capacities for reasoning, it is not supportive of autonomy to disclose them. Yet the positive dimension of practical autonomy can require that the patient has some understanding of the procedure's type and the general way in which this is likely to impact the risk calculus. This provides the patient with an indication that their treatment is subject to distinct considerations and provides them with an opportunity to deepen their understanding of the more detailed implications of this classifier if they so wish.

465 Chan, 'Legal and Regulatory Responses to Innovative Treatment' (2013) 21(1) Medical Law Review p. 92, 94.

466 ibid 94. Uncertainty was also emphasised as a characteristic in the English case of *Simms v Simms* [2003] Fam 83.

467 'Lack of approval does not necessarily mean that the drug is dangerous or ineffective, but it should raise a concern about safety that the patient should weigh in deciding whether to consent to the treatment since it has not been proven safe and effective for the prescribed purpose. Since most off-label uses are not supported by scientific evidence, they may be ineffective or even detrimental': Johns, 'Informed Consent: Requiring Doctors to Disclose Off-Label Prescriptions and Conflicts of Interest' (2007) 58(5) Hastings Law Journal p. 967, 1015.

468 Beck and Azari, 'FDA, Off-Label Use, and Informed Consent: Debunking Myths and Misconceptions' (1998) 53(1) Food and Drug Law Journal p. 71, 84.

469 Price, 'Remodelling the Regulation of Postmodern Innovation in Medicine' (2005) 1(2) International Journal of Law in Context p. 121, 137.

470 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 169, 176.

On similar grounds, it is suggested that such a categorical disclosure obligation should extend to the use of AI. It has generic features that are closely analogous to those associated with innovative treatments. These features may not always be present, or bear a direct relation to the intervention's risk profile, but they are prone to have this effect and this may influence the balance of the patient's considerations. At the same time, as Schoenberger rightly remarks,[471] it would be too much to demand of the patient to understand many of the underlying technical details.

One comparable ground is provided by ML-interventions' lesser reliance on established scientific knowledge. The dominant ML techniques driving AI will not rely on scientific knowledge in the same way that a human doctor might. Rather than basing recommendations on the validated testing of proffered hypotheses, it was seen that the algorithm's judgment stems from the establishment of probabilistic correlations.[472] Such correlations may align well with the state of scientific knowledge, but they may also offer completely new insights. This is one of the promises of medical AI after all. Even where AI outputs are seemingly generated according to established medical wisdom, they may latently be subject to certain confounders, causing the AI not to recognise a 'true signal' or causal relationship, and relying instead on a correlated but causally insignificant indicator, such as the presence/absence of a ruler in an image.[473]

Moreover, one cannot ignore the added difficulties in scientifically validating the use of AI. The multidimensional factors that these machines consider, as well as the multiple dimensions of uncertainty that surround their deployment, make its assessment according to the 'gold standard' of randomised control trials very difficult. Assessing even offline AI has been said to be more akin to the 'evaluation of highly complex health care delivery interventions'.[474] Where randomised control trials have been carried out, there have still been calls for caution based on the size and diversity of the sample.[475] Once one moves to online AI such difficulties

471 Schönberger, 'Artificial Intelligence in Healthcare' (2019) 27(2) International Journal of Law and Information Technology p. 171, 188.

472 Bennett and Doub in Luxton, *Artificial Intelligence in Behavioral and Mental Health Care* (2016) 30.

473 Narla and others, 'Automated Classification of Skin Lesions' (2018) 138(10) The Journal of Investigative Dermatology p. 2108.

474 Angus, 'Randomized Clinical Trials of Artificial Intelligence' [2020](11) The Journal of the American Medical Association p. 1043.

475 ibid and see Chapter 2 for a more comprehensive analysis of these issues.

are compounded by an evolving model. This fits itself to the population it serves and it would be self-defeating to test it every time it changed the nature of its functioning. Ultimately, as with innovative treatments, one should note that procedures implementing AI can be expected to have a more tenuous relationship with scientific knowledge and optimal forms of validation.

On top of this, one must account for the uncertainty stemming from the black box nature of any individual decision. The patient is faced with a decision that cannot be easily comprehended and they cannot know what factors contributed to it, how they were weighed or according to what criteria. As a consequence, there is a risk that the procedure does not confer the benefit that the individual patient would want or expect it to confer or that they are exposed to harms, which are only associated with certain (unidentified) populations that the AI is ill-suited to. Such uncertainties can be expected to persist even if the AI is validated and found accurate in a general sense.[476]

Of course, these relations to scientific knowledge and the degrees of uncertainty that arise are partly dependent on the features of any given ML device. They will be impacted *inter alia* by: the ML techniques being used, how they have been validated and by the information that is made available to users by the developer. Uncertainties about the factors going into a decision could be influenced by the type of interpretability or explainability mechanisms that are implemented, with the latter compounding uncertainties and adding a potential for misleading or incomplete responses (regarding which, see the next section). As was seen in Chapter 2, evaluations can also offer more or less detailed breakdowns of differential impacts on population groups.

Nevertheless, the general trends that AI exhibit in this regard are grounds for singling out ML devices as a type of tool whose challenges are comparable to those of innovative treatments. As a class they have a peculiar relation to forms of scientific knowledge and they incorporate significant types of uncertainty. In this way, the use of ML in medical care is indirectly related to the risk assessment of patients regarding their physical well-being. Understanding AI-based treatment as an accessible categorisation of risk would provide patients with relevant information that furthers their

476 Abràmoff and others, 'Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices' (2018) 1 NPJ Digital Medicine p. 1.

positive autonomy, without overloading and hindering their decision making.

2. Informational manipulation

Sometimes an AI will challenge autonomy because, even where the technology does not determine a choice directly (as in Section II.A.), it exerts an influence on the patient's decision and this influence is manipulative: it causes them to reason in a theoretically non-rational manner and/or to act in practically non-rational ways.[477] This influence can be conceived of as a nudge, a concept that was popularised by Thaler and Sunstein.[478] They define this as 'any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives'.[479]

Although this work does not purport to offer a novel analysis of nudges, it is worth amending and clarifying this definition for our purposes. First, we may say that it is not just any aspect of the choice architecture that has normative implications. Rather, as Hansen and Jespersen have argued, it applies to 'attempts at influencing choice' that are 'directed towards any well-defined consistent end'.[480] This distinguishes targeted interferences that are liable to align individual behaviour with external aims from mere accidental environmental noise.[481] Second, particularly in the healthcare sphere it should be self-evident that we are not only concerned with

477 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 80.
478 Thaler and Sunstein, *Nudge* (The Final Edition 2021).
479 ibid 8.
480 Hansen and Jespersen, 'Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy' (2013) 4(1) European Journal of Risk Regulation p. 3, 9-10.
481 Hansen and Jespersen frame this distinction more broadly in terms of intentionality: ibid 10. If intentionality were necessary, then it would be problematic, as it is questionable whether the relevant kind of intentional agency applies to AI: Froese and Ziemke, 'Enactive Artificial Intelligence: Investigating the Systemic Organization of Life and Mind' (2009) 173(3-4) Artificial Intelligence p. 466. Yet I do not believe that this step is crucial for distinguishing random behavioural influences of the environment from external manipulations, especially as autonomy violations can be non-agential under our procedural theory of autonomy: Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 61.

economic incentives, but with much broader benefits and costs.[482] Third, there is a sense that the influence of the nudge is in some way separated from the provision of reasons that can be deliberated upon. While nudges may partly appeal to deliberation and reasons, they aim to direct cognitive processes beyond this, by exploiting unconscious, hidden processes that are not necessarily in line with one's reasoned preferences.[483] It is in this sense that they undermine both the cognitive and, especially, reflective element of deliberative autonomy and further irrational or non-rational tendencies.

i. AI nudging

AI's ability to induce biases was already examined in Chapter 2. By leveraging its unique capabilities, it is predicted that AI will increase the influence that design-induced psychological mechanisms have on making individualised decisions about patient care.

Crucially, it must be noted at this stage that even in the case of physician mediation, various AI will be given the opportunity to influence patient behaviour directly. This is significant as, if the AI's presented information would only be used by the physician, then it would be only the physician's choice that would be affected. For example, as the Acumen Hypotension Prediction Index Software information is only intended to reach the physician, there is no opportunity to impact the patient's decision-making. So long as the patient has sound grounds for relying on the physician (see Section II.B.) patient autonomy is not engaged.

It is where the AI is incorporated as another party into the deliberative decision-making process, that its nudges take on the quality of challenges to autonomy. Such a role is clearly envisioned for certain clinical AI, especially where they are involved in relatively broad decision-making tasks. For example, Watson for Oncology is capable of generating 'shareable individual treatment plans and patient education materials to engage the patient'.[484]

482 Hansen and Jespersen, 'Nudge and the Manipulation of Choice' (2013) 4(1) European Journal of Risk Regulation p. 3, 7.

483 ibid 13-15.

484 IBM, '5725-W51 IBM Watson for Oncology: Sales Manual' (2020) <https://www.ibm.com/common/ssi/cgi-bin/ssialias?appname=skmwww&htmlfid=897%2FENUS5725-W51&infotype=DD&subtype=SM&mhsrc=ibmsearch_a&mhq=IBM%20WATSON%20ONcology> accessed 18.3.2023.

The use of AI as a patient-facing triage device is another example where nudging becomes a tangible possibility.

ii. Impermissible manipulation

Granting their existence, it is important to indicate precisely how such influences may be challenging for autonomy. This links to the widespread claim that nudges are problematic for autonomy because they work by manipulating choice.[485] On such a view an AI that nudges will be problematic without more.

Yet this sits uncomfortably with Pugh's process-oriented account of autonomy. It is possible to be influenced without having one's deliberative or practical autonomy diminished.[486] Indeed, prompting people to determine their organ donor status is a classic example of a nudge.[487] Even if there is an appeal to non-rational mechanisms, overall this can promote one's autonomy by causing one to make a decision that reflects one's wider, authentic preferences.[488] Benign AI nudging would occasion an opportunity for the kind of evaluative judgments regarding one's preferences and rational acceptances that furthers the patient's reflective autonomy in reaching care decisions.

The task thus becomes identifying the features of *certain* nudges that make them problematic for autonomy ('manipulative') and determining whether the influences introduced into individual clinical decisions by AI exhibit these factors. The dominant features that can be found in the literature in this respect relates to goal-divergence. There is a sense that a nudge presents a greater danger for autonomy – or only presents a danger for autonomy – when it diverges from the goals of a patient. This is evident with respect to Thaler and Sunstein's claim that nudges are liberty preserving when they 'aim to influence choices in a way that will

485 Hansen and Jespersen, 'Nudge and the Manipulation of Choice' (2013) 4(1) European Journal of Risk Regulation p. 3, 5. More recently see: Ploug and Holm, 'Doctors, Patients, and Nudging in the Clinical Context--Four Views on Nudging and Informed Consent' (2015) 15(10) The American Journal of Bioethics p. 28.

486 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 79-82.

487 Thaler and Sunstein, *Nudge* (The Final Edition 2021) 269-271.

488 Hansen and Jespersen, 'Nudge and the Manipulation of Choice' (2013) 4(1) European Journal of Risk Regulation p. 3, 21.

make choosers better off, *as judged by the choosers themselves*'.[489] Hereby nudges do not merely track an individual's anticipated choice, but they seek to realise the objects of their desires, which they may otherwise be too weak-willed or inattentive to pursue.[490] One may say that they trigger a process that enhances an individual's ability (their positive freedom) to engage in the deliberative, reflective dimension of autonomy.

The link between the goals of the chooser and the nudges that they are subjected to has also been appealed to directly in the bioethical literature. Specifically, Blumenthal-Barby and Naik maintain that nudges in the individual clinical context will tend to have a comparable autonomy-maintaining or autonomy-enhancing effect because they are responsiveness to patient values. This is framed more abstractly from the libertarian paternalistic notion of tracking the objects of an individual's desires, which is already distinguished from an anticipation of their specific choices.

For these authors it is crucial that clinicians can and do 'use their understanding of patient's values and informed preferences to guide their nudges in ways that help patients to more efficiently work through competing goals'.[491] This does not entail that the nudges exactly promote the goals that the patient wants to promote. It simply entails that the patient is able to deliberate about and clarify their own objectives in response to the nudged behaviour. From the perspective of procedural autonomy this would be acceptable: it can be conducive of rationality to promote diverging goals and choices in the clinical interaction.[492]

What all of these accounts therefore require is that nudges have some positive contribution to make to the maintenance of reflective autonomy. They either track general preferences or are in some way responsive to an individual's circumstances to promote reflection and thus deliberation.

As has been elaborated in Section II.A. above, AI design will not be closely tailored to individual preferences. It will be shaped by the goal-directed behaviour of the AI and its designers. For example, the content and form of information will plausibly be connected to its pursuit of different incentives provided by the designer and by the sub-goals that the AI

489 Thaler and Sunstein, *Nudge* (The Final Edition 2021) 7.

490 ibid 7-8. For a statement of how it is possible for a disjuncture to arise between one's effective motivations and the objects one desires most see: Mele in Mele and Rawling, *The Oxford Handbook of Rationality* (2004).

491 Blumenthal-Barby and Naik, 'In Defense of Nudge-Autonomy Compatibility' (2015) 15(10) The American Journal of Bioethics p. 45, 45.

492 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 61-63.

develops to approximate to the successful performance of these tasks. If an AI's goal is the promotion of disease-free survival, then the options that it provides, and their ordering, will naturally reflect this goal. In this manner a bias is introduced into decision-making that does not align with a patient's general or specific interests. One can see how guidance issued to professionals is beginning to respond to this danger,[493] although as discussed in Chapter 2, the precise nature of this phenomenon still needs to be adequately researched and assessed.

For a nudge to be justifiable, this leaves open the possibility of it generating a prompt for reflective thinking. To the extent that the AI pursues a therapeutic goal, and its use is accompanied by human mediation, it is arguable that such thinking is a probable consequence of the technology's use. A patient who is confronted with a default therapeutic choice, but is provided with contextual information surrounding this, is able to form their own judgment in response to that prompt – whatever interests it exactly promotes.

The most problematic situations will occur where a patient cannot contextualise a prompt by reference to wider clinical reasoning. In other words, where an ML device exerts influence in favour of a non-therapeutic objective. There are expected to be instances where this will occur. It was seen that AI must account for non-therapeutic considerations, such as the clinical resources available in a particular care setting or for the reimbursement conditions of a particular healthcare institution and system in which it operates. It also stands to reason that explainable AI or online AI would covertly pursue other goals that are not strictly speaking therapeutic, such as rendering a persuasive representation of ML reasoning to the user or improving ML performance. Where such factors are incorporated into the presented information, there is the starkest disconnect to the kind of nudges that could trigger individual reflection regarding a patient's therapeutic situation.

In sum, the outlined non-obvious pursuit of certain values can preclude an AI from serving as a tool that encourages reflective thought. Instead, it becomes a tool that surreptitiously introduces certain commitments that are never reflectively assessed or endorsed. Our analysis from the previous

493 Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

chapter further suggests that it would require considerable effort to detect and counteract these.[494]

Therefore, to the extent that there are such mismatches in values between patient and AI, and the patient cannot be expected to identify and reflect upon these, there is a problem for patient autonomy. The magnitude of this interference must depend on the circumstances of the case. Yet, as a general rule, it may be thought that this challenge is on the lower end of the scale. The patient will be influenced by a number of factors after all and even if an AI nudge is real, it will be difficult to establish that it impaired a relevant decision process in a way that significantly impacted an individual's positive freedom.

III. Conclusion

In conclusion, this chapter sets out the link between the technical knowledge that was collated in Chapter 2 and the legal analysis that is to follow. Pugh's rationalist theory of autonomy has provided us with a means of doing so: identifying four categories of AI challenge that are detrimental to a procedural account of autonomy. In moving on to our legal analysis of these challenges, it is worth remembering that they are not all of a comparable scale. Disconnecting a patient from the pursuit of their goals was understood as a particularly problematic prospect. The different nature of AI risks and of AI expertise may also be understood as something that is of considerable importance to the maintenance of the patient's theoretical and practical autonomy. By contrast, the prospect of AI nudging and manipulation was categorised as an existent, but lesser interference with the patient's positive freedom.

494 For an argument elaborating how some nudges are more difficult to resist – weakening 'attention-bringing and inhibitory capacities' – see: Saghai, 'Salvaging the Concept of Nudge' (2013) 39(8) Journal of Medical Ethics p. 487.

Part II:
Autonomy in the law

Chapter 4: Autonomy in UK law

Up until now it has been argued that artificial intelligence (AI) in medicine is a technology that is capable of posing a novel threat to patient autonomy, conceived of in a bioethical and pre-legal sense. This part aims to demonstrate that those aspects of autonomy, which were shown to be violated by AI, are also aspects that are valued and protected in the medical law of the respective legal systems. This connects the outlined AI-threat to our specific legal analysis.

Focussing on the notion of autonomy in the law also serves a further purpose. Given the novelty of the problems created by AI, existing mechanisms must, at the very least, be extended to new situations. At most, the specific nature of the AI-threat may call for creative norm-altering or even norm-generating activity. Providing an accurate description of the applicable norms therefore depends on anticipating such applications and developments, while operating in a framework of mechanisms that were not conceived with these specific issues in mind.

It is hypothesised that the law's understanding of autonomy can provide significant guidance for its adaptation to this specific innovation. As argued in the introduction, this is a crucial step that bridges the gap between the limitations of doctrinal stability and the potential for structured adaptation. The following part thus focuses on locating autonomy-related concerns in the law and on investigating how these may make a practical difference to its operation.

Two fundamental questions will shape our exposition of the legal concept of autonomy. The first question that is asked is: 'what is/are the relevant function(s) that autonomy plays?' The aim is to ascertain how the term is able to influence legal norm-creation, norm-application, argumentation and reasoning. Depending on the material that the inquiry must fit, this circles around such matters as whether the concept is incorporated

directly into a legal rule, whether it is one *telos* that provides a possible avenue for interpretation of a particular norm or whether it is a legal principle.

Second, it will also be enquired: 'what is the content that the law ascribes to autonomy?' The aim is to determine how the law interacts with a polysemous concept and whether our bioethical specification offers a defensible interpretation of this interaction. In many ways this question is closely related to the matter of function. For, the content attributed to a term by legal actors will be uniquely related to the function that it plays (and the purpose it fulfils) within the legal system.[495] This suggests the necessity of addressing the question of legal function first. The question of content then has an altogether different focus. Rather than looking towards the formal role, it looks to the substantive content. Nor is the relationship between function and meaning unidirectional. The manner in which a term is understood more generally, may itself influence the parameters within which its functions can be conceived and developed by legal actors.[496]

In this chapter a systematic approach is developed to identify the function and content of the autonomy-concept within the UK jurisdiction, before going on to resolve equivalent questions in the American context. Focussing on England for now, it appears undeniable that autonomy is a notion that has manifested itself in various forms and commands considerable respect across the legal order, spanning a range of legal practice areas, and featuring both in the common law and statute.[497]

---

495 '[T]he law can appropriate philosophical materials, put them to use for its own purposes, engraft them into a distinct mode of reasoning, *and, as a result*, convert a philosophical norm or concept into a legal one. Once philosophical materials form part of a legal reasoning process, rather than part of a philosophical enquiry, they become legal norms and concepts': Wall in Phillips, Campos and Herring, *Philosophical Foundations of Medical Law* (2019) 133.

496 Poscher frames the relevance of more general concepts thus: 'to function well, the law cannot be altogether out of step with the knowledge in the society it is supposed to govern': Poscher in Hage and Pfordten, *Concepts in Law* (2009) 102.

497 This engagement with autonomy is not understood as limited to just those instances where the courts and the legislature expressly utilise that specific term. Rather, as will emerge from the following analysis, the focus is on the range of medium- and low-level considerations that legal actors have associated with autonomy, its synonyms and related notions. For instance, 'self-determination' constitutes one such synonym, which is frequently used in English law: Wicks, *Human Rights and Healthcare* (2007) 64-65. Of course, there are also terms that are related to autonomy but that, upon proper analysis, possess distinct connotations. 'Bodily integrity' would be one example. For an analysis of this concept and its relation

Before addressing the already outlined aspects, a comment will therefore briefly be made on the appropriate scope for our inquiry. Our interest lies with its manifestation in the medical context which, it will be seen, has provided a unique – and especially fertile – ground for the formulation of legal protections of autonomy. After dealing with the matters of scope, function and content, the limits of an autonomy-focussed legal analysis will also be acknowledged.

I. Scope

Since the use of 'autonomy' spans across several areas of UK law and its functions and content cannot be assumed to be consistent across them, we should begin our investigation by demarcating an appropriate scope for our investigation.

A convenient point of departure is our area of interest, medical law, and identify whether the positive law differentiates the use of the autonomy concept within this field from others. With this object in mind, it emerges that the English courts themselves have repeatedly and explicitly sought to draw such a distinction when utilising autonomy and its related concepts. Thus Lord Scarman in *Sidaway v Board of Governors of the Bethlem Royal Hospital*, when he considered 'the patient's right to make his own decision', rejected the U.S. approach of finding a fiduciary relationship between doctor and patient.[498] He stated that 'there is no comparison to be made between the relationship of doctor and patient with that of solicitor and client, trustee and cestui qui trust or the other relationships treated in equity as of a fiduciary character'.[499]

Similarly, Lord Walker in *Chester v Afshar* – a case dealing with the non-disclosure of a risk by the treating surgeon – stated that he derived 'very little assistance from analogies based on quite different facts (such as a landowner's duty to warn of the remote risk of a hiker being injured by

to autonomy see: Herring and Wall, 'The Nature and Significance of the Right to Bodily Integrity' (2017) 76(3) The Cambridge Law Journal p. 566. Such distinct connotations should not distort this investigation. Where the courts nevertheless use such concepts interchangeably – and many common concerns and modes of legal reasoning are indeed identifiable – their use will sometimes also form a proper subject for our analysis.

498 *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 884.

499 ibid 884.

a landslide or a falling tree)'.[500] And in *Re T*, Staughton LJ, ascertaining whether a patient's refusal of medical treatment had resulted from the undue influence of another, was careful to point out that: '[t]he cases on undue influence in the law of property and contract are not, in my opinion, applicable to the different context of consent to medical or surgical treatment. The wife who guarantees her husband's debts, or the widower who leaves all his property to his housekeeper, are not in the same situation as a patient faced with the need for medical treatment.'[501] Such authorities support the proposition that, with respect to autonomy, there is a distinct jurisprudence to be found in medical law that does not readily correspond to related concepts in other fields. Consequently, this analysis is focussed primarily on the medical or healthcare sphere.

Within that field itself, one must then enquire whether the UK courts have distinguished between applications of autonomy. Put simply, this does not appear to be the case. Rather, the courts have implicitly indicated that autonomy-related norms are developed in a coherent manner across a relatively widely defined sphere of health law. The landmark medical negligence case of *Montgomery v Lanarkshire Health Board* is instructive in this regard, with Lord Kerr and Lord Reed stating:

> Under the stimulus of the Human Rights Act 1998, the courts have become increasingly conscious of the extent to which the common law reflects fundamental values. As Lord Scarman pointed out in Sidaway's case, these include the value of self-determination (see, for example, *S (An Infant) v S* [1972] AC 24, 43 per Lord Reid; *McColl v Strathclyde Regional Council* 1983 SC 225, 241; *Airedale NHS Trust v Bland* [1993] AC 789, 864 per Lord Goff of Chieveley). As well as underlying aspects of the common law, that value also underlies the right to respect for private life protected by article 8 of the European Convention on Human Rights. The resulting duty to involve the patient in decisions relating to her treatment has been recognised in judgments of the European Court of Human Rights, such as *Glass v United Kingdom* (2004) EHRR 341 and *Tysiac v Poland* (2007) 45 EHRR 947, as well as in a number of decisions of courts in the United Kingdom.[502]

500 *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [93].
501 *Re T (adult: refusal of treatment)* [1993] Fam 95, 121.
502 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [8].

Montgomery itself concerned the question of whether the standard of care to be applied in cases of non-disclosure of risks in medical procedures ought to be amended. Yet, when reflecting on the value of self-determination the majority drew on the outlined range of cases that could be termed medical in a very broad sense. The factual issues that these respectively addressed were: the disclosure of risks in a surgical procedure;[503] the power of the court to compel an adult or child to undergo a blood test;[504] the ability of a local authority to add fluoride to the water supply to increase the dental health of the population;[505] the ending of life-sustaining medical treatment;[506] the legitimacy of a hospital imposing a 'do not resuscitate' order on a patient;[507] and the absence of an effective procedure to challenge refusals of medical practitioners to carry out abortions.[508]

A further basis for potential differentiation, which is conspicuously absent in this statement, relates to the source of law. The Supreme Court developed the notion of autonomy not only by reference to the common law, but also by reference to cases involving statute and the *European Convention on Human Rights* (ECHR). Although there may be a difference in degree, the protected autonomy-interest is assumed to be the same in kind across these different sources of law and coherence is sought between them.[509] Within the field of medical law, the Supreme Court consequently appears to reject differentiating between uses of the concept on the basis of such variations of sources or factual circumstances.

We may say that, in terms of the scope of our analysis, the whole area of UK medical law is of potential interest. To avoid superficiality, some selection on the basis of significance for the present purposes (i.e. how autonomy shapes the legal perception and response of AI-threats) will emerge, as it also emerges for other analyses,[510] but behind this process

503 *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871.
504 *S (An Infant) v S* [1972] AC 24.
505 *McColl v Strathclyde Regional Council* 1983 SC 225.
506 *Airedale NHS Trust v Bland* [1993] AC 789.
507 *Glass v United Kingdom* (2004) 39 EHRR 15..
508 *Tysiac v Poland* (2007) 45 EHRR 42.
509 See also: *R (on the application of N) v Mental Health Review Tribunal (Northern Region)* [2005] EWHC 587 (Admin), [2005] ACD 92 [132]. Munby J stated: 'Liberty, autonomy and bodily integrity are interests which traditionally have received a high degree of protection under the common law and are now afforded the added protections conferred by the Convention'.
510 See for example Coggon's, still broad, focus on 'capacity, rationality, life-shortening decisions, advance directives, and the Mental Capacity Act 2005': Coggon, 'Varied

stands the broadly defined field of medical law. This delineates the exploration of the function and content of the autonomy concept.

II. Function

The function of a concept is taken to designate the formal role that it plays within the law. Particularly relevant matters in this regard are: whether it possesses norm-status, its relation to other norms and its position in the legal system more generally. With regard to autonomy, we can begin by building up from the shared understanding that it is a value that motivates and is incorporated throughout many norms of medical law. This follows simply from the hallmark-status of autonomy and individuality in Anglo-American law and society and from its particular relevance to health care, where the choices at play are so significant to the individual patient.[511] Concretely, one can see this reflected in *Montgomery*'s treatment of self-determination above. This was posited as a value that the common law reflects, one that underlies aspects of it and also Article 8 ECHR. This much is taken to be common ground.

Beyond this, however, the consensus begins to breaks down. In this vein, Richards states that autonomy has 'been hailed as either a fundamental right, an important duty or a core principle' but herself argues that it is 'a poorly conceptualised term of art employed by the judiciary to support conclusions that rest on other, more complex and diverse considerations'.[512] In particular, Richards argues that the dominant role of autonomy can be (and should be) reduced to many more specific legal rights, 'such as bodily integrity, freedom from assault, ownership of property and privacy (to name but a few)'.[513]

---

and Principled Understandings of Autonomy in English Law: Justifiable Inconsistency or Blinkered Moralism?' (2007) 15(3) Health Care Analysis: Journal of Health Philosophy and Policy p. 235, 236. Similarly, Maclean's extensive examination of the role of consent in English law and its relationship to autonomy *inter alia* touches on matters of capacity, life-shortening decisions, undue influence and court orders of treatment: Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009).

511 *Natanson v. Kline* (1960) 186 Kan. 393, 406-407; Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 220.

512 Richards in Kirchhoffer and Richards, *Beyond Autonomy* (2019) 17-18.

513 ibid 18.

One can follow this sceptical position up to a point. Indeed, some of the functions ascribed to autonomy have been misleading and one should clarify what the concept is not: it has not straightforwardly been incorporated into English law as a legal right.[514] If one takes a typical definition of common law rights as 'something which an individual possesses and which may be vindicated or protected by the provision of a remedy in the event of infringement',[515] then autonomy is not of this kind. At the very least it is strongly contested whether a claim can be based simply on the fact that a patient's autonomy has been violated – what is missing is precisely the ability to have this vindicated by the provision of a remedy. This is most evident from the fact that the courts have altogether rejected a right to demand specific forms of treatment. That is, to have autonomy respected in a positive way in the healthcare context.[516] To this extent it can hardly be conceived of as a right.

514 For one example of such an ascription see: 'Individuals have a right to make important decisions affecting their lives for themselves': *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [14].

515 Meagher, 'Is There a Common Law Right to Freedom of Speech?' (2019) 43(1) Melbourne University Law Review p. 269, 272. More widely, Cane offers a definition of a right as: 'a primary legal entitlement of one person (C) that attracts a secondary legal obligation of another (D)': Cane in Nolan and Robertson, *Rights and Private Law* (2012) 46. For judicial pronouncements to this effect see: *Ashby v White* (1703) 92 ER 126, 136 ('If the plaintiff has a right, he must of necessity have a means to vindicate and maintain it, and a remedy if he is injured in the exercise or enjoyment of it; and indeed it is a vain thing to imagine a right without a remedy'); *Kingdom of Spain v Christie Manson & Woods Ltd* [1986] 1 WLR 1120 [35] ('In the pragmatic way in which English law has developed, a man's legal rights are in fact those which are protected by a cause of action. It is not in accordance, as I understand it, with the principles of English law to analyse rights as being something separate from the remedy given to the individual').

516 Lord Phillips MR has stated that if the patient 'refuses all of the treatment options offered to him and instead informs the doctor that he wants a form of treatment which the doctor has not offered him, the doctor will, no doubt, discuss that form of treatment with him (assuming that it is a form of treatment known to him) but if the doctor concludes that this treatment is not clinically indicated he is not required (i.e. he is under no legal obligation) to provide it to the patient although he should offer to arrange a second opinion': *R (on the application of Burke) v General Medical Council* [2005] EWCA Civ 1003, [2006] QB 273 [50]. Herring and Wall have also commented on the absence of such a positive obligation and concluded that 'We do not have a right to act autonomously': Herring and Wall, 'The Nature and Significance of the Right to Bodily Integrity' (2017) 76(3) The Cambridge Law Journal p. 566, 584.

Purely negative violations of autonomy are not straightforward bases for a remedy either. The Court of Appeal, in spite of using rights-language, has rejected emphatically 'the proposition that an additional, free standing, award of damages is available for the infringement of the patient's right of autonomy'.[517] This was subsequently affirmed in *Diamond v Royal Devon and Exeter NHS Foundation Trust*, where it was further held that 'there is no self-standing right to claim damages to compensate [the claimant] for the invasion of her right to personal autonomy/choice'.[518] These recent cases highlight how atypical it is to conceive of autonomy as a traditional legal right in medical law.[519]

While conceding this much to those who are sceptical of the legal role attributed to autonomy, the denial of independent-rights-status can hardly be seen as synonymous with a denial of a distinct norm-status in medical law. Dunn and others appear to assume this latter position, referring to a 'tendency for the law to proactively invoke ethical values to shape legal judgements'.[520] However, the law is then understood to leave its understanding of autonomy ill-defined ('important concerns have been expressed about the level of sophistication in the judicial application of' it) and its impact on legal reasoning is seen as much more limited (it 'needs to be deflated').[521] This highlights that, if autonomy is merely conceived as an ethical value, the meaning attributed to it by the law can be left relatively open, and its prominent role in legal reasoning, *inter alia* shaping other norms, becomes suspect.[522]

---

517 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [65].

518 *Diamond v Royal Devon and Exeter NHS Foundation Trust* [2019] EWCA Civ 585, (2019) 170 BMLR 49 [34].

519 In another context the Court of Appeal has also commented that 'we have some doubts whether autonomy and dignity can properly be described as independent common law rights rather than values or principles which inform more specific common law rights, such as the right to bodily integrity and privacy': *R (Nicklinson) v Ministry of Justice* [2012] EWHC 2381 (Admin), (2012) 127 BMLR 197 [50].

520 Dunn and others, 'Between the Reasonable and the Particular: Deflating Autonomy in the Legal Regulation of Informed Consent to Medical Treatment' (2019) 27(2) Health Care Analysis: Journal of Health Philosophy and Policy p. 110, 113.

521 ibid 113. The authors refer specifically to the value's role in shaping the disclosure rule established in *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430. Yet, it is a commitment to viewing autonomy merely as an ethical value that sets the stage for the deflation that this specific evaluation is then taken to exemplify.

522 Leading on from the argument established in: Dworkin, *Taking Rights Seriously* (1987) 37.

It is argued here that a formal role can be carved out for autonomy, one which sits between the minimalist claim that it functions merely as a value or descriptor of specific laws on the one hand and the over-ambitious statement that it is a legal right on the other. In particular, it is argued that this intermediary role is that of a legal principle. As opposed to informative values, principles are binding legal norms themselves, possessing a particular formal structure and characteristic modes of influencing other norms and judicial reasoning.[523] They are norms that play a tangible role in structuring, as well as informing, the law, but their mode of functioning is more differentiated and more flexible than that of a right.

Distinguishing characteristics of principles in the common law include: a (perceived) positive value or importance; a level of generality; a certain way of operating in conflicts with one another and with legal rules; the possible performance of a number of roles that a specific system ascribes to them.[524] The rest of this section argues, with respect to each of these features, that autonomy can be understood to fulfil them.

As already noted, it appears undeniable, even amongst its critics, that autonomy is perceived by legal actors as possessing some positive value, 'weight' or 'importance'.[525] This is so even if the precise nature and relative strength of this value are uncertain.[526] It is consequently presumed that autonomy provides reasons for pursuing a certain course of action.[527]

523 Raz, 'Legal Principles and the Limits of Law' (1972) 81(5) Yale Law Journal p. 823, 826; Eisenberg, *The Nature of the Common Law* (1988) 82.

524 MacCormick highlights especially the former two: 'If I seek to ascertain the principles of a given legal system, I ought to search for those general norms which the functionaries of the system regard as having, on the ground of their generality and positive value, the relevant justificatory and explanatory function in relation to the valid rules of the system': MacCormick, *Legal Reasoning and Legal Theory* (1994) 152-153; The others are taken from: Raz, 'Legal Principles and the Limits of Law' (1972) 81(5) Yale Law Journal p. 823, 832-834, 839-842.

525 Dworkin, *Taking Rights Seriously* (1987) 26-27.

526 Foster is one author who is highly critical of those who believe in the primacy of autonomy in health law: Foster, *Choosing Life, Choosing Death: The Tyranny of Autonomy in Medical Ethics and Law* (2009); Herring has usefully contrasted many of the claims Foster makes with the arguments prevalent in the wider academic debate: Herring, ''Choosing Life, Choosing Death, The Tyranny of Autonomy in Medical Ethics and Law, by Charles Foster'' (2010) 30(2) Legal Studies p. 330.

527 In Raz's terminology one may therefore say that autonomy is seen as a principle of obligation: Raz, 'Legal Principles and the Limits of Law' (1972) 81(5) Yale Law Journal p. 823, 835.

Considering the factor of generality next, by this we mean not only that autonomy may be used as a shorthand to refer to a number of individual, more specific norms.[528] This would indeed indicate that it could be better conceptualised in terms of individual rights and legal mechanisms. Rather, it has been said that principles exist at a level of abstraction from rule-based norms because they prescribe highly unspecific acts, which may include the more specific acts prescribed by various rules, and are justified by more general considerations.[529] This criterion emphasises that a principle imposes a general obligation that is capable of guiding a wide variety of decisions.

That medical law's utilisation of autonomy fulfils this criterion can be seen by revisiting some of the cases touched upon during the delineation of the scope of this investigation. There it was discussed how the principle of self-determination has been invoked to require a wide variety of heterogeneous and relatively specific actions, ranging from the impermissibility of forcing an adult to provide a sample of their blood, to determining the information that a clinician ought to disclose, to suggesting the permissibility of ending life-sustaining treatment. With respect to the relationship of these demands to more specific norms, one can also note that: in the former case, the provisions of an existing rule were affirmed, in the second a contrary rule was overruled and in the final instance a specific act was proscribed, where previously no clear directive had existed. It is in this sense that autonomy is a principle in a formal, logical sense.

One of the most commonly discussed and recognised features of principles is how they operate when they conflict with other laws. Dworkin famously distinguished principles from rules in virtue of the fact that the former have relative weights and can conflict, while not superseding one another.[530] Raz has critiqued and developed this feature by outlining how, at least as a matter of legal policy, 'conflicts between principles are determined by assessing their relative importance together with the consequences for their goals of various courses of action' and conflicts between principles and rules are resolved either in the same manner, or simply on the basis of

528 ibid 828.
529 ibid 839.
530 Dworkin, *Taking Rights Seriously* (1987) 26-27.

the relative importance of the conflicting laws (i.e. without also analysing them in relation to the ensuing consequences).[531]

This kind of functioning is evident with respect to autonomy in UK law. In the medical arena it can, for instance, often be observed to conflict with the principle of beneficence. Particularly where rules are unclear as to a given outcome, the question arises whether one ought to do what is for the perceived clinical benefit of the patient or give primacy to their autonomy. In such cases, the relative importance of the principles, as well as the projected consequences of supporting one over the other, are frequently examined.

One class of examples, which suggests itself, are the judgments that elaborate on the child's ability to consent to medical treatment under established common law principles. In this respect, it has been held that the parents' control over the medical treatment administered to their child is granted on the basis that this is for the benefit of the child.[532] Simultaneously, it was found that this parental right can be disapplied 'if and when the child achieves a sufficient understanding and intelligence to enable him or her to understand fully what is proposed', thereby giving 'a consent valid in law'.[533] In other words, once a child is recognised as having the capacity to make medical decisions for themselves, their autonomy must be respected. This finding was reinforced by a consideration of the consequences that may ensue if the child was not able to make such decisions for themselves.[534]

531 Raz, 'Legal Principles and the Limits of Law' (1972) 81(5) Yale Law Journal p. 823, 833. Cf. MacCormick: 'the principle sets the limits within which judicial decisions fully justified by consequentialist arguments, are legitimate': MacCormick, *Legal Reasoning and Legal Theory* (1994) 161.

532 '[P]arental rights to control a child (…) exist for the benefit of the child and they are justified only in so far as they enable the parent to perform his duties towards the child': *Gillick v West Norfolk and Wisbech Area Health Authority* [1986] AC 112, 170.

533 ibid 188-189.

534 Specifically in this case, which was concerned with the child's ability to consent to contraceptive advice and treatment, Lord Fraser considered that requiring parental control over a child's medical treatment would have significant adverse consequences. It would mean that children would be more reluctant to seek professional advice regarding contraception and thereby expose them to the risks of pregnancy and of sexually transmitted diseases: ibid 173-174. This was echoed in *R (on the application of Axon) v Secretary of State for Health* [2006] EWHC 37 (Admin), [2006] QB 539 [91].

A comparable, partially-consequentialist mode of reasoning is also evident in some cases where autonomy has interacted with specific rules. For instance, when Lord Diplock rejected altering the standard of negligence (the so-called '*Bolam* test') in *Sidaway*, which would have enhanced the protection of patient autonomy,[535] he did so on the basis of at least two factors. On the one hand he recognised the substantial weight that was due to the existing rule:

> For the last quarter of a century the test applied in English law as to whether a doctor has fulfilled his duty of care owed to his patient has been (…) the *Bolam* test. At any rate so far as diagnosis and treatment is concerned, the *Bolam* test has twice received the express approval of this House.
>
> The *Bolam* test is far from new, its value is that it brings up to date and re-expresses in the light of modern conditions in which the art of medicine is now practised, an ancient rule of common law.[536]

But he further went on to reference the detrimental consequences that would follow from a change to this rule, focussing particularly on its potential to 'encourage "defensive medicine" with a vengeance'.[537]

In other instances of principle-rule-interaction the alternative non-consequentialist mode of reasoning is evident. Indeed, in the very same case, the dissenting Lord Scarman preferred this approach, stating that a rule-change could perhaps contribute to a practice of defensive medicine developing, but that 'in matters of civil wrong or tort, courts are concerned with legal principle: if policy problems emerge, they are best left to the legislature'.[538] He consequently focussed exclusively on the relative weights of the two norms, the autonomy principle and the existing rule, and in his opinion the former outweighed the latter.

Lastly, we stumble upon the difficulty that even norms that meet all of the above criteria are not one monolithic category, but can perform a variety of further roles in the law. These include: (1) explaining and justifying existing law[539] (2) instituting norm-change (3) aiding interpretation (4)

535 This was clearly the opinion of Lord Scarman in *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 884-885.
536 ibid 892.
537 ibid 893.
538 ibid 887.
539 MacCormick, *Legal Reasoning and Legal Theory* (1994) 152-153.

generating new norms (5) creating exceptions to rules (6) grounding an action directly.[540] A principle could perform all these roles, but it need not; they are conditional in a way that the other functions are not.[541] In the context of autonomy, a good argument can be made that it performs all of these functions, except that of grounding an action.[542]

It has already been seen how some of these functions have manifested themselves in the medical context. *Montgomery* represents an autonomy-informed overhaul of the rule governing the standard of care in medical negligence[543] and in *S v S* a conception of autonomy justified and explained why the law could not compel an adult to submit to a blood test.[544]

Turning to the three remaining roles in UK law, an example of autonomy aiding interpretation in the medical context stems from the jurisprudence around Article 8 of the ECHR. For this, there is a developing line of jurisprudence which holds that the explicitly granted right to private life implicitly includes a right to personal autonomy, or self-determination. Thus, in *Pretty v the United Kingdom* the European Court of Human Rights stated: 'Although no previous case has established as such any right to self-determination as being contained in Article 8 of the Convention, the Court considers that the notion of personal autonomy is an important principle underlying the interpretation of its guarantees'.[545]

---

540 Raz, 'Legal Principles and the Limits of Law' (1972) 81(5) Yale Law Journal p. 823, 840-842. Eisenberg provides a similar list of functions in his account of common law principles: Eisenberg, *The Nature of the Common Law* (1988) 76-83.

541 For example, both MacCormick and Eisenberg grant that there are only some principles that could (sometimes) have the kind of direct force ascribed by the last function: MacCormick, *Legal Reasoning and Legal Theory* (1994) 178; Eisenberg, *The Nature of the Common Law* (1988) 82.

542 Our above exploration of *Shaw* and *Diamond* suggests that autonomy does not provide the sole ground for an action in English law and we will return to this controversial proposition in Chapter 6. But note that, in contrast with a right where the remedial aspect was an integral definitional element, the status of autonomy as a principle in no way hinges on this one role.

543 For a more general statement of this aspect see Lord Scarman's comment that: 'the mark of the great judge from Coke through Mansfield to our day has been the capacity and the will to search out principle, to discard the detail appropriate (perhaps) to earlier times, and to apply principle in such a way as to satisfy the needs of their own time': *Gillick v West Norfolk and Wisbech Area Health Authority* [1986] AC 112, 183.

544 Or 'personal liberty' as Lord Reid put it: *S (An Infant) v S* [1972] AC 24, 43.

545 *Pretty v the United Kingdom* (2002) 35 EHRR 1 [61]. Domestically one can see this interpretation also at play, for instance in: *R (Tracey) v Cambridge University*

With respect to autonomy's role in norm-generation, the case of *Rees v Darlington Memorial Hospital NHS Trust*, which will be examined more closely in Chapter 6, can be drawn upon. Here a novel conventional award was made for the loss the claimant suffered to her autonomy in a case of wrongful conception. This award fell outside of traditional categories. It was not compensatory for physical pain and suffering, nor for the economic or mental consequences flowing from such suffering, but for the lost 'opportunity to live her life in the way that she wished and planned'.[546] While invocation of autonomy did not overturn any traditional rules on compensation, the court felt obligated to develop the law and create a rule to award £15,000 for the loss of reproductive autonomy in wrongful conception cases. Finally, one can point to the case of *Chester v Afshar* to illustrate the ability of autonomy to create exceptions to existing rules. As Lord Steyn stated, they understood themselves as providing a 'narrow and modest departure' from the normal rule of but-for causation.[547]

In sum, the manner in which autonomy functions in medical law fits well with the notion of a principle. The roles that such a principle will play are unique to its position in the respective legal system and this must be distinguished from the role played by principles in the wider ethical discourse and from the role that it plays in other legal systems. We may expect the principle to inform the aforementioned aspects of legal reasoning, to operate in the outlined manner in cases of conflict and to act as a driver for further developments of the law. All this will become relevant in our analysis of individual mechanisms.

III. Substantive content

Knowing about the formal functions of autonomy will be of little assistance in the following reasoning and argumentation if one cannot give substance to the underlying concept. Knowing *that* autonomy can do something in UK law and *how* it can do it, is of little use without knowing more about the meaning that is attributed to the norm and the nature of reasons it

Hospitals NHS Foundation Trust & Others [2014] EWCA Civ 822, [2015] QB 543 [64].

546 *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309 [8], [123].

547 *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [23]-[24].

provides. This section provides the core of this content, keeping in mind that it cannot be divorced from the functional aspects of the law. Otherwise one would risk failing to account for the unique way in which the legal enterprise transforms philosophical considerations.[548]

One fruitful way to think about this task may be in terms of a 'duality of meaning'.[549] This is a theory developed by Balganesh and Parchomovsky that seeks to give content to common law concepts in a manner that reflects their usage in legal practice and accounts for the common law's propensity to 'to accommodate the process of incremental normative change over time'.[550]

Concepts used in such reasoning have one jural meaning that is oriented towards structural concerns, including the provision of coherence, stability and a functional anchoring in legal practice.[551] And they have one normative meaning that allows them to accommodate external interpretative influences, which can be related to situational goals and which seek to keep the common law in line with developing societal ideals, values and preferences.[552] Given the uncertain, open-textured nature of the jural meaning, one must supplement it with the use of a concept that allows one to apply it to a given context.[553] Legal actors consequently have to exercise a structured discretion in the application context, according to which they must choose to rely on additional factors.[554] These factors are unavoidably normative because the process of interpreting a concept involves the normative judgment 'concept *x* should mean *y*'.[555] This can result in contestable normative meanings being ascribed to one settled jural concept and to interpretive

548 'What the norm or concept requires, in the context of the dispute, and in the context of a wider web of legal standards, is a non-philosophical question. It is a legal question': Wall in Phillips, Campos and Herring, *Philosophical Foundations of Medical Law* (2019) 135.

549 Balganesh and Parchomovsky, 'Structure and Value in the Common Law' (2015) 163(5) University of Pennsylvania Law Review p. 1241, 1255-1265.

550 ibid 1255.

551 ibid 1244.

552 ibid 1243-1244. See also Sunstein for his understanding of 'incompletely theorised agreements', which also appeals to an open-ended structure that relies on external influences for full specification: Sunstein, *Legal Reasoning and Political Conflict* (1998) 35-61.

553 Balganesh and Parchomovsky, 'Structure and Value in the Common Law' (2015) 163(5) University of Pennsylvania Law Review p. 1241, 1262.

554 ibid 1259-1260.

555 ibid 1263.

changes that precede (or do not effect) doctrinal change.[556] At the same time, there is a sense in which there is a dominant or accepted meaning that fits well with established doctrine.

To apply this approach to the autonomy concept, we begin with the understanding that there is a settled jural reference to autonomy and relevant synonyms. This functions in the way that we outlined. Concurrently, as an open-textured jural concept, autonomy has no settled or agreed-upon normative content that is consistently infused into interpretations of it. In UK law this uncertainty is arguably compounded by the relatively recent prominence that the value has assumed, so that references to contested moral notions remain relatively underdetermined and inconsistent. Different conceptions of autonomy are expressly and implicitly articulated by the courts.[557] Indeed, it is conspicuous just how diverging and haphazard direct appeals to philosophical analyses of autonomy are. So that even where they do occur, they can hardly be taken to establish universal rules.[558] There clearly has not been a concerted effort by the legislature or the judiciary to incorporate a high-level, abstract conception of autonomy into the law. Consequently the stage is set for a dual meaning to emerge: the structural, jural one and the contested, wider, normative one.

Now we turn to examining whether the jural concept, and the normative meaning attributed to it, can be aligned with the bioethical theory expounded in the previous chapter. This use must not be uncontested, but it should be sufficiently supported, either explicitly or implicitly, so as to

556 ibid 1274.

557 Coggon has convincingly argued that three different conceptions of autonomy can be found in the positive medical law of England and Wales: ideal desire autonomy, best desire autonomy and current desire autonomy: Coggon, 'Varied and Principled Understandings of Autonomy in English Law' (2007) 15(3) Health Care Analysis: Journal of Health Philosophy and Policy p. 235, 240. Cf. also Herring and Wall who identify autonomy with the thicker (second) notion, which requires one 'to identify standards, preferences and values and to have your own actions and events in your life conform to those standards, satisfy those preferences and realise those values': Herring and Wall, 'The Nature and Significance of the Right to Bodily Integrity' (2017) 76(3) The Cambridge Law Journal p. 566, 575-576.

558 Coggon, 'Varied and Principled Understandings of Autonomy in English Law' (2007) 15(3) Health Care Analysis: Journal of Health Philosophy and Policy p. 235, 235-236. For example, in *Chester* an off-hand reference was made to the definition of autonomy offered by Ronald Dworkin, without attempting to establish it as a wider legal standard or reference point: *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [18].

accord the requisite priority to the legal documentation of the principle's use.[559]

This brings one to our rationalist, procedural notion of autonomy, according to which the law should apply autonomy considerations in a way that recognises: (1) the importance of the theoretical quality of reasoning about beliefs (2) the particular significance of preferences and acceptances that make up the patient's character system, and (3) a variable obligation to promote the positive freedom of individuals by supporting their theoretical and reflective capabilities and by providing them with different classes of information that are necessary to make clinical choices.

A. Rationality

One facet of our account of autonomy stated that the process of decision-making ought to be receptive to the norms of theoretical rationality. This is arguably the most challenging dimension to attribute to the jural autonomy concept, given that English law seemingly eschews a standard of rationality as an indicator of autonomy in the clinical sphere.[560]

The courts have repeatedly stated that there is a need to respect the patients' 'choice to make decisions that others, including the court, might regard as unwise, irrational or harmful to their own interests'.[561] They endeavour to respect the patient's apparent desires without judging on the wisdom of their choices. The force of such considerations can be gleaned from the judiciary's refusal to second-guess decisions that entail even the most serious consequences, such as the risk of grave harm befalling a moth-

559 Robertson treats the priority accorded to legal documentation as a limitation on the use of policy factors in private law reasoning: Robertson in Robertson and Tang, *The Goals of Private Law* (2009) 270. We will return to his further arguments in Section IV. below.

560 Recognised by Pugh himself: Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 183.

561 *Diamond v Royal Devon and Exeter NHS Foundation Trust* [2019] EWCA Civ 585, (2019) 170 BMLR 49 [13]. A similar statement is seen much earlier in *Re T (adult: refusal of treatment)* [1993] Fam 95, 99.

er and her foetus.[562] So it initially appears that the law does not impose conditions on the historical process that led up to an individual's choices.[563]

However, these arguments are misdirected in so far as they are based on the anti-paternalist concern explored in Chapter 3.[564] The courts are expressing the worry that, by imposing a standard of rationality, many individuals will be denied the opportunity to make meaningful choices for themselves.[565] In the form expressed above, this concern posits a much more demanding interpretation of rationality than the one espoused by our adopted theory of autonomy.

It is worth recalling that, under this theory, the individual remained the primary source for the identification of their interests and, even more so, for the balancing of these interests. One should not dismiss the rationality criterion on the basis that it would deny individuals the ability to make choices that are deemed unwise by others. Pugh's theory leaves open the possibility for such disagreements and we will return to the significance of individual value judgments in the assessment of the reflective dimension below.

A more appropriate yardstick by which to measure the autonomy of the patient, is by reference to whether the law has adhered to those fairly minimal aspects touched upon in Chapter 3. That means, the patient's ability to gather appropriate evidence, to weigh it for themselves and to incorporate this into a process of reasoning.

From this perspective it seems clear that, at least in recent times, the courts' view of the patient has shifted to incorporate these abilities. Most

562 Referencing the principle of self-determination, Judge LJ stated: 'how can a forced invasion of a competent adult's body against her will even for the most laudable of motives (the preservation of life) be ordered without irremediably damaging the principle of self-determination?': *St George's Healthcare NHS Trust v S* [1998] 3 WLR 936, 953. This case was also cited with approval by Lady Hale in *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [115].

563 This is most clearly so where an individual has formed a desire to refuse a physical interference with their body: Herring and Wall, 'The Nature and Significance of the Right to Bodily Integrity' (2017) 76(3) The Cambridge Law Journal p. 566, 582.

564 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 199.

565 Purshouse encapsulates this in an analysis of a specific aspect of English law, stating that such an approach: 'would allow the interference of a person's decisions in their best interests (to ensure they comply with an objective rule or to ensure they comply with their own thought-through values). In other words, these definitions of autonomy are indistinguishable from autonomy's polar opposite, paternalism': Purshouse, 'How Should Autonomy Be Defined in Medical Negligence Cases?' (2015) 10(4) Clinical Ethics p. 107, 110-111.

notable in this respect is *Montgomery*, where the Supreme Court objected to a default view of patients as 'medically uninformed and incapable of understanding medical matters' or as 'wholly dependent upon a flow of information from doctors'.[566] Instead, they were viewed as consumers who make choices.[567] For these choices, they gathered 'information about symptoms, investigations, treatment options, risks and side-effects via such media as the internet'.[568] This meant they had to sift through sources of variable quality, identifying reliable ones, and had to then integrate these into their wider reasoning with other sources, including: materials provided by hospitals, patient support groups and, of course, the information received from their doctors.[569]

Patients are consequently seen as individuals with abilities that are not undemanding. Yet, this is not to be equated with ascribing a certain intellect to them or a certain preference.[570] Rather, it posits an ability to engage in the process of medical decision making that approximates closely to the cognitive dimension of autonomy developed in the previous chapter. As will be discussed below, the courts further recognise that realising patient autonomy involves *inter alia* placing a patient in an informational context where they can exercise these capabilities.

B. Individual reflection

Under our normative approach, a process of theoretical reasoning, as well as an awareness of certain foundational facts and values, were deemed prerequisites for any meaningful exercise of autonomy. At the same time, a reflective element was adduced to signify the way in which autonomous decisions must be the patient's own. The patient was able to form commitments that were particularly significant to their self-determination because they were resilient to challenge, long-lasting and cohered with the other elements of their character. These commitments ultimately shape a form of decision-making that is receptive to, and representative of, one's own personal interests.

566 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [76].
567 ibid [75].
568 ibid [76].
569 ibid [76].
570 Contrast the view of Lord Diplock in *Sidaway*, which reserved the disclosure of certain information for 'highly educated men of experience': *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 894-895.

To show that UK law understands autonomy in a way that fits with this dimension, one can point to two related aspects. First, the respect that it accords to the patient's balancing of even fundamental interests. Second, that such respect is especially due where the patient invokes commitments that are deeply held.

Regarding the first element, the law emphasises that patients can align their actions with their goals, even when the highest stakes are involved. For example, as Coggon has stated, a life-saving treatment may be refused under UK law, citing *Re C* and *Ms B*.[571] This is surely beyond doubt.

More pertinently for the recognition of a reflective component to autonomy, we should pause to consider how the law approaches cases where it is in doubt whether patients have the capacity to make such serious decisions. For, they arguably place great emphasis on the fact that the individual has been able to reflect and balance their interests for themselves. This requirement can be found in section 3(1)(c) of the *Mental Capacity Act* (MCA) 2005, enshrining the previous common law position under *Re C*.[572] In effect this means that, for the law to accept a patient's life-or-death decision, it will require that patient to provide some evidence that they are capable of balancing, or that they have in fact balanced, the relevant interests in their care.[573] This targets the law's protection of medical decisions towards a reflective dimension of autonomy.

Coming to the second element, one can identify a trend in the case law to primarily protect long-lasting, defensible preferences. Given that the courts often defer to the patient's current desires, which are arguably assumed

571 Coggon, 'Varied and Principled Understandings of Autonomy in English Law' (2007) 15(3) Health Care Analysis: Journal of Health Philosophy and Policy p. 235, 239.

572 *Re C (adult: refusal of treatment)* [1994] 1 WLR 290.

573 Compare in this respect: *Trust A v H (an adult patient)* [2006] EWHC 1230 (Fam), [2006] 5 WLUK 706 and *Heart of England NHS Foundation Trust v JB* [2014] EWHC 342 (COP), (2014) 137 BMLR. 232. In the former case, the patient's mental impairment meant that she was simply unable to weigh the need for life-saving surgery, which was supported by numerous concerns of hers, with countervailing considerations. In the latter case, the patient merely exhibited a tendency to minimise risks of inaction, which was a normal human way of dealing with a serious medical decision. It did not amount to any incapacity in her ability to weigh information.

to constitute such commitments,[574] the strongest evidence for this can be found where the court protects the individual's own higher-order objectives over their current wishes.

This is reflected in *Re MB*, where a pregnant woman consented to the necessary caesarean section in principle, but refused it in the moment because of her needle-phobia. The court understood actions against the woman's momentary decision, which may be for irrational reasons or based on no reason at all, to constitute an interference with her autonomy.[575] However, it also framed an exception, which was based on a lack of capacity to decide, and applied this to the woman's needle-phobia. This was considered to be so dominant in the woman's thinking that she was not able to make a decision at all.[576] Arguably, this finding was driven by the woman's own deeply desired result: a caesarean section and the safe delivery of her baby.[577] In virtue of this, *Re MB* could be distinguished from a case where a woman declined the treatment and gave no reason at all.[578] It can be argued that the court, while denying the woman's capacity for autonomous decision-making altogether, was merely prioritising her deep commitments – which were entirely coherent with her character and presumably long lasting – over preferences that the woman herself recognised as anomalous and undesirable.

A similar distinction emerges in instances where an external influence has been found to override an individual's will in a particular moment, so that 'consent or refusal of consent may not be a true consent or refusal'.[579] Specifically, in *Re T*, although the patient expressed a refusal of blood transfusions, this was not determinative and could not have been determinative in light of the influence brought to bear on her by her mother.[580] The judges emphasised particularly how the patient's apparent refusal had come 'out of the blue' and how it was influenced by religious beliefs that were not

574 Often this dimension doesn't come through because of the alignment between immediate desires and best desires: Keren-Paz in Barker, Fairweather and Grantham, *Private Law in the 21st Century* (2017) 426-427.

575 *Re MB* [1997] EWCA Civ 3093 [30], [60].

576 ibid [30]-[31].

577 ibid [30], [36]. Although conceding its tangential relevance, the court also referred to the favourable reaction of similarly situated patients after the fact: ibid [31].

578 It is also notable that the court found it necessary to state that 'panic, indecisiveness and irrationality in themselves do not as such amount to incompetence, but they may be symptoms or evidence of incompetence': ibid [30].

579 *Re T (adult: refusal of treatment)* [1993] Fam 95, 122.

580 ibid 110-111.

shown to be an established part of the patient's character.[581] In short, an inability to demonstrate the significance of the patient's choice to her wider belief system once again led to a situation where it was not equated with an autonomous decision.

To a lesser extent, one can also see the described trend in the retrospective causation analysis undertaken in negligence claims for lack of informed consent. As will be discussed at length in Chapter 6, this requires a claimant to show how they personally would have acted, had they been given the relevant information at that time. In this respect, as Turton has noted: 'it is generally best desire autonomy that can be effectively protected rather than current desire autonomy since the patient will have difficulty persuading a court on the balance of probabilities that they would have refused treatment if such a decision cannot be explained by their wider values and priorities'.[582] In medical decisions this has enabled courts to refer to the 'cautious and conservative nature' of patients,[583] their well-documented overwhelming desire to be cured,[584] their non-emphasis of cosmetic factors and their recorded hesitancy to be involved in an experimental, uncertain procedure.[585] These concerns illustrate the relevance of the patient's more general character to the individual decision.

Indeed, at times it appears that the courts are on the verge of departing from an individual assessment, almost conducting a generalised assessment of reasonableness.[586] This would violate the first element of our analysis. which demands that the individual be afforded pride of place in the balancing of their own interests. However, given that no such departure from the subjective assessment has explicitly occurred, the courts' causation analysis

581 ibid 109-112, 118-120.

582 Turton, 'Informed Consent to Medical Treatment Post-Montgomery: Causation and Coincidence' (2019) 27(1) Medical Law Review p. 108, 114.

583 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [28]. In another case it was accepted that a claimant was a cautious person, which was evidenced by their rejection of surgery in the past: *Thefaut v Johnston* [2017] EWHC 497 (QB), [2017] 3 WLUK 328 [84].

584 *C v Colchester Hospital University NHS Foundation Trust* [2018] 2 WLUK 850 [61].

585 *Mills v Oxford University Hospitals NHS Trust* [2019] EWHC 936 (QB), (2019) 170 BMLR 100 [209], [219].

586 *Diamond v Royal Devon and Exeter NHS Foundation Trust* [2019] EWCA Civ 585, (2019) 170 BMLR 49 [9], [19]-[22]; *Thefaut v Johnston* [2017] EWHC 497 (QB), [2017] 3 WLUK 328 [54], [84]; *Ollosson v Lee* [2019] EWHC 784 (QB), [2019] 3 WLUK 562 [158]. Although the distinction is fluid, all these cases still focussed in on the individual position of the patient, as required in *Montgomery*, and thus cannot be taken as an approach that is incompatible with reflective autonomy.

remains focused merely on the wider character of the individual patient. To this extent the common law therefore incorporates the outlined reflective dimension within its protection of the autonomy interest.

Further, it can be argued that UK law pays particular respect to the patient's central commitments because it recognises that interferences with such desires are more significant. For example, in *Rees v Darlington Memorial Hospital NHS Trust*, it was a denial of 'the opportunity to live her life in the way that she wished and planned' – an interference with clearly established, long-term objectives – that weighed particularly heavily with the House of Lords and inspired the exercise in norm generation referred to above.[587] It is further notable that the patient's beliefs and values are considered separately under section 4(6)(b) MCA 2005. The application of this requirement has prompted Pattinson to note that an incapacitated individual's 'longstanding beliefs and values have particular weight' in the determination of their best interest.[588]

In sum, it is undeniable that the UK courts have accommodated a reflective component within their reasoning on patient autonomy. This dimension has manifested itself in numerous ways and across different areas of law.

C. Positive and negative freedom

That UK law values both negative and positive freedom and promotes the practical dimension of autonomy to some degree can be established relatively straightforwardly. With respect to negative freedom, one can refer again to the case of *Re MB*. In this case Butler-Sloss LJ cited authority for the emphatic proposition that: 'A mentally competent patient has an absolute right to refuse to consent to medical treatment for any reason, rational or irrational, or for no reason at all, even where that decision may lead to his or her own death'.[589] Although such a right must be qualified in certain ways, as our analysis of the reflective component illustrates, the

587 *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309 [8].

588 Pattinson, *Medical Law and Ethics* (Sixth Edition 2020) 150, referring to: *Newcastle upon Tyne Hospitals Foundation Trust v LM* [2014] EWHC 454 (COP), (2014) 137 BMLR 226.

589 *Re MB* [1997] EWCA Civ 3093 [17].

fundamental commitment of the courts to the protection of the patient's negative freedom is deeply entrenched.

The legal position with regard to the protection of a patient's positive freedom is, by comparison, much more ambivalent. The courts will neither intervene to force medical professionals to provide a procedure they do not wish to provide,[590] nor will they interfere in resource-allocation decisions that may frustrate a desire for a certain treatment, unless this frustration would also violate some pertinent public law duty.[591]

Crucially for our purposes, a different route has been taken in relation to the facilitation of patient-decision making through information provision. The common law has endeavoured to protect this dimension. *Webster v Burton* highlights the special status of this category. It was presciently stated that the patient 'cannot force her doctor to offer treatment which he or she considers futile or inappropriate. But she is at least entitled to the information which will enable her to take a proper part in that decision'.[592]

More widely, there are a variety of ways in which UK law associates autonomy with conditions on patient understanding. Such a criterion has featured in cases dealing with the capacity for autonomous action,[593] those addressing the validity of consent[594] and those specifying the doctor's standard of care.[595] The law's engagement with this criterion indicates that

590 *An NHS Trust v L* [2013] EWHC 4313(Fam), (2014) 137 BMLR 141 [78]; *Portsmouth NHS Trust v W* [2005] EWHC 2293 (Fam), [2005] 4 All ER 1325 [32]-[36], citing: *Re J (a minor) (child in care: medical treatment)* [1993] Fam 15, 26-27.

591 The following cases all dealt with issues of resource allocation and duties to render positive assistance to patients in light of Article 8 of the European Convention on Human Rights: *R v North West Lancashire Health Authority ex p A* [2000] 1 WLR 977; *R (Condliff) v North Staffordshire Primary Care Trust* [2011] EWCA Civ 910 (Admin), [2011] 4 WLUK 189; *McDonald v Kensington and Chelsea Royal Borough of Kensington and Chelsea* [2011] UKSC 33, [2011] 4 All ER 881.

592 *Webster v Burton Hospitals NHS Foundation Trust* [2017] EWCA Civ 62, (2017) 154 BMLR [97].

593 For example, in *Gillick* Lord Scarman made 'the child's right to make his own decision' dependent upon reaching 'a sufficient understanding and intelligence to be capable of making up his own mind on the matter requiring decision': *Gillick v West Norfolk and Wisbech Area Health Authority* [1986] AC 112, 186.

594 See for example: *R v Melin*, where reference was made to 'what was known and understood by the complainants concerned': *R v Melin* [2019] EWCA Crim 557, [2019] QB 1063 [33].

595 'The patient is entitled to consider and reject the recommended treatment and for that purpose to understand the doctor's advice and the possibility of harm resulting from the treatment': *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 904.

if a patient lacks a relevant aspect of understanding regarding a medical decision and, more especially, if they were not furnished with adequate information and opportunity to gain such an understanding, then there is a potential violation of autonomy.

In particular, the law insists on medical professionals generating opportunities for understanding. In the case of *Al Hamwi v Johnston* the judge rejected as too onerous the notion 'that the clinician's duty is to ensure that the information given to the patient is understood'.[596] Rather, the duty arising in negligence required only that '[c]linicians should take reasonable and appropriate steps to satisfy themselves that the patient has understood the information which has been provided'.[597] Similarly, in cases dealing with the reality of consent, the issue is often framed as the doctor having sufficiently informed the patient of relevant aspects.[598] Perhaps this feature is influenced by the nature of the legal process, where – as was made clear in *Al Hamwi* – the obligations of the defendant are one of the most significant concerns. However, it also appears to be an acceptable specification of a procedural conception of positive freedom, which does not seek to guarantee understanding, but rather to facilitate informed decision making.

In addition, it is conspicuous how much the courts have, more recently, emphasised aspects relating to the process of information disclosure, including especially those of deliberation and dialogue. In *Bell v Tavistock and Portman NHS Foundation Trust* the Divisional Court mentioned repeatedly the level of information that was provided for a controversial and sensitive form of treatment, as well as the nature of the discursive and iterative dialogue through which it was conveyed.[599] In *Montgomery* too it was stated

596 *Al Hamwi v Johnston* [2005] EWHC 206 (QB), [2005] Lloyd's Rep Med 309 [69].

597 ibid [69]. Cf. also *Smith v Tunbridge Wells Health Authority* [1994] 5 Med LR 334, 339: 'When recommending a particular type of surgery or treatment, the doctor, when warning of the risks, must take reasonable care to ensure that his explanation of the risks is intelligible to his particular patient. The doctor should use language, simple but not misleading, which the doctor perceives from what knowledge and acquaintanceship that he may have of the patient (which may be slight), will be understood by the patient so that the patient can make an informed decision as to whether or not to consent to the recommended surgery or treatment'.

598 *Chatterton v Gerson* [1981] QB 432, 443.

599 *Bell v Tavistock and Portman NHS Foundation Trust* [2020] EWHC 3274 (Admin), (2021) 177 BMLR 115 [37], [39], [98]. These findings were not criticised in the successful appeal: *Bell v Tavistock and Portman NHS Foundation Trust* [2021] EWCA Civ 1363, [2022] 1 All ER 416 [22].

that 'the doctor's advisory role involves dialogue'[600] and Lord Kerr and Reed cited with approval guidance of the General Medical Council that required doctors to '[w]ork in partnership with patients. Listen to, and respond to, their concerns and preferences. Give patients the information they want or need in a way they can understand'.[601] It was also noted that such understanding would not be furthered 'by bombarding the patient with technical information which she cannot reasonably be expected to grasp'.[602] UK law therefore places certain standards on the facilitation of patient decision making, including at least a process of cooperative reasoning, dialogue and adequately focussed disclosure.

Finally, one must consider the requirement that autonomous decisions are based on certain beliefs that are accurate representations of the world. As will be explored in more depth in Chapter 6, the courts appear to implicitly establish a hierarchy of information, according to which the most significant beliefs of the patient must actually be true.[603] This includes an accurate understanding of the broad nature and purpose of their treatment.[604] Similarly, certain motivations of the professional are seen as highly relevant for an individual's understanding, covering at least financial and sexual motivations for the acts performed.[605] Below these considerations, there lies the information that a doctor should provide about certain risks and consequences of a given procedure The judiciary has clearly stated that it regards this information as an important facet of a patient's decision, but it does not afford it the same kind or level of protection as the two other categories.[606]

600 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [90].
601 ibid [77].
602 ibid [90].
603 Although not relating to information disclosure, the significance of holding certain basic beliefs is also clearly evident in the law dealing with the mental capacity to consent to treatment. See: *Re MB* [1997] EWCA Civ 3093 [3]; *Trust A v H (an adult patient)* [2006] EWHC 1230 (Fam), [2006] 5 WLUK 706 [23].
604 *Chatterton v Gerson* [1981] QB 432, 443. Pugh himself draws this distinction in the English legal context: Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 165.
605 See *Appleton v Garrett* (1997) 34 BMLR 23 and *R v Williams* [1923] 1 KB 340 respectively.
606 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430. Note in this regard, how *Montgomery* has expanded the number of factors that are deemed significant even within this one category. As will be examined in Chapter 6, under the previous standard a doctor generally needed only to disclose those risks that one responsible body of medical opinion would deem necessary.

IV. Limitations

To realistically assess how the principle of autonomy influences the application of specific legal mechanisms to the problems posed by the use of AI, one must also be candid about the limitations involved in the operation of this principle. This is necessary in spite of the suggestions by some academics that considerations of autonomy do, or ought to, predominate in medical law's reasoning. For instance, Heywood and Miola have stated that:

> The approach taken by the Supreme Court [in *Montgomery v Lanarkshire Health Board*] is the same as that adopted by the House of Lords in *Chester v Afshar* and indeed the High Court of Australia in *Rogers v Whitaker*. This entails identifying the purpose of the law as being the protection of autonomy and assessing the current legal regime to see if it meets that aim. In all three of these cases a majority of judges found that it did not and so modified the law adequately to protect autonomy (…) Future courts should look at the approach of the House of Lords and now Supreme Court, and consider cases in the same way: does the current law adequately protect patient autonomy and, if not, what needs to be changed to allow it to do so?[607]

There are two significant limitations that such an approach seems to ignore, but which are deemed important to recognise in subsequent chapters. On the one hand, we have elaborated on the fact that legal principles are liable to conflict with other principles and values. The more general demands of the law are not necessarily coherent and commensurate and cannot be brought under just one head, such as autonomy.

In particular, as we move to consider individual mechanisms, the normative demands of these rules cannot be discounted. If a rule has enough weight, or if the consequences of departing from it would be sufficiently dire, then the best legal argument may be that it ought to be upheld, even if this is to the chagrin of the autonomy principle. This is especially true in the private law context, where the relevant informed consent obligations will be seen to be overwhelmingly located. Here it has been noted that a tendency to use wider societal and political reasoning is restricted by

607 Heywood and Miola, 'The Changing Face of Pre-operative Medical Disclosure: Placing the Patient at the Heart of the Matter' (2017) 133((Apr)) Law Quarterly Review p. 296, 320.

institutional factors. These encompass: a need to give considerable priority to legal documentary forms, a convention of placing greater emphasis on consistency and doctrinal stability and the recognised need of doing justice to both parties.[608]

It is surely correct that if AI threatens the specified aspects of a patient's autonomy, especially in a wide-ranging or grave manner, then this provides an impetus for examining the legal situation in light of the underlying principle. But the limits imposed by the nature of specific rules and other principles remain. These conflicts must be assessed in each individual scenario, framed at least in part in terms of the considerations and arguments offered here.

V. Conclusion

In conclusion, this chapter has maintained that UK medical law incorporates a conception of autonomy that bears a multitude of forceful affinities to the outlined procedural approach. This represents one cogent normative specification of an open-ended jural concept and enables autonomy to be used as a legal principle in the following analysis of specific legal mechanisms. The autonomy principle demands that an autonomous decision-making process has taken place. This includes the exercise of basic rational, cognitive capabilities, as well as a reflection of one's own, personal commitments. It also means that the patient must be protected from outside interferences and have their decision-making ability facilitated by relevant actors – especially through the provision of information in the medical context. Lastly, it must also be acknowledged that this principle is not the sole goal of the law. A principle will conflict with different norms and with consequentialist forms of reasoning and its application in individual circumstances must always be tempered by the limitations that these considerations legitimately impose.

608 Robertson in Robertson and Tang, *The Goals of Private Law* (2009) 269-279.

Chapter 5: Autonomy in U.S. law

Within the United States considerations of patient autonomy have a long-established history, both at the state and the federal level. Through the judicial reshaping of its inherited doctrines, the U.S. was perhaps the first common law jurisdiction to develop a legally significant understanding of patient autonomy in medical decision making. To unpick the meaning that can be attributed to this understanding, this chapter follows the mode of analysis adopted in the British context, determining: within what confines the autonomy concept ought to be considered, what function and substantive content can be attributed to it and what limiting factors deserve particular attention.[609]

I. Scope

The first point of call is to circumscribe the boundaries within which the relevant understanding of autonomy is to be found in the positive legal order of the U.S. and, specifically, California. When doing this one cannot ignore the fact that '[t]he law governing American health care arises from an unruly mix of state and federal agencies and from a jumble of statutes and common law doctrines conceived, in the main, without medical care in mind'.[610] Such fragmentation of healthcare law raises some difficulties in demarcating the area of inquiry; difficulties that are compounded by the lack of a single institution that, *via* its treatment of the autonomy concept, could provide an authoritative statement on the relevance of these different legal materials.[611] Questions touching on patient self-determination arise

609 As in the English context, I do not restrict my analysis based on any particular terminology, given that terms such as 'autonomy', 'self-determination' and 'bodily integrity' are often used interchangeably to denote similar substantive concerns. See for example: *Planned Parenthood of Southeastern Pennsylvania v. Casey* (1992) 505 U.S. 833, 857 and *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1188-1189. It is the substance of these concerns that directs the enquiry.

610 Bloche, 'The Invention of Health Law' (2003) 91(1) California Law Review p. 247, 249-250.

611 A similar fragmentation existed in England and Wales. Yet we saw there how the United Kingdom's Supreme Court signalled the relevance of disparate components

against the backdrop of a federal system that encompasses over 50 jurisdictions with divided institutional competencies on pertinent constitutional, statutory and common law issues.

A. Jurisdictional scope

Due to the fragmented nature of the United States' common law, a selection was made in the introduction to focus on the tort law of California. Yet, this does not mean that the operation of a legal concept, which informs more specific mechanisms, must be restricted to an analysis of one state alone. It is a well-established feature of legal reasoning in the U.S. that general principles are laid down corresponding to 'the general average of the results reached by the courts of various states'.[612] In this way the judge or scholar 'aims at finding the best solution of a problem on the footing of examples from many jurisdictions [and they] will tend to concentrate on recent trends'.[613] If autonomy is a 'general principle' in these senses, then our analysis should engage with the more disparate legal landscape of the U.S.

This does not require a comprehensive, detailed accounting of each jurisdiction, however. Rather, prominent trends in federal and cross-state case law will be relevant to some aspects of our analysis. Most especially to the determination of the conceptual scope of autonomy, which delineates the outer boundary of a broad consensus, and to the function of the autonomy concept, which is the dimension that draws on the aforementioned commonality in legal reasoning. In so far as the specific substance

---

by using them to shape the construction of its understanding of autonomy. By contrast, the United States Supreme Court stated *Cruzan by Cruzan v. Director, Missouri Dept. of Health* (1990) 497 U.S. 261, 277-278: 'State courts have available to them for decision a number of sources – state constitutions, statutes, and common law – which are not available to us. In this Court, the question is simply and starkly whether the United States Constitution prohibits Missouri from choosing the rule of decision which it did'. This usefully highlights the formal position in America, which prevents the Supreme Court from performing a similar signalling role, but it will also be seen *infra* how the legal reasoning of the Supreme Court is indirectly influenced by the materials that were formally dismissed in *Cruzan*.

612 Goodhart, 'Case Law in England and America' (1930) 15(2) Cornell Law Review p. 173. Note that here I do not use the term 'principle' in the sense of a legal norm, which was elaborated in the previous chapter and will be drawn upon again below.

613 Cross, *Precedent in English Law* (Third Edition 1979) 18. This type of reasoning will be elaborated upon when considering autonomy's function *infra*.

of the autonomy principle and the limitations that are imposed on it are concerned, the focus must, first and foremost, be on California's law, which may offer its own, distinct specifications for each.

Beginning then with the question of whether autonomy is a concept for which a coherent approach is sought across the jurisdictional boundaries of U.S. law – spanning different state and federal courts – this must be affirmed beyond doubt.[614] In relation to individual autonomy, one might even say that the approaches of different legal actors are particularly interwoven.[615] This is evidenced by the U.S. Supreme Court's early recognition of the common law's formative role in the value's conception. So that in 1891, when considering whether a plaintiff seeking damages could be forced to undergo a surgical examination, it was found:

> No right is held more sacred, or is more carefully guarded by the common law, than the right of every individual to the possession and control of his own person, free from all restraint or interference of others, unless by clear and unquestionable authority of law.[616]

614 For some examples of scholars approaching the topic in this way: Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219; Appelbaum, Lidz and Meisel, *Informed consent: Legal theory and clinical practice* (Second Edition 2001) pt II; Donnelly, *Healthcare Decision-Making and the Law: Autonomy, Capacity and the Limits of Liberalism* (2010) chapter 2; Faden, King and Beauchamp, *A History and Theory of Informed Consent* (1986) chapter 2.

615 'Constitutional development is often broadly rooted in common law principles; in [medical privacy cases] the public law is somewhat in advance of the private': Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 277. In a non-medical context, Bloustein has argued forcefully for the recognition of a connection between tortious and constitutional understandings of privacy, stating that 'there is a common thread of principle and an identical interest or social value which runs through the tort cases as well as the other forms of legal protection of privacy': Bloustein in Schoeman, *Philosophical Dimensions of Privacy* (2009) 181.

616 *Union Pac. R. Co. v. Botsford* (1891) 141 U.S. 250, 251-253. Following Sunstein, one may see the common law as functioning as a baseline for constitutional reasoning in this area. Whereby 'courts used common law principles to define the judicial role in public law cases', e.g. 'finding coercion only in cases of government infringement of common law rights': Sunstein, *After the Rights Revolution: Reconceiving the Regulatory State* (1993) 210-220. For a wider application of this theory to the modern Supreme Court see especially: Shell, 'Contracts in the Modern Supreme Court' (1993) 81(2) California Law Review p. 431 and Farber and Frickey, 'In the Shadow of the Legislature: The Common Law in the Age of the New Public Law' (1991) 89(4) Michigan Law Review p. 875.

This guiding role of the common law in constitutional reasoning has become more prominent as the court has been obliged to deal with the more nuanced aspects of medical decision making. For example, when *Cruzan v. Director, Missouri Department of Health* considered the compatibility of restrictions on end-of-life decision making with the Due Process Clause of the 14th Amendment, several judges dwelt extensively on the common law's respect for self-determination and its requirements of informed consent.[617] And again, in a very different situation in *NIFLA v. Becerra* – considering the imposition of disclosure obligations on primary care clinics – all members of the Court explicitly recognised the legitimacy of imposing informed consent requirements on health care providers where this served to facilitate medical decision-making and patient self-determination.[618] The majority appeared particularly content to defer to the common law in this area. They highlighted its long-established role, before rejecting the manner in which California sought to extend the scope of informed consent to justify its statute.[619]

Ample instances can also be found where state courts have referenced the importance attributed to autonomy in the federal and state constitutions, whether or not these were directly in point. The New Jersey Supreme court in *Matter of Conroy* prefaced its purely common law analysis of the right to decline life-sustaining treatment, with an exposition of the Constitutional privacy right, as elaborated in *Griswold v. Connecticut* and *Roe v. Wade*.[620] Faced with a similar case the California Court of Appeal held that '[t]he right of a competent adult patient to refuse medical treatment has its origins in the constitutional right of privacy', citing both the state and the federal constitution, but only after having already inferred similar rights from the State's common law and statute.[621]

---

617 For example, Chief Justice Rehnquist, delivering the opinion of the Court: *Cruzan by Cruzan v. Director, Missouri Dept. of Health* (1990) 497 U.S. 261, 267-278. See also Justice Brennan, with whom Justice Marshall and Justice Blackmun joined, dissenting: ibid 305-306.

618 *National Institute of Family and Life Advocates v. Becerra* (2018) 138 S.Ct. 2361. See Justice Thomas delivering the opinion of the Court: ibid 2373-2375. See also: Justice Breyer for the dissenting justices: ibid 2385-2386.

619 ibid 2373-2374. In part this rejection seemed to turn on the fact that the disclosure fell outside the situations where disclosure had traditionally been mandated. See also the minorities' criticism of such reasoning: ibid 2386.

620 *Matter of Conroy* (1985) 98 N.J. 321, 348.

621 *Bartling v. Superior Court* (1984) 163 Cal.App.3d 186, 195. See also: *Thor v. Superior Court* where the California Supreme Court supported its proposition that 'the right

Perhaps the most significant impact of the constitutional recognition of autonomy interests on common law reasoning has been seen in the realm of so-called wrongful birth claims.[622] These are negligence actions where parents seek damages after they have not been properly informed about, or there has been a misdiagnosis of, a foetus' condition that would have caused the parents not to continue the pregnancy to birth. The indirect and yet substantial influence of *Roe v. Wade*[623] on the recognition of these purely private law claims, prior to its recent overturning in *Dobbs v. Jackson Women's Health Organization*,[624] has been noted by courts and commentators alike.[625] Whereas 'public policy' militated against the recognition of this cause of action before *Roe*,[626] courts considering wrongful birth claims after that decision argued that it brought about a shift in their background considerations, favouring the facilitation and protection of women's autonomy.[627]

---

to refuse medical treatment is equally "basic and fundamental" and integral to the concept of informed consent' *inter alia* by reference to the aforementioned U.S. Supreme Court case of *Cruzan*: *Thor v. Superior Court* (1993) 5 Cal.4th 725, 735-736.

622 One can also compare the impact of *Griswold v. Connecticut* (1965) 381 U.S. 479 on the legal recognition of wrongful conception claims: *Custodio v. Bauer* (1967) 251 Cal.App.2d 303, 317-318; *Troppi v. Scarf* (1971) 31 Mich.App. 240, 253-254.

623 *Roe v. Wade* (1973) 410 U.S. 113.

624 *Dobbs v. Jackson Women's Health Organization* (2022) 597 U.S. 215.

625 In addition to the cases *infra*, see: Haqq, 'The Impact of Roe on Prenatal Tort Litigation: On the Public Policy of Unexpected Children' (2020) 13(1) Journal of Tort Law p. 81; Harris, 'Statutory Prohibitions on Wrongful Birth Claims & Their Dangerous Effects on Parents' (2014) 34(2) Boston College Journal of Law & Social Justice p. 365, 374; Gold, 'An Equality Approach to Wrongful Birth Statutes' (1996) 65(3) Fordham Law Review p. 1005, 1015.

626 *Gleitman v. Cosgrove* (1967) 49 N.J. 22.

627 Whether this view of *Roe v. Wade* was ever correct is another question. The court in *Roe* stated: 'The decision vindicates the right of the physician to administer medical treatment according to his professional judgment up to the point where important state interests provide compelling justifications for intervention. Up to those points, the abortion decision in all its aspects is inherently, and primarily, a medical decision, and basic responsibility for it must rest with the physician': *Roe v. Wade* (1973) 410 U.S. 113, 165-166. For a critical analysis see *Daly, 'Reconsidering Abortion Law: Liberty, Equality, and the New Rhetoric of Planned Parenthood v. Casey' (1995) 45(1) American University Law Review p. 77*. Daly argues that *Roe* and its progeny afforded primacy over the abortion decision to the medical professional, not the woman. Siegel and Greenhouse have made a similar point referring to *Roe*: 'the Court explained and justified its holding in language that depicted doctors as the responsible and authoritative decisionmakers, with women as patients subject

In *Robak v. U.S.* it was mused that 'State courts have been quick to accept wrongful birth as a cause of action since *Roe v. Wade*, because it is not a significant departure from previous tort law'[628] and, significantly, because a similarity to other medical malpractice actions was not marred by 'political and moral questions concerning abortions [that] the Supreme Court has already settled'.[629] In *Smith v. Cote* the Supreme Court of New Hampshire conceded that 'we believe that *Roe* is controlling; we do not hold that our decision would be the same in its absence', adding later: 'we are bound by the law that protects a woman's right to choose to terminate her pregnancy'.[630] In California the Supreme court – while not appealing directly to *Roe* – cited with approval such out-of-state actions (including, but not limited to, *Robak*) and apparently conceded the validity of wrongful birth actions on the strength of their authority and on the basis of two rulings of the Court of Appeal.[631] In this manner, the primacy that the U.S. Supreme Court afforded to the ability to make abortion decisions has clearly permeated state courts' reasoning in accepting a novel type of negligence claim, even if it is difficult to see how a concrete constitutional argument could have been constructed to require its creation or maintenance.[632]

This leads one to the final aspect of the intertwined nature of autonomy-related reasoning in the U.S. Namely, that in many cases dealing with the

---

to their guidance (…) the Court figured the doctor as the agent responsible for abortion decisions and the criteria guiding those decisions as medical': Siegel and Greenhouse, *Before Roe v. Wade: Voices that Shaped the Abortion Debate Before the Supreme Court's Ruling* (2010) 255. This made *Roe* a surprising basis for mandating private law causes of actions against physicians who are performing their constitutionally mandated role: controlling the abortion decision.

628 *Robak v. U.S.* (7th Cir. 1981) 658 F.2d 471, 476.

629 ibid 476.

630 *Smith v. Cote* (1986) 128 N.H. 231, 239-242.

631 *Turpin v. Sortini* (1982) 31 Cal.3d 220, 225-227. This approach was described and affirmed in: *Foy v. Greenblott* (1983) 141 Cal.App.3d 1, 8.

632 Arguments to this effect were rejected in: *Hickman v. Group Health Plan, Inc.* (Minn. 1986) 396 N.W.2d 10, 13-15; *Edmonds by James v. Western Pennsylvania Hosp. Radiology Associates of Western Pennsylvania P.C.* (Pa. Super. Ct. 1992) 414 Pa.Super. 567, 575-576; *Dansby v. Thomas Jefferson University Hosp.* (Pa. Super. Ct. 1993) 424 Pa.Super. 549, 553-555. See also Kelley's analysis, which emphasises 'Refusal to recognize a cause of action for wrongful birth (…) does not constitute governmental interference with the woman's right to contraception or abortion. The Constitution does not require state courts to grant tort recovery for private interference with the exercise of constitutional rights': Kelley, 'Wrongful Life, Wrongful Birth, and Justice in Tort Law' (1979) Fall(4) Washington University Law Quarterly p. 919, 959.

concept, courts applying state law have also drawn generously upon the judicial pronouncements of their sister jurisdictions to determine what patient autonomy means and requires. Some decisions that are frequently cited, and often cited together, in the area of informed consent law span the following states: *Schloendorff v. Society of New York Hospital* (a decision of the Court of Appeals of New York Court),[633] *Salgo v. Leland* (The California Court of Appeal for the First District),[634] *Canterbury v. Spence* (a federal decision, applying the law of the District of Columbia),[635] *Natanson v. Kline* (the Supreme Court of Kansas)[636] and *Cobbs v. Grant* (the Supreme Court of California).[637]

A comparable trend is also identifiable in specific applications of autonomy arguments, such as the right to die. In this field Meisel has elaborated how the decisions of the Supreme Court of New Jersey, including *Matter of Quinlan,*[638] *In re Conroy*[639] and *Matter of Jobes,*[640] have played an important role in shaping a consensus in this area of the law, which holds across States and which has been reinforced by the United States Supreme Court's influence.[641] At the same time, prominent decisions from

633 *Schloendorff v. Society of New York Hospital* (1914) 211 N.Y. 125.
634 *Salgo v. Leland Stanford Jr. University Bd. of Trustees* (1957) 154 Cal.App.2d 560.
635 *Canterbury v. Spence* (D.C. Cir. 1972) 464 F.2d 772.
636 *Natanson v. Kline* (1960) 186 Kan. 393.
637 *Arato v. Avedon* provides a straightforward example of such reasoning, citing all of the aforementioned cases: *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1183, fn. 5.
638 *Matter of Quinlan* (1976) 70 N.J. 10.
639 *Matter of Conroy* (1985) 98 N.J. 321.
640 *Matter of Jobes* (1987) 108 N.J. 394.
641 Meisel, 'The Right to Die: A Case Study in American Lawmaking' (1996) 3(1) European Journal of Health Law p. 49. Meisel elaborates on courts' and lawyers' process of decision-making: 'Almost without exception, when deciding its first right-to-die case, each court has felt compelled to recount the development of the law in other states even though the law of other states is not binding precedent on courts outside that state' and '[lawyers advising clients] try to obtain a composite picture of the law across the US to determine if there is any uniformity or at least a majority position': ibid 59-60. He also traces the influence of these cases on other jurisdictions, such as Pennsylvania, ibid 61. Lastly, regarding the role of the United States Supreme Court he comments: 'The role played by the United States Supreme Court in the development of the right to die in American law has been dramatic, but late in coming. By the time the Court entered the arena, the consensus was very well developed. However, the action by the Supreme Court has helped to solidify and ultimately extend the reach of the consensus": ibid 63.

other states – such as *In Re Gardner*,[642] *Satz v. Perlmutter*,[643] *Superintendent of Belchertown v. Saikewicz*[644] and *Rasmussen v. Fleming*[645] – have also contributed materially to the legal understanding of autonomy in this context.[646]

All in all, given these interrelations, it appears that an exposition of legal reasoning with the autonomy concept is not restricted to one state jurisdiction, but can (and partially must) be situated in this wider context, to give a fair accounting of its meaning and significance.

B. Conceptual scope

The remainder of this section establishes the outer bounds of this broad consensus by illustrating how the U.S. courts have considered autonomy in connexion with medical decision making.

As a rule, this has not proceeded by explicitly delineating autonomy as a health law concept. Much the opposite: both state and federal courts have been more open than their transatlantic counterparts in associating patient autonomy with understandings of agency and liberty drawn from other areas of law. For instance, in an interesting point of contrast with the British position, American courts and commentators have relied substantially on the duties owed in fiduciary and commercial relationships to supplement considerations of individual agency and to define a physician's informational responsibilities.[647] Similarly, the freedom afforded to patients

642 *In re Gardner* (Me. 1987) 534 A.2d 947.

643 *Satz v. Perlmutter* (Fla. 1980) 379 So.2d 359.

644 *Superintendent of Belchertown State School v. Saikewicz* (1977) 373 Mass. 728.

645 *Rasmussen by Mitchell v. Fleming* (1987) 154 Ariz. 207.

646 *Thor v. Superior Court* refers to these and many more cases, interspersing them with analyses of the already mentioned informed consent case law, to reach such conclusions as 'Because health care decisions intrinsically concern one's subjective sense of well-being, this right of personal autonomy does not turn on the wisdom, i.e., medical rationality, of the individual's choice': *Thor v. Superior Court* (1993) 5 Cal.4th 725, 734-737. In *Robak v. U.S.* the 7th Circuit of the U.S. Court of Appeals stated that 'In the absence of any direct precedent from the state involved, a federal court applying state law should consider decisions of its sister states on the same issue': *Robak v. U.S.* (7th Cir. 1981) 658 F.2d 471, 475.

647 *Bowman v. McPheeters* (1947) 77 Cal.App.2d 795, 800-801; *Berkey v. Anderson* (1969) 1 Cal.App.3d 790, 804-805; *Moore v. Regents of University of California* (1990) 51 Cal.3d 120, 128-134. Cf. also *National Institute of Family and Life Advocates v. Becerra* (2018) 138 S.Ct. 2361, 2374, where the majority placed physician's informa-

in their healthcare decisions is often described merely as one aspect of a more general liberty interest or even supplanted in specific circumstances by other rights, such as the right to make family decisions.[648]

Going forward, the impacts of such broader doctrines will be considered when they impinge on the role of an independent conception of medical autonomy. But there is no indication that they eclipse the possibility of an independent notion of patient autonomy in the first place. Two less conspicuous modes of reasoning in the American courts betray their recognition of the unique normative demands that the health law background places on their assessments of patient autonomy.

The first type of reasoning consists of the courts' reference to the distinct nature of medical decision making, which then *implicitly* shapes the operation of patient autonomy. Both state and federal courts have proved sensitive to the unique circumstances, moral weight and personal significance of medical decisions and this has shaped their legal consideration of patient self-determination. The Californian Supreme Court has had occasion to be particularly explicit in this respect, emphasising that '[a]lthough an aspect of personal autonomy, the conditions for the exercise of the patient's right of self-decision presuppose a therapeutic focus'[649] and elsewhere '[medical ethics] is a necessary component and complement of [patient autonomy] and should serve to enhance rather than constrict the individual's ability to resolve a medical decision in his or her best overall interests'.[650] Clearly,

---

tional duties squarely in the context of other professional relationships. Schultz has stated that 'The duty to disclose for purposes of informed consent is a specific instance of such a fiduciary duty, yet, for several reasons, a generic fiduciary duty to disclose sometimes more effectively vindicates patient interests in autonomy than do narrower duties that have crystallised under ordinary rules of medical consent'. This exemplifies Schultz's analysis of fiduciary obligations as both one influence on the realisation of patient choice in other causes of action and as a cause of action, furthering this end, in its own right: Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 259-264.

648 See the opinion of the U.S. Supreme Court in *Planned Parenthood of Southeastern Pennsylvania v. Casey* (1992) 505 U.S. 833, 884: 'The doctor-patient relation does not underlie or override the two more general rights under which the abortion right is justified: the right to make family decisions and the right to physical autonomy'.

649 *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1188-1189. To be noted are also the connection to moral ideals drawn in the decision and the emphasis of the 'overall medical context': ibid 1185.

650 *Thor v. Superior Court* (1993) 5 Cal.4th 725, 743. Further expressions of this view, and concomitant practical consequences, were also expressed: 'Given the well- and long-established legal and philosophical underpinnings of the principle of

the Court believed that the doctor's obligations to ensure the patient's informed consent is concerned with a particular conception of an individual's autonomy, which is primarily at stake in healthcare decision making, and it explicitly recognised that medical ethics have a positive role to play in shaping the legal understanding of that conception.

Such reasoning is not unique to California. The Supreme Court of Maine identified the special importance of the concept of personal autonomy 'in the realm of medical care' in *In Re Gardner*, rooting it in the common law doctrine of informed consent, as well as in political and bioethical analyses of patient refusal to continue life-sustaining care.[651] In the New Jersey case of *Hummel v. Reiss* the dissenting Justice Handler insisted that the relevant clinical decision 'clearly involved not only complicated medical judgements about the condition of both [the mother] and the fetus, but also difficult moral questions involved in weighing the value of the mother's life against the potential life of the child. (...) Because the decision involved profound moral and personal issues, it was one that only [the mother] could make'.[652]

A similar tenor emerges from many voices in the U.S. Supreme Court.[653] In *Cruzan*, the dissenting Justice Brennan, picked up on the personal and moral dimensions of the medical decision-making process: '[t]he right to be free from unwanted medical attention is a right to evaluate the potential benefit of treatment and its possible consequences according to one's own

---

self-determination, as well as the broad consensus that it fully embraces all aspect of medical decisionmaking by the competent adult, we conclude as a general proposition that a physician has no duty to treat an individual who declines medical intervention' ibid 738. See also *Conservatorship of Drabick*, which stated: 'Under California law, however, human beings are not the passive subjects of medical technology. The line of decisions beginning with *Cobbs v. Grant* and continuing with *Barber*, *Bartling*, and *Bouvia* compel this conclusion. These cases recognize that medical care decisions must be guided by the individual patient's interests and values. Allowing persons to determine their own medical treatment is an important way in which society respects persons as individuals': *Conservatorship of Drabick* (1988) 200 Cal.App.3d 185, 208.

651 *In Re Gardner* (Me. 1987) 534 A.2d 947, 950-951.

652 *Hummel v. Reiss* (1992) 129 N.J. 118.

653 It is only mentioned indirectly in the text *infra* but *Roe v. Wade* provided some support for an understanding of a broad medical autonomy interest. In this respect, Daly provides an analysis of how the women in that case and its progeny were treated primarily as medical patients, rather than as people with more complex individual and social circumstances: Daly, 'Reconsidering Abortion Law' (1995) 45(1) American University Law Review p. 77. This was bolstered by an expansive definition of 'health' in the companion case of *Doe v. Bolton*: ibid 88.

values and to make a personal decision whether to subject oneself to the intrusion'.[654] The challenged rule was seen as particularly objectionable, due to the fact that it allowed for the transformation of 'human beings into passive subjects of medical technology'.[655] Justice O'Connor (giving a concurring judgment) also referred both to the patient's 'deeply personal decision to reject medical treatment', and their 'interest in directing [their] medical care', implicitly distinguishing it from the 'freedom of personal choice in matters of … family life'.[656] In this manner, the Justices in *Cruzan* emphasised the distinctive way in which medical decisions could be highly personal, engage an individual's values and render them vulnerable to atypical forms of external interference.

Even the U.S. Supreme Court majority in *Planned Parenthood v. Casey* – which preferred to rest its principal, now overturned, holding on a familial conception of liberty – repeatedly integrated *Roe v. Wade*'s relation to medical treatment into its reasoning.[657] Initially it was critically observed that the case may be read 'as a rule (whether or not mistaken) of personal autonomy and bodily integrity with doctrinal affinity to cases recognising limits on governmental power to mandate medical treatment or to bar

654 *Cruzan by Cruzan v. Director, Missouri Dept. of Health* (1990) 497 U.S. 261, 309-310.

655 ibid 325. The Justices were also explicit that there can be no distinction between different types of medical treatment: ibid 306-307. Cf. *Thor v. Superior Court* finding much the same: 'both courts and commentators generally reject attempts to draw distinctions between, for example, "ordinary" and "extraordinary" procedures, or "terminal" and "nonterminal" conditions, or "withholding" and "withdrawing" life-sustaining treatment. (…) Rather, effectuating the patient's freedom of choice remains the ultimate arbiter': *Thor v. Superior Court* (1993) 5 Cal.4th 725, 736-737.

656 *Cruzan by Cruzan v. Director, Missouri Dept. of Health* (1990) 497 U.S. 261, 289-291, citing *Cleveland Bd. of Educ. v. LaFleur* (1974) 414 U.S. 632, 639-640. This distinction can be gleaned from the fact that Justice O'Connor found that this freedom could 'also' (i.e. in addition to the patient's interest in directing his medical care) be in play where proxies, who are often family members, take decisions for the patient: *Cruzan by Cruzan v. Director, Missouri Dept. of Health* (1990) 497 U.S. 261, 289-291.

657 Although the *Dobbs* majority overruled *Roe* and *Casey* because it deemed abortion to be 'sharply' distinguishable on the basis of the critical moral question involved, it referenced the relevant constitutional case law against enforced medical procedures without apparent disapproval. Nor would a medical autonomy concept possess the 'high level of generality' that the majority apparently found objectionable: *Dobbs v. Jackson Women's Health Organization* (2022) 597 U.S. 215, 256-257. Citing the same case law, the dissenting opinion is more explicit in championing constitutional protections of bodily integrity and restrictions on the government's ability to interfere with medical decisions: ibid 379.

its rejection'.[658] In spite of this negative introduction, the majority could not avoid associating *Roe* with the right to terminate medical treatment in its broader analysis.[659] They recognised that a doctor's informational requirements in assisting the abortion decision were, as far as the Constitution was concerned, set by reference to the general requirements imposed on clinical procedures.[660] Justice Blackmun (concurring in part) likewise categorised abortions as relating primarily to decisions of 'reproduction and family planning'.[661] Nevertheless, his analysis of women's right to bodily integrity focused on analogies with surgical interventions and emphasised the connection between pregnancy, its health consequences and risks – factors that feature prominently in the law of informed consent and which frame abortion as a matter of medical decision making.[662] Although *Casey* rejected the medical context (and especially the doctor-patient relationship) as *the* basis for a woman's freedom to end her pregnancy, it was still recognised as an important facet of that freedom, making legally significant contributions.

Such a partial invocation of medical autonomy was also evident in the case of *NIFLA v. Becerra*. Here the Supreme Court adjudged upon a regulation of the California legislature that sought to promote the decision-making capabilities of, especially low-income, women by inter alia requiring the disclosure of publicly funded abortion services to them. This focused the Court's reasoning on the relationship between healthcare provider and patient, with Justice Thomas adopting with approval the quotation from *Wollschlaeger v. Governor of Florida* that '[d]octors help patients make

658 Ultimately it is clear that there was a preference for the right being categorised as the familial type, see: *Planned Parenthood of Southeastern Pennsylvania v. Casey* (1992) 505 U.S. 833, 884.

659 ibid 858-859.

660 ibid 859, citing *inter alia*: *Matter of Quinlan* (1976) 70 N.J. 10.

661 *Planned Parenthood of Southeastern Pennsylvania v. Casey* (1992) 505 U.S. 833, 927-928.

662 ibid 926-927. He also noted: 'Just as the Due Process Clause protects the deeply personal decision of the individual to refuse medical treatment, it also must protect the deeply personal decision to obtain medical treatment, including a woman's decision to terminate a pregnancy', positing a broader understanding of medical autonomy, which includes a positive dimension: ibid 927, fn. 3. But see also Spindelman's criticism of such reasoning: Spindelman, 'Are the Similarities between a Woman's Right to Choose an Abortion and the Alleged Right to Assisted Suicide Really Compelling' (1996) 29(3) University of Michigan Journal of Law Reform p. 775, 814, fn. 151.

deeply personal decisions and their candor is crucial'.[663] Indeed the full quotation from *Wollschlaeger* demonstrates the distinct significance that the Court has attributed to the facilitation of autonomous healthcare decisions in a free speech context:

> Health-related information is more important than most topics because it affects matters of life and death. Doctors help patients make deeply personal decisions, and their candor is crucial. If anything, the doctor-patient relationship provides more justification for free speech, not less.[664]

However, the outcome of *Becerra* – an indication that California's notice requirement would be subject to, and not survive, strict scrutiny – suggests that the facilitation of personal decisions by the patients was not a decisive influence on the majority's reasoning.[665]

The dissenting minority's approach therefore also warrants consideration. Their support of the notice requirement, as well as their analysis, arguably represented a more sincere attempt to accommodate the normative demands of health care and the patient-doctor interaction. Justice Breyer relied on the Court's distinct jurisprudence on the regulation of the medical profession[666] and the bioethical concerns raised by the case (speech involving 'health, differing moral values and differing points of view').[667] Moreover, he went on to explicitly critique the majority's reliance on a narrow conception of 'medical procedure' as lacking 'moral, practical, and legal force', given the need to obtain the patients' informed consent and the 'health considerations' that would favour disclosure in this instance.[668]

These cases showcase the acknowledgements of state and federal courts that, in the final analysis, the demands of individual autonomy take a

663 Although he then appears to categorise this as a broader aspect of professional relations: *National Institute of Family and Life Advocates v. Becerra* (2018) 138 S.Ct. 2361, 2374-2375.

664 *Wollschlaeger v. Governor, Florida* (11th Cir. 2017) 848 F.3d 1293, 1328. This is difficult to reconcile with the manner in which the majority in *Becerra* then appeared to categorise the patient-doctor interaction as a normal aspect of professional relations: ibid 2374-2375.

665 Montanez, 'Pregnant and Scared: How NIFLA v. Becerra Avoids Protecting Women's Reproductive Autonomy' (2019) 56(3) San Diego Law Review p. 829, 849-851.

666 *National Institute of Family and Life Advocates v. Becerra* (2018) 138 S.Ct. 2361, 2382-2383.

667 ibid 2383.

668 ibid 2386.

unique shape where individuals make healthcare decisions. This is so regardless of whether the concept is partially subsumed under more abstract or related interests.

The second type of consideration reinforces this impression by focussing on the courts' selection of precedent for their analogical reasoning. This reveals a tendency to ascertain the value of autonomy by reference to a class of cases dealing with medical care and medical decision making, even if this is not made explicit and sometimes even at the expense of circumventing purported limitations on their reasoning.

A telling illustration of this is provided by the U.S. Supreme Court's approach in *Gonzales v. Carhart*. Here the majority relied on the balancing between personal autonomy and state interest that was undertaken *Washington v. Glucksberg* [669] (an assisted suicide case) to judge on the legality of a restriction on abortion.[670] It has been argued that such an analogy was possible primarily because both scenarios involved 'a medical procedure that necessarily implicates a life interest, though arguably different life interests, and uncertainty'[671] and, one might hasten to add, therefore also involves fundamental questions of patient autonomy. On the strength of this analogy the Supreme Court upheld a regulation that would have been much more problematic from the perspective of a narrower consideration of the Court's abortion case law or a broader doctrinal lens that focused on the demands of the right of privacy, considered under a standard of strict scrutiny.[672]

More generally, numerous examples of federal and state cases have already been adduced that draw analogies ranging from cases of informed consent, to those dealing with the refusal of medical treatment and those focussing on reproductive freedom.[673] These forms of legal argumentation

669 *Washington v. Glucksberg* (1997) 521 U.S. 702.

670 *Gonzales v. Carhart* (2007) 550 U.S. 124, 156-160.

671 Coyle, 'Gonzales v. Carhart: Justice Kennedy at the Intersection of Life Interests, Medical Practice and Government Regulations Comment' (2008) 27(2) Temple Journal of Science, Technology & Environmental Law p. 291, 309.

672 ibid 311-313. See also the dissenting Justices Ginsburg, Stevens, Souter and Breyer in *Gonzales v. Carhart* (2007) 550 U.S. 124, 171-174.

673 For example, Meisel has stated that informed consent law 'does not automatically decide right-to-die cases. However, courts have generally held that the law of informed consent stands for the proposition and reflects the principle that competent patients have a right of self-determination in medical matters': Meisel, 'The Right to Die: A Case Study in American Lawmaking' (1996) 3(1) European Journal of Health Law p. 49, 60.

highlight a unique understanding of autonomy *in medical decision making*. Within these bounds it seems possible to identify a coherent and relatively rich concept that can serve as a normative standard according to which the law's application to medical AI should be measured.

II. Function

With the UK background in mind, the function of autonomy in American law can be arrived at *via* a less circuitous route than was pursued there. The American courts' use of terminology with respect to self-determination is broadly similar to that employed in the UK, with it being described as a right,[674] a value,[675] and a principle.[676]

Taking these in order, it is relatively clear that autonomy cannot be seen as an independent right afforded to individual patients. In spite of the occasional rhetorical flourish, it has been correctly stated that within American law 'autonomy has never been recognized as a legally protectable interest'.[677] Broadly this insight can be based on a similar mode of reasoning as was utilised in the previous chapter: autonomy is only vindicated partially and imperfectly through remedies aimed primarily at related interests.[678] For the U.S. this reasoning can be bolstered further by considering the rights-landscape into which autonomy interests had to be integrated, which evinces a strong antipathy towards positive rights.[679]

674 *Canterbury v. Spence* (D.C. Cir. 1972) 464 F.2d 772, 784.

675 *In re Gardner* (Me. 1987) 534 A.2d 947, 950.

676 There are 'well- and long-established legal and philosophical underpinnings of the principle of self-determination': *Thor v. Superior Court* (1993) 5 Cal.4th 725, 738.

677 Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 219.

678 ibid 276. See also the discussion *infra* regarding the ability of the autonomy principle to ground an action directly.

679 ibid 277; see also: Currie, 'Positive and Negative Constitutional Rights' (1986) 53(3) The University of Chicago Law Review p. 864. The latter also places the constitutional jurisprudence in a wider common law context: ibid 866-867, fn. 13. Writing specifically in relation to *Cruzan*, Spindelman has further noted that this case 'did not hold that a competent individual has a common law or constitutional right to obtain any form of medical treatment she wishes. Therefore, in light of the limitations lower courts have placed on the right to informed consent, there is no basis for the position that anyone, including the terminally ill, has a common law right, much less a Fourteenth Amendment liberty interest, to receive medical assistance from a doctor to commit suicide. Any claim to the contrary would constitute a vast

In California specifically, this difficulty is exemplified by the fact that the autonomy interests under the constitutional right to privacy are stated negatively: 'interests in making intimate personal decisions or conducting personal activities without observation, intrusion, or interference'. Furthermore, even these negative interests are restricted, as they protect '*certain intimate and personal decisions* from government interference' and they do not 'create any *unbridled* right of personal freedom of action that may be vindicated in lawsuits against either government agencies or private persons or entities'.[680]

Conversely, it is almost banal to claim that patient autonomy is a value that has influenced the American legal system. The view of the concept as something of value is embodied in numerous legislative and judicial pronouncements that endeavour to translate its moral ideal into concrete norms.[681] Yet, a purely extra-legal, policy-based form of reasoning is hardly a satisfactory explanation for the considerable influence that we have already begun to see the autonomy concept exert upon U.S. law.

This naturally leaves us to consider again whether autonomy could be conceived of as a legal principle. Recall in this regard that, under the theories adopted in the previous chapter, a principle is identifiable by reference to a number of factors, including a (perceived) positive value and a level of generality. On the basis of the case law already examined, both of these criteria appear to be straightforwardly fulfilled. Autonomy has shaped multifaceted areas of the law and evidently is considered an important goal to pursue. What has not been sufficiently established, is the concept's manner of interacting with other norms and with forms of consequentialist

---

departure from the Supreme Court's past jurisprudence that has never held that a competent individual has an affirmative constitutional right to obtain medical treatment': Spindelman, 'Are the Similarities between a Woman's Right to Choose an Abortion and the Alleged Right to Assisted Suicide Really Compelling' (1996) 29(3) University of Michigan Journal of Law Reform p. 775, 813-814.

680 *Hill v. National Collegiate Athletic Assn.* (1994) 7 Cal.4th 1, 35-36 (emphasis added).

681 See e.g. *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1184-1185, commenting on the interaction between the moral ideal of patient autonomy and the derived legal standards. See also: California Probate Code section 4650, subdivision (b): 'Modern medical technology has made possible the artificial prolongation of human life beyond natural limits. In the interest of *protecting individual autonomy*, this prolongation of the process of dying for a person for whom continued health care does not improve the prognosis for recovery may violate patient dignity and cause unnecessary pain and suffering, while providing nothing medically necessary or beneficial to the person' (emphasis added).

reasoning, as well as the typical functions that these interactions perform within the legal system.

That the interest in autonomy interacts with other principles is well-illustrated by the constitutional jurisprudence that, as the previous section has shown, utilises the common law's understanding of this concept. Here the courts have consistently referred to the need to evaluate patient autonomy by reference to other fundamental interests – above all the 'protection of life'.[682] Autonomy protection is balanced against other deeply held, legally recognisable commitments. Given the close relationship to the constitutional right to privacy, and the analysis demanded there – which considers a particular facet of the interest and entails a deference to state justification[683] – this line of jurisprudence can also be understood to stand for the proposition that the realisation of autonomy must sometimes yield to rule-specific considerations.

To examine a less constrained relationship with conflicting principles, one should further consider the application of autonomy alongside other general norms of the common law. Such an application can be identified, for example, in the way the courts have carved out an emergency exception to the necessity of obtaining patient consent. They have limited the protection afforded to patient autonomy to the extent that there is a sufficiently serious danger to the patient's health and life.[684] By balancing the principles of beneficence and autonomy, they have carved out an exception that is a matter of degree, requiring an assessment of the patient's interest in making their own decision and the gravity of the danger that must be addressed.[685]

682 '[T]his Court's post-Roe decisions accord with Roe's view that a State's interest in the protection of life falls short of justifying any plenary override of individual liberty claims': *Planned Parenthood of Southeastern Pennsylvania v. Casey* (1992) 505 U.S. 833, 835.

683 *People v. Privitera* (1979) 23 Cal.3d 697, 703-705.

684 *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1268.

685 In *Burchell v. Faculty Physicians & Surgeons of Loma Linda University School of Medicine* the gravity of the risk to the patient did not qualify the procedure as an emergency intervention – 'it would be a stretch to characterize a "low risk" associated with taking a more conservative approach, or "speculation" about possible risks, as evidence of an emergency, requiring the surgeon to act despite a lack of express consent Jury was within reason to find that there was no life-or health-threatening situation that justified Barker's decision to perform an operation substantially beyond the scope of Burchell's express consent': *Burchell v. Faculty Physicians & Surgeons of Loma Linda University School of Medicine* (2020) 54 Cal.App.5th 515, 525.

A blanket approach, whereby the autonomy interest of a class of individuals is uniformly judged to be outweighed by a principle of beneficence, can also be found in the common law. Specifically in the rule that minors are not generally able to consent to medical treatment.[686] The purpose of this rule 'was to protect the health and welfare of minors, safeguarding them from the potential overreaching of third parties or the improvidence of their own immature decisionmaking'.[687] Across these precedents one can therefore see that patient autonomy has been weighted alongside other legal principles to reach a concrete determination.

Similarly, the requisite kind of interaction between autonomy and consequentialist reasoning is well-supported by precedent. A seminal Californian case, *Cobbs v. Grant*, exemplified this aspect in considering whether battery or negligence ought to be the legal mechanism by which the common law protects a patient's interest in informed decision making. It was held that:

> Although this is a close question, either prong of which is supportable by authority, the trend appears to be towards categorizing failure to obtain informed consent as negligence. That this result now appears with growing frequency is of more than academic interest; it reflects an appreciation of the several significant consequences of favoring negligence over a battery theory[688]

Inter alia these consequences included the availability of punitive damages under a battery cause of action, which may not be covered by a physician's malpractice insurance, and factors implicitly related to the propagation of an unjustified number of claims – including the longer limitation period under the tort of battery and the ease with which the other requirements of the requisite action could be fulfilled.[689]

Moreover, where these consequential concerns have been deemed inapplicable, or to lack the same force, the courts have been prepared to opt for the stronger protection of patient autonomy under battery.[690] Similarly, where the autonomy violation at stake is perceived to be more important, this also supports a stronger form of protection, regardless of the

686 *Bonner v. Moran* (D.C. Cir. 1941) 126 F.2d 121, 122-123.
687 *American Academy of Pediatrics v. Lungren* (1997) 16 Cal.4th 307, 314-315.
688 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 240.
689 ibid 240.
690 *Rains v. Superior Court* (1984) 150 Cal.App.3d 933, 942.

consequential considerations.[691] In sum, the Californian courts exhibit a pattern of reasoning with the autonomy concept that is consistent with its classification as a legal principle. Its influence is dependent on its relative weight *vis-à-vis* countervailing considerations that can be found in principle and/or in consequentialist argumentation.

With this established, one can turn to examine the additional, hallmark roles that principles can play within jurisdictions. Namely, justification of existing norms, instituting norm change, aiding interpretation, generating new norms, creating exceptions to rules and grounding actions directly. Before exemplifying autonomy's ability to function in most of these capacities, it is worth dealing briefly with the last function. For, it has already been considered that autonomy is not conceived of as a right in California and the U.S. Rather, the protection of the concept has been achieved through its insertion into established causes of action and, in Chapter 7, it will be seen how the requirements of such actions limit any ability to claim for autonomy violations *per se*. Let us therefore focus on the remaining roles.

Regarding the function of justifying existing law, *Foy v. Greenblott* illustrates how patient self-determination has been drawn upon to perform just this role. California's Supreme Court considered, among other things, the extent to which a mental health facility was obligated to interfere in the reproductive decisions of an incompetent patient to prevent the birth of a child (in essence a type of wrongful birth claim). This obligation was suggested to entail a supervision of sexual contacts and the prescription of contraceptives, if necessary overriding the patient's wishes.[692] In deciding against the imposition of such a duty, the court referred to a range of statutes and case law, which were held to:

> express a public policy of maximizing patients' individual autonomy, reproductive choice, and rights of informed consent. Within the considerable range of discretion left to them, mental health professionals are expected to opt for the treatments and conditions of confinement least restrictive of patients' personal liberties. The threat of tort liability for insufficient vigilance in policing patients' sexual conduct and in second-guessing their reproductive decisions would effectively reverse these incentives and encourage mental hospitals to accord patients only

691 *Stewart v. Superior Court* (2017) 16 Cal.App.5th 87, 105-106.
692 *Foy v. Greenblott* (1983) 141 Cal.App.3d 1, 9-10.

> their minimum legal rights. Consequently, these aspects of respondents' conduct are not actionable.[693]

The *Foy* court thereby rejected a use of the autonomy concept to add to, or alter, existing norms in this instance, leading to an imposition of tort liability. However, it stated clearly that the legitimacy of a range of sources derived from their approximation to this ideal.[694]

Unlike in *Foy*, the dimensions of norm generation and alteration have played a significant role in many other contexts. One of these contexts is the development of the doctrine of valid consent and informed consent that will be discussed in-depth in Chapter 7. As will be examined there, the judiciary has been more than prepared to change existing standards to meet the demands of patient autonomy. Whether this takes the form of an innovation in the breach analysis of negligence – supplementing the reasonable defendant standard with the reasonable patient standard – or the imposition of a legal presumption regarding the medical professional's intention in battery – that substantial deviations from patient consent were intentional.

Accompanying the concept's use in rule creation, there has naturally also been a significant reliance on patient autonomy to interpret these novel norms. In *Truman v. Thomas* the Supreme court determined whether disclosure obligations applied to a situation that was not directly covered by precedent: where a procedure had been refused, rather than consented to. By appealing to patient autonomy – specifically to the fact that the significance of the information to patient decision making was the same in both scenarios[695] – it was held that the leading case (*Cobbs v. Grant*) clearly required the physician to advise the patient also in this situation.[696]

A similar role for the value can be found in statutory interpretation. For example, in *Daum v. Spinecare Medical Group, Inc.* the court paired an analysis of statutory disclosure requirements with the common law's analysis of a 'patient's right of self-decision' in order to determine the standard for the form and content of disclosure regarding the use of an

693 ibid 11-12.

694 Consider also *Barber v. Superior Court* (1983) 147 Cal.App.3d 1006, 1015-1016. Here the general right of a patient to control their medical treatment justified, but also exceeded, statutory provisions dealing with end-of-life treatment.

695 *Truman v. Thomas* (1980) 27 Cal.3d 285, 292-293.

696 ibid 292-293.

investigational device.[697] Such cases are indicative of the prominent role that autonomy has been able to play in the process of norm creation, alteration and interpretation in Californian law.[698]

Lastly, it remains to be shown that the concept is utilised to generate exceptions to rules. *Ballard v. Anderson* demonstrates such usage, but also that it is not always easy to distinguish from autonomy's role as an interpretative aid. The court determined that a minor's consent to a therapeutic abortion fell under a statute allowing for her to consent to 'medical and surgical care related to her pregnancy'.[699] This rule already had a restricted application, concerning 'a minor of any age, only if she is pregnant and unmarried and only for medical, hospital and surgical care related to her pregnancy'.[700] Yet, after asserting that a therapeutic abortion fell within the requirements of this rule, the court also found 'an additional limitation implicit in each of the medical emancipation statutes: the minor must be of sufficient maturity to give an informed consent to any treatment procedure'.[701] The Californian Supreme Court rested this insight on general considerations stemming from the informed consent doctrine and especially highlighted the minor's possession of a requisite understanding, which will be seen to constitute an important substantive element of California's autonomy principle. Ultimately, it can be noted that, once *Ballard* had interpreted a statutory provision to cover a certain situation, it then created an exception to this application by reference to characteristics that stemmed not from the rule itself, but rather from a concern for patient autonomy.

In summation, it is argued that the operationalisation of the autonomy concept in the U.S. and California fits well with a conceptualisation of it as a legal principle. It is neither just an abstract non-legal value, nor a concrete, assertable right. It is a general norm to which the law attributes importance, it has a certain mode of interacting with other norms and consequentialist forms of reasoning and it has a set of typical roles that the Californian common law has attributed to it.

697 *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285, 1303-1305.
698 See also: *Bouvia v. Superior Court* (1986) 179 Cal.App.3d 1127, 1139-1140. Holding that the right to control medical care, specifically to refuse life-sustaining treatment, was wider than the one granted by statute – which was limited to a certain group of persons.
699 *Ballard v. Anderson* (1971) 4 Cal.3d 873, 882-883.
700 ibid 882-883.
701 ibid 883.

III. Substantive content

We now turn to the meaning that, specifically California, has imbued patient autonomy with. Given the emphasis in the previous section on the legally determined function of the concept, it is important to give credence to its legal nature and to the role that the law has in the provision of a definition. Without more, one cannot simply take over the bioethical theory espoused in Chapter 3. Faced with the same challenge, we will turn to the same theory as was relied upon in the British analysis. This called for a distinction between autonomy's jural and its normative meaning.[702] That such a distinction is apposite emerges from the above analysis. It touched upon a number of jural reference points for autonomy and its analogues, but it did not indicate that these derive from an alignment with a particular philosophical position.[703]

As a consequence, our task is not to prove that the law has included any high-level, abstract theory within its precepts. Rather, it is to identify support in the legal material for the argument that the identified bioethical theory provides one viable, useful means of conceptualising its approach to patient autonomy. Again, it bears emphasising that this conceptualisation need not be uncontested, but it should be sufficiently strongly entrenched to play the requisite guiding role. Moreover, this also implies that the normative meaning of autonomy should provide sufficiently concrete indicators for the courts. Pugh's conceptualisation broadly fulfils these requirements. It lays down standards that find considerable support in the law and are concrete enough to provide guidance, while also leaving space for jural considerations to influence their specification.

The standards to be assessed in this way are: (1) a cognitive element, identifying autonomy by reference to a patient's engagement with rationality and reasons in their decision making (2) a reflective element, giving pride of place to the individual's own values, especially their long-established, defensible commitments (3) a practical element, referencing the positive and negative dimensions of freedom that shape an individual's ability to act – including a recognition that the patient's autonomy must be facilitated and that a certain kind of disconnect with the true state of

702 Balganesh and Parchomovsky, 'Structure and Value in the Common Law' (2015) 163(5) University of Pennsylvania Law Review p. 1241.

703 In California specifically, the reliance on such positions is extremely limited and haphazard. This is exemplified in *Thor v. Superior Court* (1993) 5 Cal.4th 725, 734-735; *People v. Privitera* (1979) 23 Cal.3d 697, 729, fn. 8.

affairs (a failure to hold decisionally necessary beliefs) is a particularly grave violation of autonomy.

A. Rationality

The first standard of Pugh's theory identifies autonomous decision making with deliberation that follows the norms of theoretical rationality, at least to a minimal extent. This is arguably the most controversial aspect of his theory when considered from the perspective of Californian judicial pronouncements on medical decision making. As in the UK, the courts have rejected an explicit rationality requirement in their recognition of a patient's capacity to make autonomous decisions.

In this vein, Justice Arabian stated in *Thor v. Superior Court* that, since 'health care decisions intrinsically concern one's subjective sense of well-being, this right of personal autonomy does not turn on the wisdom, i.e., medical rationality, of the individual's choice'.[704] And, later in the same case, the Californian Supreme Court rejected a judicial intervention to assess a prisoner's capacity to make a rational choice because this 'tends to denigrate the principle of personal autonomy, substituting a species of legal paternalism for the medical paternalism the concept of informed consent seeks to eschew. "Rationality" is for the patient to determine'.[705]

Writing extra-judicially, Justice Arabian expanded upon this holding, explaining that requiring a rational decision-making ability of the patient, in addition to comprehension, would constitute a drastic and severe invasion into their personal autonomy.[706] Such statements are a paradigmatic example of the anti-paternalistic opposition to an association between autonomy and rationality.[707] Once this association is made, it is believed that it will give licence to a denial of decisional autonomy, effectively allowing professionals to overrule it.

However, the Californian manifestation of this concern relies on factors that appear to be, more properly, subsumable under the reflective component of decisional autonomy. The ability of an individual to pursue a

704 *Thor v. Superior Court* (1993) 5 Cal.4th 725, 736-737.
705 ibid 747-748.
706 Arabian, 'Informed Consent: From the Ambivalence of Arato to the Thunder of Thor' (1994) 10(3) Issues in Law & Medicine p. 261, 287.
707 Pugh, *Autonomy, Rationality, and Contemporary Bioethics* (2020) 183.

subjective sense of well-being is not precluded by a relation of autonomous decisions to minimal standards of theoretical rationality – including the ability of inductive reasoning or a reliance on evidence in the formation of one's beliefs. This is the extent of Pugh's particular conception of rationality. *Pace* the above statement, such matters are not for the patient to determine and Justice Arabian recognised this. In the aforementioned peace of extra-judicial writing, he went on to state 'if the illness of a severely disturbed patient precludes his making a rational decision regarding treatment using only the information a reasonable person would require, then this level of information will not be sufficient to effectuate individual autonomy'.[708] An ability to use and reason with information, in a manner that allows one to direct one's decisions accordingly, is an element of autonomy.

Consequently, while the courts clearly reject a thick conception of rationality, which would allow professionals to disregard personal commitments and idiosyncrasies (a matter we must return to in a moment in our consideration of the reflective component), this does not entail a commitment to the view that there is no theoretical component to decisional autonomy. Other statements indicate that just such a commitment has shaped the courts' standard understanding of medical decision making. For instance, in *Moore v. Regents University of California* the Supreme Court mused that:

> medical treatment decisions are made on the basis of proportionality—weighing the benefits to the patient against the risks to the patient. As another court has said, "the determination as to whether the burdens of treatment are worth enduring for any individual patient depends upon the facts unique in each case," and "the patient's interests and desires are the key ingredients of the decision-making process."[709]

Framed in these terms, there is no opposition between the patient's subjective assessment of their interest and an insistence on a (fairly minimal) rational process of decision making. The former is to be realised through the latter.

In other instances the courts have not been content to require only a modest degree of theoretical rationality in reasoning, but have made determinations that approximate to the more demanding rationality re-

708 Arabian, 'Informed Consent: From the Ambivalence of Arato to the Thunder of Thor' (1994) 10(3) Issues in Law & Medicine p. 261, 287.

709 *Moore v. Regents of University of California* (1990) 51 Cal.3d 120, 130; citing: *Barber v. Superior Court* (1983) 147 Cal.App.3d 1006, 1018-1019.

quirement rejected in *Thor*. For example, in the already discussed case of *Ballard v. Anderson* the court created the requirement that a minor must possess a degree of understanding and maturity to be deemed capable of making the requisite autonomous decision.[710] More starkly still, in *Stewart v. Superior Court* the Court of Appeal found it necessary to clarify that the decision maker there was 'not an uneducated patient objecting to a procedure without explanation'.[711] Rather, they were a registered nurse who was able to point to alternative explanations for a diagnosis (engaging in inductive reasoning) and wished to receive a second opinion (seeking out further evidence on which to base a decision).[712]

In short, the aspects of decision making to which *Stewart* afforded particular respect have a strong affinity to the theoretical rationality standard. It thereby provides further support for the argument of this section. Yet, the court also did more than this. By referring to the patient's education and the professionally informed explanation provided for the decision, it intimated that a patient's choice would be granted more deference if it is of a certain intellectual calibre and/or aligns with an accepted body of opinion. To see why this particular pronouncement goes too far, and does not provide a basis for generalisation,[713] it is now important to assess the reflective dimension of autonomy.

B. Individual reflection

The previous section showcased that one overriding concern in judicial appeals to self-determination was the ability of an individual to strike their own path – to determine what is valuable and direct their decisions accordingly. Under our theoretical approach this intuition was expanded upon. While an acceptance of certain basic values and desires was deemed necessary for any meaningful exercise of autonomy, it was essential to respect an

710 *Ballard v. Anderson* (1971) 4 Cal.3d 873, 883.

711 *Stewart v. Superior Court* (2017) 16 Cal.App.5th 87, 105-106. Similarly *Bouvia v. Superior Court* referenced the fact that the patient 'is intelligent, very mentally competent' and that they 'earned a college degree': *Bouvia v. Superior Court* (1986) 179 Cal.App.3d 1127, 1136.

712 *Stewart v. Superior Court* (2017) 16 Cal.App.5th 87, 105-106.

713 See in this regard both: *Barber v. Superior Court* (1983) 147 Cal.App.3d 1006, 1015 and *Bartling v. Superior Court* (1984) 163 Cal.App.3d 186, 194. These cases implicitly reject the proposition that the right to make an autonomous medical decision is limited to a class of educated individuals.

individual's prerogative to weigh even these facets. It was further possible to identify a subset of an individual's commitments that were regarded as particularly intertwined with their personality and therefore central to their autonomy. Namely, acceptances and preferences that were long-lasting, cohered with the patient's wider character system and were resilient to challenge. Ascertaining the law's respect for the individual weighting of, even fundamental, interests and ascertaining whether it accords especial significance to the individual's acceptances or preferences on the basis of these factors will indicate that the law values autonomy *inter alia* for the kinds of reasons adduced under our concept.

In the first respect, Californian case law has repeatedly and emphatically asserted that it is for the individual alone to determine the proper balancing of their interests:

> The weighing of [disclosed] risks against the individual subjective fears and hopes of the patient is not an expert skill. Such evaluation and decision is a nonmedical judgment reserved to the patient alone. A patient should be denied the opportunity to weigh the risks only where it is evident he cannot evaluate the data, as for example, where there is an emergency or the patient is a child or incompetent.[714]

Moreover, deference has been accorded to the patient's evaluation even – or rather especially – where fundamental interests were recognised to be at stake. In judging on medical decisions that may cause the patient's death, California's Supreme Court held: 'Especially when the prognosis for full recovery from serious illness or incapacitation is dim, the relative balance of benefit and burden must lie within the patient's exclusive estimation: "That personal weighing of values is the essence of self-determination."'[715] Framed as an aspect of the reflective dimension – rather than a factor opposing a rationalist, procedural understanding of autonomy – this insight into the significance of individual evaluation can be readily subsumed under our theoretical conception.

Barber v. Superior Court provides a further elaboration of the courts' consideration of reflective autonomy. As in *Cobbs* and *Thor*, the court accepted that the patient has a paramount role in balancing the values and disvalues of a clinical decision for themselves. Yet, also in line with our

714 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243-244.

715 *Thor v. Superior Court* (1993) 5 Cal.4th 725, 739; citing *In re Gardner* (Me. 1987) 534 A.2d 947, 955.

theoretical analysis, the court referenced the necessity of adducing certain basic values, before leaving it to the patient to factor these into their own desires and beliefs:

> the determination as to whether the burdens of treatment are worth enduring for any individual patient depends on facts unique to each case, namely, how long the treatment is likely to extend life and under what conditions. "[S]o long as a mere biological existence is *not considered the only value*, patients may want to take the nature of that additional life into account as well." (…)
> Of course the patient's interests and desires are the key ingredients of the decision making process.[716]

In this manner the courts have accepted that, even if there are fundamental values that must bear on relevant autonomous decisions, the attribution of importance to these is for the individual, not for the professional and not for the court.[717] Furthermore, it is the prerogative of the patient to define their own interests.

The next question that arises is whether the legal material also supports the view that, among this class of individual interests, it is especially a patient's deeper commitments – which are long-held, cohere with their character and which would be defended or defensible if challenged – that receive protection in the law. Such an approach is not without controversy. Judicial pronouncements, such as those above, are often framed in terms of an unconditional respect for a patient's assessment of their interest. Nevertheless, as has been discussed, the courts' treatment of the concept has not always been consistent and, even a strict adherence to this position, would not preclude a differentiation in terms of the seriousness of an intervention.

California's common law has had the opportunity to refer to the outlined kinds of factors under the negligence cause of action, the relevant aspects of which will be discussed in-depth in Chapter 7. For present purposes it is necessary to anticipate one forceful, rule-specific limitation that has been imposed on the law's ability to protect the reflective dimension of

716 *Barber v. Superior Court* (1983) 147 Cal.App.3d 1006, 1019 (emphasis added), citing: President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, 'Deciding to Forego Life-Sustaining Treatment: A report on the Ethical, Medical and Legal Issues in Treatment Decisions' (Washington, DC 1983) 32-39.

717 See also: *Bartling v. Superior Court* (1984) 163 Cal.App.3d 186, 193; *Bouvia v. Superior Court* (1986) 179 Cal.App.3d 1127, 1140.

autonomy under this tort. Namely, by relying heavily on the figure of the reasonable patient in the causation stage of the analysis, the cases render it irrelevant what information the individual would have responded to on the basis of their own balancing of interests and their wider commitments.[718] All that is asked is: what information would a reasonable person in the patient's position have acted upon to avoid a relevant form of harm?

Crucially, there appears to be a subset of cases that are prepared to overcome these rule-specific limitations and to offer protection, if not to the finer aspects of the patient's weighting of their own interests, then at least to their deeply held commitments. In *Hernandez ex rel. Telles-Hernandez v. U.S.* a district court applying Californian law determined the issue of causation – here concerning whether the plaintiff would have opted for a caesarean section over a vaginal birth – by reference to the plaintiff's 'emphasis on prenatal care and her desire to deliver her baby without the use of medication'.[719] The individual mother's commitment to the health of her baby (above all other factors) was clearly an established one, as evidenced by her recourse to prenatal care, and a robust objective, as demonstrated by her decision not to take medication and endure substantial hardship in order to secure it.

A consideration of individual circumstances in the causation analysis, and a connection of this to the reflective element of autonomy was also on display in *Wilson v. Merritt*. Here the plaintiff was wheelchair-bound and suffered from adhesive capsulitis in his shoulder.[720] He was pursuing physical therapy options with some success before being recommended a manipulation under anaesthesia.[721] He claimed that his physician had not adequately informed him of the risks of this latter procedure, which in fact resulted in a torn rotator cuff and a fractured shoulder, and had he been so informed he would not have undergone the procedure and would not have suffered the damage.[722] The Court of Appeal, overturning the trial court's determination that the evidence was insufficient for the plaintiff to succeed on the causation element of their claim, held:

718 A somewhat comparable reliance is also placed on a patient's reasonableness in the breach stage, although for reasons to be explored in Chapter 7 there appears to be a greater degree of leeway for autonomy-based reasoning under this element.

719 *Hernandez ex rel. Telles-Hernandez v. U.S.* (N.D. Cal. 2009) 665 F.Supp.2d 1064, 1078-1079.

720 *Wilson v. Merritt* (2006) 142 Cal.App.4th 1125, 1129.

721 ibid 1138-1139.

722 ibid 1138.

> A jury reasonably could determine that an adult paraplegic who was suffering some problems with stiffness and flexibility, but was functional in his then current condition, who was seeing some improvement in his condition through physical therapy, who had suffered devastating damage from surgery in the past, and who was so concerned about the potential risks associated with the recommended procedure that he took his mother with him to question the medical doctor on the topic, would indeed turn down the opportunity for the procedure if informed that it could result in a loss of his remaining mobility due to a torn rotator cuff or a fractured bone.

This summation of the *Wilson* court's causation analysis showcases the remarkable extent to which it was prepared to rely on factors personal to the plaintiff and which, arguably, derived their normative significance from autonomy's reflective dimension. The patient's decision did not merely involve a balancing of interests. A subjective preference against a clinical intervention that could impair mobility was taken to be firmly anchored in the plaintiff's system of beliefs and desires, not least because of his history and the limited mobility he already possessed in light of a past surgical intervention. The robustness of this commitment was further evidenced by his questioning of the doctor and the wish to have another party present during the process.

Both *Hernandez* and *Wilson* highlight that individual motivations that are associated with an exercise of reflective autonomy have been treated as particularly significant and have been able to have a distinct impact on the reasoning of courts.[723]

It is further notable that in cases dealing with the right to die under constitutional privacy protections, Californian courts have felt it necessary to point to long-standing commitments of the patient. In this vein, *Barber v. Superior Court* asserted that a surrogate decision maker for an incapacitated person, should first consider their past wishes (if they were expressed).[724] This at least implies an acceptance of the continuity and coherence of character that underlies the coherentist interpretation of reflective autonomy.

723 See also the analysis in *Morgenroth v. Pacific Medical Center, Inc.* (1976) 54 Cal.App.3d 521, 534-535. The patient's fundamental interests in an active lifestyle were discussed as important factors, but were taken to support the disputed diagnostic intervention. It therefore did not allow for autonomy considerations to play an independent role.

724 *Barber v. Superior Court* (1983) 147 Cal.App.3d 1006, 1021.

More directly in point, in *Bouvia v. Superior Court* and *Bartling v. Superior Court* the judges identified a protracted pattern of behaviour, where the patient expressed the relevant preference – to die, rather than to go on living in their condition. This was used as a justification for their right to refuse medical treatment.[725] Such a pattern did not have to be unbroken,[726] but in both cases it was framed as a long-standing, robust commitment that was entirely consistent with the patient's wider beliefs (e.g. regarding the non-improvement or deterioration of their condition and the consequences of continued existence) and their desires.[727] So, in these instances too, the courts have indicated that the reflective dimension of autonomy influences the weight that the law will attribute to reasoning with the autonomy principle. This supports the conclusion that Californian common law operates with a conception of autonomy that includes this dimension.

C. Positive and negative freedom

That the law includes ideas of positive and negative freedom (elements of the practical dimension of autonomy) is arguably the least contentious aspect to prove. The citation from *Union Pacific Railway Co. v. Botsford*, adduced under the first section of this chapter, bears testament to the esteem in which negative liberty is held under the common law: 'No right is held more sacred, or is more carefully guarded by the common law, than the right of every individual to the possession and control of his own person, free from all restraint or interference of others'.[728] This sentiment has been reiterated in many settings, including the informed consent one.[729]

725 Respectively: *Bouvia v. Superior Court* (1986) 179 Cal.App.3d 1127, 1135-1136; *Bartling v. Superior Court* (1984) 163 Cal.App.3d 186, 190-193.

726 '[T]he doctors and Glendale Adventist questioned Mr. Bartling's ability to make a meaningful decision because of his vacillation (…) The fact that Mr. Bartling periodically wavered from this posture because of severe depression or for any other reason does not justify the conclusion of Glendale Adventist and his treating physicians that his capacity to make such a decision was impaired to the point of legal incompetency': *Bartling v. Superior Court* (1984) 163 Cal.App.3d 186, 192-193.

727 *Bouvia v. Superior Court* (1986) 179 Cal.App.3d 1127, 1143-1144; *Bartling v. Superior Court* (1984) 163 Cal.App.3d 186, 191.

728 *Union Pac. R. Co. v. Botsford* (1891) 141 U.S. 250, 251-253.

729 'The purpose underlying the doctrine of informed consent is defeated somewhat if, after receiving all information necessary to make an informed decision, the patient is forced to choose only from alternative methods of treatment and precluded

Moreover, when it comes to the positive freedom, it is particularly the informed consent context that has provided a framework for a legal understanding of this dimension to be developed and (partially) realised. With commendable clarity it was stated in *Cobbs v. Grant* 'the patient's consent to treatment, to be effective, must be an informed consent. And (…) the patient, being unlearned in medical sciences, has an abject dependence upon and trust in his physician for the information upon which he relies during the decisional process'.[730] Similarly it was stated in *Thor v. Superior Court* that 'Doctors have the responsibility to advise patients fully of those matters relevant and necessary to making a voluntary and intelligent choice'.[731] Under the Californian conception of autonomy it is thus beyond doubt that a patient's process of practical reasoning must be facilitated.

One may also point to more nuanced elaborations of this standard, whereby the necessity of assisting the patient in their decision-making process is emphasised and it is their need that serves as the relevant gauge. For instance, in *Daum v. SpineCare Medical Group, Inc.* it was stated that 'the medical profession must conform its methods of disclosure to the needs and understanding of patients'.[732] Likewise, certain conditions on the categories of information to be disclosed have been developed to maintain an effective facilitation of medical decision making.[733] The courts are mindful that the positive dimension of practical autonomy would be threatened not only by a deficit of information, but also by a flood of data that overwhelms the decision-making process. The appropriate, selected yardstick is the patient's ability to direct their actions effectively.

The final element of our theoretical approach that was found relevant to our analysis of AI, and which must be identified in U.S. law, is the recognition that the holding of certain true beliefs is a prerequisite for autonomous action in medicine. Without assisting the patient to acquire these beliefs, it is not possible for them to exercise their autonomy.

Although the common law does not explicitly address the necessity of holding such beliefs, their existence is acknowledged implicitly in its tiered

---

from foregoing all treatment whatsoever. We hold that the doctrine of informed consent—a doctrine borne of the common-law right to be free from nonconsensual physical invasions—permits an individual to refuse medical treatment': *Rasmussen by Mitchell v. Fleming* (1987) 154 Ariz. 207, 216.

730 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 242.

731 *Thor v. Superior Court* (1993) 5 Cal.4th 725, 742-743.

732 *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285, 1304-1305.

733 *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1185-1186.

approach to information disclosure. Specifically, as will be seen in Chapter 7, the common law has asserted that a strong battery cause of action is appropriate where the patient has not given any consent at all. This situation exists *inter alia* where the patient lacks critical information about the procedure, such that the one which is in fact performed is 'substantially different'.[734] Here a necessary true belief was absent, the interference with the positive freedom of the patient was particularly grave and the legal response must be equally forceful. Lesser failures of facilitation are, by contrast, addressed under the aforementioned categories of negligence and are considered aspects of reasonable, rather than necessary, disclosure.

In sum, there are substantial grounds for concluding that our conception of procedural autonomy is present in California's common law and, to a significant extent, shapes the courts' reasoning with patient autonomy. This provides the basis for its application in our consideration of specific legal mechanisms.

IV. Limitations

Alongside the force of the autonomy principle, whose form and content has been considered above, a legal analysis must also respect the normative and structural factors that provide countervailing impulses and limit a principle's relevance. It is self-evident that in a legal system one cannot argue in a vacuum.

This raises the issue of the strength of relevant restrictions. One may recall, for instance, the observation in the previous chapter that courts exhibit a particular reticence in the private law sphere to amend settled norms by reference to more abstract considerations, especially social and political objectives or values.[735] As we have not conceived of autonomy as a mere political, extra-legal value, this reticence should be somewhat less of a hindrance.

Nevertheless, it is worth recapitulating that many factors have a recognised potential to limit the realisation of the autonomy principle. These included: other principles, which in the medical context notoriously encapsulates maleficence, beneficence and justice,[736] specific rules and the

734 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 239.
735 Robertson in Robertson and Tang, *The Goals of Private Law* (2009) 266.
736 Beauchamp and Childress, *Principles of Biomedical Ethics* (Fifth Edition 2001).

restrictions stemming from their underlying doctrine, and consequentialist reasoning. These all have the ability to impose severe restrictions on any argumentation with the autonomy principle.

In the final analysis, the issue remains that much is dependent on the relative weight or importance of considerations – including an assessment of the precise nature of the autonomy violations at issue – that have been particularised for a specific situation. The nature of the relationship between the autonomy principle and its limitations will therefore become fully apparent only in our detailed analysis in Chapter 7. Simultaneously, it should already be anticipated that the American courts, in spite of their use of powerful rhetoric and their ability to inspire significant developments, have arguably maintained a relatively stringent approach that is prepared to circumscribe autonomy's influence by reference to rule-specific factors.

V. Conclusion

As indicated in the introduction of this thesis, the more fragmented nature of U.S. common law presents distinct challenges. These have emerged in the delineation of the autonomy concept. To provide a firm basis for argumentation it was not possible to restrict this analysis to just Californian common law. In particular, an examination of federal law and the law of other states provided insights into how the conception of autonomy was operationalised within legal reasoning, and within what scope.

Regarding the question of content, it was possible to focus on the Californian legal system. This presented the most relevant specification of a contestable concept for our subsequent, targeted analysis. Under this head we determined that, while there were some uncertainties and inconsistencies, substantial support could be found for the elements of procedural autonomy that were outlined in Part I. Moreover, here too the limitations on the deployment of this principle had to be taken seriously. As we will come to see in the next part, U.S. law clearly recognises the force of doctrinal and rule-specific restrictions on principled, autonomy-based argumentation.

Part III:
Informed consent and artificial intelligence in medicine

Chapter 6: UK tort law

This part asks whether the law imposes an obligation on relevant parties to obtain the consent of patients to the involvement of artificial intelligence (AI) in their care, to inform patients of this involvement and, crucially, to advise them of relevant features of the technology in such a way that meets the autonomy challenges outlined in Chapter 3.

The most direct way in which these questions arise under UK law is *via* the common law causes of action of battery and negligence. These have gained an established role in protecting patient autonomy, either through imposing consent requirements on most forms of medical treatment or through recognising legal obligations to provide patients with information that is deemed relevant to clinical decision making.[737]

737 This combined role is often demarcated by the use of the more general label of 'informed consent'. The English courts initially proved reluctant to adopt this American terminology and it was disputed whether, if it were to find application, it would be more suited to denote liability in battery or negligence. Compare *Freeman v Home Office (No 2)* [1984] QB 524, 555 with the statement in *Davis v Barking, Havering and Brentwood Health Authority* [1992] Lexis Citation 2495, that 'the concept of "informed consent" cannot be made to fit the cause of action for negligence; its place lies in trespass to the person'. Contrast, the recent statement in the context of negligence that: 'Whatever uncertainty there may have been in the past, the requirement of informed consent to medical treatment is now a fundamental and settled principle of the law in both England and Wales and Scotland': *Gallardo v Imperial College Healthcare NHS Trust* [2017] EWHC 3147 (QB), [2017] 12 WLUK 198 [1]. For an analysis of the 'informed consent' terminology and the law of negligence see: Maclean, 'The Doctrine of Informed Consent' (2004) 24(3) Legal Studies p. 386.

Both of these mechanisms are protected through the imposition of *ex post* sanctions on those who breach their legal duties.[738] If a patient can make out either tort, then they are entitled to monetary damages. Courts may also be asked to make a declaration on whether a given clinical procedure or type of intervention would constitute a battery and thus whether the medical professionals concerned are legally permitted to proceed with the relevant actions.[739] This will technically proceed as an administrative law procedure, but the claim will turn on whether the substantive requirements of the relevant civil law claim are made out.

Following the evaluation of the common law, a specific aspect of the *UK General Data Protection Regulation* and the *Data Protection Act* 2018 will also be addressed. Admittedly this falls outside of the immediate remit laid down in Chapter 1, which focussed on the outlined causes of action that have shaped the evolution of this area of the law. It is nevertheless examined as it marks a targeted legislative intervention, which addresses one specific shortcoming in the common law analysis of ML instrumentalisation. In so far as this approach is to be critiqued in the final chapter, it must be done with an understanding of the functioning of this statutory mechanism.

I. Battery

Consent functions to legitimise an act that would otherwise constitute a battery – a criminal offence and a civil wrong under English law.[740] Judicial pronouncements on battery have established the necessity of consent for most forms of medical treatment and the conditions under which such

738 Battery also constitutes a criminal offence, as some of the cited cases *infra* demonstrate. However, as the mechanisms' requirements are largely consistent, our primary framing will be under the law of tort.

739 This may be at the level of an individual decision: *Re F (mental patient: sterilisation)* [1990] 2 AC 1; *Airedale NHS Trust v Bland* [1993] AC 789. The public law procedure of judicial review may also be utilised to request a formal declaration from the court stating what a relevant private law mechanism requires in certain circumstances. This may affect the legality of a policy to adopt a medical intervention more widely. For a recent summary of the appropriate deployment of this approach see: *Bell v Tavistock and Portman NHS Foundation Trust* [2021] EWCA Civ 1363, [2022] 1 All ER 416 [66]-[90].

740 *Chatterton v Gerson* [1981] QB 432, 442.

consent is valid.[741] These will be drawn upon to determine the information that must be shared with a patient to ensure that the consent they have given to a procedure involving AI is valid.

A. Limitations flowing from the battery doctrine

Before engaging with these specific questions, an assessment of the tort's broader requirements is indispensable in order to understand the limitations upon its operation in the healthcare sphere. To some degree the courts have shown flexibility in this respect, accommodating the unique demands of the clinical situation. For instance, the usual restrictions placed on an individual's ability to consent to certain types of bodily injury have been relaxed in so far as 'reasonable surgical interference' is concerned.[742] Nevertheless, other conditions stemming from the tort's origin as a mechanism for the prevention of violence persist, without providing much scope for argumentation.[743]

Most especially, the clinical interaction complained of must involve some form of direct or immediate interference with the claimant's body.[744] This severely limits the role that battery's consent requirement can play in facilitating patient control of AI use in their treatment. Not only will consent not be pertinent to many decisions not to treat, which involve no relevant contact, but battery's consent and information requirements will likely not be applicable to indirect forms of treatment, such as the prescription of drugs.[745] Therefore, while some such decisions are likely to benefit from the assistance of intelligent decision-support systems, battery provides no basis on which the doctor is required to inform the patient or obtain their consent for this use.

741 Of course, there are instances where consent cannot be obtained and other justifications are in place to allow for the emergency treatment of an unconscious patient, but these are less relevant for our analysis of AI autonomy violations.

742 *Attorney-General's Reference (No. 6 of 1980)* [1981] QB 715, 719.

743 Brazier and Lobjoit in Erin and Bennett, *HIV and AIDS: Testing, Screening, and Confidentiality* (2001) 185-186.

744 *Scott v Shepherd* (1773) 3 Wils KB 403.

745 Maclean does construct a well-reasoned argument as to why battery should be applicable to drug prescriptions, but he notes how unlikely this is to succeed in English courts: Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009) 150–152. See also the analysis of Seabourne who reaches a similar conclusion: Seabourne, 'The Role of the Tort of Battery in Medical Law' (1995) 24(3) Anglo-American Law Review p. 265, 270-271.

Battery further requires an intentional action by the defendant.[746] In the past this had generated some uncertainty: must the relevant intention be directed towards inflicting an injury on another? Or is it sufficient that the application of force (the touching) was envisaged?[747] However, as has been settled in *Wilson v Pringle*, the latter interpretation is the correct one: 'It is the act and not the injury which must be intentional. An intention to injure is not essential to an action for trespass to the person. It is the mere trespass by itself which is the offence'.[748] A patient's action will not be barred under battery because a professional's intention did not go beyond the touching.

There is also a potential mandate under UK law that, although no intention to injure is required, the behaviour of the defendant must be hostile in nature.[749] This factor would almost certainly exclude ordinary clinical situations (including those involving AI) from its scope.[750] Yet it is only a 'potential requirement', since its meaning is extremely ill-defined and several judgments have strongly indicated, or have even been decided on the basis, that it has been abrogated, although it is yet to be directly overruled.[751] As requiring hostility in medical interactions would effectively bar the battery action from offering any meaningful protection to patient autonomy, this principle is argued to provide a further, compelling reason for rejecting the questionable validity of this requirement.

In sum, before one even begins to analyse the nature of battery's consent and information requirements, there are broader difficulties that limit the application of this norm to the use of AI in medical treatment. UK law has arguably maintained an emphasis on rule-specific factors that are separate from, and serve as restrictions upon, the autonomy principle. Most significantly, it is relatively certain that, if a patient is not touched during AI use

746 'The least touching of another in anger is a battery' represents an early indication of this requirement: *Cole v Turner* (1704) 6 Mod 149.

747 *Wilson v Pringle* [1987] QB 237, 248-249.

748 ibid 249.

749 ibid 246-248.

750 '[T]he requirement of hostility, (…) unless it is interpreted in a sense so weak that it collapses into the question of intentional action, is likely always to be a stumbling block to plaintiffs in medical battery': Seabourne, 'The Role of the Tort of Battery in Medical Law' (1995) 24(3) Anglo-American Law Review p. 265, 272.

751 Lord Goff's well-known rejection of this requirement in *Re F* was *obiter dicta* (although it is interesting that this rejection was specifically based on the 'libertarian principle of self-determination' in medical treatment): *Re F (mental patient: sterilisation)* [1990] 2 AC 1, 72-73. See also the total omission of this requirement in *Re B (adult: refusal of medical treatment)* [2002] EWHC 429 (Fam), [2002] 2 All ER 449.

or with a view to carrying out an AI intervention, then the professional will have no obligation to advise the patient of their reliance on this technology or its implications.

B. Battery and the nature of valid consent

If the above requirements are fulfilled, then the claimant must next establish that they did not consent to the touching.[752] Where no consent whatsoever has been given, this claim can be straightforwardly made out.[753] This situation is rare, however. Moreover, where an AI/ML device is involved, it has been argued that there is generally a wider intervention to which the patient has agreed. To ground a successful battery action in such circumstances it is necessary to determine a deficiency in the claimant's given consent: either its scope was too limited or, relatedly, it was given without knowing about a certain dimension of the clinical intervention.

The doctrine that shapes this area of the law emerged from *Chatterton v Gerson*, where Bristow J held that consent to a medical treatment must not only be in form but in reality.[754] This required the patient to be 'informed in broad terms of the nature of the procedure'.[755] A similar definition of the relevant test can be found in *Re T*, where it was said to be enough if the 'patient knew in broad terms the nature and effect of the procedure to which consent (or refusal) was given'.[756]

While providing the patient with sufficient information on their treatment is one important prerequisite for the validity of their consent, it is clear from such statements that battery is intended to set a low bar for disclosure. Information will not have to be provided on many specific features of medical interventions. For many of these subsidiary aspects the

752 While the Court of Appeal did not address this matter, McCowan J had agreed with counsel at first instance on this point: *Freeman v Home Office (No 2)* [1984] QB 524, 539. For a critique of this position and an overview of how Canada and Australia have framed the matter differently see: McHale in Laing and McHale, *Principles of Medical Law* (Fourth Edition 2017) 424.

753 *Border v Lewisham and Greenwich NHS Trust* [2015] EWCA Civ 8, (2015) 143 BMLR 182 [19]. See also *Cull v Butler* where a hysterectomy was carried out against the express wishes of the patient: *Cull v Butler* [1932] 1 BMJ 1195.

754 *Chatterton v Gerson* [1981] QB 432, 442-443.

755 ibid 443.

756 *Re T (adult: refusal of treatment)* [1993] Fam 95, 115.

courts explicitly state a preference for a claim under the negligence cause of action.[757]

Weighty rule-specific considerations are used to support this approach. Foremost among them is a distaste for subjecting medical professionals acting in good faith to this form of liability.[758] However, commentators have also drawn the logical inference that in those cases where a battery claim is nonetheless successful the courts are responding to a particularly grievous interference with patient autonomy.[759] That is, the weight of the principle is great enough in those situations to direct legal reasoning towards offering a strong response. Such a differentiated approach further corresponds to the claims made in the theoretical analysis in Chapter 3.

The remainder of this section categorises the significant deficiencies in consent which have given rise to a battery action, yielding three classes: the nature of the procedure, the identity of the actor and the purpose or motivation for the act. Shortcomings in the patient's understanding of these dimensions have proved sufficiently substantial so as to fall outside of their broad consent to diagnosis or treatment. Bearing in mind that AI autonomy violations must be of a comparable gravity to those found in the existing analyses, we turn to evaluate whether any of AI's challenges can be subsumed under these classes.

1. Nature of the procedure

The first class of cases is constituted by a change in the nature of the procedure. This was the type of case explicitly elaborated and examined in *Chatterton*. On its face, the judgment appears tailored to ensure that recourse to a claim in battery is only available for clear-cut misunderstandings of the fundamental physical nature of the procedure. Thus, the court in *Chatterton* referred to the case of a boy who is circumcised instead of re-

757 *Chatterton v Gerson* [1981] QB 432, 442; *Hills v Potter* [1984] 1 WLR 641, 653; *The Creutzfeldt-Jakob Disease Litigation* (1995) 54 BMLR 1, 4-6; *Davis v Barking, Havering and Brentwood Health Authority* [1992] Lexis Citation 2495.

758 '[I]t would be deplorable to base the law in medical cases of this kind on the torts of assault and battery': *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 883.

759 Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009) 195-196.

ceiving an expected tonsillectomy.[760] Similarly, in *Devi v West Midlands Regional Health Authority* the sterilisation of a woman was conducted without prior discussion during a postpartum dilation and curettage operation.[761] The Court of Appeal accepted that the action for damages proceeded on the basis of an 'assault', even if the issue was not expressly dealt with. The sterilisation was treated as altering the nature of the procedure to such a degree that the patient's consent had been violated.[762]

One should also consider *Davis v Barking Havering and Brentwood Health Authority*. Here the court asked whether an aspect of a wider procedure required separate 'sectional' consent.[763] This is arguably another way of framing the issue of whether a patient knew enough about the overall nature of what was to be done, for their consent to be valid.[764] However, by enquiring as to the necessity of separate consent, McCullough J provided more nuanced insights into battery's analysis of this question. In particular, while accepting that separate consent must be obtained in some instances, for example to separate surgeries, the judge objected forcefully to the prospect of complex procedures being broken down into their individual steps and consent being demanded for each of these. Such an approach would give battery an unduly prominent role in medical law. Rather, these were 'details' that did not go to the validity of consent.[765]

Obiter statements in the judgment further conveyed the kinds of acts that would be considered to be aspects of a single procedure: the attachment of an ECG to the chest of an unconscious patient, the insertion of a tube into the unconscious patient's trachea and an injection of morphine while the patient is coming around after the surgery.[766] *Davis* itself was decided on the basis that an injection of a local (caudal) anaesthetic could not be separated from the patient's consent to the administration of a general anaesthetic, even though the former carried specific risks that she had not

760 *Chatterton v Gerson* [1981] QB 432, 443.
761 *Devi v West Midlands Regional Health Authority* [1981] Lexis Citation 1417.
762 Cf. *Abbas v Kenney* (1995) 31 BMLR 157.
763 *Davis v Barking, Havering and Brentwood Health Authority* [1992] Lexis Citation 2495.
764 Maclean, 'Consent, Sectionalisation and the Concept of a Medical Procedure' (2002) 28(4) Journal of Medical Ethics p. 249.
765 *Davis v Barking, Havering and Brentwood Health Authority* [1992] Lexis Citation 2495.
766 ibid.

anticipated.[767] A comparison can also be drawn to *R v Mental Health Commission, ex p X*, where the court relied on the common law's delineation of the concept of consent. It was held that a patient must understand the likely effects of a treatment, but that this must not amount to a detailed understanding of physiological processes.[768]

All in all, these cases suggest that an understanding of the fundamental physical components of a procedure are sufficient for valid consent. It is not necessary to know about all the processes involved in an intervention, whether these have subsidiary physical manifestations or not. Moreover, the UK courts have been clear that the disclosure of the specific risks of a procedure is not something that can properly ground a claim in battery.[769] Failing to disclose these matters will not vitiate consent.

In so far as ML devices are concerned, it is envisaged that their operation will be closely connected to methods of diagnosis, treatment and prognosis (either human or technological) and the physical manifestation of the procedure will, for the most part, remain singular, apparent and unchanged. A general disclosure of AI use, treating it as a separate procedure, therefore cannot be maintained. Furthermore, although it has been argued that AI may lead to changes in the general risk profile of an intervention, these are matters to properly be considered under negligence.

At the same time, the basis for the courts' delineation of a procedure's nature remains ambiguous and there are indications that it is not entirely restricted to its physical components. Arguably, the effects of an intervention, with their wider significance, must also be conveyed. In *R v Mental Health Commission, ex p X* Stuart-Smith LJ mandated that real consent entails that a patient is aware of the likely effect of a treatment, which here involved the 'effect of reducing male sexual drive', and as stated did not depend upon an awareness of the underlying physiological processes.[770] This more general association between the nature of an intervention and the

767 Grubb, 'Battery and Administration of Anaesthetic: Davis v. Barking, Havering and Brentwood Health Authority' (1993) 1(3) Medical Law Review p. 389.

768 *R v Mental Health Commission, ex p X* (1988) 9 BMLR 77, 86-87.

769 *R v Richardson* [1999] QB 444, 450; *Hills v Potter* [1984] 1 WLR 641, 653; *Chatterton v Gerson* [1981] QB 432, 442-443. Consider also *The Creutzfeldt-Jakob Disease Litigation* as a particularly strong example of this position. There appeared to be absolutely no disclosure of *any* risk of the clinical intervention, but this was still not seen as a basis for a battery claim: *The Creutzfeldt-Jakob Disease Litigation* (1995) 54 BMLR 1, 3-4.

770 *R v Mental Health Commission, ex p X* (1988) 9 BMLR 77, 83, 86.

wider non-physical significance to the patient can also be found in *Abbas v Kenney*. Here the court considered that the consent to a procedure which removed the patient's reproductive organs was valid, when the patient had agreed to 'the least extensive surgery possible, consistent with saving her life' and 'understood that [the surgeon's] main brief was to save her life, and if possible to save her fertility'.[771] Although the physical manifestations of the surgery were clearly relevant, it was primarily important that the patient was aware of the significance of the procedure for the purposes of saving her life and for maintaining her fertility.

The resulting uncertainty has allowed commentators to argue that the significant implications of a procedure for the patient, shape its nature. This has been extended *inter alia* to the effects of diagnostic tests. A particularly vigorous debate has addressed the question of whether an analysis of a patient's blood for its HIV status falls within a patient's general consent to 'some blood tests'.[772] The physical nature of this interaction remains unchanged whether the additional analysis takes place or not.[773] Moreover, the fact 'that one of the tests is for HIV does not alter the general nature of the procedure as part of a process of therapeutic diagnosis'.[774] The focus on the physical manifestations of the intervention, as well as the wider clinical processes underlying it, find support in the bulk of the case law analysed above.

771 *Abbas v Kenney* (1995) 31 BMLR 157, 165-166.

772 To my knowledge no case dealing with this issue ever came before the UK courts. Yet the Public Health Laboratory Service did carry out HIV tests on anonymised blood samples, even though they had been obtained from patients for other purposes: Coghlan, 'Could HIV tests land doctors in court?' (19.1.1994) <https://www.newscientist.com/article/mg14119100-600-could-hiv-tests-land-doctors-in-court/> accessed 9.3.2021; Brazier and Lobjoit in Erin and Bennett, *HIV and AIDS* (2001). It appears to have been the fact that several professional bodies sought legal advice on the matter which sparked the wide-ranging academic debate: Sherrard and Gatt, 'Human Immunodeficiency Virus (HIV) Antibody Testing: Guidance from an Opinion Provided for the British Medical Association' (1987) 295(6603) British Medical Journal p. 911; Kennedy and Grubb, 'Testing for HIV Infection: The Legal Framework' (1989) 86(7) Law Society Gazette 30-35; Keown, 'The Ashes of Aids and the Phoenix of Informed Consent' (1989) 52(6) The Modern Law Review p. 790; Grubb and Pearl, *Blood Testing, AIDS, and DNA Profiling: Law and Policy* (1990).

773 Grubb and Pearl, *Blood Testing, AIDS, and DNA Profiling: Law and Policy* (1990) 6.

774 Keown, 'The Ashes of Aids and the Phoenix of Informed Consent' (1989) 52(6) The Modern Law Review p. 790, 796.

On the other hand, it may be thought that the patient must have knowledge 'of the underlying purpose of the procedure (ie HIV testing) which is vital in order for the patient to have the necessary understanding of the *quality* of the touching'.[775] This appeals to the courts' reference to the wider significance and effects of a procedure. To concretise their argument for disclosure, its proponents have exhibited a pronounced reliance on the differential implications for patient autonomy. Kennedy and Grubb state explicitly that 'the consequences for a patient that flow from a positive HIV test (or in some instances even a negative test) are so serious for him that a court would consider it to be contrary to public policy to regard consent to testing as including a procedure with such far reaching implications for the patient'.[776] Grubb and Pearl also highlight the 'grave and adverse personal and social consequences' that accompany a diagnosis of HIV. Further they cite: the widespread discrimination that such a finding may result in (especially since it cannot always be ensured that confidentiality will be maintained in practice), the effect on the patient's ability to obtain life insurance coverage and the fact that the patient would be deprived of the opportunity to receive appropriate counselling before their diagnosis.[777]

All of these factors indicate that undertaking a HIV diagnosis without the patient's understanding seriously impairs their autonomy. As explored in Chapter 3, it determines an aspect of their medical care, prevents them from reconsidering their position in a reflective manner and directs their care according to external non-subjective values. Invoking these dimensions of autonomy to ascertain the significance of a procedure, is what subsumes the hypothetical HIV-testing scenario under the existing case law.

Moreover, as Keown has argued: if consent is found necessary for HIV-testing, then it would have far-reaching consequences. For instance, one would expect it to be required for many other tests, such as those used in the diagnosis of cancer.[778] This is what, in turn, connects this line of argumentation with our analysis of ML use.

As noted in Chapters 2 and 3, ML devices can be used to pursue broad, ill-defined goals and in a way that is partially determinative of the clinic-

775 Grubb and Pearl, *Blood Testing, AIDS, and DNA Profiling: Law and Policy* (1990) 6.

776 Kennedy and Grubb, 'Testing for HIV Infection' (1989) 86(7) Law Society Gazette 30-35.

777 Grubb and Pearl, *Blood Testing, AIDS, and DNA Profiling: Law and Policy* (1990) 6-7.

778 Keown, 'The Ashes of Aids and the Phoenix of Informed Consent' (1989) 52(6) The Modern Law Review p. 790, 797.

al decision. The most relevant example that was given, is AI's ability to reach serious, surprising diagnoses. Without more information, it will not be obvious to the patient that a relevant evaluation is intended or what condition(s) will be tested for. In doing so, the ML device may be striving to accomplish an end that is not the patient's own and on which they have had no influence.

It can even be argued that the autonomy violation is more serious in the novel, AI scenario. The machine does not only provide a positive/negative result for one well-defined condition but may run a more complex diagnosis for a number of illnesses. As such, it is capable of generating a range of surprising insights and the infringement of the patient's autonomy is greater because they are deprived of the opportunity to understand and control a decision that is substantively wider and more nuanced. In such situations there is a strengthened argument that the nature of the procedure is changed by the reliance of the ML device.

A further factor that has emerged from the discussions surrounding blood analysis, and which should certainly not be discounted, concerns the timing of the doctor's intention to proceed with HIV testing. Regardless of one's understanding of the nature of the procedure, this must be determined at the time of the direct and intentional intervention. [779] Therefore, if AI use is to vitiate the patient's consent for the reasons outlined above, then the doctor who is interacting with the patient must have formed an intention to proceed with an AI analysis before engaging in the requisite contact.

This will place many instances of AI use outside of battery's scope. For instance, although pathology represents a field where the prospects for implementing AI are particularly promising, pathologists would generally have no contact with the patient and would only form an intention to analyse material *via* a certain methodology involving AI after it has already been obtained by another professional. In other scenarios it will also not be clear that AI use was contemplated before contact – a subsequent analysis of the patient's data may be an easy thing, requiring no additional testing. Even if a claim will not be barred here, this requirement will nevertheless create evidential difficulties for the patient.

In sum, it is possible to claim that subjecting a patient to certain unsolicited AI analyses can alter the nature of a procedure and vitiate the validity of their consent. These are analyses where the AI has a degree of independ-

779 Grubb and Pearl, *Blood Testing, AIDS, and DNA Profiling: Law and Policy* (1990) 21.

ence and is capable of arriving at conclusions with significant implications for the patient. The argument from the case law, in combination with the autonomy principle, has been stated to be stronger than that in the established debate on HIV testing. Nevertheless, the argument must remain of uncertain strength given the limited jurisprudence on the matter and the substantial ambiguity in the precedents there are.

2. Identity of the professional

Another class of cases indicates that a patient's consent can be invalidated if they are not informed of the identity of a (purported) professional carrying out a procedure on them.[780] In general, these are not cases where one person has pretended to be someone else or where one doctor performs a procedure when the patient expected it to be performed by another. These could be seen as true instances of mistaken identity. Rather, liability has arisen primarily where the claimant believed the defendant to have certain attributes, qualifications or a certain status, when in fact they did not. These are instances of confused identity. The following will consider how both of these categories are to be applied to AI.

Regarding true cases of mistaken identity, *Michael v Molesworth* is an illustrative case from the medical sphere. Here a claimant succeeded in suing a house surgeon for battery because an apprentice had carried out the relevant surgery to practise his skill, even though this was done entirely competently.[781] However, it is unlikely that this case would serve as precedent in the context of the modern National Health Service (NHS). As McHale has noted, the NHS standard contract reflects the fact that there is generally no expectation that one particular professional will be responsible for one's care in an NHS hospital and it would likely require exceptional circumstances for the patient to elevate such a factor to a condition for their

780 In one of the cases within this class, the court's finding was framed solely in terms of the 'quality' of the act, explicitly distinguishing the 'identity' of the actor: *R v Tabassum* [2000] Lloyd's Rep Med 404. However, this framing does not appear to be based on a substantive distinction. Rather, it can be explained by a need to distinguish the decision in *R v Richardson* [1999] QB 444. As we will see, the courts now appear to accept that arguments can proceed straightforwardly under the head of identity, even where purely the status, attributes or qualifications of the 'professional' are concerned: *R v Melin* [2019] EWCA Crim 557, [2019] QB 1063.

781 *Michael v Molesworth* (1950) 2 BMJ 171.

consent and legitimate treatment.[782] Similarly, it appears to be implicit in the more recent case of *R v Richardson* that true mistakes as to identity may only occur where an individual is quite literally in error about who it is that is interacting with them.[783]

Following this line of argumentation, there will generally be little scope for mistakes of identity to occur in relation to AI. It has not been claimed that ML devices possess legal personhood and it has been argued that they will not fully replace human involvement in any given type of medical treatment. It is true that, similar to what occurred in *Michael*, an AI could take over one aspect of a procedure that involves multiple human professionals. In Chapter 2 this was subsumed under the category of 'devices that partially replace pre-existing cognitive capabilities'. Yet, as this kind of substitution falls far short of the replacement or impersonation of a known individual, a claim in battery is not envisaged.

It is more promising to argue that AI involvement may lead to confusions of identity and that they would vitiate consent on this basis. Several cases go directly towards this issue. In *R v Tabassum* the defendant purported to show a number of women how to examine their own breasts for the purposes of detecting breast cancer.[784] The women believed that he had the requisite medical training to do this. However, while he did have minimal experience in this field – having prepared several leaflets for self-examination and having learned about several diseases during his work as a medical representative – he had neither relevant training nor qualifications to give such instructions.[785] The centrality of the defendant's identity to the finding

782 The patient would have to reach an agreement with the relevant organisation and professional that the consent is conditional on the treatment being undertaken by a particular consultant: McHale in Laing and McHale, *Principles of Medical Law* (2017) 437. See also Brazier and Cave, *Medicine, Patients and the Law* (Sixth Edition 2016) 133. A different view is expressed by Donnelly, who draws on 'Lord Hope's obiter statement in *Chester* that a patient has a right to be informed "as to whether, and if so when and by whom, to be operated on"': Donnelly, *Healthcare Decision-Making and the Law: Autonomy, Capacity and the Limits of Liberalism* (2010). Yet the influence that such an *obiter* statement, which was made in a case dealing solely with negligence, can have on the law of battery is to be doubted.

783 Otton LJ cited a definition of identity as 'the condition of being the same' and considered that cases of mistaken identity must be of the same nature as mistakes that could lead a woman to have sexual relations with a man that she believes to be her husband: *R v Richardson* [1999] QB 444, 459-450.

784 *R v Tabassum* [2000] Lloyd's Rep Med 404 [4].

785 ibid [18].

that the women's consent was indeed vitiated in these circumstances is palpable, even if the court relied on a notion of 'qualitative deception' to decide the issue.[786] According to one commentator, the patients' consent was vitiated because 'the Court of Appeal considered that the women's consent was to "breast examination for the purposes of preparing a medical software package *by a medically qualified person*" rather than simply "breast examination for the purposes of preparing a medical software package"'.[787]

This convoluted approach appears to have resulted from a desire to distinguish the decision from the aforementioned *R v Richardson*. In *Richardson* a claim was rejected that a dentist's status or attributes – specifically her disqualification from practicing – went to her identity, vitiating the consent of her patients.[788] In spite of the different terminology, these decisions are not easy to reconcile. Aspects of a defendant's identity are relevant to the courts' consent analysis, but the relevant arguments have partially been couched in terms of the quality of the act.

Fortunately, the recent case of *R v Melin* has clarified the legal position. The Court of Appeal upheld a finding that the consent of the claimant had been vitiated by aspects of the defendant's identity. To reach this result, the Court had no need to rely on the reasoning of *Tabassum*, but rather proceeded by giving a narrow interpretation to *Richardson*'s restrictive comments on identity.[789]

Specifically, the issue in this case was whether the consent of several women to a series of Botox injections was vitiated by the fact that they believed the defendant to possess medical qualifications, when in fact he did not. It was decisive for the affirmation of the defendant's liability that his identity was inextricably bound up with his claimed status as a doctor. And the two were inextricably bound up with one another because that status was operative in the complainant's mind when they gave their consent.[790] The significance of certain attributes for the purposes of giving a valid consent was thereby expressly acknowledged. Citing *Smith, Hogan and Ormerod's*

786 ibid [37]-[38]. See also *R v Dica*, which assessed the 'quality of the act' to determine whether there had been consent to sexual intercourse and to the risk of the transmission of HIV where the defendant had not disclosed his positive status: *R v Dica* [2004] EWCA Crim 1103, [2004] QB 1257 [38]-[39].

787 Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009) 162 (my emphasis).

788 *R v Richardson* [1999] QB 444.

789 *R v Melin* [2019] EWCA Crim 557, [2019] QB 1063 [34]-[35].

790 ibid [33], [40].

Criminal Law it was stated that 'it could be that the attribute is actually more important than the identity. For example, would a patient visiting a general practitioner and being told that a new doctor is taking the surgery be more concerned as to the "status" of the person or his "identity"'.[791]

It was also held that the legal requirements for the level of qualifications necessary to undertake a particular procedure, were not determinative of the factual inquiry related to identity.[792] Therefore, for the purposes of the consent enquiry in *Melin*, it did not matter that Botox injections were not legally required to be undertaken by medically qualified practitioners. What mattered was that, as a matter of fact, the patient's consent was predicated on the medical qualifications of the individual treating her and that she was mistaken to this extent.

The issues addressed in these cases will become relevant to those situations where AI partially replace human cognitive capabilities. Especially where such devices reduce the level of human expertise that is brought to bear on a given decision. The example of IDx-DR was provided in Chapter 2 in this respect. The key questions that are posed in light of the outlined legal context are this: is an individual taking advantage of such systems to carry out clinical tasks obligated to inform their patients of the fact that they are extending their expertise by relying on the aid of an AI? Are they misleading the patient about possessing a relevant status, qualification or attribute if they do not?

Several grounds can be advanced to support of an affirmative answer. In the first instance, as *Tabassum* and *Melin* illustrate, the treating professional's expertise and qualifications will often be a factor that is operative in patients' minds. By contrast, the possibility that a healthcare worker would only have the expertise to deal with their condition as a result of technical assistance is currently unlikely to occur to most patients.[793] It is therefore arguable that in circumstances where patients would expect a procedure to be carried out only by a professional with a certain qualification and/or level of expertise, they are only consenting on the basis of this assumption. If they are not informed that the professional does not themselves possess

791 ibid [30], citing: Ormerod and Laird, *Smith, Hogan, & Ormerod's Criminal Law* (Fifteenth Edition 2018) 672.

792 *R v Melin* [2019] EWCA Crim 557, [2019] QB 1063 [33].

793 Distinguish the ordinary situation where the human has the requisite capabilities but merely relies on technical assistance (e.g. an X-Ray machine) as a tool that provides the information upon which these capabilities are exercised.

these attributes, there is a *prima facie* case that this may vitiate their consent.

Melin has further made clear that this analysis is not prejudiced by the wider legal framework. So that, even if an AI is an approved medical device and if professionals are legally permitted to undertake more specialised tasks with its help, this will not mean that the patient's consent to such a procedure will automatically be informed and valid. This is a question of fact.

Another supportive aspect of the existing law is the repeated insistence that what ultimately matters is not the source of the mistake, but the fact of its occurrence.[794] The ancillary and novel nature of the AI technology make it likely that the patient would be mistaken about its involvement as a result of an omission to inform them, not because of positive misrepresentation or intentional fraudulent conduct. The cases are now clear that the presence or absence of such conduct on behalf of the defendant is not determinative of whether the patient's consent has been vitiated. The decisive factor remains whether there was a mistake of the requisite kind, which was operative in the patient's mind. In consequence, so long as the initial analogy to cases of confused identity is deemed convincing, the wider legal requirements (allowing AI use) or the absence of fraud will not defeat a claim in battery.

What, then, are the factors that make the drawing of this analogy contentious? First, there are grounds for distinguishing medical practice with AI assistance from the claims that have successfully asserted confused identity. Most notable is the fact that, in the established judgments, the defendants evinced a fundamental lack of relevant medical expertise and qualifications. As MacLean has commented, in a case like *Tabassum* the only relevant factor appears to have been 'that the accused was not medically qualified, which emphasises the courts' reluctance to find doctors liable for battery'.[795] In other words, given the nature of the tort and the case law on confused identity, the fact that a user of AI is likely to have *some* medical

794 'The common law is not concerned with the question whether the mistaken consent has been induced by fraud on the part of the accused or has been self induced': *R v Richardson* [1999] QB 444, 450; 'it would be undesirable for the law to treat all false or fraudulent representations as vitiating consent': *R v Melin* [2019] EWCA Crim 557, [2019] QB 1063 [29].

795 Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009) 162.

experience and qualification is a distinction that weighs against a successful claim in battery

The exact force of this consideration is difficult to pin down. On the one hand, it appears hard to argue that the total lack of qualification is part of the *ratio* of these cases. For, in *Tabassum* the defendant possessed some rudimentary forms of healthcare experience and in *Melin* some consideration was given to the differing levels of expertise that the defendant claimed to possess.[796] On the other hand, it is conspicuous that in *Richardson* – the one case dealing with a qualified individual, albeit practising without a licence – a claim in battery was rejected. It seems that the broadly hostile position that the courts have adopted in relation to the application of battery to professionals cannot easily be discounted.

Second, there is the fact that the AI in question will be designed with the purpose of replacing and thereby raising the expertise of the utilising professional to the requisite specialist level. As our analysis of medical AI demonstrates, the technology is intended to compensate for a lack of human experience or qualification. By contrast, the outlined judgments only deal with situations where the individuals, in spite of holding themselves out as having a certain level of expertise, did not themselves possess it and did not and could not acquire it through assistance.

This raises the question of whether the courts will consider that the *relative* proficiency of the professional remains the crucial aspect, or whether the patient should only be informed if there are variations in the *aggregate* levels of expertise that are brought to bear on their treatment. This is an open question under the existing jurisprudence. Consequently, it is at this juncture that one must consider the differentiated approach that the British courts have taken to responding to interferences with patient autonomy.

Our analysis from Chapter 3 suggests that the different nature of AI-generated knowledge poses a problem for a patient's procedural autonomy. In a sense, even when an AI achieves functioning comparable to that of a human expert, it can never represent the same kind of expertise – it never provides the same aggregate level of skill. This deficiency will be more pronounced in relation to some AI uses than others.

Nevertheless, unless there is a substantial unenvisaged issue with the AI's operation (effectively subjecting the patient to unskilled treatment in much

796 The defendant had variously been introduced as a doctor in the Turkish army who had specialised in facial surgery, a cosmetic surgeon and a nurse: *R v Melin* [2019] EWCA Crim 557, [2019] QB 1063 [14], [16].

the same way as the individuals in *Tabassum* and *Melin*), the resulting autonomy violation would be on a much smaller scale. The patient's ability to reason with the requisite information (their theoretical rationality) may be impaired, but, without more, they are not prevented from achieving their goals through the human-AI hybrid form of assistance.

Ultimately, in cases where there is no aggregate loss of expertise, the weight of the autonomy principle is more limited and, in spite of it receiving increasing attention, it is unlikely to outweigh the entrenched factors that have deterred the courts from imposing liability on professionals in cases of mistaken and confused identity. It would take a considerable development of the case law to establish liability for qualified medical staff. Attention to the autonomy principle is arguably insufficient to bring this about.

3. Non-therapeutic motivations

The final category of cases to be addressed are those that deal with the non-therapeutic motivations of professionals (or purported professionals). In particular, one can see that in cases where the primary motivator for offering the patient an intervention has been sexual or financial, rather than clinical, the courts have had no trouble entertaining claims that the patient's consent was vitiated and a battery perpetrated – even by qualified medical professionals.[797]

From the perspective of legal policy this is understandable and commendable. A concern that an injustice is being done to the defendant is evidently misplaced in such circumstances, nor is it plausible to maintain that a finding of battery would have far-reaching, negative implications on medical practice. However, since these cases are only tangentially related to AI use, they will be dealt with only briefly.

It is true that, in several respects, non-clinical motivations will be relevant to AI decision-making. This was highlighted in Chapter 3, in our discussion of AI's goal-directed action. Irrespective of how concerning these hidden and ancillary motives may be, a claim in battery is unlikely to

797 *R v Williams* [1923] 1 KB 340 and *Appleton v Garrett* (1997) 34 BMLR 23 respectively. The defendant in the former was a singing instructor who had sexual intercourse with his student under the pretence of an operation that would improve her breathing, while the latter concerned a qualified dentist who carried out unnecessary procedures for financial reasons.

succeed because of the aforementioned targeting of this mechanism. In the existing case law, the motive in question is one that is highly personal to the relevant professional, shaping the nature of their actions and providing the basis for their liability. By contrast, where an AI is involved, the non-clinical motive will often be unknown to the AI user, especially if it is sufficiently problematic. Even if such a purpose is known, it will, as was highlighted in Chapter 2, be difficult to assert that divergent AI goals then actually drove the medical decision or supplanted the therapeutic objective of the human professional.

In sum, this class of case seems the least likely to establish liability in battery for non-disclosure because the actor who is targeted by battery (the professional) and the actor who is covertly introducing the non-medical motive into the individual medical decision (the AI and/or the AI developer) come apart. The analogy with this category of case law must break down and it cannot constitute a basis for the duty to disclose AI involvement.

C. Summation

Following the above analysis, there would be one class of case where the tort of battery requires the disclosure of AI-specific information in medical treatment. A duty to disclose arises where the AI alters the nature of the relevant procedure, by introducing a potential, which is not easily foreseen, to diagnose serious conditions in the patient or to bring about other significant effects without their knowledge. If this use of AI and its implications are not disclosed to a patient, then their consent will arguably become invalid and a claim in battery will lie – subject to the fulfilment of the other outlined requirements. This finding is supported by a differentiated application of the autonomy principle.

Regarding the other classes of information, the claimant will have to turn to the negligence action. However, it must be borne in mind that negligence presents a much more restricted means of vindicating patient autonomy. Unlike the tort of battery, negligence requires that damage eventuated and was caused by the defendant's act. Whereas a touching is actionable *per se* and damages can be claimed for all direct consequences in battery, bringing a claim in negligence ordinarily calls for a patient to have suffered physical

harm and it imposes further limitations in terms of causation and the remoteness of damages.[798]

II. Negligence

To construct an obligation to advise patients of AI characteristics under negligence, one must begin with an analysis of the tort's broader elements. As with battery, these will shape its operation in relation to information provision.

Negligence is concerned with the obligation of individuals and organisations to exercise due care and skill in their interactions with others. Specifically, the elements of the action are: 'a duty of care, a breach of that duty and consequent damage'.[799] Giving due weight to the fact that 'consequent damage' concerns both an element of damage and an element of causation, our analysis can be split into four categories. First, an actionable form of damage must be suffered by the claimant. Second, the defendant must owe a duty of care of the requisite scope to the claimant. Third, this duty of care must be breached. Fourth, this breach must have caused the damage. It will be seen that each of these elements has been impacted by the principle of patient autonomy in the disclosure context. In addition, it will be seen that it has influenced the monetary compensation, the 'damages' recoverable under the tort.

It should therefore come as no surprise that, as has been discussed above, negligence constitutes the focus of judicial and academic pronouncements on the tortious obligation of a medical professional to obtain the patient's *informed* consent. The common law of England has developed negligence to impose broad informational duties on individuals and organisations, duties that go far beyond the requirements of obtaining a patient's valid consent to treatment.

Indeed, the common law has been prepared to adapt and stretch its doctrinal integrity to accommodate the principle of patient autonomy. Some argue that this has been to the breaking point, questioning the extent to

798 For a more general discussions of the advantages and benefits of the two claims and their relation to one another see: Feng, 'Failure of Medical Advice: Trespass or Negligence?' (1987) 7(2) Legal Studies p. 149; Brazier, 'Patient Autonomy and Consent to Treatment: The Role of the Law,' (1987) 7(2) Legal Studies p. 169.

799 *Burton v Islington Health Authority* [1993] QB 204, 224.

which informed consent actions can still be classed as negligence actions.[800] The following assesses whether this flexibility would provide an adequate form of protection in the face of the new challenges posed by ML technologies.

It should be highlighted that this action lies in tort. Claims that will not be examined further in relation to English law are those of 'contractual negligence' and claims based on contract more generally. This stems partly from the fact that, in the great majority of cases, AI will be applied to the treatment of patients in the context of the NHS, where there is no contract between patient and provider.[801] It is also informed by the view that the courts will be reluctant to imply that contractual agreements raise the negligence-based standard of care.[802] So it is assumed that a separate consideration of contracts would contribute little to our analysis.

A. Actionable damage

Although often neglected in practical and theoretical argumentation, the requirement that actionable damage must be suffered by the claimant is a precondition for liability in negligence.[803] Unlike battery, negligence is not a tort that is actionable *per se*. As it has been said: damage is the gist of the action.[804] This requires the eventuation of legally specified kinds of harm. Under UK law personal injury, property damage, economic loss

800 Purshouse, 'The Impatient Patient and the Unreceptive Receptionist: Darnley v Croydon Health Services NHS Trust [2018] UKSC 50' (2019) 27(2) Medical Law Review p. 318, 329; Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 381.

801 '[T]he arrangement between doctor and patient under the aegis of the NHS is statutory rather than contractual': *Reynolds v Health First Medical Group* [2000] Lloyd's Rep Med 240, 242. See also *Pfizer Corporation v Ministry of Health* [1965] AC 512, 535-536, 544-545, 548, 552-553, 571. Here the House of Lords outlined how the existence of a statutory obligation, in this case for the provision of prescription drugs, negates the existence of a contractual relationship in the NHS context.

802 Pattinson, *Medical Law and Ethics* (Sixth Edition 2020) 52-53, citing *ARB v IVF Hammersmith* [2018] EWCA Civ 2803, [2020] QB 93 and *Thake v Maurice* [1986] QB 644.

803 '[D]amage is an essential element in a right of action for negligence': *Dulieu v White & Sons* [1901] 2 KB 669, 673. See also Nolan, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review p. 59, 59-61.

804 *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 883-884.

and psychiatric harm currently constitute the well-established categories.[805] By contrast, the courts have been consistent in rejecting distress, inconvenience or discomfort as actionable types.[806]

Without identifying the actionable damage that provides the basis for the claim first, one risks confusing the analysis of later elements.[807] This is particularly true for the current analysis of the law's response to clinical AI's challenges. A traditional category of harm, specifically physical injury, may sometimes accompany problematic uses of ML devices in this context. But it is ultimately an ancillary possibility. In comparison, as will be elaborated, individual autonomy is not a traditionally recognised category of legally cognisable injury. Yet, the coherent protection of the outlined principle will depend upon the judiciary's preparedness to compensate its violation. This will have repercussions throughout the assessment: whether autonomy is compensable as such can affect analyses of duty, breach, causation and damages.

1. Personal injury

It is beyond contention that personal injury is a legally recognised form of damage under UK negligence law. Often this will mean that the law is in a position to address grievous autonomy violations and to offer a redress that is consistent with this value.[808]

In the case of the challenges posed by AI this can be concretised. In a relatively ordinary medical non-disclosure case, it may happen that the use of a specific AI tool poses a risk to the patient that is not properly disclosed

805 Albeit the latter two are subject to additional conditions and thus actionable only in some circumstances, see: Nolan, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review p. 59, 60-61.

806 'If his negligence has caused me neither injury to property nor physical mischief, but only an unpleasant emotion of more or less transient duration, an essential constituent of a right of action for negligence is lacking': *Dulieu v White & Sons* [1901] 2 KB 669, 673.

807 Nolan, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European Tort Law p. 259, 264-265; Purshouse, 'Judicial Reasoning and the Concept of Damage: Rethinking Medical Negligence Cases' (2015) 15(2-3) Medical Law International p. 155, 163-164.

808 '[N]egligence protects autonomy as a second-order value because the kinds of injuries that ground negligence claims almost inevitably have a negative impact on the plaintiff's ability to live the life she would choose to live': Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 357.

(as will be specified in our analysis of the breach element) and that this leads to a situation where the patient suffers precisely that harm, which they were not warned against. The example cited in Chapter 3 was the *Acumen Hypotension Prediction Index Software* giving a false positive reading, which results in an unnecessary intervention. This will straightforwardly interfere with the patient's ability to direct their lives according to their values and, potentially permanently, prevent them from living their life as they had planned.

Alternatively, it may be the case that an ML-related autonomy violation would have weighed heavily enough in a patient's reasoning to alter their conduct. The hypothetically altered decision may have avoided a harm that then eventuated incidentally, potentially entirely unrelated to the AI's functioning. Here too a personal injury claim will protect the patient's autonomy to some extent.

Ultimately, the patient's autonomy can be protected through negligence in such situations because there is a physical harm which flows from the defendant's non-disclosure. This will have repercussions both for the breach element, as certain forms of non-disclosure are more easily associated with the eventuation of physical harm, and for the causation element of negligence, as only certain forms of non-disclosure will be significant enough to bring about a different practical decision.

2. Loss of autonomy

Unlike physical injury, it is not a straightforward task to identify whether the tort of negligence views autonomy as a separate head of damage. As outlined, it is clear that the law recognises multiple categories of damage. However, there is no generally determinative ground for identifying these.[809] Following Nolan, we may say that the technical legal nature of the term 'damage' gives rise to some circularity: the defendant's action must simply violate one of 'various rights (to bodily integrity, property and so on) which are protected against negligent interference'.[810]

809 Nolan, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European Tort Law p. 259, 265-267.

810 ibid 268. This can be related to the position that 'the concept of damage is at the end of the day not a factual, but a normative concept': *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [45].

Established legal doctrine must therefore play a leading role in identifying whether a diminishment of autonomy is a harm that is recognised at law. It would be difficult to make this argument without the support of prior legal documentation. This ensures the condition's role as a controlling and stabilising factor for the negligence action.[811]

In evaluating the relevant documentation, one can begin with the position that there is no well-established common law tradition according to which loss of autonomy is perceived as actionable damage in negligence.[812] Yet this tradition is itself mutable. Purshouse provides the historical examples of harms that have been removed from the accepted catalogue, including: the seduction of a daughter, or the enticement of a wife.[813] Conversely, the UK courts have been prepared to develop a class of recognised psychiatric harm.[814] Since the categories of loss remain fluid, and especially given that the interest in autonomy has gained relatively recent prominence in the bioethical, societal and legal discourse, the absence of a fortified common law position recognising autonomy damage should not be treated as prohibitive for the recognition of a relevant injury. A more recent case of the United Kingdom's highest court, has arguably brought about just such a development, acknowledging individual autonomy as an independent injury and permitting a claimant to bring an action on this basis.

This is the judgment of the House of Lords in *Rees v Darlington Memorial Hospital NHS Trust*.[815] The House of Lords began by rejecting one form

811 Purshouse, 'Autonomy, Affinity, and the Assessment of Damages: ACB v Thomson Medical Pte Ltd [2017] SGCA 20 and Shaw v Kovak [2017] EWCA Civ 1028' (2018) 26(4) Medical Law Review p. 675, 687. Note how such an argument fits with our general observations of private law reasoning in the common law in Chapter 4.

812 Purshouse, 'How Should Autonomy Be Defined in Medical Negligence Cases?' (2015) 10(4) Clinical Ethics p. 107, 107-108; 'the right to make an informed choice is not a right that is traditionally protected by the tort of negligence': *Duce v Worcestershire Acute Hospitals NHS Trust* [2018] EWCA Civ 1307, (2018) 164 BMLR 1 [88].

813 Purshouse, 'Judicial Reasoning and the Concept of Damage' (2015) 15(2-3) Medical Law International p. 155, 158.

814 Compare *Dulieu v White & Sons* [1901] 2 KB 669 and *Victorian Railway Commissioners v Coultas* (1888) 13 App Cas 222. See generally: Law Commission, 'Liability for Psychiatric Illness' (Law Commission Consultation Paper No 137, 1995) 4, 8. Likewise, Priaulx argues that 'the action for wrongful conception (…) clearly demonstrates the law of tort's ability to embrace a widening ambit of harms under its cloak': Priaulx, 'Joy to the World! A (Healthy) Child Is Born! Reconceptualizing Harm in Wrongful Conception' (2004) 13(1) Social & Legal Studies p. 5, 6.

815 *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309.

of actionable damage. Namely, maintaining the position that had been laid down in *McFarlane v Tayside Health Board*[816] that a claimant who had been negligently sterilised was not, when they became a parent, entitled to the costs of raising their child. These costs were not an actionable form of damage.[817] At the same time, the parent was considered the victim of a legal wrong and it was maintained that this wrong would not be adequately remedied by an award that covered only the immediate personal and economic consequences of pregnancy and birth.[818] To respond to this wrong, the majority in *Rees* therefore departed from the earlier *McFarlane* decision – adding 'a gloss' to it, as it was described in the case[819] – and awarded a conventional sum of £15,000 to the parent.[820] Understanding the nature of this sum is crucial for the purposes of establishing whether loss of autonomy constitutes a standalone head of damage.

In order for the award to support the proposition that autonomy is a new actionable head of loss, which can coherently support the remainder of the negligence action, two things must be true: (1) the award must be compensatory (2) the compensatory award must be for a suitable interest in autonomy.[821] The legal position with regard to both will be examined in turn, considering the interpretations and developments that can be found in the subsequent case law.

i. The nature of the award

First, the sum must constitute compensation for a damage (often described as a loss).[822] This has proved somewhat controversial, as it has been argued

816 *McFarlane v Tayside Health Board* [2000] 2 AC 59.

817 I follow Purshouse in this view: Purshouse, 'Judicial Reasoning and the Concept of Damage' (2015) 15(2-3) Medical Law International p. 155, 160-166.

818 *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309 [8].

819 ibid [7], [17].

820 ibid [8], [10], [17], [19]. This picked up upon Lord Millet's suggestion in McFarlane (which was not adopted by the majority) to award a lesser sum of £5,000: *McFarlane v Tayside Health Board* [2000] 2 AC 59, 114.

821 For a similar distinction in the analysis of *Rees*, see the decision of the Singaporean Court of Appeal in: *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [111].

822 As was pointed out, Nolan outlines the technical nature of the term, in particular that it cannot be treated as equivalent with terms such as harm, injury or loss: Nolan, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European

that the award in *Rees* is best conceived of as a vindication of the right to autonomy.[823]

Vindication has a fundamentally expressive – rather than compensatory – function, seeking to affirm certain interests or rights of the claimant.[824] The aim is 'to make it clear to the world, or more precisely to the two parties, that the wrong was a wrong and should never have happened'.[825] If this is the case, then the award and its relation to negligence become highly suspect. The vindication of a right does not presuppose any damage and this calls into question a limitation that is central to this broad tort, with its focus on the provision of compensation for careless behaviour.[826] Any argument that the decision in *Rees* ought to be affirmed or even extended would correspondingly be weakened.

Those who would claim that the conventional award in *Rees* was vindicatory can point to several supporting factors. Quite explicitly, some of their Lordships framed the award by employing the language of rights.[827] However, in our discussion of the autonomy concept we have already

---

Tort Law p. 259, 266. I will continue to use these terms interchangeably, in line with common usage, without intending to imply that a legal damage must always entail a factual loss or injury.

823 Witzleb and Caroll, 'The Role of Vindication in Torts Damages' (2009) 17(1) Tort Law Review p. 16, 38. The same is implicit in Foster's claim that: 'Previously the claimant came into court and said: "The defendant owes me a duty. He has breached it. I have suffered loss. I want compensation." Now the claimant can say instead: "I have a right. It has been violated by you. You should not have violated it. I may or may not have suffered real loss, but I want compensation': Foster, *Choosing Life, Choosing Death: The Tyranny of Autonomy in Medical Ethics and Law* (2009) 127-128.

824 Varuhas, 'The Concept of 'Vindication' in the Law of Torts: Rights, Interests and Damages' (2014) 34(2) Oxford Journal of Legal Studies p. 253, 258-260.

825 Smith, 'Duties, Liabilities, and Damages' (2012) 125(7) Harvard Law Review p. 1727, 1753-1754. Mulligan adopts this approach in the UK context: Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 63.

826 Nolan frames this in terms of maintaining potential defendants' freedom of action: Nolan, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review p. 59, 79. Varuhas also makes the restricted nature of negligence clear: '[Negligence's non-vindicatory nature] explains the focus or emphasis within negligence on actionable damage, the nature of the defendant's conduct, and compensation for factual harm, as well as the generalized nature of the concepts governing liability': Varuhas, 'The Concept of 'Vindication' in the Law of Torts' (2014) 34(2) Oxford Journal of Legal Studies p. 253, 260.

827 For example Lord Millett argued that the 'right to limit the size of their family' is 'increasingly being regarded as an important human right which should be protected

highlighted the imprecision often involved in such statements and that, accordingly, they can hardly be taken to be determinative. More convincingly one can refer to Lord Bingham's statement that 'The conventional award would not be, and would not be intended to be, compensatory. It would not be the product of calculation (…) It would afford some measure of recognition of the wrong done'.[828] It is also noticeable that the dissenting Lord Hope despaired at the fact that the award deviated from common law principle and that the majority's position lacked 'any consistent or coherent ratio'.[829] In combination with the rights language employed, these statements support the view that the orthodox objective of negligence law, compensation, was not the purpose of *Rees'* award.

Another argument to this effect turns on the relationship between the two outlined cases of *McFarlane* and *Rees*. *Rees* expressly affirmed the earlier decision, and this is said to imply that the claimant did not suffer any other compensable loss – that is, loss going beyond that immediately associated with pregnancy.[830] By contrast, if *Rees* found that, exceptionally, one should vindicate the individual's right to autonomy through damages – this issue would lie outside of *McFarlane*'s scope. In other words, a rights-based analysis accords due respect to precedent and presents a more coherent approach.

Beyond an immediate analysis of this case law, commentators have further relied on doctrinal reasons that are related to the fixed nature of *Rees*' award. A compensatory award would usually aim to put the claimant back into the position that they would have been in had the wrong not been perpetrated.[831] As such, it is rather anomalous to hold that there would be a fixed sum for very different kinds of experiences (some of which Lord Bingham outlined).[832] It is hard to maintain that the same award is just

by law': *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309 [123].

828 ibid [8].

829 ibid [74].

830 Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 64-65.

831 *Livingstone v Rawyards Coal Co* (1880) 5 App Cas 25.

832 'The spectre of well-to-do parents plundering the National Health Service should not blind one to other realities: that of the single mother with young children, struggling to make ends meet and counting the days until her children are of an age to enable her to work more hours and so enable the family to live a less straitened existence; the mother whose burning ambition is to put domestic chores so far as possible behind her and embark on a new career or resume an old one. Examples

or appropriate, even if one considers only the relatively narrow band of wrongful conception cases.[833] In comparison, a vindicatory award can readily explain the fixed nature of the conventional award: 'it compensates for the interference with one's autonomy interest, quantum being held constant on the basis that normative damage is assessed objectively'.[834]

Ultimately, given the undeniable difficulty of reconciling vindicatory damages with negligence, the most convincing proponents of the outlined reading of *Rees* argue that it responded to a vindicatory impulse.[835] This leaves an autonomy-based award as something anomalous and exceptional. As proponents appear to admit, this would make it hard to apply to novel situations, even where a strong analogy can be drawn with *Rees*.[836] For our present circumstances it would mean that AI's autonomy infringements would only become actionable *per se* if they led the claimant to have a child that they did not wish to.

This conclusion must be rejected however. Apart from the countervailing statements adduced above, and some admitted variation in their Lordships' statements, the overwhelming tenor of the majority's judgment in *Rees* was compensatory. Lord Bingham spoke directly in terms of the 'real loss suffered' and went on: 'a parent, particularly (even today) the mother, has

---

can be multiplied. To speak of losing the freedom to limit the size of one's family is to mask the real loss suffered in a situation of this kind. This is that a parent, particularly (even today) the mother, has been denied, through the negligence of another, the opportunity to live her life in the way that she wished and planned.': *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309 [8].

833 Witzleb and Caroll, 'The Role of Vindication in Torts Damages' (2009) 17(1) Tort Law Review p. 16, 38. Mulligan provides a summary of the relevant criticisms: Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 64.

834 Varuhas, 'The Concept of 'Vindication' in the Law of Torts' (2014) 34(2) Oxford Journal of Legal Studies p. 253, 269.

835 ibid 269; adopted by Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 65.

836 See especially Mulligan's attempt to extend *Rees* to the relatively similar circumstances of mistakes in reproductive treatment: Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 69, 71. Varuhas and Mulligan both appear to agree that direct and systematic protection of autonomy in such cases would require a standalone action: Varuhas, 'The Concept of 'Vindication' in the Law of Torts' (2014) 34(2) Oxford Journal of Legal Studies p. 253, 270; Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 72-76.

been denied, through the negligence of another, the opportunity to live her life in the way that she wished and planned'.[837] Lord Scott similarly supported 'a conventional sum to compensate', placing 'a monetary value on the expected benefit of which [the claimant] was, by the doctor's negligence, deprived'.[838] Lord Millet framed the matter in terms of 'A modest award [to] compensate for the very different injury to the parents' autonomy'.[839] And it is this characterisation that has primarily been echoed in the literature.[840]

The adduced doctrinal criticisms also seem manageable: a fixed award can be given for a compensated damage in order to render it objective.[841] Doing so may appear anomalous and unjust, but it is not fundamentally subversive or unheard of. In this respect, Purshouse notes how a £200 lump sum was awarded for 'loss of expectation of life' in *Benham v Gambling*,[842] as the calculation in individual cases proved troublesome.[843] He goes on to note that, on a similar account, such an award may be appropriate for autonomy violations where calculating individual losses may cause chaos.[844]

Another commentator has also considered the outcome in *Rees* to be explainable 'on orthodox terms as a fixed amount for non-pecuniary loss'.[845] In support they adduced a further, later case considering this type of approach: *Shaw v Kovac*.[846] Here, the Court of Appeal understood the conventional award to 'mark the injury and the loss', being 'designed to

837 *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309 [8].

838 ibid [148].

839 ibid [125].

840 Nolan, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review p. 59, 79-80; Purshouse, 'Judicial Reasoning and the Concept of Damage' (2015) 15(2-3) Medical Law International p. 155, 166-169; McGregor and others, *McGregor on Damages* (Twenty-First Edition 2021) para 40-300.

841 Varuhas, 'The Concept of 'Vindication' in the Law of Torts' (2014) 34(2) Oxford Journal of Legal Studies p. 253, 269.

842 *Benham v Gambling* [1941] AC 157.

843 Purshouse, 'Judicial Reasoning and the Concept of Damage' (2015) 15(2-3) Medical Law International p. 155, 167. Cf. Todd, who emphasises the point that this is truly unusual and that only the cited case of *Benham* had previously made such an award: Todd, 'Common Law Protection for Injury to a Person's Reproductive Autonomy' (2019) 135 Law Quarterly Review p. 635, 652.

844 Purshouse, 'Judicial Reasoning and the Concept of Damage' (2015) 15(2-3) Medical Law International p. 155, 167.

845 McGregor and others, *McGregor on Damages* (Twenty-First Edition 2021) para 40-300.

846 ibid para 40-300.

compensate the claimant for the loss of the opportunity to live her life in the way she had wished and planned'.[847] In short, the unorthodox, anomalous nature of the conventional award must not be overstated.

The need for consistency with *McFarlane*, and the degree to which a vindicatory approach ensures such consistency, should also not be exaggerated. Clearly the House of Lords in *Rees* saw itself as adding to that judgment and it is no less consistent to recognise a distinct form of loss, which remained underappreciated in *McFarlane*, than to recognise a novel need for the vindication of a right.[848] In both cases the core tenet of *McFarlane* – the denial of the costs for raising the child – stands. This is precisely how the Singaporean Court of Appeal in *ACB v Thomson Medical Pte Ltd* conceived of the award. The claim for compensating autonomy does not fix 'on the liabilities arising out of the care of *the unplanned child* (which is the gravamen of the objection against the award of upkeep) but on *the independent interests of* ***the parents*** *which have been transgressed as a result of the negligent act*'.[849] Clearly, the much larger problem for this evaluative dimension is not the inconsistency that would arise with *McFarlane* if the award were treated as compensatory, but the incoherence that would follow from making a vindicatory award in a tort that is fundamentally structured to be compensatory.[850]

Purshouse also points to another fundamental issue for the rights-based approach: it is difficult to conceive of a right that would need vindication, without first determining that negligence has provided a cause of action.[851] This objection can be concretised by reference to the aforementioned case of *Shaw v Kovac*, where the Court of Appeal held that the common law would require a clear and fundamental right before awarding vindicatory damages.[852] Specifically, drawing on the Supreme Court's judgment in *R*

847 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [80].

848 Consider especially in this regard Lord Steyn's statement that the majority in *McFarlane* had considered and rejected a conventional award – even if it was not explicitly discussed by them: *Rees v Darlington Memorial Hospital NHS Trust* [2003] UKHL 52, [2004] 1 AC 309 [41].

849 *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [108].

850 Nolan persuasively outlined this dimension, arguing that a vindicatory award would be a fundamental challenge to negligence: Nolan, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review p. 59, 79-80.

851 Purshouse, 'Should Lost Autonomy be Recognised as Actionable Damage in Medical Negligence Cases?' (2015) 65-66, commenting on the fact that there is no independent right to autonomy under orthodox negligence actions.

852 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [50]-[55].

(Lumba) v Secretary of State for the Home Department,[853] Davis LJ highlighted the necessity of an egregious violation of constitutional rights for such an award.[854] Whatever developments have taken place, the autonomy interest is still not of a type that would meet these demanding requirements.

In sum, it would be much more controversial to recognise a right to autonomy, which could provide the basis for a vindicatory impulse, than to acknowledge that autonomy is a head of loss that can be compensated according to established principles. The compensatory interpretation represents an incremental evolution that is consistent with the case law and is situated historically and structurally within negligence's wider requirements. In the following, this interpretation of the conventional award will be assumed. As referred to above, this is likely to allow for the wider relevance of this head of damage – including situations involving medical AI. Yet, the feasibility of such an argument must ultimately rest on the nature of the loss of autonomy that is being compensated.

ii. The autonomy interest

The other feature of the damage that must be addressed is its characterisation as an interference with autonomy. *Rees* itself leaves little doubt that the damage was conceptualised as the parents' – specifically, there, the mother's – loss of autonomy. Several quotes from the judgment have already been adduced to this effect, whether one refers to the parents' ability to live their life in the way they planned (as stated by Lord Bingham),[855] or more directly to a denial of their autonomy (as outlined by Lord Millett).[856]

853 *R (Lumba) v Secretary of State for the Home Department* [2011] UKSC 12, [2012] 1 AC 245.
854 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [53].
855 Rees v Darlington Memorial Hospital NHS Trust [2003] UKHL 52, [2004] 1 AC 309 [8].
856 ibid [123].

That this is the type of damage upon which *Rees* was based is also widely reiterated in the commentary[857] and echoed in the cases.[858]

The truly controversial prospect is the scope of the autonomy head of damage. Different positions have been defended in this regard. Some would restrict the new head of loss to the very specific circumstances of *Rees*.[859] Others have argued that it was the specific loss of reproductive autonomy that underlay the award.[860] Others again find that it was explicitly an award for interference with autonomy, albeit it was only a first step towards defining this loss in a manner that would be operational within the law.[861]

857 Nolan, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review p. 59, 78; Purshouse, 'Liability for Lost Autonomy in Negligence: Undermining the Coherence of Tort Law?' (2015) 22(3) Torts Law Journal p. 226, 229-233; Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 67-68; 'Certainly, interference with autonomy can sometimes be compensated in damages': Todd, 'Common Law Protection for Injury to a Person's Reproductive Autonomy' (2019) 135 Law Quarterly Review p. 635, 647. See also Priaulx, 'Joy to the World! A (Healthy) Child Is Born! Reconceptualizing Harm in Wrongful Conception' (2004) 13(1) Social & Legal Studies p. 5, 16.

858 '[The] search for an award to compensate for the 'real loss' culminated in the recognition, in *Rees*, of a novel head of damage: that for a loss of autonomy': *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [107]. See also the claimant's argument in *Khan v Meadows* that the wrongful birth claim at issue 'should not be characterised as pure economic loss but as a mixed claim which combined her loss of autonomy through the continuation of the pregnancy and psychiatric damage incidental to her son's disability as well as her claim for the cost of caring for [him]': *Khan v Meadows* [2021] UKSC 21, [2022] AC 852 [21].

859 Davis LJ appeared to take this approach in: *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [78]-[79].

860 '[I]nterference with reproductive autonomy should be supported as a principled head of damage': Todd, 'Common Law Protection for Injury to a Person's Reproductive Autonomy' (2019) 135 Law Quarterly Review p. 635, 648. Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 371-374. If this is the opinion one takes, that autonomy does or can provide the basis for more specific heads of damage, then the subsequent discussion on AI can also be read in this light. Given the multifaceted nature of its capabilities, there are likely to be circumstances where the technology's use touches on the reproductive choices of parents in artificial treatment. As this argument is not accepted here, this aspect will not be pursued further.

861 Keren-Paz in Barker, Fairweather and Grantham, *Private Law in the 21st Century* (2017). The court in *ACB* interpreted *Rees* in this way: 'This search for an award to compensate for the "real loss" culminated in the recognition, in *Rees*, of a novel head of damage: that for a loss of autonomy': *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [107].

It is arguably the third position that can command the most support from *Rees*. Although the factual circumstances before the court led to a framing of the autonomy interest by reference to reproductive choice and parenthood, the language that was used is clearly applicable to a wider sphere of decision making. For instance, the notion that individuals may be deprived of the ability to shape and live their lives in the way they want, has wider, almost pervasive, relevance to medical decisions. This is precisely what was encapsulated under the reflective dimension of decisional autonomy in Chapter 3. Indeed, looking back to *McFarlane* Lord Millet explicitly stated that the parents have 'lost the freedom to limit the size of their family. They have been denied an *important aspect* of their personal autonomy'.[862] These analyses understand reproductive freedom as particularly significant *instantiations* of autonomy – as important, as capable of determining the course of a life – they do not purport to establish a separate *type*.

Furthermore, a significant indicator that the head of damage must be wider than merely reproductive autonomy was provided by the House of Lords only a few years later in *Chester v Afshar*.[863] As has been noted in Chapter 4, this case concerned the failure of a surgeon to disclose a specific risk to a patient before a surgery was conducted on her spine. The risk eventuated and the patient claimed in negligence , even though she admitted that, had she been warned of the risk, she would still have opted for the surgery at a later date.[864] Although a single rationale is difficult to discern, the majority's reasoning appears to have accepted that there was a breach of duty (in the form of a failure to obtain informed consent) that was connected to the personal injury suffered, just not *via* traditional causation principles.[865] However, given the significance of patient autonomy, their Lordships provided for a limited exception to orthodox causation

862 *McFarlane v Tayside Health Board* [2000] 2 AC 59, 114 (my emphasis).

863 *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134.

864 ibid [7].

865 Referring to a commentator's discussion of a comparable Australian case, Lord Steyn stated: 'Professor Honoré was right to face up to the fact that *Chappel v Hart* — and therefore the present case — cannot neatly be accommodated within conventional causation principles. But he was also right to say that policy and corrective justice pull powerfully in favour of vindicating the patient's right to know': *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [22].

principles to allow for recovery.[866] Ultimately, the patient was compensated for the full amount of her personal injury.

It is hard to avoid the conclusion that the real damage that was suffered in this case was to the patient's autonomy.[867] The House of Lords identified that there was an injury to autonomy, stemming from the physician's failure to sufficiently facilitate the making of a significant medical decision, and sought to provide compensation for it. Autonomy must therefore be a head of damage that is wider than the circumstances in *Rees* or the reproductive autonomy of potential parents.

In light of this, the discussion of causation between the breach and the personal injury should have been superfluous. As Keren-Paz has noted, the patient suffers an interference with autonomy (the damage) by being denied an opportunity to consent (in an informed manner) and there is no issue in finding a causal connection with the breach here.[868] The absence of a conventional award and the overcompensation for the personal injury in *Chester* then appear puzzling – why did the court not simply apply *Rees* to this situation?[869]

Several reasons may be adduced. Simply the fact that the eventuation of damage is rarely examined at any length and often subsumed under the other elements of negligence played its role. The cases' purported focus on, respectively, the quantification of damages and the causation of harm obscured any common ground between them.[870] This confusion was argu-

866 Lord Steyn maintained that the patient's 'right of autonomy and dignity can and ought to be vindicated by a narrow and modest departure from traditional causation principles': ibid [24]. To similar effect Lord Hope stated: 'The function of the law is to protect the patient's right to choose. If it is to fulfil that function it must ensure that the duty to inform is respected by the doctor. It will fail to do this if an appropriate remedy cannot be given if the duty is breached and the very risk that the patient should have been told about occurs and she suffers injury': ibid [56].

867 Amirthalingam, 'Causation and the Gist of Negligence' (2005) 64(1) The Cambridge Law Journal p. 32, 34-35. Clark and Nolan outline the advantages of this view: Clark and Nolan, 'A Critique of Chester v Afshar' (2014) 34(4) Oxford Journal of Legal Studies p. 659, 684-685, 688-689. See also: Keren-Paz, 'Compensating Injury to Autonomy in English Negligence Law: Inconsistent Recognition' (2018) 26(4) Medical Law Review p. 585, 592-593.

868 Keren-Paz, 'Compensating Injury to Autonomy in English Negligence Law' (2018) 26(4) Medical Law Review p. 585, 592-593.

869 Clark and Nolan, 'A Critique of Chester v Afshar' (2014) 34(4) Oxford Journal of Legal Studies p. 659, 684; *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [114].

870 Purshouse, 'Judicial Reasoning and the Concept of Damage' (2015) 15(2-3) Medical Law International p. 155, 157.

ably amplified by the fact that a considerable form of personal injury had occurred in *Chester*, representing a traditional, readily accepted category of damage. It may have appeared convoluted to engage in a reframing of this loss. It has also been argued that the harm suffered in *Rees* was undercompensated because of its gendered nature.[871] Without such specific factors in play in *Chester*, the House of Lords approached the case differently and compensated the loss of autonomy more appropriately.

Given the diverging paths taken by the UK's highest court in these cases, and without a clear rationale, it is arguably not possible to resolve this inconsistency in the law. However, from a perspective that takes the autonomy principle seriously, the most cogent interpretation of *Chester* is that the House of Lords built upon *Rees*. The court thereby recognised an injury in the diminishment of the patient's autonomy, yet opted to compensate it with damages for personal injury.[872]

This indicates that UK law provides a patient with the ability to bring a claim on the basis of an interference with their general autonomy interest. However, it must also be recognised that a definition of this interest has not been forthcoming and lower courts have since expressed a reticence to invoke it. In particular, the Court of Appeal judgments of *Shaw v Kovac* and *Duce v Worcestershire Acute Hospitals NHS Trust* purport to limit the relevance of *Rees* and *Chester* severely. Both decisions refused to recognise an autonomy loss *per se* as a form of damage.[873] Moreover, they purported to limit both cases to their facts or, at least, to a very specific fact pattern.[874]

---

871 Priaulx, 'Joy to the World! A (Healthy) Child Is Born! Reconceptualizing Harm in Wrongful Conception' (2004) 13(1) Social & Legal Studies p. 5, 10-15.

872 Amirthalingam has argued that the 'the physical injury, which turned on the patient's response, merely went to the quantification of the loss': Amirthalingam, 'Causation and the Gist of Negligence' (2005) 64(1) The Cambridge Law Journal p. 32, 33-34. Keren-Paz argues for a similar resolution of this issue, although our approaches as to how the new head of damages should then be defined within acceptable bounds differ: Keren-Paz in Barker, Fairweather and Grantham, *Private Law in the 21st Century* (2017) 428-437. Note also how crucial the difference between actionable damage and damages becomes in this analysis.

873 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [58]-[74]; *Duce v Worcestershire Acute Hospitals NHS Trust* [2018] EWCA Civ 1307, (2018) 164 BMLR 1 [88].

874 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [61], [79]; *Duce v Worcestershire Acute Hospitals NHS Trust* [2018] EWCA Civ 1307, (2018) 164 BMLR 1 [87]-[90]. The latter case especially highlighted the exceptional nature of *Chester v Afshar* and the role that policy factors played in determining an unorthodox approach.

From a purely formal perspective these cases cannot overturn the above decisions – they remain bound by the higher courts' judgments.[875] If *Rees* and *Chester* do carve out an autonomy interest that is actionable, as it has been argued, then the lower courts' refusal to apply it constitutes a practical problem. The Court of Appeals' findings would be more troublesome for the present analysis if they appealed to broader, persuasive normative grounds for rejecting the autonomy claim. But such arguments were not explicitly provided. The *ratio* of these cases has been interpreted to be very narrow.[876] Furthermore, the broader, *obiter* argumentation is undoubtedly of a questionable calibre.[877]

In fact, the strongest normative case for rejecting the view that an independent autonomy injury was created, is provided by the aforementioned case of the Singaporean Court of Appeal: *ACB*. Although this only constitutes persuasive authority in England, the court arguably detailed criticisms that were implicitly touched upon in *Shaw* and which have carried great weight with British commentators.[878]

In essence, one can identify two kinds of objections in this judgment: one conceptual argument, that goes towards the nature of the legal (non-)identification of autonomy, and one class of more principled objections: going towards the compatibility of a moral-/political- concept with the structure of the common law, and specifically the requirement of damage in negligence. The assumption is that a narrower aspect of that concept (such as reproductive autonomy) better satisfies the requirements

875 Leggatt LJ highlighted that the Supreme Court may need to reconsider this area of the law: *Duce v Worcestershire Acute Hospitals NHS Trust* [2018] EWCA Civ 1307, (2018) 164 BMLR 1 [92].

876 For a full argument see: Keren-Paz, 'Compensating Injury to Autonomy in English Negligence Law' (2018) 26(4) Medical Law Review p. 585, 599-602. Similarly Todd notes that Shaw 'holds at least that the notion of injury to autonomy as a new head of loss in a personal injury claim should not be supported, but that says nothing about the implications of the decision in *Rees*': Todd, 'Common Law Protection for Injury to a Person's Reproductive Autonomy' (2019) 135 Law Quarterly Review p. 635, 648.

877 Purshouse, 'Autonomy, Affinity, and the Assessment of Damages' (2018) 26(4) Medical Law Review p. 675, 681-684.

878 ibid 684-687; Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 364-371; Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 69-70.

of conceptual clarity and (partially as a result) can better fulfil its function as a head of damage.[879]

Here we need not devote too much time to the basis for the conceptual objection, given Chapter 4's extensive treatment of the issue. It was seen that different types of autonomy are discussed in the legal sources, as well as in the wider literature, and/or are incorporated into them. What is notable about the specific criticism of Phang JA, which is echoed and amplified by Nolan,[880] is the sense that arriving at a sufficiently settled definition is impossible for autonomy.[881] The positions reflected in the law are simply too diverse and philosophically and politically contested.[882]

Curiously, there is also a sense that, if a sufficiently settled legal definition could be reached, then it would be a highly individualistic conception of autonomy that would contribute to the second, principled objection.[883] In particular, it would reinforce the argument that one cannot objectively assess whether a damage has eventuated, given the inherent subjectivity of autonomy.[884] I will address these different objections in turn, in light of the procedural principle of autonomy that has been adopted throughout this work.

To begin with, one must acknowledge that, undoubtedly, there is not one uncontested concept of autonomy in UK law. Yet, this does not mean that the common law could not adopt a defensible understanding of its nature and influence, providing an adequate degree of fit with the legal material.

879 On the interrelation between the two types of criticism: *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [119].

880 Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 364-366.

881 *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [116]-[119].

882 ibid [119].

883 Purshouse explicitly notes that in English law 'it is the current desire version of autonomy (…) that the courts presently use in a number of contexts': Purshouse, 'How Should Autonomy Be Defined in Medical Negligence Cases?' (2015) 10(4) Clinical Ethics p. 107, 111. This approach was relied on in *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [120].

884 Both Purshouse and ACB note that an autonomy violation may (objectively) leave a person no worse, no different, or even better, off: Purshouse, 'Liability for Lost Autonomy in Negligence' (2015) 22(3) Torts Law Journal p. 226, 237; *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [120]. Similarly in *Shaw*, it was found objectionable that a patient could recover for damage to autonomy even where the operation 'was a complete success': *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [71]. Nolan provides a convincing refutation of this argument, citing both the fact that damage in negligence does not require being worse off and that it ignores the possibility of the autonomy interest itself being the damage: Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 364.

Chapters 3 and 4 offered one coherent concretisation in this regard. Conceived of as a legal principle, a procedural conception of autonomy was argued to command considerable support in the legal system – particularly with a view to the medical context and the issues of informed consent. As we will return to below, adopting such a definition does not allow for the delineation of a qualitatively distinct concept of reproductive autonomy.

To respond to the objection that autonomy does not fit easily into the damage element, an established legal category, one ought to begin by rejecting the assumption that the law must rely on a purely subjective account of autonomy. This is precisely the kind of conception that generates the problems to which the proponents of a restrictive approach then object. Keren-Paz has already made the argument that autonomy is not self-evidently so subjective as to prevent the identification of an interference.[885] Verifiable cases of significant violations patently do exist, and this is obscured by both Phang JA's and Purshouse's reference to an allegedly indivisible class of autonomy violations that do not meet a *de minimis* threshold of seriousness.[886] The task therefore becomes the identification of a yardstick by which the law can limit itself to addressing only these interferences.

As has been pointed out by Chico, an understanding of autonomy that incorporates objective or ideal elements is much more suited to this task.[887] This is true of our theoretical account. Autonomy violations were to be assessed by reference to objective, or objectively verifiable, elements, including: the need for certain classes of information (necessary true beliefs) and the centrality of the patient's established commitments and character under the reflective dimension of autonomy. To make out the significance of an interference, the patient should be able to adduce these.

885 Keren-Paz in Barker, Fairweather and Grantham, *Private Law in the 21st Century* (2017) 432-433.

886 *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [120]; Purshouse, 'Liability for Lost Autonomy in Negligence' (2015) 22(3) Torts Law Journal p. 226, 234-242.

887 Especially: 'the English courts would be more willing to recognize that a novel interest, namely autonomy, might form the basis for legally recognized harm where the interest is amenable to limitation by the courts. An ideal conception of autonomy as imbued with rationality, as opposed to a liberal conception which rests simply on capacity and independence, would allow greater analytical purchase on what autonomy actually consists in, thereby providing potential boundaries to the legal recognition of the interest. In this way, the courts would be able to prioritize those interferences with autonomy which they perceive to be the most deserving of legal recognition': Chico, *Genomic Torts: The English Tort Regime and Novel Grievances* (2010) 47-49.

Moreover, as Keren-Paz has argued, the different consequences of interferences for a patient's practical decision will also shape its significance: if a patient would have made a different decision (as in *Rees*) then the interference is arguably stronger, if they were denied the opportunity to reflectively affirm that position (as in *Chester*) it is arguably weaker.[888] Such an approach tracks the distinction drawn in Chapter 3 between the practical element of autonomy and the decisional one. Ultimately, our theories' components, and specifically English law's understanding of them, seek to provide a yardstick for the objective identification of significant autonomy violations. If this is accepted, then the necessary legal specification of autonomy damage is not an insurmountable hurdle.

The courts' and commentators' treatment of reproductive autonomy also suggests as much. For, there is a sense that *Rees* was defensibly decided for the purposes of remedying a specific injury to reproductive autonomy.[889] Hence it is puzzling how critics envision the more specific aspect of autonomy to overcome the conceptual and objectivity objection. Does it really assist in an objective evaluation to say that autonomy can 'underlie'[890] or 'be used as a justification for'[891] a specific head of damage. A definitional choice is still needed. This should be made consistently and it should comport with the requirements of the law.

The only difference in cases involving reproductive decisions is that judges and commentators can be relatively confident that there has been an interference with the patient's deepest desires,[892] their long-term plans,[893] etc. There is a relatively well-documented process (sometimes involving a contraceptive intervention) leading up to their decision and capturing their commitments. If this is the approach to autonomy, which the present author would endorse as an outgrowth of its reflective dimension, then a limitation to reproductive choice is unsatisfactory and, beyond allusions to

888 See Keren-Paz's analysis on this issue: Keren-Paz, 'Gender Injustice in Compensating Injury to Autonomy in English and Singaporean Negligence Law' (2019) 27(1) Feminist Legal Studies p. 33, 44-45.

889 The court in *ACB* were not bound to follow *Rees*, and distinguished it in other ways, but still argued for an award based on an aspect of autonomy: *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [125]-[130].

890 ibid [115].

891 Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 371.

892 Keren-Paz in Barker, Fairweather and Grantham, *Private Law in the 21st Century* (2017) 427.

893 ibid 415.

practicality, no rationale for it is forthcoming.[894] As Austin has highlighted: where a relevant clinical decision involving non-disclosure is brought before a court, the claimant's position must be presented, justified and analysed in a way that avoids attributing hindsight to them.[895] If they can do this outside of the reproductive context, then it should not be for the law to say that there loss is not sufficiently 'objective'.

If an objective delineation of autonomy is possible, then one must further consider its interaction with established rule-specific categories. For example, commentators have pointed out the challenges of disentangling a violation of autonomy from other kinds of damage – the concern here is the prospect of double recovery.[896] However, double recovery is only objectionable if autonomy is not seen as a harm in itself, which proponents of autonomy would deny, or if courts mistakenly overcompensate it, as occurred in *Chester*.[897] In addition, this consequence is not inevitable. As Keren-Paz has highlighted (and Nolan submits) it is open to the court to amend *Chester* and to refuse additional recovery for autonomy where an existing head of damage is already compensated.[898] We will return to this discussion below, in our assessment of available damages.

In the final analysis it is therefore argued that our development of the autonomy principle affirms the position that the English negligence action provides a compensatory award for certain, significant and verifiable, violations of patient autonomy. Nevertheless, in light of the House of Lords'

894 '[W]hereas previously autonomy was only protected indirectly by the law of negligence (via claims for personal injury and so forth), increasingly loss of autonomy appears to be being accorded recognition as a form of actionable damage in its own right. (…) Since autonomy is a very important interest, this development is to be welcomed, although if negligence liability is to be kept within acceptable limits, protection can only be accorded, as at present, to certain derivative autonomy interests – freedom of movement, reproductive autonomy and so forth – rather than to autonomy in the round': Nolan, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review p. 59, 87.

895 Austin, 'Correia, Diamond and the Chester Exception: Vindicating Patient Autonomy?' (2021) 29(3) Medical Law Review p. 547, 559.

896 Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 366. See also *ACB v Thomson Medical Pte Ltd* [2017] SGCA 20 [123]-[124].

897 Clark and Nolan, 'A Critique of Chester v Afshar' (2014) 34(4) Oxford Journal of Legal Studies p. 659, 684.

898 Keren-Paz, 'Compensating Injury to Autonomy in English Negligence Law' (2018) 26(4) Medical Law Review p. 585, 590 (fn. 37), 600-601. Much depends on how the autonomy concept is shaped and limitations can still be imposed at the duty of care stage: Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 367.

inconsistent rulings and the Court of Appeal's hesitancy to recognise and implement such an approach, its strength is not beyond question. Proactive steps by the courts, most likely even an intervention by the UK's Supreme court, would be necessary to clarify the legal position. At the same time, given how recently the autonomy interest has emerged as a significant influence on UK jurisprudence,[899] it should also not be surprising that a concrete approach has yet to be fully specified. For now, our autonomy concept offers one defensible interpretation of the still open-textured legal position.

3. Summation

Going forward it will be argued that UK law can compensate a patient for significant forms of autonomy violations, as well as for the personal injury, that they suffer as a result of AI use. At points it will be necessary to distinguish between the two forms of damage, since it is beyond doubt that a claim for personal injury is more likely to succeed. As our discussion of *Chester* has demonstrated, the recognition of an autonomy interest will also shape the other elements of the analysis, including causation. Overall, loosening the law's insistence on physical damage will allow for a more coherent remedy to be provided in response to violations of the patient's informed consent.

B. Duty of care

A duty of care specifies those relationships where purely inadvertent and unintentional conduct constitutes an appropriate basis for liability. As such it is an integral element in the delineation of the scope of liability of a negligence action. This is encapsulated by the fact that the existence of a duty of care is based primarily upon established categories of case, where a class of defendants has already been found to owe a duty with respect to a certain class of claimants. If the relevant constellation falls within such a category, then this is normally the end of the analysis: a duty of care is

899 'As recently as the 1970s, for example, commentators framed the issue of "informed consent" in the medical context as a question of dignity, rather than autonomy': Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 382.

established. Alternatively, if the court is confronted with a situation that falls outside of existing classifications, then a duty of care may still be made out, but only *via* incremental development. Here the duty is extended beyond existing categories on the grounds that it would be fair, just and reasonable to do so.[900]

In the following we will consider the duties owed by two classes of defendants with respect to AI use: medical professionals and healthcare institutions. These are the actors who are anticipated to instrumentalise AI in the English healthcare system and who are likely to be the primary targets for negligence actions concerning the provision of medical information. Parties that also owe some informational duties of care to the patient, but which will not be investigated further below, are AI developers and manufacturers. Rather than conducting a separate analysis of such duties, this aspect of the designers' responsibility will be considered only in so far as it shapes professional and institutional breaches.[901] This is justified by the fact that the manufacturer's informational, facilitative role to the patient is mediated by healthcare professionals and institutions and that their obligations are more narrowly defined: to ensure that the user of a product is adequately warned of its risks.[902] It is not arguable that these duties can be extended to cover the unique properties of ML devices.

900 *Caparo Industries Plc v Dickman* [1990] 2 AC 605. That this is the correct order of consideration, focussing first on the established situations where a duty is owed and then potentially seeking to expand upon these incrementally, was stated *obiter* in *Michael v Chief Constable of South Wales Police* [2015] UKSC 2, [2015] AC 1732 [102]-[111] and affirmed in *Darnley v Croydon Health Services NHS Trust* [2018] UKSC 50, [2019] AC 831 [15]-[16].

901 This approach finds the strongest support in existing case law 'Where the treatment involves a medical device there is a duty on the doctor to apprise themselves of the implications of the use of that device together with any risks associated with its use. That again is part of the clinician's duty to the patient but its importance is underlined by the doctor's role as a learned intermediary between the producer of the product and the patient. The producer has no relationship with the patient who relies on the doctor to advise them of the risks associated with it': *AH v Greater Glasgow Health Board* [2018] CSOH 57, (2019) 169 BMLR 120 [61]. *AH* also cited another recent authority that dealt with the disclosure of information required from the manufacturer of a prosthetic hip. Here the court held that 'Parliament has determined that, in relation to such products, information about risks is best relayed to, considered by, and applied and passed on to the patient by the treating surgeon, who must advise that patient as to intervention choices, and seek and obtain that patient's informed consent the particular chosen implant procedure': *Wilkes v DePuy International Ltd* [2016] EWHC 3096 , [2018] QB 627 [107].

902 Goldberg, *Medicinal Product Liability and Regulation* (2013) 58-61.

1. Medical professionals

Regarding medical professionals, it is trite law that they owe duties to their patients.[903] And it will be seen that, beyond any doubt, these duties encompass demanding informational components to those they are diagnosing, advising and treating.

If a medical professional utilises an ML device, then an enquiry into their informational responsibilities does not call for a reconsideration of the existence of this duty. Rather, it problematises the actions that are necessary, with respect to that technology, to ensure compliance with it. It is primarily a question of whether that duty was breached.

UK courts do not appear to have limited this obligation to a particular class of medical professionals. They have considered whether the treatment team as a whole has obtained the necessary consent from the patient. For example, in *Lybert v Warrington Health Authority* the patient's gynaecologist was found negligent for a failure to warn, although he was not the only individual whom the claimant had 'sought to criticise' and although he was 'entitled to assume, or at least to expect, that an initial warning had been given somewhere along the line'.[904] Similarly, it was stated in *Montgomery v Lanarkshire Health Board* that:

> a wider range of healthcare professionals now provide treatment and advice of one kind or another to members of the public, either as individuals, or as members of a team drawn from different professional backgrounds (with the consequence that, although this judgment is concerned particularly with doctors, it is also relevant, *mutatis mutandis*, to other healthcare providers).[905]

Consequently, where ML devices are utilised in the care process, there may be an obligation on several professionals to obtain the patient's informed consent and, in this respect, no straightforward delineation can be drawn between them at the duty stage.[906]

---

903 *Barnett v Chelsea and Kensington Hospital Management Committee* [1969] 1 QB 428.

904 *Lybert v Warrington* (1995) 25 BMLR 91, 95.

905 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [75] (my emphasis).

906 As will be discussed under the breach element, liability must of course still depend on the knowledge that the professional in question did have or should reasonably have had.

2. Healthcare institutions

Regarding healthcare institutions, the position is admittedly more complex since there are several routes by which they can be held legally responsible. The nature of organisational liability, and its potential significance in situations where organisations opt to rely on AI, therefore bears more detailed elaboration.

One route for establishing organisational liability in negligence takes the form of vicarious liability. At least since *Cassidy v Ministry of Health* and *Roe v Minister of Health* it has been established that healthcare providers can be held liable for the actions of their medical and non-medical staff in this fashion.[907] In such an instance it is the relationship between two identifiable individuals, such as healthcare professional and patient, that founds the duty. The organisation is then held strictly liable for the negligence of the tortfeasor because the two are connected in the requisite manner.[908] This aspect of liability should be noted for its practical relevance but need not be dwelled on here: if informational duties are owed by professionals in respect of AI, then there is no reason the technology would prevent the employer from being held vicariously liable as well.

In addition to vicarious liability, liability may alternatively be established on the basis that an institution owes a direct duty of care to the patient. For this mechanism, accountability does not depend on the mediation of an identifiable human professional with a relevant duty.[909] For example, it is settled law that a healthcare institution can be held directly liable where there is some organisational or managerial failure; where the negligence can be ascribed to 'no one and nothing but the system itself'.[910] Thus in *Bull v Devon Area Health Authority* the system for summoning obstetricians

907 *Cassidy v Ministry of Health* [1951] 2 KB 343; *Roe v Minister of Health* [1954] 2 QB 66.

908 Deakin, 'Organisational Torts: Vicarious Liability Versus Non-Delegable Duty' (2018) 77(1) The Cambridge Law Journal p. 15, 17-18.

909 Syrett distinguishes between direct and non-delegable duties: Syrett in Laing and McHale, *Principles of Medical Law* (Fourth Edition 2017) 379. I have in mind the former.

910 *Bull v Devon Area Health Authority* (1989) 22 BMLR 79, 100. In a similar vein Lord Phillips MR, in another case, spoke of 'A duty to use reasonable care to ensure that the hospital staff, facilities and organisation provided are those appropriate to provide a safe and satisfactory medical service for the patient': *A (a child) v Ministry of Defence* [2004] EWCA Civ 641 [32].

for the provision of maternity services was found wanting and a negligence claim against the hospital was made out.[911]

As addressed in Chapter 2, AI implementation will generally be premised on a degree of expert mediation. Typically, informational duties would therefore fall to be consider under the duties owed by human professionals. Primary institutional liability will only gain relevance in a restricted set of circumstances. Namely, where the institution deploys AI to determine an aspect of the patient's medical care. The example offered for this in Part I. was the use of AI in triage. ML devices would be giving patients medical and/or organisational data – such as a preliminary assessment of their condition and information on which further healthcare services to access (if any) and how to access them. If any informational duties are to be asserted here, then the role of the corporate person becomes paramount.

A recent case of the UK Supreme Court, *Darnley v Croydon Health Services*, suggests that institutions do owe some such duties.[912] Specifically it was held that providing the patient with information on the timeframe within with which they could expect to be seen in an accident and emergency department fell within an NHS Trust's primary duty.[913] This case emphasised the corporation's general responsibility to ensure that there is an adequate system of information provision in place. In this respect, Lord Lloyd-Jones JSC expressed agreement with a dissenting judge of the Court of Appeal, McCombe LJ, indicating that: 'The failure to impart the reality of the triage system to the claimant on his arrival was, on the facts of this case, a breach of duty by the hospital'.[914] It was seemingly irrelevant that the actionable misinformation – that the patient would be seen in four-to-five hours by a doctor, rather than in 30 minutes by a triage nurse – was the result of an individual receptionist's mistake and it was explicitly stated that the duty could be discharged *via* medical staff, non-medical staff, pamphlets or notices.[915]

911 *Bull v Devon Area Health Authority* (1989) 22 BMLR 79.

912 *Darnley v Croydon Health Services NHS Trust* [2018] UKSC 50, [2019] AC 831. The Supreme Court also drew on an earlier Court of Appeal finding that an ambulance driver was liable for provided misleading assurances concerning the arrival time of an ambulance: ibid [18]-[20], citing *Kent v Griffiths (No.3)* [2001] QB 36.

913 *Darnley v Croydon Health Services NHS Trust* [2018] UKSC 50, [2019] AC 831 [16] - [17], [21].

914 ibid [13], [19].

915 ibid [26]-[27].

Consequently, an institution owes a direct duty to patients that covers one aspect of information provision. However, it would be a mistake to equate this with the obligation of medical professionals to obtain informed consent.[916] At the risk of conflating the duty and breach analyses, it is worth clarifying already at this stage that the obligation defined in *Darnley* is not suited to the task of alerting the patient to the unique aspects of AI treatment.

Such a limitation emerges from the Supreme Court's focus on information with the potential to cause physical injury.[917] The court explicitly framed the duty to be 'one to take reasonable care not to cause physical injury to the patient'.[918] Similarly, it was held that 'it is not the function of reception staff to give wider advice or information in general to patients', the NHS Trust must simply 'take care not to provide misinformation to patients'.[919] It appears that it was the close relationship between the provision of misleading information and a direct danger to the patient's health that shaped the Supreme Court's finding.

In the case of an ML triage system, such a relationship is arguably lacking, or exists only to a very limited extent. Indeed, it is where a patient is not using an AI, that it is arguable that a healthcare institution must alert the patient to the availability of such an automated triage system, which could enable faster access to the necessary care. In this sense the ML device would assume the role of the triage nurse to whom the patient in *Darnley* ought to have been alerted.[920] By contrast, it cannot be maintained that a patient who uses an AI, or is prepared to do so, must be advised of its specific risks or even its risk profile – never mind its ability to influence value judgments or its relationship to human expertise. To enable the patient to avoid suffering physical harm it would clearly suffice to warn them not to treat the AI output as definitive and, if concerned, to seek further medical assistance. As with AI manufacturers' and developers' duty to warn, such an obligation is a patently unsatisfactory basis for the promotion of patient autonomy and informed decision making.

916 Purshouse, 'The Impatient Patient and the Unreceptive Receptionist' (2019) 27(2) Medical Law Review p. 318, 329.

917 For a similar framing of the matter see: Armitage, Charlesworth and Percy, *Charlesworth & Percy on Negligence* (Fifteenth Edition 2022) para 10-178.

918 *Darnley v Croydon Health Services NHS Trust* [2018] UKSC 50, [2019] AC 831 [16].

919 ibid [19].

920 ibid [26].

Limited support for a marginally more extensive institutional disclosure duty can be found in the earlier case of *Lybert v Warrington Health Authority*. Here the patient had not been warned of the risk that a sterilisation procedure performed on her may not have been effective and Otton LJ was prepared to find the health authority primarily liable in the circumstances.[921] It was held that there was 'a duty upon those responsible for the conduct of this unit to ensure that there was a proper and effective system for giving a proper warning at some stage during her time as a patient'.[922]

This duty is arguably broader than the one found in *Darnley*. It requires the affirmative provision of a class of information which did not present an immediate danger to the patient's health. However, it is again noticeable just how narrowly the duty is framed. It essentially constitutes a duty to warn of specific risks and therefore still maintains a strong connection to the patient's physical well-being. As will be seen below, a duty to disclose AI's specific risks does not in any way begin to address the technology's novel autonomy challenges. Furthermore, to fulfil this obligation, it was sufficient to avoid 'having a lax system in place that allowed the plaintiff to "slip though the system"'.[923] In consequence, *Lybert* does not do much to advance the argument that institutions bear responsibility for facilitating the patient's informed decision making. Instead, they primarily bear responsibility for the protection of their health.

In conclusion, there is some basis for finding that, in England, institutional liability does extend to minimal informational obligations, at least in so far as they are relevant to the protection of the patient's physical well-being. If appropriately developed, direct institutional liability could provide one useful legal mechanism for imposing obligations on institutions, which are utilising relatively independent ML models, to let patients know that they are being processed by an AI with certain goals, capabilities etc. Given the recent changes to the breach element in non-disclosure cases, there is a possibility that a recalibration may occur in the future. However, currently there are no existing indications that the autonomy principle has shaped the evolution of this duty. Quite the opposite: it is a concern for the patient's beneficence that has determined a narrow framing of relevant obligations. In light of this, it would be difficult to construct a disclosure

921 *Lybert v Warrington* (1995) 25 BMLR 91, 92.
922 ibid 95.
923 Grubb, 'Failed Sterilisation: Duty to Provide Adequate Warning' (1995) 3(3) Medical Law Review p. 297, 298.

obligation that could ensure the protection of the patient's autonomy in the AI context.

3. Summation

For most uses of AI, the duty we should consider is that of healthcare professionals providing their services to a patient with the aid of an ML device. Here it is well established that informational duties are owed to patients. The existence of a duty is certainly not called into question by reliance on a new technology, although it may shape its discharge. By contrast, in the scenario where an AI device functions systematically as an information conveyer, determining an aspect of the patient's care, it has been argued that the institutional provider owes only limited direct duties of care and that these are not suitable vehicles for the protection of the patient's autonomy.

C. Breach

When a duty has arisen, the next question goes towards the circumstances under which the relevant party will fall short in complying with it. It is a defining aspect of negligence liability that the defendant is not strictly liable but must, in their actions, meet a certain standard of care. This standard has long been defined by reference to the reasonable person:

> Negligence is the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do.[924]

In other words, in order to succeed a claimant must usually show that a defendant fell short of the standards expected of this objective reasonable person.

Given the difficulty of judging the reasonableness of the actions of specialist actors, the British courts have traditionally provided a deferential test

924 *Blyth v Birmingham Waterworks Co* (1856) 156 ER 1047, 1049.

to make this assessment in clinical negligence.[925] Under the seminal case of *Bolam v Friern Hospital Management Committee* it is enough for a medical professional to show that they acted in accordance with one responsible body of medical opinion to demonstrate that they met the reasonableness standard.[926] Following *Bolitho v City and Hackney Health Authority* one may add the caveat, in line with the reasonableness standard, that the professional position must withstand logical scrutiny by the courts.[927]

The specific question for our purposes is: what kind of information must be provided about the involvement of AI in a patient's care, in order to avoid breaching a healthcare professional's duty of care? Here, the relevant obligations have been, and continue to be, subsumed under the ordinary negligence action.[928] However, the value of patient autonomy has brought about an alteration in the standard of care governing cases that concern the patient's informed consent. This section defines this standard, identifies its requirements in operation and applies these to medical AI/ML.

1. The informed consent standard

The prevailing standard of care in nondisclosure cases was, for a long time, the same as the aforementioned *Bolam* standard applicable to medical negligence in general, establishing a uniform approach. This was the finding in *Sidaway v Board of Governors of the Bethlem Royal Hospital*.[929]

925 As has been argued: this standard 'does not represent a departure from the ordinary standard of the reasonable person, but is merely an attempt to apply that standard in the light of the fact that judges lack the knowledge and expertise usually required to choose between competing professional opinions': Nolan, 'Varying the Standard of Care in Negligence' (2013) 72(3) The Cambridge Law Journal p. 651, 654-655.

926 *Bolam v Friern Hospital Management Committee* [1957] 1 WLR 582.

927 *Bolitho v City and Hackney Health Authority* [1998] AC 232, 243.

928 It has been argued however that the novel emphasis on the autonomy of the patient in the definition of the standard of care (see *infra*) should be recognised as having created a *sui generis* cause of action: Nolan, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European Tort Law p. 259, 376-382; Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 71-74. Even if this would be a preferable way to frame the courts' anomalous approach to medical non-disclosure cases, this is not purported to represent current practice.

929 'In English jurisprudence the doctor's relationship with his patient which gives rise to the normal duty of care (…) has hitherto been treated as single comprehensive duty covering all the ways in which a doctor is called upon to exercise his skill and

Only in 2015 did the Supreme Court, in *Montgomery v Lanarkshire Health Board*, overrule this established norm on the strength of the autonomy principle.[930] By appealing to the patient's need for material information, this case established a patient-centred standard of disclosure, under which the courts are 'developing separate rules to those governing the rest of medical negligence'.[931]

i. The meaning of reasonable disclosure

Montgomery v Lanarkshire Health Board sets the current standard for the steps that a health professional must take to advise a patient about the nature of their treatment. In an oft-cited passage, the majority held:

> An adult person of sound mind is entitled to decide which, if any, of the available forms of treatment to undergo, and her consent must be obtained before treatment interfering with her bodily integrity is undertaken. The doctor is therefore under a duty to take reasonable care to ensure that the patient is aware of any material risks involved in any recommended treatment, and of any reasonable alternative or variant treatments. The test of materiality is whether, in the circumstances of the particular case, a reasonable person in the patient's position would be likely to attach significance to the risk, or the doctor is or should reasonably be aware that the particular patient would be likely to attach significance to it.[932]

This statement encapsulates the remarkable shift in the influence attributed to patient autonomy in negligence actions dealing with issues of informed consent. A concern for this principle led to a transformation from a defendant friendly standard of care – where breach was analysed primarily by

---

judgment (…) This general duty is not subject to dissection into a number of component parts to which different criteria of what satisfy the duty of care apply, such as diagnosis, treatment, advice (including warning of any risks of something going wrong however skilfully the treatment advised is carried out.)': *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 893. For an application of the *Bolitho* caveat to disclosure, see also: *Pearce v United Bristol Healthcare NHS Trust* (1999) 48 BMLR 118.

930 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430.

931 Purshouse, 'The Impatient Patient and the Unreceptive Receptionist' (2019) 27(2) Medical Law Review p. 318, 329.

932 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [87].

reference to the actions of a responsible body of professional opinion – to one where the standard of care is set 'by reference to *the reasonable person in the position of the claimant*'.[933] And, even more remarkably, by reference to the information that the specific patient deems significant.

In defining this novel approach, the Supreme Court employed rhetoric that, as outlined in Chapter 4, aligns with our understanding of procedural autonomy. Recognising an important prerequisite for the exercise of the patient's practical autonomy, the court framed patients as individuals with the power to make medical decisions. They are 'widely treated as consumers exercising choices' and 'widely regarded as persons holding rights'.[934] These choices and these rights are not subject to the discretion of the medical profession but are located within a bureaucratic framework, subject to public law supervision.[935] All in all, the patient is properly regarded as having a decision to make, an action to perform, and the court evidently recognised that it is their unique commitments that shape this decision.

Moreover, it is the medical profession that bears responsibility for facilitating the process leading up to this decision. In this respect, Lord Kerr and Lord Reed appealed to the positive, facilitative dimension of practical autonomy when they rejected indications from *Sidaway* that the disclosure standard could be determined by the patient's preparedness to ask specific questions:

> the more a patient knows about the risks she faces, the easier it is for her to ask specific questions about those risks, so as to impose on her doctor a duty to provide information; but it is those who lack such knowledge, and who are in consequence unable to pose such questions and instead express their anxiety in more general terms, who are in the greatest need of information. Ironically, the ignorance which such patients seek to have dispelled disqualifies them from obtaining the information they desire.[936]

Clearly, these remarks are based on the fact that the patient is in need of assistance to realise their values in the medical contexts – obtaining 'the in-

933 Arvind and McMahon, 'Responsiveness and the Role of Rights in Medical Law: Lessons from Montgomery' (2020) 28(3) Medical Law Review p. 445, 476 (my emphasis). Despite the reference to a treatment interfering with the patient's bodily integrity, it is clear that the concern for autonomy takes a wider form in the context of negligence.

934 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [75].

935 ibid [75].

936 ibid [58].

formation they desire'. The statement also implicitly delineates between the function that different classes of information play in a patient's autonomous decision making. Some information is necessary for the patient to concretise, and therefore to decide and act upon, their preferences – without this they cannot even ask the necessary questions. More information may then be provided to improve the position of patients who already have some knowledge, i.e. enough to ask pertinent questions about their care.[937] This suggests that autonomy requires a tapered form of information disclosure, as discussed in Chapter 3.

Simultaneously, the Supreme Court viewed the patient, not as someone who was uncritically reliant on professional guidance, but as an agent with the capabilities to participate meaningfully in the clinical process. Most stringently, the majority was clear that a view of patients as 'medically uninformed and incapable of understanding medical matters' could not be 'the default assumption on which the law is to be based'.[938] The laws on the labelling of pharmaceutical products and on the provision of information sheets assumed as much.[939] More widely, societal developments had also done their part to solidify the fact that patients were not 'wholly dependent upon a flow of information from doctors'.[940] Patients obtained information about 'symptoms, investigations, treatment options, risks and side-effects' *via* media like the internet (where reliable sources could be distinguished from unreliable ones), they formed support groups and had recourse to materials provided by healthcare providers.[941]

Put briefly, patients were not only competent to make medical decisions in the barest, minimal sense, but were to be understood as actors that engaged meaningfully in the decision-making process. They gathered evidence, evaluated the reliability of sources, applied them to their own circumstances, etc. This is arguably the dimension that was captured under our theoretical approach by the reference to rationality. Moving from the abstract to the more specific, one therefore also sees that a procedural

937 See in this regard: *Ollosson v Lee* [2019] EWHC 784 (QB), [2019] 3 WLUK 562 [156].
938 It is also noticeable that the Supreme Court distinguished this capability from the much more demanding characterisation posited by Lord Diplock in *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 894-895, which classed only certain patients as 'highly educated men of experience': *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [76].
939 ibid [76].
940 ibid [76].
941 ibid [76].

autonomy principle has a significant role to play in the law's understanding of a medical professional's disclosure obligations. Especially as it will be seen in the next sections that UK law still leaves many concrete areas of the standard undefined.

ii. The operationalisation of reasonable disclosure

The evaluated rhetoric of *Montgomery* suggests a strong affinity with the procedural autonomy concept and it therefore appears that the developed standard could offer a response to the ML challenges, which were conceived under this umbrella. Indeed, certain commentators assume that the disclosure required by the law is now directly related to that principle.[942] Hereby, a professional advising a patient would be liable, not in virtue of their failure to meet a duty of care in disclosure, but in virtue of their failure to achieve the outcome of protecting the patient's autonomy.[943]

Some of the rhetoric used in *Montgomery* offers support to this argument. For instance, in defending the new duty, the judgment of Lord Kerr and Lord Reed framed it as, fundamentally, relating to what 'the respect for the dignity of patients requires'.[944] Likewise Lady Hale cited Jonathan Herring, who asked: 'whether there was enough information given so that the doctor was not acting negligently *and* giving due protection to the patient's right of autonomy'.[945]

Under this interpretation, reasonable disclosure would be equated with autonomy-compliant disclosure. An operationalisation of this standard would simply require the courts to elaborate upon patients' decision-making needs; the professional would be obligated to meet these.[946] For example, as *Montgomery* itself stated, this requires a degree of comprehensib-

942 An overview can be found in: Nolan, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European Tort Law p. 259, 380.

943 ibid 378-379.

944 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [93].

945 ibid [108] (my emphasis), citing Herring, *Medical Law and Ethics* (Fourth Edition 2012) 170.

946 As a result, 'the Supreme Court arguably abandoned negligence analysis altogether, in that for liability to arise it may no longer need to be shown that the defendant acted unreasonably in all the circumstances of the case': Nolan, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European Tort Law p. 259, 378.

ility that is not achieved 'by bombarding the patient with technical information which she cannot reasonably be expected to grasp'.[947]

However, when the Supreme Court directly addressed the doctrine underlying the newly fashioned disclosure obligation, it clearly understood itself to be operating within a 'traditional framework of negligence'.[948] Holding that 'the doctor's duty of care takes its precise content from the needs, concerns and circumstances of the individual patient, *to the extent that they are or ought to be known* to the doctor'.[949] In operation therefore, a connection to the reasonable beliefs and actions of the professional is maintained. An untempered protection of the autonomy principle is not foreseen.

In particular, regarding the types of disclosure with which the *Montgomery* case was concerned – disclosure of risks and alternatives – more concrete indicators were provided. Namely, the legal standard required one to ask whether the professional acted reasonably in identifying and providing material information. A patient had to be informed of the risks and benefits that the professional 'anticipated' and of the alternatives that were deemed 'reasonable'.[950] Regarding the resultant communication of this information, Lord Kerr and Lord Reed also stated that they were imposing 'a duty on the part of doctors to *take reasonable care to ensure that a patient is aware* of material risks of injury that are inherent in treatment'[951] and that it was the 'aim' of the doctor's advisory role to engage in a dialogue to impart such information.[952]

This approach, defining categories of information as appropriate subjects of disclosure, was a further, less noticeable manner in which the UK's highest court sought to reconcile their unorthodox focus on the patient's need with negligence's narrower analysis of the reasonableness of the defendants' actions. By implication, it was only certain types of information that a doctor could realistically expect the reasonable patient to require. As was noted above, this may be interpreted to coincide with the finding that a patient's autonomy is not served by bombarding them with information. Unfortunately, it was left relatively open what such categories of disclosure

947 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [90].
948 ibid [82].
949 ibid [73].
950 ibid [90].
951 ibid [82] (my emphasis).
952 ibid [90].

are, beyond the firmly established classes of anticipated, i.e. known, risks and reasonable alternatives.

Below we will take a casuistic approach, one focussed on these and other categories that the established case law has considered. However, also recognising the openness of the standard laid down by the Supreme Court, we will appreciate the demands of the autonomy principle. This arguably has the potential to reshape certain dimensions of the medical professional's obligation to act with reasonable care, specifically in the AI context.

What ought to be mentioned before moving on, and which was clear even before *Montgomery*, is that negligence imposes broader obligations than battery to convey information to the patient. A doctor's duty is not limited to interferences with the claimant's bodily integrity. They may be negligent for the information they provide on the drugs that they prescribe[953] and for the information they give regarding decisions not to treat at all.[954] Nor is the assessment of the professional's reasonableness limited to disclosure that precedes a specific intervention, but encompasses the disclosure that they ought to make at later stages as a part of the ongoing relationship with the patient.[955]

iii. Summation

The 'material information' standard of disclosure laid down in the seminal case of *Montgomery* appeals to the figure of the reasonable patient, but also to the reasonable professional and to the information that such a professional ought to know a specific patient would attach significance to. In laying down this standard, the court left open many questions, but

953 In *Kennedy v Frankel* a breach of duty was found for a failure to warn about certain side effects of an oral dopamine agonist: *Kennedy v Frankel* [2019] EWHC 106 (QB), [2019] 1 WLUK 216 [50]. See similarly: *Blyth v Bloomsbury Health Authority* [1987] 2 WLUK 77 (although no breach of duty was found).

954 *Webster v Burton Hospitals NHS Foundation Trust* [2017] EWCA Civ 62, (2017) 154 BMLR 129. In *Mordel v Royal Berkshire NHS Foundation Trust* full information had to be given about a treatment that was positively declined: *Mordel v Royal Berkshire NHS Foundation Trust* [2019] EWHC 2591 (QB), (2020) 172 BMLR 106.

955 *Spencer v Hillingdon Hospital NHS Trust* [2015] EWHC 1058 (QB), [2015] 4 WLUK 354; *Gallardo v Imperial College Healthcare NHS Trust* [2017] EWHC 3147 (QB), [2017] 12 WLUK 198.

evidently gave great weight to the kinds of factors that were argued to shape the principle of procedural autonomy.

In the following, the challenges associated with clinical ML devices will be analysed from the perspective of established categories of information disclosure found in English law. Where the law is open-textured or uncertain, the principle of autonomy and the materiality test will be appealed to.

2. The risks of medical AI

Montgomery affirmed the general position that a patient must be advised of the risks associated with a procedure.[956] Its novelty arose from the standard it set in determining which risks are disclosable: material risks that a doctor knows or ought to know a reasonable patient, or this particular patient, would attach significance to. In this section the goal is to determine whether AI's autonomy challenges pose distinct risks and/or whether they give rise to interrelated factors that must be disclosed.

i. Specific risks

If one considers whether an AI challenge constitutes a specific disclosable risk, analogous to the risk of shoulder dystocia in *Montgomery*, then one should begin with the judiciary's usage of the concept. The courts have not explicitly defined risk for the informed consent context, but they have assumed the relevance of two functions: 'the degree of likelihood of [the event] occurring and the seriousness of the possible injury if it should occur'.[957] The judgments also indicate that medical evidence is relevant to the definition of a risk, going towards both of these dimensions.[958]

A concrete example of the approach can be given by reference to the already discussed decision of *Chester v Afshar*. Here there was a recognised 1% to 2% risk of neurological damage during an operation on the claimant's

956 This had already been clear, even under less generous interpretations of the *Bolam* standard: *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 894-895.

957 ibid 888.

958 *Montgomery* cited Lord Scarman's approach, ibid 888, with apparent approval: *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [48], [83].

spine.[959] Although this was relatively small, the consequences of its occurrence were correspondingly serious: including the risk of paralysis.[960] The duty to disclose this specific risk, with a small probability but grave significance, was not in dispute.[961]

In relation to AI this means that, where the professional knows that the technology alters the severity or probability of a specific physical harm befalling the patient, a straightforward case can be made for disclosure. And some of the AI risks, outlined in Chapters 2 and 3, will satisfy this test. For instance, it was seen that the technology will assist in providing critical care in certain settings, such as for the identification of brain haemorrhages. In addition, even if a device exhibits good initial accuracy (going towards the likelihood of an adverse event occurring), we saw that such assessments are likely to require adjustments for real-life clinical settings. Consequently, where AI is related to a sufficiently likely and sufficiently serious event, it poses a type of risk that the doctor ought to know would be significant to the reasonable patient, or sometimes, to the specific, individual patient.

However, some distinct problems still emerge for cases involving AI-generated risks of this kind, given that negligence requires the defendant to meet a certain standard in their behaviour. *Inter alia* a defendant professional must have been aware of the absent information, or ought to have been aware of it, in order to be liable for its non-disclosure: 'A doctor cannot be held negligent for failing to make a disclosure of matters of which he had no knowledge, unless that knowledge can be imputed to him (ie, unless it can be concluded that these were matters he ought to have known about)'.[962] As was discussed in Chapter 2, the nature of AI's specific risks is less likely to be known or quantifiable in this way. Particularly as techniques for scientific validation are more difficult to apply and evaluate.

Another issue arises in relation to the concern, expressed in *Montgomery*, that the patient can be overwhelmed by disclosure of technical information and, specifically, that risk disclosure should not be broken down to percentages.[963] It is possible to provide the patients with too much information about a procedure and its risks. This is illustrated by *Ollosson v Lee*. It was

959 *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [11].
960 ibid [39].
961 ibid [55].
962 *AH v Greater Glasgow Health Board* [2018] CSOH 57, (2019) 169 BMLR 120 [55]; *Duce v Worcestershire Acute Hospitals NHS Trust* [2018] EWCA Civ 1307, (2018) 164 BMLR 1 [42].
963 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [89]-[90].

sufficient that it was communicated to the patient that there was 'a risk of long term persisting pain which could range from mild to severe' and that, in terms of magnitude, this risk was 'small, adding that it was greater than the rare/remote risks of early or late failure'.[964] A more detailed accounting of the risk's size and effects did not have to be provided.

In relation to the clinical deployment of ML technologies, these findings suggest that there will be specific risks raised by AI that are material to patient autonomy, but do not need to be disclosed separately. More generally it suggests that professionals need to be mindful of not overburdening patients with technical information on AI and the way in which its functioning shapes these risks. This analysis buttresses the observations made in Chapter 3 that the disclosure of specific risks is not the correct vehicle for conveying the unique challenges posed by AI to patients.

ii. Risk-relevant status

Beyond AI's alteration or generation of specific risks, it was also claimed that the technology constitutes a risk-relevant characteristic of an intervention. This characteristic was likened to that involved in the deployment of innovative procedures or of unlicensed or off-label uses of devices. When combined with the autonomy principle, this analogy provides an argument for the disclosure of the technology's use and status.

In establishing the relevance of such a risk-related status under UK law, it is notable that, in defining the disclosure standard, *Montgomery* deemed a multiplicity of features relevant to the materiality of a risk:

> the assessment of whether a risk is material cannot be reduced to percentages. The significance of a given risk is likely to reflect a variety of factors besides its magnitude: for example, the nature of the risk, the effect which its occurrence would have upon the life of the patient, the importance to the patient of the benefits sought to be achieved by the treatment, the alternatives available, and the risks involved in those alternatives.[965]

This marks the adoption of an open-ended approach by the Supreme Court. Certainly, it indicates that a patient must be advised of a broad

964 *Ollosson v Lee* [2019] EWHC 784 (QB), [2019] 3 WLUK 562 [151]-[152].
965 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [89].

range of factors that can reasonably be anticipated to contribute to their assessment of a risk.

One such factor, which has been given considerable attention by legal commentators, is the innovative nature of a procedure. This is not merely an aspect of the seriousness or magnitude of risks that the patient is facing in their case. Instead, this classification conveys generic information to the patient. For example, it indicates the lack of scientific testing of a procedure, whereby one is not proceeding according to the ordinary standards of validated scientific knowledge.[966] Relatedly, there is an assumption that there is greater degree of uncertainty in such situations about the nature and degree of physical harms to which the patient is exposed.[967]

Mills v Oxford University Hospitals NHS Trust provides a rare judicial examination of the issue. The case applied *Montgomery's* materiality test to demand disclosure of the fact that an operation involved the use of 'a novel technique, still in its evolution and not well established'.[968] In the course of this, the court emphasised particularly the uncertainty involved in such innovative procedures. Stating: 'if [the surgeon] had referred to the series [of novel operations] he had undertaken up to that point, it would have been important to emphasise that it was uncertain what the complication rate for a longer series would be and, as with any new technique, the risks and benefits were not yet clear'.[969] The absence of adequate validation of the professional's performed procedure was also brought up in the claimant's successful argument: the surgery was not 'a standard well tested technique for resection of a brain tumour'.[970] It appears that *Mills* accepted a procedure's wider risk-related status as a piece of information

966 '[I]n offering innovative treatment, the physician is working on a hunch or scientific theory that has not been adequately investigated or researched': Chan, 'Legal and Regulatory Responses to Innovative Treatment' (2013) 21(1) Medical Law Review p. 92, 94.

967 'Such medical procedures are administered for the benefit of a specific patient but have uncertain outcomes because they have not been adequately tested': ibid 94. The 'great uncertainty' of a proposed innovative treatment was also a dominant theme in: *Simms v Simms* [2002] EWHC 2734, [2003] Fam 83.

968 *Mills v Oxford University Hospitals NHS Trust*, [2019] EWHC 936 (QB), (2019) 170 BMLR 100 [197], [201].

969 ibid [204].

970 ibid [15].

that is material to the reasonable patient or at least material to certain risk-averse individuals.[971]

Beyond *Mills*, commentators have also regarded the uncertainty generated by innovative treatments as a significant, reasonable concern. In this respect, Cockburn and Fay have noted that: 'While by definition it is impossible to quantify unknowns as statistical risks, ensuring the patient realises that s/he is effectively rolling the dice by undergoing innovative treatment is an important aspect of obtaining consent'.[972] Similarly, O'Neill surmises that 'Although under UK law, surgeons cannot be expected to warn of unforeseeable unknown risk, unknown harm is foreseeable in the case of innovative medical treatment' and one can require the disclosure of this factor.[973]

Given the limited authority on the matter of innovative interventions, it is also worth considering the relevance of another status explored in Chapter 3. Namely, the unlicensed or off-label use of medical products and devices. While there is no exact connection to an intervention's risk-profile, these categorisations were also understood as something significant for the patient's risk-related decision-making.

Jones v Taunton and Somerset NHS Foundation Trust supports this analysis in the British context. This case concerned the novel, unlicensed drug (Nifedipine), which had not been subject to 'convincing double-blind studies' for the relevant use, and its risks and benefits were not fully determined.[974] Nevertheless, it was becoming the drug of choice over a licensed drug that was associated with marked side-effects. The question before the court was whether the prescription of this drug, in these circumstances was negligent.[975]

---

971 ibid [215]. Contrast the prior situation under the *Bolam* standard, which was incorrectly maintained in *Grimstone v Epsom and St Helier University Hospitals NHS Trust* [2015] EWHC 3756 (QB), [2015] 12 WLUK 749. On the problematic nature of the case, see especially Austin's commentary: Austin, 'Grimstone v Epsom and St Helier University Hospitals NHS Trust: (It's Not) Hip to Be Square' (2018) 26(4) Medical Law Review p. 665.

972 Cockburn and Fay, 'Consent to Innovative Treatment' (2019) 11(1) Law, Innovation and Technology p. 34, 46.

973 O'Neill, 'Lessons From the Vaginal Mesh Scandal: Enhancing the Patient-Centric Approach to Informed Consent for Medical Device Implantation' (2021) 37(1) International Journal of Technology Assessment in Health Care e53, 1-5, 4.

974 *Jones v Taunton and Somerset NHS Foundation Trust* [2019] EWHC 1408 (QB), [2019] 6 WLUK 193 [135].

975 ibid [135].

Consequently, *Jones* is not strictly speaking concerned with informed consent obligations. Still, it supports the approach taken to the classification of risk-related statuses in *Mills*. For one, the judge endorsed the relevance of the unlicensed nature of a drug to the clinical assessment of risks. He acknowledged that a licensed drug should not be abandoned until the risks and benefits of an unlicensed one have been established.[976] But he was also clear that this relationship was not determinative of relevant assessments: 'the fact that Nifedipine was unlicensed is merely one factor, and I find, not a strong factor, in the light of the other evidence'.[977] In addition, it was mentioned that the patient's informed consent must be obtained with respect to the unlicensed nature of the drug and that this was a prerequisite for the finding that the drug's use was non-negligent.[978]

Even if the innovative or off-label nature is not determinative of a risk/benefit balance, the case law therefore suggests that it provides the patient with important information that they should be provided with. Such a view gains added support from the General Medical Council's (non-legally binding) guidance. This specifically indicates that the doctor should share information with the patient on 'whether an option is an innovative treatment designed specifically for their benefit'.[979] Here the principles of informed consent do not require a detailed explanation of the (lack of) scientific evidence or approval, or the mechanisms associated with this, but they do require the disclosure of the procedure's type and the general way in which this is likely to impact the risk calculus.

Such a disclosure obligation may further be modelled on the finding in *Webster v Burton Hospitals NHS Foundation Trust*, in which a specific risk was identified. Namely, 'the increased risk of perinatal (the period around birth) mortality, including *ante partum* (before delivery) mortality' from a rare combination of conditions.[980] However, this risk had to be inferred from a 'small (or extremely small) statistical base'.[981] While the court was therefore not strictly speaking dealing with an innovative procedure, this

976 ibid [139]-[140].
977 ibid [135].
978 The defence expert accepted that prescription was non-negligent on the supposition that the patient was told 'that it is an unlicensed drug' and was given a full explanation: ibid [131].
979 General Medical Council, 'Decision Making and Consent' (London 2020) para 13.
980 *Webster v Burton Hospitals NHS Foundation Trust* [2017] EWCA Civ 62, (2017) 154 BMLR 129 [38].
981 ibid [39].

case also exemplifies judicial dealings with a form of uncertainty that exists at a high level of generality. Crucially for the present argument, the suitable informed consent standard was stated at a corresponding level of generality: 'that there was "an emerging but recent and incomplete material showing increased risks of delaying labour in cases with this combination of features'.[982] *Webster* arguably strikes a reasonable balance between the patient being provided with sufficient information to make a practical decision and not providing them with too much information, thereby overwhelming their decision-making capabilities. This balance is crucial for the preservation of procedural autonomy.

It is argued that, on the basis of the factors discussed in Chapters 2 and 3, an argument can be constructed for UK law requiring AI characteristics to be disclosed at a comparable level of abstraction. In particular, the patient should be told that there may be limitations in the way in which an ML output can be connected to established scientific knowledge and that, relatedly, the use of AI involves substantial uncertainties, similar to those associated with novel or unlicensed procedures. The provision of such information would avoid the aforementioned issues with the disclosure of specific risks. While health professionals may not know the precise nature of the dangers inherent in AI deployment, they are (or certainly should be) aware that an ML device is being used in the patient's care and that certain structural uncertainties accompany this.

In the final analysis, while some statements in existing case law, in combination with our normative autonomy-based analysis, suggest that such an obligation can plausibly be formulated under UK common law, its strength must necessarily remain uncertain given the limited nature of precedent.

3. Alterations of expertise

Under English law, the analysis of obligations to disclose (the lack of) professional expertise appears to have been couched almost exclusively in terms of battery. The limitations of such an approach were assessed above. Particularly regarding AI, it is not straightforward to subsume the technological supplementation of human capabilities under the existing case law.

982 ibid [40].

This raises the question of the role that negligence can play in this realm. Theoretically, the greater emphasis placed by this mechanism on *informed* consent, rather than merely valid consent, should grant patients a more comprehensive protection. The finding in *Jones v Royal Devon and Exeter NHS Foundation Trust* grants some support to this position. In this case, there was a discrepancy between the surgeon that the patient believed would operate on her spine and the one who in fact did. The patient had been admitted under the care of a particular surgeon and was given to understand that her operation was scheduled to be performed by him.[983] She was only told of the change in her surgeon's identity as she was going into theatre, at which point she could not effectively provide her consent.[984] Under these circumstances the court found that there was a breach of the defendant's informed consent obligations.[985]

At the very least, *Jones* recognises that a change in the specific identity of a medical professional can defeat the claimant's consent. This was in spite of the patient being told that it could not be guaranteed that her procedure would be performed by one particular person, but only that it would be a team member with appropriate experience.[986] Moreover, the claimant had withdrawn the argument that the surgeon who performed the procedure was not adequately qualified.[987] As such, the court arguably found that, regardless of comparable expertise, a breach of informed consent obligations arises where the personal identity of a known professional changes. This would be similar to the position assumed by the earlier battery cases. As AI would not generally bring about such changes, being deployed by an identifiable human individual, *Jones* would not demand the disclosure of changes in the expertise of that professional.

Other comments in the case, however, suggest that it was a shift in expertise that was the operative factor about which the claimant should have been informed. It was emphasised that the desired professional 'enjoyed a very high reputation both locally and nationally' and that he was

983 *Jones v Royal Devon and Exeter NHS Foundation Trust* [2015] Lexis Citation 3571 [28], [31].

984 ibid [37]. We will return to the possibility of giving informed consent under certain pressures below. What is crucial here, is that there was no effective disclosure to the patient of the surgeon's identity and/or experience.

985 ibid [37].

986 ibid [23].

987 ibid [12].

a surgeon of considerable seniority.[988] Furthermore, the patient's GP had made the medically informed recommendation that it would be preferable for her to wait (in spite of considerable pain) to be operated on by this surgeon.[989] Lastly, the patient made her own assessment of the surgeon's reliability, based upon the operations she knew he had performed on her acquaintances.[990] In light of these statements, it appears that the court found a breach of duty because the patient had reason to believe that one level of professional expertise was being brought to bear on her care (that of a senior, very experienced surgeon) when in fact a professional with a lower, albeit adequate, capability was involved.

This line of argument would be extendable to AI. In some circumstances a patient may have grounds for believing that a human professional involved in their care is an expert with a high degree of skill, as they are performing a demanding procedure. This impression may be misleading if an ML device is in fact supplying a substantial amount of the relevant capabilities. Even though these capabilities may be adequate, Chapter 3 illustrates their different nature.

Given the ambiguity in *Jones* regarding the significance of shifts in expertise *per se*, rather than changes in human identity, two further arguments can be advanced. First, that the more recent battery analyses, holding that individuals may attach significant weight to the status and expertise of their carers, and limited *dicta* in other negligence actions, reinforce the recognition of related forms of actionable non-disclosure in negligence. Second, *Montgomery's* test of material information and the weight of the autonomy principle indicate the correctness of the same conclusion.

Regarding the findings of the battery analysis, one can recall the case of *R v Melin*. Here the court strongly affirmed the significance that a patient would attach to the experience-related status of their professional. In particular, the example was cited of a new doctor taking over a general practice and it was implied that their status, not their identity, would be the most important piece of information for the patient to know.[991] This evinces a novel receptiveness to the patient's concerns about the professional's expertise, one which had previously been obscured by a focus on quite drastic changes of identity.

988 ibid [6], [65].
989 ibid [29].
990 ibid [64].
991 *R v Melin* [2019] EWCA Crim 557, [2019] QB 1063, citing Ormerod and Laird, *Smith, Hogan, & Ormerod's Criminal Law* (Fifteenth Edition 2018) 672.

To supplement this finding, there is also limited *dicta* in the lower courts that reasonable patients may attach significance to the automated, rather than human, nature of an important decision. Specifically, in considering the patient's refusal to enter a clinical trial that involved a randomised computer-based selection of participants, the judge in *C v Colchester Hospital University NHS Foundation Trust* noted: 'Mr C, understandably, refused to enter into that study (…) he wanted a clinician to dictate and determine when he was to be operated upon and not to be dealt with, as he saw it, by computer'.[992] In effect this amounts to a recognition that the full automation of an important clinical decision constitutes a legitimate concern for a patient.

Second, one can consider the role of the informed consent standard delineated in *Montgomery*, which had not yet explicitly been applied in *Jones*. As was noted, this has been interpreted to impose a broad duty to disclose material information to the patient. That is, information which would be significant to the reasonable patient, or which the doctor knows or ought to know is significant to the particular individual. The outlined findings suggest a willingness on behalf of the courts to recognise that shifts in expertise, and to a lesser extent a shift between human and computerised clinical decision making, are significant factor for a patient. Consequently, it can be argued to be a natural step, under *Montgomery's* only relatively recently redefined approach, to deem certain shifts in expertise to be material information.

In the case of ML devices, this would mean disclosing certain instances where the technology is covertly substituting human capabilities, which are expected to be of a high standard, to a meaningful degree. For example, both the battery case law and the *dicta* cited above are referring to substantial, meaningful replacements of expertise; what is significant to the patient is a change in 'status' or the comprehensive automated randomisation of a decision. Similarly, the distinctions drawn in Chapters 2 and 3 illustrate that a full automation of a process is much more problematic from the perspective of patient autonomy than the enhancement of a human decision.

Relevant differences also emerge when one considers whether the cumulative amount of human expertise, which is brought to bear on a decision, is decreased or whether the AI is merely providing a 'second opinion'. To the extent that *Montgomery* mirrors such autonomy concerns, the courts should apply the fact-sensitive materiality standard to determine exactly

992 *C v Colchester Hospital University NHS Foundation Trust* [2018] 2 WLUK 850 [17].

those kinds of substitutions that would be deemed significant by patients. Focussing on instances where there is a partial replacement of pre-existing human capabilities, as outlined in Chapter 2, could serve as a useful starting point for this purpose.

In sum, *Jones* indicates that, where a patient has an expectation that a high degree of expertise will be brought to bear on their care, not disclosing shifts in this expertise may constitute a breach of duty. Drawing on an analysis of other case law and the *Montgomery* materiality standard, we have added the caveat that relevant shifts must be sufficiently substantial. Where AI significantly reduces the overall level of expertise brought to bear on a procedure – substituting itself for the decision making of an experienced human – a breach may consequently occur. It is more questionable whether lesser shifts, or indeed simply a shift from human to computerised capabilities, must also be disclosed to the patient.

4. Information concerning the choice of goals

In the context of AI, the most striking effect on the patient's autonomy is perhaps not its risk-profile or its alteration of the balance of human-machine expertise, but the novel way in which it sets and pursues its own objectives. This generates two related issues that were outlined in Chapter 3 and touched upon under the battery analysis above.

First, certain AI can partially determine an aspect of the patient's care and therefore pose a particularly significant challenge to a patient's ability to select the objectives that they wish to realise. Second, ML devices can exhibit a propensity to direct human decision making through nudging. These were problematic to a lesser extent – subjecting patient decision-making to systematic external influences. It is argued here that legal categories dealing with alternative procedures and with external influences on patient choice can partially respond to these interferences with patient autonomy.

i. Understanding choices

One class of disclosure cases seeks to identify those procedures that a patient must be informed about in complex healthcare interactions to

make appropriate choices in their care. As outlined, *Montgomery* expressly requires that the patient be made aware of 'any reasonable alternative or variant treatments'.[993] The reference to treatment should not be read restrictively. Even prior to the development of the patient-centred standard of care, the courts had required the disclosure of alternative modes of diagnosis.[994] Accordingly, the term procedure will be used as a catch-all term for both treatment and diagnostic scenarios.

One ought to begin by establishing the kinds of choices that the law has protected by requiring the disclosure of alternative options. One type of case is relatively straightforward in this regard: where one readily available alternative represents a distinct risk-benefit balance to the professional's chosen intervention. This normally constitutes the most important factor that the courts have considered in determining whether an alternative represents a choice that must be discussed with the patient.[995] *Montgomery* itself bears testament to this fact. In determining the necessity of disclosure, Lord Kerr and Lord Reed considered that there was a stark difference between the substantial risk of shoulder dystocia occurring during vaginal delivery and the miniscule risks of harm resulting from a caesarean section.[996]

Birch v University College London Hospital NHS Foundation Trust relied on a similar analysis. Here the doctors carried out their investigation of the patient's symptoms with an invasive cerebral catheter angiogram, when the performance of an MRI scan constituted another non-invasive option.[997] A choice was made in terms of sequencing. Both procedures were modes of assessing the patient for a similar range of conditions and the angiogram was more accurate but also riskier in some respects.[998] The professionals opted for the former *instead of* the latter and this had implications for the specific risks that the patient was exposed to.

---

993 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [87].

994 *Birch v University College London Hospital NHS Foundation Trust* [2008] EWHC 2237 (QB), (2008) 104 BMLR 168.

995 '[J]udgements made about information about risk that should be disclosed to a patient will often have a direct impact on the range of alternative or variant treatments that a patient should be offered accordingly': Dunn and others, 'Between the Reasonable and the Particular' (2019) 27(2) Health Care Analysis: Journal of Health Philosophy and Policy p. 110, 112.

996 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [94].

997 *Birch v University College London Hospital NHS Foundation Trust* [2008] EWHC 2237 (QB), (2008) 104 BMLR 168 [39].

998 ibid [50]-[52].

Such a straightforward analysis can also be applied to choices between procedures with risk-relevant characteristics. Consider the aforementioned judgment in *Mills v Oxford University*. The case concerned the use of a novel surgical technique over a pre-existing standard procedure. Neither the experts nor the court questioned the claim that the standard procedure was a relevant, alternative choice.[999] This was accepted as a matter of course, in spite of certain clinical factors that favoured the new procedure, including: decreased short-term mortality and better cosmetic results.[1000] Apparently the uncertainty surrounding the comparative risks and benefits of the new technique over the older one sufficed to frame the professional's choice as being between material alternatives.

This solidifies the argument that the patient must be given information on the AI's risk-related characteristics and be offered a choice regarding the technology's use. It is one alternative in this respect. However, this approach would only grant the patient information and control over the ML device's objectives in so far as they are related to risks of physical harm.[1001] The necessity of maintaining this connection was made clear

999 *Mills v Oxford University Hospitals NHS Trust*, [2019] EWHC 936 (QB), (2019) 170 BMLR 100 [195]-[197].

1000 ibid [215].

1001 This depends somewhat on how closely one associates the anticipated benefits of a procedure with the risks of action or inaction. For example, Veatch has noted that 'almost any medical procedure will involve a mixture of potential benefits and potential harms. Almost any procedure or therapeutic agent has potential side effects. Even determining that an effect is a "side" effect rather than a "benefit" involves value judgments that are complex': Veatch, 'Doctor Does Not Know Best' (2000) 25(6) The Journal of Medicine and Philosophy p. 701, 706. See similarly Feng, who holds that the nature-risk 'distinction is also fallacious as it assumes that there is an inherent difference in terminology and substance between nature of treatment and risks inherent in treatment (…) If there is a duty in the tort of negligence on the part of a doctor to inform his patient of risks inherent in treatment there must surely be the same duty on the doctor to disclose to his patient the patient's ailment, treatment (including nature of treatment), benefits of treatment, and other matters connected with medical advice': Feng, 'Failure of Medical Advice' (1987) 7(2) Legal Studies p. 149, 156-158. The courts' approach to the disclosure of benefits remains ambivalent however. In *Montgomery*, the benefits and the seriousness of the condition were considered alongside risks – apparently going towards the nature of the risk, rather than constituting a consideration in their own right: *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [89]-[90]. Lord Scarman in *Sidaway* seems to treat them as separate but complementary: 'The doctor's duty can be seen, therefore, to be one which requires him not only to advise as to medical treatment but also to provide his patient with the information needed to enable the patient to consider

in *Plant v El-Amir*. Here the limited likelihood of achieving the patient's highly valued outcome (to be able to read again) through an eye operation was only deemed disclosable to the extent that it impacted her readiness to run the risks of the procedure.[1002]

To contend that the disclosure of AI's ability to pre-determine choices is independently necessary, one can appeal to the argument of Section II.A.: that negligence does not strictly require the eventuation of physical injury and recognises harms to autonomy. But, under the breach element, one must bolster this argument by additionally examining the rationale underlying the disclosure of alternatives.

This rationale can arguably be found in the disconnect that an advising professional may cause between a patient's decision making and the realisation of their goals, echoing the autonomy challenge conceptualised in Chapter 3. In the UK legal context, this was put most forcefully by Lady Hale in *Montgomery*, when she concretised the majority's observation that a choice between alternatives will often depend upon non-clinical considerations.[1003] She observed that the doctor's decision to withhold information interfered with the patient's evaluative judgment – itself evincing the view that vaginal delivery was 'in some way morally preferable to a caesarean section'.[1004] By making this judgment the doctor was, precisely, 'depriving the pregnant woman of the information needed for her to make a free choice in the matter'.[1005] Turton has similarly emphasised this in her analysis of *Montgomery*, stating that: 'The duty is not limited to enabling the patient to decide which risks she is willing to accept, but is also focused on protecting her right to make that decision in line with her own values'.[1006]

---

and balance the medical advantages and risks alongside other relevant matters': *Sidaway v Board of Governors of the Bethlem Royal Hospital* [1985] AC 871, 886. Similarly, Lord Hope stated in Chester that 'whether the risk is worth running for the benefits that may come if the operation is carried out': *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134 [86]. Given the significance of certain choices for autonomy, independently of involved risks, this section considers this aspect in isolation.

1002 *Plant v El-Amir* [2020] EWHC 2902 (QB), [2020] All ER [82]-[87].

1003 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [82], [109].

1004 ibid [114].

1005 ibid [114].

1006 Turton, 'Informed Consent to Medical Treatment Post-Montgomery' (2019) 27(1) Medical Law Review p. 108, 129.

A later first instance decision, *C v Colchester Hospital University NHS Foundation Trust*, also bears out this analysis of the considerations underlying the disclosure of alternative options. In this case the claimant had been subjected to a course of treatment that involved chemoradiotherapy (CRT) in order to shrink his tumour, before being subjected to a life-changing surgery to remove it completely. The patient argued that he had not been informed that a proportion of people were cancer-free after undergoing only CRT and, had he been so informed, he would have refused surgery.[1007] Walden-Smith DJ found that CRT in combination with no further treatment was not an alternative to surgery, at least not for a patient who had made it clear that they wanted to be cured.[1008] Given the patient's desire and the state of medical knowledge, it was fallacious to argue that CRT constituted a separate type of treatment. The information the patient lacked in no way disconnected them from the choice to pursue their admitted objective.

In consequence, one can see that the information about alternatives is envisaged to place patients in a position where they can make a decision that realises their personal values. Only certain deficiencies will altogether prevent them from occupying this position.

In the AI context, Chapter 3 identified a comparable deficiency in one particular subset of cases: where the technology had a relatively wide discretion to select non-personalised goals and partially determined a clinical decision for the patient. This pre-empts them from making a choice by reference to their own considerations, even if a professional is not the one deploying an evaluative standard or foreclosing the patient's options.

Two ML scenarios were taken to exemplify sufficiently serious interferences of this type. Namely, ML uses that pre-determine a patient's choice to undergo a serious diagnostic procedure and AI uses that triage a patient's access to health care.[1009] Given the outlined case law, it is arguable that information about these objectives would be considered significant by the reasonable person in the patient's position and by many specific patients.

Under *Montgomery*'s standard it is therefore arguable that, when a patient is being subjected to these AI uses, they must be made aware that they are committing themselves to the pursuit of non-personalised objectives

1007 *C v Colchester Hospital University NHS Foundation Trust* [2018] 2 WLUK 850 [36].
1008 ibid [47].
1009 Assuming for the sake of argument that one can identify a person with a relevant duty for the triaging uses of AI.

– whether these possess specific or generalised risk characteristics or not. This is material information. At the same time, such a disclosure is then closely aligned with the discussion of alternatives. The aim is to give the patient a choice between courses of action and one must consider the wider limitations that the law has placed on the disclosure of different options.

The jurisprudence encapsulates these limitations under the already mentioned concept of *reasonable* alternatives.[1010] This qualifier represents a concession to a standard orientated towards the defendant.[1011] It also coincides with the intuition that, if there is no realistic decision for the patient to make, then a failure to advise them causes no possible violation of their practical autonomy. Consequently, even where an ML's non-personalised commitments amount to material information, a duty to disclose will not be breached if the patient did not have a reasonable choice to make on whether to submit to these commitments or not.

The legal standard for the determination of reasonableness is somewhat uncertain in the wake of *Montgomery* because, while introducing the terminology, the case stopped short of specifying its meaning in any detail. The resulting controversy has been discussed in *AH v Greater Glasgow Health Board*, with Lord Boyd distinguishing between two approaches that arguably constitute the extreme positions on a spectrum. At one end of possible interpretations, a reasonable alternative would be one that is determined by the *Bolam* test – i.e. one that 'no ordinarily competent clinician, exercising ordinary skill and care, would have failed to offer' – and, at the other end, it would be one that draws on *Montgomery's* patient-friendly

1010 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [87], [90]; *Bayley v George Eliot Hospital NHS Trust* [2017] EWHC 3398 (QB), [2017] 12 WLUK 670 [56]; *AH v Greater Glasgow Health Board* [2018] CSOH 57, (2019) 169 BMLR 120 [39].

1011 For example, a doctor cannot be expected to disclose (material) information if they did not know and did not reasonably have to be aware of it. This proved fatal in *Bayley* where 'physicians would not have been familiar with the stenting procedure for this condition in 2008' … 'I am not satisfied that a reasonably competent vascular surgeon would or ought to have known about the availability or potential use of this treatment in the second half of 2008': *Bayley v George Eliot Hospital NHS Trust* [2017] EWHC 3398 (QB), [2017] 12 WLUK 670 [64], [99]. Cf. *Webster*, where it was the doctor's carelessness that caused his lack of knowledge and thus did not exclude the information from consideration, he 'had failed to inform himself about the implications of the rare combination [of conditions]': *Webster v Burton Hospitals NHS Foundation Trust* [2017] EWCA Civ 62, (2017) 154 BMLR 129 [38].

approach – determined by what 'a patient might find reasonable after a full discussion of all the treatments whether or not these are available'.[1012]

The former approach would be unusually limiting in the wake of *Montgomery*. For AI the application of this standard would mean that, if at least one body of medical opinion believed that the patient should only be offered the AI-related procedure and this view withstood logical analysis, then this would be the end of the court's assessment. No disclosure obligation would lie.

However, the courts have not generally embraced such a drastic limitation, which would 'simply be reinstating [the *Bolam* standard] by the back door'.[1013] What emerges from the cases is that clinical suitability, including availability (matters to be determined with the assistance of medical experts) will indeed perform an important gatekeeping role in the determination of which options are reasonable.[1014] For example, in *Malik v St George's Hospital Trust* the court accepted that a non-surgical intervention was not a realistic option *inter alia* because of the long-waiting times involved in this method of treating the patient's acute needs.[1015]

To this extent the approach is moved more towards the *Bolam*-end of the continuum, but this does not constitute the final assessment. In light of the force of the autonomy principle as asserted in *Montgomery*, the courts will not simply defer to one body of medical opinion. Patient-specific factors,

1012 *AH v Greater Glasgow Health Board* [2018] CSOH 57, (2019) 169 BMLR 120 [8].

1013 Sutherland, 'The Law Finally Reflects Good Professional Practice: Montgomery v Lanarkshire Health Board' [2015](123) Reparation Bulletin p. 4, 7-8. See: *Bayley v George Eliot Hospital NHS Trust* [2017] EWHC 3398 (QB), [2017] 12 WLUK 670 [61].

1014 *AH v Greater Glasgow Health Board* [2018] CSOH 57, (2019) 169 BMLR 120 [43]. See also: *McCulloch v Forth Valley Health Board* which applied *Ah Glasgow* and rejected the necessity of disclosing alternatives not indicated in the circumstances of the case: *McCulloch v Forth Valley Health Board* [2021] CSIH 21, [2021] 3 WLUK 569 [40]. Sutherland has convincingly concluded that *AH* 'accepted that doctors were entitled to filter information given to patients on the basis of what other doctors considered reasonable but concluded that a doctor could not withhold information about a reasonable alternative treatment and the risks associated with that on the basis of their own preference': Sutherland, 'Montgomery: Myths, Misconceptions and Misunderstanding' [2019](3) Journal of Personal Injury Law p. 157, 166.

1015 *Malik v St George's University Hospitals NHS Foundation Trust* [2021] EWHC 1913 (QB), (2021) 181 BMLR 135 [45]-[93]. Compare *Bayley*, where the court did consider alternative that were available privately and in Europe, suggesting a somewhat more patient-friendly interpretation of reasonableness: *Bayley v George Eliot Hospital NHS Trust* [2017] EWHC 3398 (QB), [2017] 12 WLUK 670 [49].

such as the patient's definition of their clinical need and the relevance of individual value judgments to the selection of alternatives, will also define which alternatives are reasonable.[1016]

For our two case studies of disclosable AI, based upon their pre-emption of patient choice, this reasonableness limitation is predicted to limit the informed consent obligations for AI's use in triaging. There will be forceful arguments that, in so far as alternative methods of categorising the patient's need are not immediately forthcoming, there will not be a clinically suitable (and therefore reasonable) alternative for the relevant task. Nor will it usually be possible to appeal to the autonomy principle to overcome this limitation. For, any default triaging system – e.g. one broadly aligned with medical urgency or with a first come, first served approach – may be expected to be equally unreceptive to the patient's personal objectives. Consequently, disclosure of this AI use and its, potentially uncertain, objectives will not be necessary, even where it makes certain decisions for the patient.

By comparison, a patient is currently envisaged to have available, suitable alternative options when deciding whether to undergo a diagnostic AI analysis. This includes the option of non-testing or of seeking out a narrower form of analysis. Indeed, even if other testing procedures were not immediately available, the significance of this affirmative choice for the patient's autonomy would still demand its disclosure.[1017]

The only challenge that could emerge for the disclosure of this aspect is arguably a lack of professional awareness: an alternative cannot be reasonable if the professional ought not to have been aware of it. As was mentioned above, unlike a traditional situation involving the non-disclosure of alternatives, ML devices do not involve the professional imposing their preference on the patient. It is the use of the tool that has the consequence of determining a choice. This choice, and the evaluative criteria underlying it, may be obscure to differing degrees, even for the user.

However, this is not anticipated to be a substantial problem for the kinds of disclosure advocated for here. As discussed in Chapter 2, a professional deploying the technology must have a broad awareness of its purposes and – in the case of AI diagnosis – they can reasonably foresee that a patient may not wish to confront the possibility that they are suffering from

1016 *Webster v Burton Hospitals NHS Foundation Trust* [2017] EWCA Civ 62, (2017) 154 BMLR 129 [31]; *Bayley v George Eliot Hospital NHS Trust* [2017] EWHC 3398 (QB), [2017] 12 WLUK 670 [96].

1017 The discussion of this autonomy challenge under the battery mechanism should also be considered relevant to this assessment.

a serious condition. Anticipating the AI's determination of this choice is possible and therefore it must be disclosed.

In sum, the existing UK approach to the disclosure of alternative options suggests that there is a concern to maintain a connection between patients' values and the decisions made in their care. Some instances where AI possess a wide discretion to determine treatment objectives must be disclosed to the patient under this head. One may say that the selection of the AI tool amounts to a pre-determination of their alternatives. Simultaneously, this is not a uniform duty and – in seeing how far it extends – one ought to consider whether the AI denied the patient a true choice that could be anticipated by the professional.

ii. AI's lesser influence on the pursuit of objectives

As the previous section has argued, the foisting of external evaluative judgments upon the patients can constitute an interference with the patient's autonomy that can trigger disclosure obligations under the law of negligence. Yet, up until now, we have considered only a particularly strong interference: where a patient is not given any choice at all whether to pursue certain goals in their care. Another related, lesser violation was seen to occur in relation to AI where it subjected the patient to external influences that were not easily accessible to rational evaluation: nudges.

English law certainly recognises the existence of external psychological pressures and that these can contribute to circumstances where it is not possible for the patient to provide their informed consent. For example, in *Thefaut v Johnston* the judge found that a conservation between surgeon and patient on the day of the surgery was not sufficient for obtaining their informed consent to the procedure because:

> this is neither the place nor the occasion for a surgeon for the first time to explain to a patient undergoing elective surgery the relevant risks and benefits. At this point, on the very cusp of the procedure itself, *the surgeon is likely to be under considerable pressure of time* (to see all patients on the list and get to surgery) and *the patient is psychologically committed to going ahead*. There is *a mutual momentum towards surgery*

> *which is hard to halt*. There is no "adequate time and space" for a sensible dialogue to occur and for free choice to be exercised.[1018]

There are arguably two related objections encapsulated in this quote. First, under circumstances of acute time pressure the professional is not able to facilitate the kind of comprehension that is necessary for informed consent. This is not our main concern here. Second, there is a mutually reinforcing momentum between patient and doctor, which may be described as a bias impacting them, that results in an impaired process of decision making. In short, *Thefaut* recognises that the presence of biases can defeat the patient's informed consent. This has subsequently been affirmed.[1019]

In *Thefaut's* circumstances there was an available way to avoid the presence of this bias of course. It was open to the surgeon to lead the discussions with the patient at an earlier date, or under different circumstances, where time-pressure would not seriously impact either of their decision making. The extent to which the creation of these circumstances, i.e. the exclusion of such pressures, is a matter of degree has since been exemplified in *Ollosson v Lee*. While important information about a risk was only imparted to the patient on the day of a procedure, the court found that 'the information provided at the surgery was in an unpressurised situation, with time to reflect, and against a background where Mr Ollosson had arrived with some knowledge that there was a risk'.[1020] In consequence, the patient did give 'properly informed consent'.[1021]

These cases stand for the proposition that a doctor is obligated to create circumstances where the patient can meaningfully appreciate the information they are given for a clinical choice. This means *inter alia* that they are not influenced by biases to an unacceptable extent when giving their consent. Given the distinction of *Thefaut* in *Ollosson* it is probable, as well as understandable, that the bar for such biases will not be set too low. Biases are endemic in all decision making and they cannot be avoided. Arguably, the moral position that the doctor allegedly assumed in *Montgomery* could

1018 *Thefaut v Johnston* [2017] EWHC 497 (QB), [2017] 3 WLUK 328 [78] (my emphasis).

1019 *Hassell v Hillingdon Hospitals NHS Foundation Trust* [2018] EWHC 164 (QB), (2018) 162 BMLR 120 [53]. See also the aforementioned case of *Jones v Royal Devon and Exeter NHS Foundation Trust* [2015] Lexis Citation 3571.

1020 *Ollosson v Lee* [2019] EWHC 784 (QB), [2019] 3 WLUK 562 [157].

1021 ibid [158].

acceptably have influenced their advice, but they 'must not put pressure on the patient to accept their advice'.[1022]

The same can be observed in cases dealing with the legal figure of undue influence, according to which the validity of consent may be undermined by an external influence that persuades 'the patient to depart from her own wishes, to an extent that the law regards it as undue'.[1023] It must be borne in mind that the pressures exerted to obviate consent altogether must be correspondingly higher than those that may affect a patient's ability to give an informed consent. Nevertheless, it is instructive that in *U v Centre for Reproductive Medicine* the Court of Appeal accepted the argumentation of the first instance court that undue influence requires 'something more than pressure (…) it does not matter how strong the persuasion was so long as it did not overbear the independence of the patient's decision'.[1024] The law is clearly accustomed to distinguishing between acceptable and unacceptable external pressures on patient decision making.

It is hardly arguable that the kinds of AI-induced biases identified in our technical analysis amount to external pressures that undermine the validity of consent, overbearing the independence of a patient's decision. By contrast, it may be arguable that they can impact the patient's *informed* consent under the doctrine elaborated in *Thefaut*, although the biases of ML technology require a slightly different analysis. For, as outlined in Chapter 2, it is not clear that the patient's wider position can be changed to eliminate a relevant bias, akin to the possibility in *Thefaut* of simply making the relevant disclosures earlier. Some degree of AI nudging appears inevitable and acceptable – much like the position regarding a doctor's advice.

As NHS documentation has acknowledged, the best way to counteract biases in the use of ML technology under such conditions is to educate individuals 'to recognise their inherent biases, and understand how these affect their use of AI-derived information'.[1025] In other words, where pa-

1022 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [78], citing General Medical Council, 'Consent: Patients and Doctors Making Decisions Together' (London 2008) para 5.

1023 *Re T (adult: refusal of treatment)* [1993] Fam 95, 121.

1024 *U v Centre for Reproductive Medicine* [2002] EWCA Civ 565, [2002] Lloyd's Rep Med 259 [19].

1025 Nix, Onisiforou and Painter, 'Understanding Healthcare Workers' Confidence in AI' (2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in

tients may be affected by AI biases, *Montgomery's* and *Thefaut's* application of the informed consent standard may translate into an obligation to explain potential AI biases to the patient to facilitate, as far as is reasonably practicable, an unbiased decision.

This still leaves the problem that ML devices would not ordinarily be in a position to influence patients to a sufficiently problematic degree. Nudging a patient into making a certain evaluative choice regarding their care is not a natural analogy to acute time pressure, which impairs their appreciation of a material risk. Yet, a more detailed analysis of problematic pressures is not forthcoming in the case law.

Drawing on the reflective dimension of autonomy it could be argued that even slight, uncertain influences (such as nudging) are sufficiently problematic where it relates to certain non-therapeutic interests of the patient. These would serve not as prompts to clarify decision-making, but as a means to subvert it. However, in the UK it is a complicating factor that these obligations have not arisen under the negligence action. Rather, similarly to the above analysis of expertise-related disclosure, successful claims have been made under the stronger battery action.[1026] The success of such actions under negligence must be seen as an unlikely prospect.

Moreover, in light of the physician's own, potentially limited, understanding of ML biases – a knowledge of which is not necessarily linked to responsible use – it may also be difficult to prove that a reasonable defendant has fallen short of the informed consent standard in not discussing this aspect of AI use. In consequence, although nudging may be a challenge to patients' procedural autonomy, it is not one against which the law of informed consent can provide straightforward protection.

5. Summation

In summation, the breach element of the informed consent action in UK law was redefined less than a decade ago. As such, it is understandable that uncertainties persist in the operationalisation of the standard that was laid down. Nevertheless, arguments can be advanced that the following classes of information must be imparted to the patient: that the AI possesses a

-ai> accessed 11.11.2022. This document referred specifically to practitioners but the solution to relevant biases is, as discussed in Chapter 2, not evidently effected by an individual's level of experience.

1026 *Appleton v Garrett* (1997) 34 BMLR 23.

risk-relevant status, that there have been substantial changes to the human expertise which the patient expected to be applied to the clinical decision, and that a significant choice may be pre-determined if the patient agrees to a seemingly innocuous use of AI.

D. Causation

The causation test connects the outlined elements of a relevant defendant's breach of duty to the actionable damage. Traditionally under UK law the claimant must prove that, but for the breach, they would not have suffered the damage.[1027] *Montgomery* applied this test to the informed consent context, considering what the particular patient would have done had they been advised of the information which the breach of duty has obscured from them.[1028] In other words, post-*Montgomery* the courts apply a subjective test for causation.[1029]

Given the findings on actionable damage in UK negligence, one must distinguish two possible applications of this test. In so far as an autonomy violation itself can be counted as the injury suffered by a patient, it is sufficient to show that a requisite, serious breach of the principle has occurred. This is arguably supported by the evaluated case of *Chester v Afshar*. There, given that the non-disclosure of a material piece of information had sufficiently impacted the patient's autonomy, it was not necessary to establish a traditional form of but-for causation. As Amirthalingam has stated, this approach finds 'a causal link between the negligence and the loss of the right to informed consent, but it does not establish causation with respect to the physical injury'.[1030] The causation element simply asks whether these shortcomings did impact facets of decision making that are important to the individual patient. As the three arguably actionable breaches regarding ML have themselves been framed as possessing a degree of significance, they do not call for a separate analysis at this stage.

1027 *Barnett v Chelsea and Kensington Hospital Management Committee* [1969] 1 QB 428.

1028 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [96]-[105].

1029 Turton, 'Informed Consent to Medical Treatment Post-Montgomery' (2019) 27(1) Medical Law Review p. 108, 109.

1030 Amirthalingam, 'Causation and the Gist of Negligence' (2005) 64(1) The Cambridge Law Journal p. 32, 33-34.

By contrast, if a claim is made for the physical injury that is suffered attendant upon an autonomy violation, then it must be established that, but for the failure to disclose a piece of information to the patient, they would have made a decision that averted the physical injury. This was the kind of claim advanced in *Montgomery* itself, where it was found that the patient would have opted for a caesarean section had she been informed of the possibility of shoulder dystocia eventuating. On a balance of probabilities the baby would then have suffered no harm.[1031]

In making this finding the courts have had recourse to the kinds of considerations that evidence a patient's longer-held, serious commitments. In other words, those that may be taken as evidence of how they would have exercised the reflective dimension of autonomy. A vivid example is given by *Shaw v Kovac*, where the patient's hypothetical decision to refuse an operation was ascertained by reference to the fact that 'He was mentally alert (…) having the cautious and conservative nature of someone who had been born in the Scottish Highlands'.[1032] By contrast, in *C v Colchester Hospital University NHS Foundation Trust* it was found that the patient would not have taken the risk of not having a surgery, given their well-documented, overwhelming desire to be cured. Causation could therefore not be established.[1033]

In spite of this test's affinity with one element of autonomy, a focus on decision-making related to physical harm would nevertheless limit the protection offered to this principle. It would require a patient to demonstrate that established elements of their character would have led to their rejection of a procedure involving AI use that, in turn, would have avoided a physical injury. Whether this is possible will depend on factors incidental to autonomy.

It is also unclear in how far the courts overcome this evidentiary difficulty by appealing to an objectified patient, rather than to the individual patient's personal commitments. *Diamond v Royal Devon and Exeter NHS Foundation Trust* arguably blurred this line by endorsing an objectively rational decision and then finding that the patient would not have behaved irrationally.[1034] It will be seen in the American analysis that such a position

1031 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430 [104].
1032 *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [28].
1033 *C v Colchester Hospital University NHS Foundation Trust* [2018] 2 WLUK 850 [61].
1034 *Diamond v Royal Devon and Exeter NHS Foundation Trust* [2019] EWCA Civ 585, (2019) 170 BMLR 49 [9], [19]-[22]; Austin, 'Correia, Diamond and the Chester

would restrict the protection of the individual patient's autonomy even further.

The *Chester v Afshar* autonomy-based exception to ordinary causation principles is not envisaged to provide much assistance to such claims. Subsequent courts have indicated that its analysis was exceptional – only assisting the patient when an injury is intimately connected with a failure to warn and in so far as it would have led the patient to defer or reject a relevant procedure.[1035] As AI's autonomy challenges have not been argued to have any distinct effect on the timing of a patient's decision, *Chester* will not meaningfully expand the scope for personal injury claims that are related to autonomy violations.

Overall, one may make two findings. To the extent that a claimant can claim successfully for a violation of their autonomy interest, the causation element should serve to emphasise that this interference must be of a sufficiently serious type to impact their process of decision making. This is entirely congruous with the outlined autonomy principle and it is not envisaged to separately restrict the identified instances where a patient can recover for undisclosed features of ML use.

Where, however, the patient brings (or must bring) a claim with respect to a personal injury they have suffered, they must go to the additional, not undemanding, lengths of showing that their decision would have been altered so as to avoid the harm.[1036] Although the courts still consider aspects of the patient's reflective autonomy to make out this claim, it imposes a further burden that can restrict the success of autonomy-based claims. Whether this will be the case depends very much on the kind of evidence that is available to patients regarding their wider character and foundational commitments.

Exception: Vindicating Patient Autonomy?' (2021) 29(3) Medical Law Review p. 547, 558. Note also that this understanding of rationality (what Austin refers to as an ideal desire approach) goes beyond Pugh's conception of theoretical rationality discussed in Chapter 3, which left the weighting of relevant interests to the individual.

1035 *Correia v University Hospital of North Staffordshire NHS Trust* [2017] EWCA Civ 356, [2017] 5 WLUK 285 [24], [28].

1036 For an analysis of this see: Austin, 'Correia, Diamond and the Chester Exception: Vindicating Patient Autonomy?' (2021) 29(3) Medical Law Review p. 547.

E. Awarding damages

Under UK tort law the fundamental aim of an award of damages is to put the claimant in the position that they would have been in had they not suffered the tortious injury.[1037] However, as Jones has eloquently stated regarding non-pecuniary loss, including not just injuries to intangible interests like autonomy, but also instances of personal injury that impair the patient's physical capacities:

> The award is inevitably an arbitrary one. In practice the courts adopt a tariff or "going rate" for specific types of injury in an attempt to achieve some degree of consistency between claimants with similar injuries and to provide a basis for the settlement of claims.[1038]

In the examined cases, where an autonomy injury has been deemed a legally cognisable loss, the arbitrariness of this award has been singled out as a particular focus of criticism.[1039]

If the real loss suffered in these situations is to the personal autonomy of the individual, as it has been argued, then it is open to the courts to make conventional awards, akin to that awarded in *Reese*, in other non-disclosure cases.[1040] Ordinarily this will be a more appropriate response than compensating for a personal injury, which was not strictly speaking caused by the defendant, as occurred in *Chester*.[1041] By comparison, if autonomy harms and other recognised categories of harm, such as personal injury, are caused simultaneously by a defendant's breach then it may be a justifiable step for the law – in order to maintain coherence and simplicity – to determine an award for autonomy protection through the rules applicable to that other category of injury.[1042]

1037 *Livingstone v Rawyards Coal Co* (1880) 5 App Cas 25, 39.

1038 Jones, *Medical Negligence* (Sixth Edition 2021) para 12-003.

1039 Keren-Paz, 'Gender Injustice in Compensating Injury to Autonomy in English and Singaporean Negligence Law' (2019) 27(1) Feminist Legal Studies p. 33, 45; Priaulx, *The Harm Paradox: Tort Law and the Unwanted Child in an Era of Choice* (2014) 76.

1040 Maclean, *Autonomy, Informed Consent and Medical Law: A Relational Challenge* (2009) 189.

1041 Keren-Paz, 'Gender Injustice in Compensating Injury to Autonomy in English and Singaporean Negligence Law' (2019) 27(1) Feminist Legal Studies p. 33, 45.

1042 Keren-Paz, 'Compensating Injury to Autonomy in English Negligence Law' (2018) 26(4) Medical Law Review p. 585, 590 (fn. 37), 600-601. This is arguably what

To the extent that any special issues arise from our definition of the autonomy principle and its application to the informed disclosure of ML characteristics, it is arguable that these concern the amount to be awarded in a standalone autonomy claim. The principle does not provide a general yardstick with which to determine an appropriate amount of damages. At most, the award made for breaches of informed consent should, as Jones has stated, be consistent with the awards made in similar cases. In other words, *Rees's* £15,000 sum should serve as a reference point for particularly serious autonomy violations, akin to the denial of an individual's right to opt for or against parenthood. Here, AI's pre-determination of significant diagnostic choices provides the closest analogy.

Reductions in professional expertise or the non-disclosure of AI's risk related status arguably have lesser impacts and thus justify lesser autonomy-based awards. As the analysis in Chapter 3 has indicated, they impair but do not preclude the exercise of procedural autonomy – even where the relevant decisions have serious consequences. Ultimately one may therefore doubt whether a theoretically available award of damages, which ought to be acknowledged in such circumstances, would amount to an effective form of protection in practice, given the minimal awards that are likely to be made and the likely cost of bringing a tort claim.[1043]

III. The UK General Data Protection Regulation

A separate in-depth examination of data protection law was rejected in Chapter 1. Nevertheless, one would be remiss not to mention that a subset of obligations under the *United Kingdom General Data Protection Regulation* (UK GDPR) and the associated *Data Protection Act* (DPA) 2018 could supplement one particular shortcoming identified under the common law causes of action.[1044] Namely, the lack of any detailed institutional disclosure

occurred in *Montgomery*. Nolan appears to accept the possibility of this position too: Nolan, 'Negligence and Autonomy' [2022](2) p. 356, 366-367.

1043 Clark and Nolan, 'A Critique of Chester v Afshar' (2014) 34(4) Oxford Journal of Legal Studies p. 659, 685. Contrast the argument in *Shaw v Kovac* that awards of relatively modest amounts may spur claims, in the hopes of forcing a settlement: *Shaw v Kovac* [2017] EWCA Civ 1028, [2017] 1 WLR 4773 [82].

1044 Following the UK's separation from the European Union, the UK adheres to the United Kingdom General Data Protection Regulation. This was implemented through: Section 3 European Union (Withdrawal) Act 2018 and the Data Pro-

obligation regarding AI use. Indeed, the provisions that will be identified as relevant to this circumstance are inapplicable to other situations that have been argued to constitute the great majority of clinical situations that involve human mediation of ML devices. It will be seen that, for UK GDPR's requirements to shape relevant disclosure obligations, there must be 'a decision based solely on automated processing'.[1045]

One can therefore focus on institutional AI use. The example cited above concerned a hospital using an ML device to triage patients on the basis of the information provided by them. This would uncontroversially render the healthcare institution a data controller under article 4(7) UK GDPR with relevant obligations while they process the patient's data.[1046] One such obligation is to obtain the patient's consent for 'significant decisions based solely on automated processing'.[1047] They also have a duty to inform the patient 'at the time when personal data are obtained' of 'the existence of automated decision-making [and provide] meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'.[1048]

---

tection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019.

1045 Article 22(1) UK GDPR; Section 14(1) DPA 2018. As has been previously discussed, human mediation of AI decisions must be a matter of degree and this is recognised under the GDPR too: 'To qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture.': Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2018) 20. However, so long there is more than a token acceptance of AI recommendations, it is unclear what kind of human oversight would be insufficient. In relation to our identified types of medical ML devices, this limits the practical relevance of UK GDPR and associated legislation to the purely institutional use of AI.

1046 Article 4(2) UK GDPR. See also the example of a GP surgery on the website of the Information Commissioner's Office: Information Commissioner's Office, 'What are 'controllers' and 'processors'?' (17.10.2022) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/controllers-and-processors/what-are-controllers-and-processors/#1> accessed 18.3.2023.

1047 Section 14(1), (3)(c) DPA 2018, deriving from Article 22 UK GDPR. For a detailed analysis of the configuration of this requirement in the triaging context, see: Mourby, Ó Cathaoir and Collin, 'Transparency of Machine-Learning in Healthcare: The GDPR & European Health Law' (2021) 43 Computer Law & Security Review.

1048 Article 13(2)(f) UK GDPR. See similarly Article 14(2)(g) UK GDPR, where data is obtained from a third party, rather than the patient, and Article 15(1)(h) UK GDPR, seemingly conferring an *ex post* right of access to meaningful information.

To be clear, these would be obligations outside of the common law. It would not be equivalent, or give rise, to a duty of care in negligence.[1049] Instead, medical AI uses must be brought under the statutory obligations and their requirements must be interpreted. As argued in Chapter 4, the autonomy principle has been relied upon also for these purposes in UK law. This provides an additional, forceful reason to examine this aspect of the UK GDPR: a reliance on principle is arguably appropriate and necessary given the substantial ambiguity involved in the GDPR's specification of 'meaningful information' and the lack of case law on the definition of this element.

In defining the requirement one can say, as a point of departure, that the data controller will necessarily have to disclose the fact that it is using an automated decision system. In our example, a patient must be told that a triage decision will be made by an ML device without meaningful human involvement. So much seems uncontroversial.[1050] It also represents an improvement *vis-à-vis* the common law position from the perspective of patient autonomy, given that only a limited duty to warn had been constructed for institutions under negligence.

It is also evident from the statutory text that more information must be provided: the logic and envisaged significance of the decision. However, it is hotly contested what information falls under these heads.[1051] Without authoritative or persuasive legal sources on the issue, a case of uncertain

1049 Under the earlier Data Protection Act 1998 the imposition of a co-extensive duty in negligence had been forcefully rejected: *Keith Smeaton v Equifax plc* [2013] EWCA Civ 108, [2013] 2 All ER 959 [75].

1050 E.g. Hamon and others, 'Impossible Explanations?: Beyond Explainable AI in the GDPR From a COVID-19 Use Case Scenario' (FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, 03.03.2021 - 10.03.2021) 557; Mourby, Ó Cathaoir and Collin, 'Transparency of Machine-Learning in Healthcare' (2021) 43 Computer Law & Security Review, 5-6.

1051 Most of these debates are conducted regarding the General Data Protection Regulation under EU Law (EU GDPR), but they bear obvious relevance to the UK's scheme, which has taken over the relevant, outlined aspects. See: Selbst and Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) International Data Privacy Law p. 233; Wachter, Mittelstadt and Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) International Data Privacy Law p. 76; Custers and Heijne, 'The Right of Access in Automated Decision-Making: The Scope of Article 15(1)(h) GDPR in Theory and Practice' (2022) 46 Computer Law & Security Review p. 105727; Hoeren and Maurice Niehoff, 'Artificial Intelligence in Medical Diagnoses and the Right to Explanation' (2018) 4(3) European Data Protection Law Review p. 308; Goodman and Flaxman, 'European Union Regula-

strength can be made for a form of disclosure that responds to the AI challenges, as identified under the concept of autonomy in UK law.

In this respect, the ability of AI to independently pursue objectives falls most naturally under the significance of anticipated consequences. As outlined above, a patient ought to be made aware of the identifiable outcomes associated with the AI's purposes – particularly where they are surprising and significant. To a lesser degree, this obligation may also be framed to include the specific risks posed by AI, although these would have to be sufficiently likely to qualify as 'envisaged' and sufficiently severe to class as 'significant'. Arguably, this would be a very high bar indeed, one which is unlikely to be satisfied by many AI.

Regarding the logic of AI functioning, a case can be made that a patient should be made aware of the difference between AI expertise and human, professional expertise in a general sense. It has been argued that AI functions without necessarily relying on the same body of scientific knowledge as their human counterparts, and that this is capable of generating decision-relevant uncertainty.

Where a patient has to interact directly with an AI, the framing of UK GDPR may also require forms of disclosures that are additional to those described for negligence. Arguably these should impart that ML devices may provide explanations that are not immediately connected to the underlying functionality – a pervasive issue identified regarding explainable AI in Chapter 2. The legislation's reference to 'logic' should, at a minimum, require information about the different model types in play: one reaching a decision and one explaining that decision – designed not to replicate the decision-making process but also to be understandable. The patient is then in a better position to assess whether they are being influenced towards reaching a certain decision and they can better accommodate the AI output within their own evaluative framework. Similarly, where the kinds of features and criteria used by the AI are known, there is an argument that these must be disclosed to the patient. This can be framed as a general *ex ante* disclosure,[1052] required by UK GDPR under the autonomy principle.

---

tions on Algorithmic Decision-Making and a "Right to Explanation"' (2017) 38(3) AI Magazine p. 50.

1052 Framing it as an aspect of the obligation under Article 13 EU GDPR: Wachter, Mittelstadt and Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) International Data Privacy Law p. 76, 82-83.

To conclude, the statutory requirements under DPA 2018 and UK GDPR require an institution that is relying solely on an AI/ML device for certain medical purposes, such as triaging, to give the patient notice that this is occurring. On top of this, there must be disclosure of: potential, significant outcomes; the general differences between human and AI functioning; the presence of different models with different objectives; known criteria and features relied on by the AI. If such disclosure is not made, then a claim may be possible for compensation, provided that the patient has suffered material damage or immaterial damage, including distress.[1053]

IV. Conclusion

In summation, UK law has been argued to provide three mechanisms through which AI's autonomy challenges can be partially addressed. There is scope for argumentation that the tort of battery would, in circumstances where there is direct contact with the claimant and where an intention to use the AI is formed before such contact, obligate a professional to disclose the broad purpose of that device. This includes its ability to make certain serious choices.

The negligence mechanism also supported such disclosure. In addition, weaker arguments could be advanced that advice must be proffered regarding AI's risk-relevant status and substantial shifts in human expertise – at least where the patient had some expectation that the person undertaking their care is a specialist. For all such disclosures the negligence action imposed conditions that could severely restrict the availability of claims. For example, it was not possible to argue that institutions possess duties of care with demanding informational components.

By comparison, the force of other limiting elements was uncertain. Under the authority of the UK's highest court, it was possible to argue that (1) UK law recognised a significant autonomy violation, without more, as an actionable form of damage and (2) the significance of an autonomy violation at the breach stage would bring about a relaxation of the test at the causation stage. Subsequent case law has not sought to affirm or extend such pronouncements, however. In consequence, negligence has arguably been interpreted to provide a coherent framework for autonomy protection,

1053 Article 82 UK GDPR; Section 168 DPA 2018.

especially under the influence of the procedural autonomy principle, but how far this is accepted in practice is in serious doubt.

Finally, it was argued that a targeted legislative intervention has provided a complementary form of protection to battery and negligence, by imposing informational obligations on institutions. This would require the disclosure of unmediated AI use and potentially, albeit much less apparently, AI's more specific autonomy-related characteristics.

Chapter 7: Californian tort law

In the United States of America, as in England, it is not the abstracted principle of patient autonomy, but its implementation through specific common law mechanisms that will determine the informational duties owed to patients regarding medical uses of artificial intelligence (AI). Both battery and negligence were identified as the modalities with the most potential in this respect. It is the task of this chapter to evaluate the capacity of these torts to meet the outlined, novel autonomy challenges introduced by machine learning (ML).

In order to conduct this analysis in the requisite depth, and with a sufficiently nuanced understanding of the applicable rules-based framework, the tort law of California has been selected to serve as a case study. California provides an instructive example partly because the pronouncements of its courts have had a significant impact on the development of informed consent law in the other states – setting persuasive precedent and inspiring innovation – and partly because its rules typify wider American trends.[1054] These trends have led to a more stringent adherence to the doctrinal underpinnings of tort law, representing a notable contrast with the analysis of the previous chapter.

As outlined in Part II, the autonomy principle will nevertheless have an integral role to play in our assessment. To gauge the common law's response to novel problems one must not limit oneself to the established rules, but one must also be capable of anticipating their potential for development. In the following, I therefore analyse how California's torts of battery and negligence can be applied to the challenges of medical AI's implementation, while allowing for realistic developments in line with the principle of patient of autonomy.

Generally, these torts will have a *post facto* role: awarding compensation for violations that have already occurred. In so far as battery is concerned,

1054 These will be pointed out in the course of the analysis. Examples include: tight restrictions on the individuals on whom informed consent duties can be imposed, a narrow approach to legally compensable injuries in negligence and the (determinative) recourse to the reasonable person in shaping negligence's breach and causation requirements.

a close connection to the patient's right to privacy will, for a restricted subset of medical decisions, also enable a constitutional claim to be brought on the permissibility of certain actions. Although AI's autonomy problems are expected to remain outside of the scope of this right (as discussed in Chapter 5), it should be mentioned that even here the outlined actions and elements will provide a central reference point and, as such, deserve close attention.[1055]

I. Battery

A battery constitutes both a tort and a crime. Although our focus is on the former, the Californian courts have generally operated under the assumption that the statutory, criminal definition can also be applied to the common law tort.[1056] A preliminary understanding can therefore be gleaned from California Penal Code section 242, which states: 'A battery is any willful and unlawful use of force or violence upon the person of another'.[1057] An individual's consent to a tortious battery will generally provide a defence to the one perpetrating the wilful use of force or violence.[1058]

With a preliminary definition in hand, one can understand why Californian courts, representing a theme in the United States more generally,[1059] utilised the battery cause of action early on to establish the legal requirement that a patient's consent is necessary for medical treatment. Such claims tended to involve invasive surgeries that were performed without authorisation. For instance, in *Valdez v. Percy* the plaintiff authorised the surgeon to remove an enlarged axilla gland but alleged that she had not consented to the procedure which was then performed during the same surgery: the removal of her right breast. The Court of Appeal held that,

1055 See for example: *Barber v. Superior Court* (1983) 147 Cal.App.3d 1006.

1056 *McChristian v. Popkin* (1946) 75 Cal.App.2d 249, 260; *Fraguglia v. Sala* (1936) 17 Cal.App.2d 738, 742. However: 'the torts of assault and battery are not defined by statute, and the court is afforded the opportunity to extend the concept of the tort beyond the limits placed on the corresponding crime by its statutory definition': *California Jurisprudence* (Third Edition 2022), Assault and Other Willful Torts § 27.

1057 California Penal Code section 242.

1058 Witkin, *Summary of California Law* (Eleventh Edition 2022), Torts § 457.

1059 Faden, King and Beauchamp, *A History and Theory of Informed Consent* (1986) 120-125; Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 224-226.

without obtaining the plaintiff's consent, these actions would constitute an actionable battery.[1060]

On the basis of *Valdez* it was subsequently asserted in *Estrada v. Orwitz* that '[t]here can be no doubt that an action based on the theory that an operation is performed without consent of the patient charges an assault and battery, and that negligence has nothing to do with such an action'.[1061] Hence, when the plaintiff argued that certain teeth had been extracted without consent, the court determined that they had charged a battery.[1062] Many other cases could be adduced to support this proposition.[1063] Collectively they establish that, to avoid a liability in battery, a medical treatment must ordinarily be based on the patient's consent.

This has also connected the battery cause of action to the protection of patient autonomy. Both courts and commentators have recognised that battery serves a distinct role in this endeavour. In particular, it recognises the patient's ability to exercise control over their body as a significant instantiation of the right to make autonomous decisions.[1064] In *Thor v. Superior Court* the Californian Supreme Court adduced the outlined battery case law as the first piece of evidence for the common law's long standing tradition of protecting the principle of patient autonomy.[1065] The principle was not only seen as intertwined with the informed consent doctrine ('a corollary'),[1066] but it was also seen to further the patient's ability to undertake their own balancing of relevant interests.[1067] The recent case of *Stewart v. Superior Court* similarly drew on federal and state law to define 'the right to personal autonomy' that was denied a patient by the doctors' decision to sign the consent form for them.[1068]

1060 *Valdez v. Percy* (1939) 35 Cal.App.2d 485, 491-492.
1061 *Estrada v. Orwitz* (1946) 75 Cal.App.2d 54, 57. Note that the usage of the term 'assault' in this statement represents a common ambiguity.
1062 ibid 57.
1063 *Weinstock v. Eissler* (1964) 224 Cal.App.2d 212; *Hundley v. St. Francis Hospital* (1958) 161 Cal.App.2d 800; *Keister v. O'Neil* (1943) 59 Cal.App.2d 428.
1064 'Patient autonomy was initially identified with and subsumed under an interest in physical security, protected by rules proscribing unconsented touch': Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 224.
1065 *Thor v. Superior Court* (1993) 5 Cal.4th 725, 735.
1066 ibid 735.
1067 ibid 734-736.
1068 *Stewart v. Superior Court* (2017) 16 Cal.App.5th 87, 104-107.

Moreover, battery's role in the protection of this interest has also been recognised to have a constitutional dimension in California. Specifically, this stems from the constitutional guarantee of Privacy under Article I, section 1 of the California Constitution. This provides a right to 'give or withhold informed consent with respect to a proposed medical treatment'.[1069] As was already addressed in Chapter 5, analyses of both the California and United States Constitution place heavy reliance on the common law reasoning and development in the shaping of this right. As such, it does not require separate analysis, but it does highlight the preeminent part that the battery cause of action plays in the protection of one of the patient's most fundamental interests. This strengthens the legitimacy of its principled application and development in a relevant healthcare context.

With this fundamental framework in place, two dimensions of the battery cause of action should constitute our focus. First, while battery is associated with these autonomy objectives, the Californian interpretation is subject to rule specific considerations that must be addressed. Second, one must consider the information that must be disclosed to obtain a valid consent to battery.

A. Limitations flowing from the battery doctrine

Within the context of medical treatment, *Ashcraft v. King* provides a summation of battery's requirements that builds upon the above definition and which is frequently utilised by the courts: 'A battery is any intentional, unlawful and harmful contact by one person with the person of another'.[1070] This states the three elements that must be considered before engaging with the nature of the patient's consent: the requisite contact; the unlawful or harmful nature of that contact (in spite of the 'and' formulation, it will be seen that fulfilling one of the conditions suffices); and the intent of the person touching the other.

1069 *Foy v. Greenblott* (1983) 141 Cal.App.3d 1, 11.

1070 *Ashcraft v. King* (1991) 228 Cal.App.3d 604, 611. Although there is a general problem of courts using varying, partially conflicting, definitions of these elements, this represents a widely accepted one in the clinical sphere.

1. Contact

Regarding the element of contact, it is clear that absent any touching, as may occur when a physician decides not to treat the patient, there can be no battery and an action lies, if at all, only in negligence.[1071] This was stated in *Scalere v. Stenson*: 'Under a battery theory the doctor's failure to disclose the risks and benefits of non-treatment would not be actionable because there was no unconsented touching'.[1072] Thus, where an AI leads to a decision not to treat, as where a mere analysis of data provides a false negative diagnosis, a battery cause of action will be unavailable. This is manifestly a restriction unrelated to the patient's autonomy.

Beyond this, one must ask what type of contact between doctor and patient fulfils the contact requirement. Californian case law has set a low threshold in this respect. It objects to 'touching of any kind'.[1073] For instance, the touching of the plaintiff's hands, arms and shoulder post-surgery were enough to sustain a relevant claim in *So v. Shin*.[1074] Requiring such direct, albeit minimal, physical contact is one natural interpretation of the statement that a 'mere touching' is required by, as well as sufficient for, a battery.[1075]

As has already been suggested in the British analysis, this would further limit the circumstances in which a battery claim corresponds to AI's autonomy challenges. It would mean that battery could not occur in a class of actions that may be characterised as indirect physical contacts.[1076] In particular, it encompasses the scenario where the doctor gets the patient to

1071 Regarding U.S. tort law generally see: Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 229-232; Dobbs, Hayden and Bublick, *Dobbs' Law of Torts: Practitioner Treatise Series* (Second Edition 2022) § 33.

1072 *Scalere v. Stenson* (1989) 211 Cal.App.3d 1446, 1455.

1073 *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260. The California Supreme Court also endorsed the statement that '"It has long been established that 'the least touching' may constitute battery. In other words, force against the person is enough; it need not be violent or severe, it need not cause bodily harm or even pain, and it need not leave a mark"' in: *People v. Shockley* (2013) 58 Cal.4th 400, 404-405.

1074 *So v. Shin* (2013) 212 Cal.App.4th 652, 671-672.

1075 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 240.

1076 Simons, 'A Restatement (Third) of Intentional Torts' (2006) 48 Arizona Law Review p. 1061, 1077.

ingest forms of medication,[1077] which is not obviously a less serious decision than many forms of medical treatment involving physical touch.

In California no decision directly confronts this point. *Obiter* statements were made in *Freedman v. Superior Court* that an action in battery should not be barred by the manner in which contact was made. Specifically: 'we would not suggest distinctions in terms of the physical nature of the touching (punctures or cuttings being batteries and ingestion of medication not, for instance)'.[1078] This was intermingled with a discussion of the informed consent necessary for treatment, suggesting that responding to grave deviations from the patient's authorisation to treatment was the primary concern.[1079]

Some support for this position can also be gleaned from out-of-state case law, which purported to follow the principles underlying a leading Californian case on informed consent. In *Mink v. University of Chicago* the District Court perceived that the mechanics by which a treatment was administered were indistinguishable according to the principle underlying *Cobbs v. Grant*.[1080] It was held that a battery claim should be permitted for such indirect contact: '[h]ad the drug been administered by means of a hypodermic needle, the element of physical contact would clearly be sufficient. We believe that causing the patient to physically ingest a pill is indistinguishable in principle'.[1081]

Since the question at issue, whether or not there was contact sufficient for battery, goes directly towards the question of whether the patient's bodily integrity was threatened, *Mink's* extension of 'touching' to indirect contacts must be supported by an alternative justification. Namely, the principle of autonomy. This would explain why the doctrinal validity of the court's approach remains highly contested.[1082] It is because, by accepting the relevance of the principle of patient autonomy, the court interpreted the

1077 ibid 1077.
1078 *Freedman v. Superior Court* (1989) 214 Cal.App.3d 734, 740, fn. 2.
1079 ibid 740, fn. 2.
1080 *Mink v. University of Chicago* (N.D.Ill. 1978) 460 F.Supp. 713, 716-718. *Cobbs* constitutes a seminal Californian informed consent case, as was discussed in Chapter 5 and will be discussed further below.
1081 ibid 718.
1082 See: Simons, 'A Restatement (Third) of Intentional Torts' (2006) 48 Arizona Law Review p. 1061, 1077.

battery cause of action in a way that departed from the more traditional interpretation, shaped by narrower value of bodily integrity.[1083]

Ultimately, whether Californian courts would follow the same reasoning remains an open question, but there is *obiter* support for this position and it is arguably in line with the autonomy principle underlying the Californian case law that inspired *Mink* itself. As such, we will proceed on the assumption that only a very narrow band of medical cases (those involving non-treatment) is excluded by the contact requirement. All manner of positive treatment plans, which include AI assistance, would fall to be considered under this element of battery.

2. Unlawful nature

The second condition for a successful battery claim is that the touching was unlawful, harmful or – to use a term that has frequently been treated as synonymous – offensive.[1084] This restriction bears a strong connection to the patient's autonomy. Conduct must be offensive as defined by law and in the clinical sphere California's courts have consistently equated an offensive touching with one that was without, or exceeded, the patient's consent.[1085]

Thus, the court in *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* stated that 'Although typically a battery is a violation of a person's wishes to avoid bodily contact that is hostile, aggressive or harmful, the tort is committed if there is unwanted intentional touching of any kind. (…) For example, a person is entitled to refuse well-intentioned medical treatment.'[1086] *Rains v. Superior Court*, to which *Conte* referred, found that an intentional deviation from consent was sufficiently objectionable – as, by implication, was 'a physician's good faith effort to effect a cure by

1083 This is something that one alternative explanation for the application of battery to medication, is not able to account for as neatly. Namely, that it constitutes an extension of the doctrine that certain non-bodily contacts (e.g. the touching of some object connected to the body) is sufficient for battery. This would be much less controversial, see: Ezra, 'Smoker Battery: An Antidote to Second-Hand Smoke' (1989) 63(4) Southern California Law Review p. 1061, 1092-1093.

1084 *Rains v. Superior Court* (1984) 150 Cal.App.3d 933, 938; *Barbara A. v. John G.* (1983) 145 Cal.App.3d 369, 375.

1085 *People v. Miranda* (2021) 62 Cal.App.5th 162, 175.

1086 *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1266.

exceeding the precise treatment to which consent was given'.[1087] Likewise, a contact has repeatedly been described as unlawful for the purposes of clinical battery where it has not been consented to.[1088]

This means that this limitation, including its application to medical AI, will be co-extensive with the nature of the consent that the law requires. If the consent provided to a contact is invalid or non-existent, then the contact is also offensive. The conditions under which consent is deemed insufficient will be dealt with in Section I.B.

3. Intention

The final limitation on the battery cause of action, which cannot be related to the objective of protecting the patient's ability to make a medical decision for themselves, is the requirement that the doctor must have had a requisite intention. This is a strong rule-specific limitation that flows from battery's status as a paradigmatic intentional tort, a characteristic that distinguishes it from the tort of negligence.[1089] Continuing uncertainty persists, however, regarding the intention that is in fact required by U.S. tort law doctrine. Is it merely to do the act in question, such as inserting a scalpel? Is it to inflict some kind of offense or harm? Or is it an intention to act without, or in excess of, the consent of the patient?[1090]

Californian case law appears to represent a microcosm of this uncertainty, referring to a mixture of these intentions. For example, in *Freedman v. Superior Court* it was held that 'battery requires no showing of "scienter" or any intent to do wrong—only an intent to cause the harmful unconsented touching'.[1091] In other words, the first position is supported: the

1087 *Rains v. Superior Court* (1984) 150 Cal.App.3d 933, 941.

1088 *Ashcraft v. King* (1991) 228 Cal.App.3d 604, 611; *Fluharty v. Fluharty* (1997) 59 Cal.App.4th 484, 497; *Daley v. Regents of University of California* (2019) 39 Cal.App.5th 595, 602; *Piedra v. Dugan* (2004) 123 Cal.App.4th 1483, 1495.

1089 Dobbs, Hayden and Bublick, *Dobbs' Law of Torts: Practitioner Treatise Series* (Second Edition 2022) § 28.

1090 Moore provides an overview of the confusion in the case law, as well as the lack of clarity in the Second Restatement of Torts: Moore, 'Intent and Consent in the Tort of Battery: Confusion and Controversy' (2012) 61(6) American University Law Review p. 1585, 1597-1606.

1091 For this proposition, see also *People v. Miranda* (2021) 62 Cal.App.5th 162, 175-176: 'Battery is a general intent crime (…), which means it requires the intent to do the act involved, not an intent to cause a resulting harm'.

physician must only intend to carry out the act that is classified as offensive. By stark contrast, in *Austin B. v. Escondido Union School District* it was found that if the touching is lawful 'it is appropriate and, indeed required, that the jury be instructed that to be liable for battery, a defendant must intend to harm or offend the victim'.[1092]

In medical cases there is generally no intent to cause harm, given the aim of the therapeutic endeavour to confer a benefit on the patient. This means that there must be a direct intention to act unlawfully, i.e. without consent or other legal authorisation.[1093] Merely the intent to do the procedure is insufficient.

Indeed, this was the position adopted by the Court of Appeal in the medical case of *Piedra v. Dugan*. In response to the plaintiff's claim that the doctor had violated a condition that they had placed on their consent, the court asked whether it could be maintained that the doctor had intentionally violated that condition.[1094] Here, given that the treating doctor did not know of this condition, and that knowledge of it could not be imputed to him,[1095] no battery was found. Agreeing with the trial court, it was ultimately held that the doctor had not 'intentionally rendered a treatment that had not been consented to'.[1096] Conversely, one may consider the aforementioned case of *Ashcraft*, where the jury could 'infer an intent to wilfully disregard plaintiff's conditional consent'.[1097] This was borne out by the fact that the defendant, having acknowledged the plaintiff's condition to use only family donated blood in their operation, proceeded to use blood from the hospital's general supply.

Taken together, these cases stand for the proposition that there are instances where the patient's consent has been exceeded, and a grave violation of their ability to make their own medical decisions may have occurred, but a battery claim will not succeed because the requisite intention on the part of the defendant is missing. Yet such instances appear limited to

1092 *Austin B. v. Escondido Union School Dist.* (2007) 149 Cal.App.4th 860, 872.
1093 ibid 872-873; *Dennis v. Southard* (2009) 174 Cal.App.4th 540, 554; *Barouh v. Haberman* (1994) 26 Cal.App.4th 40, 46-47.
1094 *Piedra v. Dugan* (2004) 123 Cal.App.4th 1483, 1497-1499. The ability of patients to impose conditions on treatment will be explored in the next section.
1095 The condition had only been expressed orally to other employees of the institution and the court found that '[t]here is no authority for imputing knowledge of Fountain Valley employees to Dr. Dugan on the claim for medical battery': *Piedra v. Dugan* (2004) 123 Cal.App.4th 1483, 1498.
1096 ibid 1494.
1097 *Ashcraft v. King* (1991) 228 Cal.App.3d 604, 613.

the giving of conditional consent – the purported violation of which was at stake in *Piedra* and *Ashcraft*.

In the great preponderance of cases, dealing with actions that exceed consent, it is suggested that the Californian courts reason primarily on the basis of the nature of the consent obtained and the gravity with which a deviation impacts the patient's ability to make their decision. For the most part, the professional's intent to deviate is then inferred, so that an argument that the scope of the consent had been inadvertently forgotten or deviated from by the defendant would gain little traction.[1098] The law presumes in such cases that the doctor could not have had a reasonable belief that they were acting within the consent of the patient and then they will also be found to have had the requisite intention for battery.

This will encompass cases of non-consent, where the patient has not consented to the procedure at all and it will also be relevant to cases where the doctor carries out a substantially different procedure to the one that the patient consented to. Thus, it was stated in *Cobbs v. Grant* that '[w]hen the patient gives permission to perform one type of treatment and the doctor performs another, the requisite element of deliberate intent to deviate from the consent given is present'.[1099] The Supreme Court reached a determination on the issue of intent by reference to the nature of the consent and also by reference to the gravity of the violation. In contrast, where 'an undisclosed inherent complication with a low probability occurs, no intentional deviation from the consent given appears' and (partially for this reason) battery is inapplicable.[1100]

It was subsequently made clear in *Dennis v. Southard* that it is indeed a *legal presumption* that the professional has the deliberate intent to deviate from consent in such cases – a presumption that was rightly summarised in the Judicial Council of California Civil Jury Instructions (CACI).[1101] Whereas, for conditional consent cases it was determined that these do not allow for an equivalent legal inference from the doctor's actions.[1102] In

1098 Moore, 'Intent and Consent in the Tort of Battery: Confusion and Controversy' (2012) 61(6) American University Law Review p. 1585, 1547-1548.

1099 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 240-241.

1100 ibid 240-241. See also *Burchell*, where the belief that there was an emergency did not defeat the requisite intent: *Burchell v. Faculty Physicians & Surgeons of Loma Linda University School of Medicine* (2020) 54 Cal.App.5th 515, 526.

1101 *Dennis v. Southard* (2009) 174 Cal.App.4th 540, 544.

1102 ibid 544. Arguably, there is uncertainty in how far intention must be proven even in the conditional consent scenario, however. The aforementioned case of *Conte*

People v. Miranda the Court of Appeal also took care to summarise the clinical case law as follows:

> The law in the medical context likewise defines the circumstances when a patient's advance consent to a procedure while unconscious means that a doctor has a reasonable belief that the procedure is lawful and thus does not commit a battery by performing it. A doctor commits a battery when deviating from the consent given to perform a substantially different procedure than the one for which consent was given[1103]

Therefore, Californian law typically only requires that the doctor has an intention to do the offensive act in question. If this act is then performed in circumstances where the law will not attribute to them a reasonable belief that they are acting within the patient's consent, then the element of battery is fulfilled. This includes the performance of procedures that are substantially different to the one consented to and it includes instances where the patient has not consented at all.

4. Summation

In sum, one can see that the courts' stretching of traditional doctrine permits the nature of the patient's consent to shape the wider requirements of the battery action and bears testament to the significance attributed to the principle of patient autonomy. These actions highlight that, when considering failures to provide information on AI use, the key issue will be the legal validity of patient consent.

B. Battery and the nature of valid consent

Due to the aforementioned role of consent in establishing the offensiveness of the act, it is incumbent upon the patient to prove that their consent was absent or tainted in such a way as to justify their claim.[1104] If they can,

---

refers primarily to the specificity of the condition on consent, rather than the doctor's intent to deviate from it: *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1269.

1103 *People v. Miranda* (2021) 62 Cal.App.5th 162, 176.

1104 *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1266.

then a battery action provides one mechanism for realising the autonomy principle's demand for the patient to be informed in their medical decision making. As will be seen, battery requires at a minimum that a patient is informed as to the basic kind of procedure they are being subjected to, the identity of the professional providing their care and certain motivations that may be involved in that professional's decision making.

Negligence will not be able to provide a full replacement for these claims, given the greater restrictions that legal doctrine imposes on this tort.[1105] Moreover, it must be added that the intervention of the Californian legislature, through the *Medical Injury Compensation Reform Act* (MICRA), substantially limits plaintiffs' prospects for recovery under a professional negligence claim, but generally not under battery.[1106] This currently makes the latter action all the more attractive to plaintiffs.[1107]

As was highlighted in the overview, battery establishes the requirement that a patient's consent is required for a medical procedure. Conversely, 'one who consents to a touching cannot recover in an action for battery'.[1108] In relation to AI, with the largely assistive functions highlighted in Chapter 2, one should however assume that some consent to a wider procedure has been given by the patient. The crucial question thereby becomes not one of consent and non-consent, but rather a question of the validity conditions that the battery mechanism imposes on a given consent. One must ask whether AI's unique characterises, with their defined impact on patient autonomy, are cognisable as challenges to the validity of consent, so as to fulfil this requirement of the battery action.

The way to frame these problems under Californian law is to ask whether AI use *per se*, or certain characteristics thereof, can (when left undisclosed) render a procedure substantially different to the one that was *prima facie* consented to. If so, and the contact requirement is fulfilled, then liability in battery will ensue. This is the first aspect to be addressed here. A second

1105 Moore, 'Intent and Consent in the Tort of Battery: Confusion and Controversy' (2012) 61(6) American University Law Review p. 1585, 1646.

1106 *Perry v. Shaw* (2001) 88 Cal.App.4th 658, 668; *Larson v. UHS of Rancho Springs, Inc.* (2014) 230 Cal.App.4th 336, 348-349; *Saxena v. Goffney* (2008) 159 Cal.App.4th 316, 324-325.

1107 'A problem that sometimes arises is when a plaintiff hoping to evade the restrictions of MICRA, will choose to assert intentional torts': *Unruh-Haxton v. Regents of University of California* (2008) 162 Cal.App.4th 343, 352-353.

1108 *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1266.

question would arise if a patient were to make their consent conditional on non-AI use.

1. Substantially different procedure

Cobbs v. Grant expressly limited the relevance of the battery action in circumstances where the patient has given some form of consent, but alleges that this was invalidated by a lack of information.[1109] California's highest court held that the mechanism should be invoked only in situations where this deficit led to the procedure being 'substantially different' to the one consented to.[1110] A legal obligation to disclose more nuanced types of information would flow, if at all, from professional negligence.[1111] Henceforth battery could be claimed only for particularly grievous interferences with medical decision making. Namely, in situations where the consent procedure does not determine the essential character of the procedure that the patient is subjected to.[1112]

The cases relate various types of shortcomings to this criterion of a 'substantially different' procedure and the AI challenges identified in Chapter 3 must be evaluated in relation to these. In the course of this analysis, some ambiguity must be tolerated as a result of the jury's prominent role in U.S. tort cases. As the jury is responsible for applying the substantially different criterion to the facts, there is often no need for a specific legal determination of the matter.[1113] Furthermore, as was stated in *Kaplan v. Mamelak*,

1109 This can be contrasted with earlier cases, where a wider role was envisioned for battery: *Dow v. Kaiser Foundation* (1970) 12 Cal.App.3d 488.

1110 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 239.

1111 ibid 240-241. This has been clearly affirmed in subsequent cases. For instance, in *Saxena v. Goffney* it was stated: '[o]ur high court has made it clear that battery and lack of informed consent are separate causes of action', and later '[p]erforming a medical procedure without informed consent is not the same as performing a procedure without any consent': *Saxena v. Goffney* (2008) 159 Cal.App.4th 316, 324-327.

1112 *Rains v. Superior Court* (1984) 150 Cal.App.3d 933, 939-941.

1113 For example, in *Kaplan v. Mamelak* it was held: '[i]n the absence of any definitive case law establishing whether operating on the wrong disk within inches of the correct disk is a "substantially different procedure," we conclude the matter is a factual question for a finder of fact to decide and, at least in this instance, not one capable of being decided on demurrer': *Kaplan v. Mamelak* (2008) 162 Cal.App.4th 637, 647. This illustrates the more general claim of Gardner that: 'US legal systems, unlike other common law legal systems, tend to use jury trials for the

the distinction laid down in *Cobbs* does not lend itself to the formulation of some overarching test.[1114] In consequence, in order to understand the application of the test, one must draw analogies to three identifiable classes of cases that have been defined by the courts and which can be analysed to provide guidance.

i. Physical nature of the procedure

The first class of cases are the ones where the physical manifestation and effects of a procedure are deemed substantially different. For instance, in *Perry v. Shaw* the plaintiff consented to a removal of excess skin, but the physician also performed a breast enlargement procedure. Following *Cobbs*, this could only be characterised as an operation to which the plaintiff had not consented.[1115] In *Burchell v. Faculty Physicians & Surgeons of Loma Linda University School of Medicine* surgery on the plaintiff's penis, during an operation where the plaintiff's consent had been given 'to have a small mass removed from his scrotum', was found to be substantially different to the consented to procedure.[1116] More generally, it has been stated: 'Consent to an operation carries with it a consent to remove an organ or body tissue which is a normal incident to the operation (…), but such a consent does not carry with it permission to remove a different organ or one which is not normally excised as an incident to the operation consented to'.[1117]

These framings already exhibit the cases' narrow focus on interventions with a markedly different physical manifestation than the one anticipated.[1118] This holds true in the more contentious, borderline, decisions. So that in *Daley v. Regents of University of California* the heart of the contro-

---

bulk of their tort litigation. It may be that trial judges tend to pass the buck to the jury on the point in question, and appellate judges then stay clear of it, with the result that the legal position is indeterminate': Gardner, *Torts and Other Wrongs* (2019) 1-2.

1114 *Kaplan v. Mamelak* (2008) 162 Cal.App.4th 637, 646-647.

1115 *Perry v. Shaw* (2001) 88 Cal.App.4th 658, 664.

1116 *Burchell v. Faculty Physicians & Surgeons of Loma Linda University School of Medicine* (2020) 54 Cal.App.5th 515, 524-525.

1117 *Rainer v. Buena Community Memorial Hosp.* (1971) 18 Cal.App.3d 240, 256-257.

1118 Sometimes reference is made to the nature of the procedure in general: 'Cobbs implies that the failure to discuss the nature of the treatment sounds in battery': *Nelson v. Gaunt* (1981) 125 Cal.App.3d 623, 634. But this does not appear to be reflected in the development of the case law.

versy turned on the significance of physical differences – 'a percutaneous surgery (with access to the organs established by a needle puncture)', as opposed to 'an open laparotomy and open hysterotomy' – rather than an assessment of the wider significance for the patient of such differences.[1119]

In *Osborn v. Irwin Memorial Blood Bank* the court was faced with the question of whether the nature of a procedure could be changed by a circumstance that was not immediately connected to the physical nature of the procedure. Namely, the plaintiffs consented to an operation after they had been led to believe that they could not give direct blood donations to their child. In fact, they could have directed such donations.[1120] This was a significant circumstance that could (and ought to) have shaped their decision making and impacted their consent. Yet the court curtly dismissed any argument on this point: for their consent to be valid it was enough for the parents to understand the mechanics of the procedure, which involved their child receiving blood transfusions from general supplies.[1121]

Consequently, the potential relevance of this category to medical AI could only emerge from the fact that reliance on the technology may itself constitute a physical change in the procedure. For instance, leading to the use of a different type of device. Moreover, one characteristic of AI use is its impact on an intervention's risk profile, which could also be argued to impact the physical effects of the procedure. However, neither of these arguments is particularly persuasive.

In the first respect, it is highly improbable that AI use (largely comprised of assisting human actors in the performance of their tasks) would be characterised as a separate, substantially different physical dimension. Californian courts have proved extremely generous in subsuming different aspects of clinical care into one overall procedure to which consent has been given. For example, in *Piedra v. Dugan*, the administration of a drug to a child was seen to be a necessary and nonelective component of a procedure that came under a general consent provided by the parents.[1122] In *Kaplan v. Mamelak* it was held to be a question for the jury whether consent to an operation

1119 *Daley v. Regents of University of California* (2019) 39 Cal.App.5th 595, 600.

1120 *Osborn v. Irwin Memorial Blood Bank* (1992) 5 Cal.App.4th 234, 246.

1121 ibid 287. See also the case of *Richmond v. Patel* (Dec. 17, 2021, No. B310903) [non-published opinion]: The nature of the procedure was determined by the physical processes taking place when a biopsy was performed, rather than the subsequent use of the plaintiff's tissue for research.

1122 *Piedra v. Dugan* (2004) 123 Cal.App.4th 1483,1491.

on one spinal disk included a consent to an operation on another.[1123] The comparatively small changes in physical procedures that are anticipated to be brought about by AI hardly exceed these, much more substantial, variations.

Regarding the risks of an intervention, there are some indications that the gravity of the procedure and its actual side effects are relevant to the determination of substantial difference.[1124] However, since *Cobbs* the courts have stopped short of invalidating consent on the basis of a procedure's risks alone (irrespective of their manifestation).[1125] For our purposes this is well illustrated by *Daum v. SpineCare Medical Group, Inc.* Here, the investigatory status of a device that was implanted into the plaintiff could not invalidate the given consent.[1126] More starkly still, in *Stone v. Foster* it was held that under *Cobbs*' categorisation a patient's complete lack of awareness that the procedure involved *any* risks, fell properly to be considered in negligence.[1127]

Although we have discussed how machine learning technologies may alter the risk profile of an intervention, akin in many ways to forms of innovative treatment, it is thus not anticipated that they will alter its physical characteristics of an intervention in a way that is demanded by the substantial difference criterion. The primarily risk-related matters that flow from the employment of this additional device and the technologies' wider

1123 *Kaplan v. Mamelak* (2008) 162 Cal.App.4th 637, 647.

1124 In *Burchell* the court appears to have considered the substantial difference criterion together with the eventuation of serious risks: 'Barker removed the mass from both the scrotum and the penis, a different and substantially more invasive procedure than had been contemplated. Burchell suffered serious side effects, some of which are permanent and irreversible': *Burchell v. Faculty Physicians & Surgeons of Loma Linda University School of Medicine* (2020) 54 Cal.App.5th 515, 518. *Berkey v. Anderson* even drew a direct connection: 'The procedure as outlined by the doctors obviously entailed much more, both as to comfort and risk. Appellant asked Dr. Anderson, "What is a myelogram; is it like the electromyograms that I have been having?" The jury could have found that this called for more than a few mollifying words which grossly understated the seriousness of the procedure': *Berkey v. Anderson* (1969) 1 Cal.App.3d 790, 804.

1125 In *Kerins v. Hartley* the mere increase in risk to the patient, generated by their surgeon's HIV-positive status, was insufficient to ground a battery claim: *Kerins v. Hartley* (1994) 27 Cal.App.4th 1062, 1077-1078. Although it must be noted that there were special considerations at play in this case: the plaintiff had allegedly imposed a condition that the surgeon should be in good health on the procedure and they were claiming for emotional distress: ibid 1066-1067.

1126 *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285, 1313.

1127 *Stone v. Foster* (1980) 106 Cal.App.3d 334, 346-347.

impacts on the non-physical nature of a procedure fall to be assessed under negligence.

ii. Identity of the professional

The next category to be examined is the identity and expertise of the treating doctor, factors external to the medical procedure.[1128] This may be affected by AI's impact on the expertise that is brought to bear on the medical decision, as well as by the influence of its determinations on the decision making of human professionals, which was argued to lead to a partial substitution of human decision making.

California provides relatively little case law on this point. From *obiter* statements in *Newhouse v. Board of Osteopathic Examiners* and *Clarke v. Hoek* it can be deduced that an unanticipated third party's intervention in a procedure, whether experienced or not, is liable to be judged a battery.[1129] Similarly, it was stated *obiter* in *Rains v. Superior Court* that fraud as to identity would be capable of giving rise to a battery if it affected the essential character of the act.[1130] Regarding a more specific characteristic of a physician (in this case their HIV positive status), *Kerins v. Hartley* indicated that a consent may be invalidated. Yet this was seemingly tied to an express condition imposed by the patient regarding their surgeon's health, rather than a matter of substantial difference.[1131]

All in all, one must say that these discussions do not provide purchase for an analogy to be drawn with AI's characteristics. An argument has not been advanced that AI constitutes a third party in itself, nor that it alters the personal identity of a treating physician, but only that it may affect the physician's level of expertise. Under U.S. tort law doctrine, and

1128 Iheukwumere, 'Doctor, Are You Experienced? The Relevance of Disclosure of Physician Experience to a Valid Informed Consent' (2002) 18(2) The Journal of Contemporary Health Law and Policy p. 373, 396.

1129 *Newhouse v. Board of Osteopathic Examiners* (1958) 159 Cal.App.2d 728, 732-733; *Clarke v. Hoek* (1985) 174 Cal.App.3d 208, 218.

1130 *Rains v. Superior Court* (1984) 150 Cal.App.3d 933, 938-940. Albeit the court attributes more weight to the (non-therapeutic) purpose of such actors, than their identity. This will be the focus of the next section.

1131 *Kerins v. Hartley* (1994) 27 Cal.App.4th 1062, 1066-1067.

consistent with *Cobbs*, this discussion fits most naturally within the realm of professional negligence.[1132]

Furthermore, any objectionable reliance on AI expertise is something that will occur in the course of a wider procedure that calls for the professional's judgment. *Conte* intimates that such a balancing exercise by the professional will support the finding that the issue should be left to be regulated under negligence. There the surgeon, in the course of a surgery, chose not to fixate a broken shoulder. This was an aspect of medical judgment that should be adjudged under professional negligence standards, so as to not unnecessarily fetter the exercise of clinical discretion.[1133] This supports our conclusion that battery does not provide a mechanism through which the patient can assert a right to be informed about the professional's relative expertise or a supplementation of it *via* AI.

iii. Non-therapeutic motivations

A medical professional who acts with a non-therapeutic intent may be found liable in battery. This is in spite of the physical mechanics of an action remaining the same. Several Californian cases have addressed this matter.

In *Rains v. Superior Court* the California Court of Appeal conceded that the 'alleged violent touching of which plaintiffs now complain are physically identical to the contact consented to',[1134] but nevertheless found that the 'nontherapeutic purpose of touching by a psychiatrist goes to the "essential character of the act itself" and thus vitiates consent obtained'.[1135] Here, the alleged non-therapeutic purpose was for the defendant psychiatrists to control the patients,[1136] but it was also considered that a desire for personal

1132 Iheukwumere, 'Doctor, Are You Experienced? The Relevance of Disclosure of Physician Experience to a Valid Informed Consent' (2002) 18(2) The Journal of Contemporary Health Law and Policy p. 373. This view is also supported by the framing of problems of 'ghost surgery' as issues for professional regulation: McDonald, *California Medical Malpractice: Law & Practice* (Revised Edition 2022) § 2:10.

1133 *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1269-1270.

1134 *Rains v. Superior Court* (1984) 150 Cal.App.3d 933, 937.

1135 ibid 941.

1136 ibid 938.

gratification or the mere pursuit of a fee could assume this role.[1137] *So v. Shin* is also instructive. The actions of an anaesthetist, in seeking to forestall an embarrassing report rather than seeking to provide patient care, were classified as a battery.[1138]

These instances are to be contrasted with *Freedman v. Superior Court*, where a patient was deceived as to the purpose of a medication administered while she was giving birth. She was made to believe that the drug was necessary to prevent infection, whereas it was in fact used to induce labour.[1139] In spite of this divergence, which can be expected to have impacted the patient's decision making significantly, the Court of Appeal found that, as the defendant's purpose was not alleged to be non-therapeutic, the relevant consent remained valid. Since the mutual end of physician and patient remained the treatment of the latter, the essential character of the treatment remained unchanged.[1140]

This approach again highlights that the factors capable of affecting consent under California's battery action, and thus requiring disclosure, have been very narrowly delimited and without reference to patient autonomy. The principle seeks to foster the patient's decision making and to allow them to weigh their interests for themselves. Allowing physicians to pursue an abstract end of beneficial treatment clearly defeats this goal, endorsing a form of medical paternalism.

More to the point for the present investigation, this broad-brush approach effectively precludes the possibility of battery being used to regulate clinical AI, especially their propensity to pursue, or to serve, purposes that may diverge from the patient's own. The law would almost certainly sanction purposes that are framed primarily in therapeutic terms – although, as will be discussed below in relation to the tort of negligence, there may be related non-therapeutic dimensions. Patient consent would not be invalidated by the pursuit of such purposes. Moreover, even if a non-therapeutic goal could be attributed to an AI or AI developer, it does not seem feasible to impute this purpose to a physician treating the patient with its help. Consequently, battery would not require the disclosure of AI's relatively independent pursuit of objectives.

1137 ibid 940-941.
1138 *So v. Shin* (2013) 212 Cal.App.4th 652, 667.
1139 *Freedman v. Superior Court* (1989) 214 Cal.App.3d 734, 737.
1140 ibid 738-739.

2. Conditional consent

In addition to the above criterion of substantial difference, the Californian battery doctrine further instantiates a relatively strong right on behalf of the patient to impose individual conditions on their consent. If such a condition is imposed and the intentionality requirement is fulfilled, then a treating physician must normally respect it or be subject to liability.[1141]

In theory this would provide patients with the power to determine that an AI should not be used in their care or, if it is, for what purposes. This may be one way for patients who are *ex ante* well informed about AI and its associated difficulties to assert control over the use of the technology.

However, this solution is hardly ideal, given the need for patient awareness and proactive behaviour regarding a modality that largely remains hidden from view. The already examined case of *Conte* further clarifies that any condition would have to be specific. The plaintiff's purported condition on his consent to shoulder surgery, that it be '*with* repair', was deemed too intangible to fulfil this requirement.[1142]

Our preceding analysis has demonstrated that clinical AI use is already incredibly varied and challenging to define. Demanding of a patient that they delimit the bounds of permissible and impermissible applications of the technology in their care appears unrealistic and ineffective. Rather, the autonomy challenges posed by AI must be solved primarily through the clinician's proactive role as a facilitator of their patient's decision-making.

C. Summation

It was seen that the structural doctrinal limitations flowing from the tort of battery were substantially reshaped to favour the principle of autonomy in the medical sphere. However, when one considered the requirements of valid patient consent, it emerged that the courts operated within narrow categories that did not bear an obvious relation to procedural autonomy considerations.

Focusing the analysis on the similarities and differences of the physical nature of the procedure and ignoring wider factors, which may be of sub-

1141 *Grieves v. Superior Court* (1984) 157 Cal.App.3d 159; *Ashcraft v. King* (1991) 228 Cal.App.3d 604.

1142 *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1269.

stantial significance to a patient's decision making, almost precludes this mechanism from performing a useful role in the regulation of AI's novel, more nuanced autonomy challenges. This impression is further supported by the fact that no reliable analogies could be drawn to cases dealing either with the identity of the professional or their extraneous motivations.

In addition, while there is a well-documented possibility for an individual to impose conditions on the use of AI in their care, it seems unlikely that many patients would be in a position to exercise this discretion effectively (i.e. in a legally recognised way). To maintain the patient's positive freedom in interactions with clinical AI, providing them with an opt-out or conditional opt-in is clearly insufficient.

Ironically, it may be precisely because Californian (and other U.S.) courts have been prepared to loosen battery's wider doctrinal restrictions in the medical context, that it became all the more important to impose rule-adjacent limitations through the case law. Limitations that delineate clearly between the deficiencies in consent that allow for a plaintiff to bring a claim in battery and those that allow them to bring a claim in negligence. Furthermore, the practical relevance of this distinction was undoubtedly heightened by California's unique legislative background, which sought to limit professional negligence actions in the healthcare sector in a number of ways.

On the basis of such limiting factors, Californian common law has opted to offer battery's strong protection only to a very narrow subset of medical cases. For our purposes this means that the burden of protecting the patient against AI/ML's violations of procedural autonomy (strong and weak) is placed squarely on the negligence action.

II. Negligence

The tort of negligence imposes obligations on individuals to act with due care in their interactions with others. Due care is usually measured by reference to the ordinarily prudent person, who serves as a standard both for the actions and omissions of individuals.[1143] We will see that, for a successful claim, a plaintiff must show that a defendant owed a relevant duty, fell short of their obligation(s) and caused some legally cognisable injury.

1143 *Fouch v. Werner* (1929) 99 Cal.App. 557, 564.

If these elements are made out, then a plaintiff is entitled to monetary compensation: 'the obligations will primarily be enforced by the traditional judicial remedy of an action for damages for their breach', although it is often additionally envisaged that these will have a 'prophylactic effect'.[1144]

As already outlined, the role of the negligence action in the enforcement of the principle of patient autonomy arises from the California Supreme Court's decision to tether the emerging doctrine of informed consent to it.[1145] Although the tort of battery retains a minimal role in this endeavour, securing the validity of a patient's consent, it is negligence that imposes the most demanding informational duties on healthcare professionals under their overarching duty to take care. More concretely, it requires one to ask whether 'the doctor in obtaining consent may have failed to meet his due care duty to disclose pertinent information'.[1146]

It should also be noted that, although Californian courts have been prepared to find the basis for a physician-patient relationship in an implied contract,[1147] the cause of action considered here remains the tortious one based in ordinary negligence. This is primarily because, while it is possible for patient and physician to enter into a contractual relationship and to specify the nature of their relationship and the obligations arising under it,[1148] the courts 'have shown an historical distaste (…) to allow verbal artfulness to obscure the fact that a tort cause of action truly underpins the plaintiff's grievance'.[1149] Therefore, in the absence of clear evidence of express agreements, the applicable standard is that to be found in general negligence.[1150]

1144 *Moore v. Regents of University of California* (1990) 51 Cal.3d 120, 178-179.

1145 *Cobbs v. Grant* (1972) 8 Cal.3d 229.

1146 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 240-241.

1147 *McNamara v. Emmons* (1939) 36 Cal.App.2d 199, 204.

1148 *Custodio v. Bauer* (1967) 251 Cal.App.2d 303, 315.

1149 McDonald, *California Medical Malpractice: Law & Practice* (Revised Edition 2022) § 1:3. See also *Tunkl v. Regents of University of Cal.*, where the court found that a hospital's contractual waiver could not stand and could not bring about a deviation from negligence's general standard of care: *Tunkl v. Regents of University of Cal.* (1963) 60 Cal.2d 92, 102-104.

1150 'It is thoroughly settled in California that "In the absence of an express contract, the physician or surgeon does not warrant cures. By taking a case he represents that he possesses the ordinary training and skill possessed by physicians and surgeons practicing in the same or similar communities, and that he will employ such training, care, and skill in the treatment of his patients (…)"': *McNamara v. Emmons* (1939) 36 Cal.App.2d 199, 205. 'To recover for breach of warranty or contract in a medical malpractice case, there must be proof of an express contract

Similarly, while the courts have been prepared to categorise the relationship between physician and patient as fiduciary in nature, this has not defined the form of the plaintiff's action for recovery. To be sure, it has helped to establish the kinds of information that must be disclosed and thus influenced the analysis under the action. This will be discussed further below, regarding the Supreme Court's decision on the disclosure of a physician's personal interest in *Moore v. Regents of University of California*.[1151] Nevertheless, the elements of the claim (including a demanding causation requirement)[1152] remain anchored in negligence.[1153]

Ultimately the claim of a patient that they were entitled to some form of information in their medical care – and that this was not provided – is to be brought in negligence, under ordinary negligence principles. Resulting obligations may sometimes be modified or influenced by the contractual and fiduciary nature of the relationship.

The elements that must be proved under these principles in California are summarised in the following formula: 'a legal duty, breach thereof, proximate causation and resulting damage'.[1154] A slight amendment to this classification is necessary, since it subsumes the requirement that there must be some legally cognisable injury under the causation requirement. Fusing these elements would not do justice to the analysis of whether patient autonomy is directly or indirectly protected through an informed consent claim. Consequently, and consistently with the approach undertaken in Chapter 6, this is the first element to be considered.

---

by which the physician clearly promises a particular result and the patient consents to treatment in reliance on that promise': *McKinney v. Nash* (1981) 120 Cal.App.3d 428, 442.

1151 *Moore v. Regents of University of California* (1990) 51 Cal.3d 120.

1152 This was highlighted by the dissenting Justice Mosk in: *Moore v. Regents of University of California* (1990) 51 Cal.3d 120, 179-180. See also: Giuffrida, 'Moore v. Regents of the University of California: Doctor, Tell Me Moore' (1991) 23(1) Pacific Law Journal p. 267, 307-309.

1153 *Jameson v. Desta* (2013) 215 Cal.App.4th 1144, 1164-1165.

1154 *Stafford v. Shultz* (1954) 42 Cal.2d 767, 774. Note that this also applies to informed consent cases and that, consistent with these elements, the burden of proof for lack of informed consent is placed on the plaintiff: 'she bears the burden of proving the elements of every legal theory she proffers in support of that negligence claim —including her informed consent theory': *Flores v. Liu* (2021) 60 Cal.App.5th 278, 297-298.

A. Damage

In the absence of a legally cognisable injury, it will not be possible for a plaintiff to bring a claim for negligence.[1155] This is not always made clear under U.S. and Californian case law and, ordinarily, the eventuation of some form of (uncontroversially accepted) personal injury will be a practically significant factor in many patients' decision to bring a legal action in the first place.[1156] However, in cases that deal with new classes of damage, or in claims where the timing of an injury must be determined to ascertain the corresponding action's limitation period, this factor has been emphasised.

In the context of professional clinical negligence, the necessity of damage is also well documented. For example, in *Hills v. Aronsohn* it was held that some legally compensable injury or 'damages' is an element of the cause of action.[1157] In *Turpin v. Sortini* the California Supreme Court aptly described the plaintiff's suffering of an injury as a 'threshold question' for the relevant negligence claim.[1158] Without passing over this hurdle an action cannot succeed.[1159]

1155 'If the allegedly negligent conduct does not cause damage, it generates no cause of action in tort': *Budd v. Nixen* (1971) 6 Cal.3d 195; 'If the allegedly negligent professional conduct does not cause damage, it generates no cause of action in tort. The mere breach of a professional duty, causing only nominal damages, speculative harm, or the threat of future harm—not yet realized—does not suffice to create a cause of action for negligence': *Van Dyke v. Dunker & Aced* (1996) 46 Cal.App.4th 446, 452.

1156 Weisbard, 'Informed Consent: The Law's Uneasy Compromise with Ethical Theory' (1986) 65(4) Nebraska Law Review p. 749, 753-754. *Looney v. Moore* is further instructive. This federal case applied Alabaman law, but also made reference to the negligence-based approach expounded in *Cobbs v. Grant*: 'Alabama common law requires an actual injury to maintain a negligence cause of action—and in the specific context of informed consent claims, so do the majority of other courts. Finally, although it is true that in cases discussing informed consent claims, the Alabama Supreme Court did not list actual injury as a required element for those claims, there was no dispute in those cases that an actual injury existed, and the court was focusing on what constitutes informed consent. Thus, we do not consider the omission of "injury" from the list as dispositive': *Looney v. Moore* (11th Cir. 2018) 886 F.3d 1058, 1069.

1157 *Hills v. Aronsohn* (1984) 152 Cal.App.3d 753, 762, fn. 7.

1158 *Turpin v. Sortini* (1982) 31 Cal.3d 220, 235-236.

1159 '[T]he obstacle to recovery in a wrongful life case is that it is impossible to determine that the plaintiff has been harmed': Kearl, 'Turpin v. Sortini: Recognizing the Unsupportable Cause of Action for Wrongful Life' (1983) 71(4) California Law Review p. 1278, 1286-1287.

The damage that is capable of giving rise to a self-standing negligence action must then be of a certain, legally determined, type.[1160] As has been intimated, personal injury is one kind of damage that triggers the threshold for liability and provides a readily available basis for most clinical negligence claims. In *Lantis v. Condon* it was further recognised that a spouse's loss of consortium constitutes a separate injury in response to a separate right.[1161] By contrast, a type of injury that has been expressly rejected in the medical malpractice suits is the illegitimacy of a child.[1162] This begs the question: 'which types of injury provide a sufficient basis for an informed consent claim?' This will be central to determining the extent to which patient autonomy can be protected against medical AI's challenges through this action. Three candidates appear most relevant in this respect: personal injury, a setback to the autonomy interest itself and emotional distress.

1. Personal injury

In *Townsend v. Turk* the Court of Appeal cited with approval the statement that: '[s]tandard negligence analysis protects an interest in physical well-being. The doctrine of informed consent injects into the established framework of negligence a concern with patient choice that would otherwise be absent'.[1163] This makes clear that the orthodox approach, both in California and the United States, is to understand informed consent as protecting the patient's right to be free from personal injury.

The case of *Warren v. Schecter* illustrates the nuances of this understanding. As is common in informed consent actions, the claim here concerned a procedure for which a surgeon had disclosed a variety of dangers but had failed to disclose one particular risk (metabolic bone disease) from

1160 I use 'damage' to refer to the type of setback of interests that gives rise to a negligence claim; distinct from 'damages' which refers to types of setback that are compensated once an action has been established. The two are sometimes treated synonymously, as in *Hills v. Aronsohn* (1984) 152 Cal.App.3d 753, 762, fn. 7. Nevertheless, the distinction is recognised as in England: Donovan, 'Is the Injury Requirement Obsolete in a Claim for Fear of Future Consequences' (1993) 41(5) UCLA Law Review p. 1337, 1342-1343.

1161 *Lantis v. Condon* (1979) 95 Cal.App.3d 152.

1162 *Alexandria S. v. Pacific Fertility Medical Center, Inc.* (1997) 55 Cal.App.4th 110.

1163 *Townsend v. Turk* (1990) 218 Cal.App.3d 278, 284, citing Schultz, 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal p. 219, 232.

which the plaintiff then, subsequently, suffered. It was held that the action only accrued when this risk materialised (when the plaintiff broke her back years later).[1164] *Warren* provides a concrete illustration of the claim that the damage which provides the basis for an informed consent negligence action is ordinarily the manifestation of a physical risk, rather than the violation of a more intangible interest in autonomy. For, in the latter case, the injury would have occurred immediately.

Where claims concern information pertaining to alternative treatments or non-treatment, this link with personal injury is more indirect and this element will consequently be more difficult to establish.[1165] A plaintiff will hardly succeed in these instances, unless it is also the case that their physical well-being was impaired by the selection of an alternative or by the pursuit of treatment over non-treatment.

This illustrates why the outlined challenges associated with ML technologies are not automatically subsumed under this classification. On occasion it will be possible to point to a specific risk that has been altered by their use. For example, the harm that results from a failure to discover a progressing disease from a false negative diagnosis. But the more abstract alteration in the understanding of risks, the alteration of expertise, or the pursuit of differing treatment goals will not necessarily be linked to an identifiable setback of an individual's interest in physical well-being.

2. Autonomy interest

If the protection of the patient from bodily harm is necessarily of limited assistance to a patient who feels that their decision-making process is affected by non-disclosure of AI features, then it is natural to enquire whether it may not be open to them to bring a negligence action directly on the grounds that their autonomy interest has been violated. Although judgments are sparse on this issue, several commentators in the U.S. have advocated for the adoption of this position.[1166]

1164 *Warren v. Schecter* (1997) 57 Cal.App.4th 1189, 1204.

1165 Meisel, 'A Dignitary Tort as a Bridge between the Idea of Informed Consent and the Law of Informed Consent' (1988) 16(3-4) Law, Medicine and Health Care p. 210, 214-215.

1166 Meisel, favours basing informed consent on a somewhat more elusive dignitarian interest, and usefully outlines the lack of precedent, positive or negative, on the issue: 'few courts have rejected such protection outright, but that is probably

Some indirect, tentative support can be gleaned from various aspects of Californian medical malpractice case law. Early *dicta* for example emphasised the need for the significance of the patient's familiarity with treatment alternatives, a need that appeared to be only indirectly connected to the eventuation of harm.[1167] Moreover, the seminal decision in *Truman v. Thomas* held that a physician could be liable for a failure to warn a patient of the consequences of opting to not undertake a diagnostic test.[1168] This arguably opened the door for a more general recognition of the patient's protected interest in informed decision making, beyond a warning of the risks to bodily integrity that would accompany a patient's consent.[1169]

There is also some evidence that causes of action for medical malpractice have received separate recognition in Californian courts due to their impact on personal autonomy. For instance, in *Zambrano v. Dorough* the Court of Appeal held that an injury that resulted in the patient's loss of reproductive capacity was of a different type than the injuries she suffered from the misdiagnosis of a tubal pregnancy (including the rupture of a fallopian tube and unnecessary surgery).[1170] The basis for this distinction was the significance of the former injury for the patient's reproductive autonomy, rather than the gravity for the physical well-being of the patient, which had already been substantially impacted.

In spite of these indicators and arguments, a standalone negligence claim for the violation of a plaintiff's autonomy has not been recognised in the courts. Indeed, the impact of *Truman* has proven relatively narrow: mandating disclosure regarding the refusal of a procedure only once it has been

---

because so few have been called on to recognize such a right': ibid 211. See also: Weisbard, 'Informed Consent: The Law's Uneasy Compromise with Ethical Theory' (1986) 65(4) Nebraska Law Review p. 749, 764 (*inter alia* suggesting 'viewing patient self-determination as a goal independent of the avoidance of physical injury'); Twerski and Cohen, 'Informed Decision Making and the Law of Torts: The Myth of Justiciable Causation' [1988](3) University of Illinois Law Review p. 607, 609 ('courts should identify and value the decision rights of the plaintiff which the defendant destroyed by withholding adequate information').

1167 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243.

1168 *Truman v. Thomas* (1980) 27 Cal.3d 285.

1169 Meisel, 'A Dignitary Tort as a Bridge between the Idea of Informed Consent and the Law of Informed Consent' (1988) 16(3-4) Law, Medicine and Health Care p. 210, 215-216.

1170 *Zambrano v. Dorough* (1986) 179 Cal.App.3d 169, 172-174.

recommended by a physician.[1171] The common law continues to require the eventuation of a specific physical harm to trigger informed consent claims. *Zambrano's* holding must similarly be classed as of little consequence for the present discussion. Another Court of Appeal decision has declined to follow it altogether, even regarding the narrow purpose of ascertaining whether a plaintiff's claim is part of the same cause of action or constitutes a separate one.[1172] Therefore, there is little prospect that *Zambrano* will alter the wider view on the threshold issue of damage.

It may be thought that another possibility is to construct a form of autonomy damage by reference to the constitutional right to privacy. One commentator has argued that, in the context of Californian wrongful birth claims, the courts assume that the injury consists in a violation of the plaintiff parents' privacy interest.[1173] However, it appears that such a move is of limited assistance. Not least, because the above claim is followed almost immediately by the undoubtedly correct insight that, *vis-à-vis* private actors like doctors, 'tort liability does not arise from the mere existence of constitutional rights'.[1174]

Much more problematic for this line of argument though, is the fact that the class of wrongful birth case law that is alluded to, simply does not bear out the claim. In the leading Californian case on wrongful birth, *Custodio v. Bauer,* the legally cognisable injuries are stated in terms of alternative and comparatively orthodox categories: emotional suffering (of either parent), physical injury and consequential economic expenses.[1175] If a constitutional right has imbued the common law with a novel autonomy interest, then it appears that this has yet to be discovered.

1171 *Scalere v. Stenson* (1989) 211 Cal.App.3d 1446, 1450; *Munro v. Regents of University of California* (1989) 215 Cal.App.3d 977, 986-988.

1172 *DeRose v. Carswell* (1987) 196 Cal.App.3d 1011, 1023-1026.It was also held in *Massey v. Mercy Medical Center Redding* that a morphine injection without informed consent was not an independent cause of action from the fall that it was allegedly administered to cover up: *Massey v. Mercy Medical Center Redding* (2009) 180 Cal.App.4th 690, 699.

1173 Goebelsmann, 'Putting Ethics and Traditional Legal Principles Back into California Tort Law: Barring Wrongful-Birth Liability in Preimplantation Genetic Testing Cases' (2010) 43(2) Loyola of Los Angeles Law Review p. 667, 686.

1174 ibid 687.

1175 *Custodio v. Bauer* (1967) 251 Cal.App.2d 303, 322-324. Note that this Californian case avoids the issues that were seen to arise around wrongful birth claims in England. Since there is not attempt to limit recovery under some heads of damage, there is no need to create a separate injury to allow for fair compensation.

The next subsection will illustrate that the categories of injury that negligence can remedy are malleable, not fixed, under American tort law. Nevertheless, without direct, supportive authority on this point – and in light of the decidedly negative trend in recognising autonomy violations as distinct actions – an argument from principle cannot find sufficient purchase here. The doctrinal limitations currently appear too strong to grant it independent protection.

3. Emotional distress

Even if there is no recovery for a violation of the patient's interest in informed decision making *per se*, Californian law has recognised that a negligence action may be founded on another type of intangible injury. Namely, the emotional distress of the patient.

The availability of this action is, by itself, not uncontroversial. Historically, American courts and commentators were content to follow the inherited English position that '[t]he mere temporary emotion of fright not resulting in physical injury is, in contemplation of law, no injury at all, and hence no foundation of an action'.[1176] This original position is represented in California by *Sloane v. Southern California Railway*. It was stated that 'mental suffering alone will not support an action, yet it constitutes an aggravation of damages when it naturally ensues from the act complained of'.[1177]

Against this background, California became the first U.S. jurisdiction to challenge the *status quo* and to recognise that a physician may be liable for the negligent infliction of emotional distress to a direct victim.[1178] In *Molien v. Kaiser Foundation Hosp.* a physician was held to be liable to the plaintiff for negligently diagnosing his wife with syphilis and advising her

1176 Throckmorton, 'Damages for Fright' (1923) 57(6) American Law Review p. 828, 835-836.

1177 *Sloane v. Southern Cal. Ry. Co.* (1896) 111 Cal. 668, 679-680. See also *Espinosa v. Beverly Hospital* for a case of medical malpractice, where the trial court was found to have properly instructed the jury: '[u]nless you find that plaintiffs suffered actual physical injury as the proximate result of defendants' negligence, they cannot recover in this case because fright or mental suffering alone will not sustain a recovery for plaintiffs'*: Espinosa v. Beverly Hospital* (1952) 114 Cal.App.2d 232, 234.

1178 Meisel, 'A Dignitary Tort as a Bridge between the Idea of Informed Consent and the Law of Informed Consent' (1988) 16(3-4) Law, Medicine and Health Care p. 210, 212-213.

to communicate this to him.[1179] The plaintiff did not suffer a physical injury as a result, but merely an extreme emotional reaction. Without more, this was found to be a compensable injury for the purposes of negligence.[1180]

Viewed in the abstract, it is imaginable that such a claim could form one means of expanding the application of informed consent requirements to medical AI. The crux of Part I's theoretical argument maintains that the patient's process of decision making is impacted by an uninformed use of ML technologies. It is conceivable that some of these uses will provoke an emotional reaction in the patient. Covert manipulation of choices, or the subjection of a medical decision to a bias regarding sensitive or protected characteristics, may serve as good examples.

However, *Molien* and subsequent decisions have imposed demanding limitations on recovery for emotional distress which make an action for this injury ill-suited to informed consent claims in general,[1181] and claims concerning medical AI in particular. The precise nature of the test(s) to be applied remains confused.[1182] Depending on the circumstances, the Californian courts have variously required: a reasonable foreseeability of emotional harm,[1183] a guarantee of genuineness of that harm,[1184] or proof that a specific injury is more likely than not to occur.[1185]

In spite of the confusion, it appears that any of these restrictions would exclude most forms of AI use. The outlined aspects of the technology may no doubt appear unnerving to patients, and more so to some than to others. Yet, in all but the most serious cases, it is not reasonably foreseeable that emotional distress will result from the technology's application.[1186] Neither

1179 *Molien v. Kaiser Foundation Hospitals* (1980) 27 Cal.3d 916.

1180 ibid 930-931.

1181 Meisel, 'A Dignitary Tort as a Bridge between the Idea of Informed Consent and the Law of Informed Consent' (1988) 16(3-4) Law, Medicine and Health Care p. 210, 212-213.

1182 Heidenreich, 'Clarifying California's Approach to Claims of Negligent Infliction of Emotional Distress' (1995) 30(1) University of San Francisco Law Review p. 277, 298-302.

1183 *Molien v. Kaiser Foundation Hospitals* (1980) 27 Cal.3d 916, 923.

1184 *Burgess v. Superior Court* (1992) 2 Cal.4th 1064, 1079.

1185 *Potter v. Firestone Tire & Rubber Co.* (1993) 6 Cal.4th 965, 997. This was applied in *Kerins.v. Hartley* to a plaintiff's fear of catching HIV from her surgeon: *Kerins v. Hartley* (1994) 27 Cal.App.4th 1062, 1073-1074.

1186 For a critical analysis of the difficulties in applying the reasonableness requirement, especially regarding plaintiff's with an existing predisposition see: Chillag, 'Negligent Infliction of Emotional Distress as an Independent Cause of Action in

can a 'genuine' emotional grievance be anticipated, nor a link to a future eventuation of a disease that is more probable than not.

4. Summation

The Californian negligence action, in line with wider trends of U.S. law, is realistically expected to require the eventuation of physical injury to the patient before they can bring a successful informed consent claim. This is not fatal to our envisaged claim for lack of informed consent because, in the high-stakes medical arena, there will be circumstances where a patient suffers physical injury as a result of another's use of AI. Depending, of course, on the fulfilment of negligence's other elements, a claim could then be made out.

Still, this limitation means that even serious intrusions into a patient's decision-making procedures will only be actionable upon the incidental occurrence of physical harm. This is far removed from a coherent protection of the patient from AI's autonomy challenges.

B. Duty of care

To succeed in an informed consent claim, a patient will further have to establish that a relevant party owed them a duty to disclose such information. To fulfil this element it would be sufficient in most circumstances to point to the general duty of care that the Californian legislature has created.[1187] However, this duty applies only to situations where a defendant actively creates a risk for another; so that, for example, not every individual has a duty to take affirmative action to assist another.[1188] It will be seen that the duty of informed consent has been similarly limited to certain individuals. In the following, I consider two classes of persons who may be argued to owe a duty to obtain the informed consent of a patient: medical professionals and healthcare institutions.

California: Do Defendants Face Unlimited Liability' (1982) 22(1) Santa Clara Law Review p. 181, 195-198.

1187 'Everyone is responsible, not only for the result of his or her willful acts, but also for an injury occasioned to another by his or her want of ordinary care or skill in the management of his or her property or person': California Civil Code section 1714, subdivision (a).

1188 *Brown v. USA Taekwondo* (2021) 11 Cal.5th 204, 213-214.

The potential duty of AI developers and manufacturers to convey information to the patient will not be explored further. In line with the law of all other states,[1189] California has adopted the learned intermediary doctrine.[1190] Where a doctor utilises a medical device in a medical environment and under medical supervision, this doctrine applies and holds unquestionably that developers' informational obligations are directed towards the doctor.[1191] As has been stated at length, such a professional utilisation is the normal situation envisaged for AI use. Here, similarly to many other medical products 'the physician stands in the shoes of the product's ordinary user'.[1192]

Moreover, even in the small subset of cases where the patient is placed in a position to make use of AI directly, in a non-clinical environment, any envisaged duty (arising either from negligence or strict product liability) is a very narrow one: to 'warn of a particular risk'.[1193] This duty does not provide a suitable vehicle for requiring information on AI's unique features and therefore need not be explored further.[1194]

1. Medical professionals

It is not disputed that a physician has a duty to obtain their patient's informed consent. Already in *Salgo v. Leland Stanford Jr. University Bd.*

1189 'Every state in the country, along with the District of Columbia and Puerto Rico, has adopted the learned intermediary doctrine in some iteration': Dearinger v. Eli Lilly and Company (Wash. 2022) 199 Wash.2d 569, 574-575.

1190 This was recently reaffirmed in: Amiodarone Cases (2022) 84 Cal.App.5th 1091.

1191 *Bigler-Engler v. Breg, Inc.* (2017) 7 Cal.App.5th 276, 319.

1192 *Gall v. Smith & Nephew, Inc.* (2021) 71 Cal.App.5th 117, 122.

1193 *Bigler-Engler v. Breg, Inc.* (2017) 7 Cal.App.5th 276, 312-317.

1194 As will be explored in the breach element below, a professional can only be held liable to disclose information that they know or (under ordinary negligence standards) ought to know. Chapter 2 considered that a reasonable use of AI is predicated on a degree of knowledge regarding the characteristics of AI that were taken to render it challenging for autonomy: its general relationship to risk characteristics, its alteration of expertise, the ability to pursue goals relatively independently and a user's propensity to be influenced by it. This knowledge does not stem from a manufacturer's duty to warn. Simultaneously, developers will realistically provide descriptions of their device. Liability will more appropriately be imposed in light of this, where they actively mislead a physician about the nature of these characteristics. As this presumes positive malfeasance by the actors, however, it will not be considered as a general problem here.

of Trustees the court offered its seminal judgment on the duty to disclose in these terms: 'A physician violates his duty to his patient and subjects himself to liability if he withholds any facts which are necessary to form the basis of an intelligent consent by the patient to the proposed treatment'.[1195] Similarly, when *Cobbs* solidified the doctrine's association with negligence, it framed the duty 'as an integral part of the physician's overall obligation to the patient'.[1196] Many subsequent judgments have placed a similar emphasis on the person of the physician.[1197]

Although there may be specific circumstances where the finding of such a relationship is controversial, AI use does not appear to add to them. In *Hale v. Superior Court* it was made clear that the patient-physician relationship can be initiated by a doctor's examination, diagnosis or furnishing of treatment.[1198] Where an ML device is used as an assistive technology, the determination of a relationship can still proceed according to such, relatively straightforward, involvement.

A further factor that must be considered here is that, in spite of relatively broad statements regarding physicians, several cases have found that not all of those who owe the patient a general duty of care, also owe a duty to obtain their informed consent. Specifically, in *Mahannah v. Hirsch* and *Townsend v. Turk* it was held that a pathologist and a radiologist respectively did not owe a duty to obtain the patient's informed consent.[1199] Information does not need to be conveyed at every treatment level and it is 'the therapist',[1200] the one actively managing the plaintiff's care,[1201] or the professional in direct contact with the patient that owes this obligation.[1202]

This will restrict a patient's entitlement to information about AI use. Where the technology is employed by a specialist, to which the patient

1195 *Salgo v. Leland Stanford Jr. University Bd. of Trustees* (1957) 154 Cal.App.2d 560, 578.

1196 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243.

1197 For example: *Conte v. Girard Orthopaedic Surgeons Medical Group, Inc.* (2003) 107 Cal.App.4th 1260, 1266-1267.

1198 *Hale v. Superior Court* (1994) 28 Cal.App.4th 1421, 1423-1424.

1199 *Mahannah v. Hirsch* (1987) 191 Cal.App.3d 1520; *Townsend v. Turk* (1990) 218 Cal.App.3d 278.

1200 *Jamison v. Lindsay* (1980) 108 Cal.App.3d 223, 230.

1201 *Townsend v. Turk* (1990) 218 Cal.App.3d 278, 286.

1202 *Mahannah v. Hirsch* (1987) 191 Cal.App.3d 1520, 1527; *Wilson v. Merritt* (2006) 142 Cal.App.4th 1125, 1135. See also *Quintanilla v. Dunkelman*, which highlighted the importance that 'the role of Dr. Dunkelman was more than merely that of a referring physician': *Quintanilla v. Dunkelman* (2005) 133 Cal.App.4th 95, 118-119.

or their data has been referred, there is no readily cognisable duty of disclosure to the patient. Even the obligation to disclose this information to the doctor will only arise if it has some bearing on their fulfilment of the ordinary, beneficence-oriented, standard of care.[1203]

Furthermore, one must investigate the situation of clinical personnel, other than physicians, who can assume a primary caregiver role, such as nurses. That these owe a duty with a professional standard is well established.[1204] The case law and statute additionally suggest that this encompasses disclosure obligations in appropriate circumstances.[1205] With respect to the former, one can point to *Massey v. Mercy Medical Center Redding* in which it appears to have been assumed that a nurse's decision to administer a morphine injection could have constituted a breach of her informed consent obligation (even though the relevant action was in fact barred by the statute of limitations).[1206] Regarding the latter, it is worth noting that the legislature has, through various forms of legislation, recognised specific situations in which a nurse will be responsible for obtaining the patient's informed consent. For example, a recent amendment to the California Business and Professions Code, namely section 2746.54, lays down a number of factors with regard to which a certified nurse-midwife must obtain informed consent.[1207]

A good argument can therefore be made that a duty encompassing informed consent requirements has been, or can consistently be, imposed on non-physician direct caregivers. This is a welcome addition in those

1203 *Townsend v. Turk* (1990) 218 Cal.App.3d 278, 287.

1204 *Fraijo v. Hartland Hospital* (1979) 99 Cal.App.3d 331, 341-342.

1205 For an example of the situation where informed consent obligations were not properly imposed on caregivers other than the physician, see: *Ermoian v. Desert Hospital* (2007) 152 Cal.App.4th 475, 510-513.

1206 *Massey v. Mercy Medical Center Redding* (2009) 180 Cal.App.4th 690, 698-699. See also *Anderson v. PIH Health Hospital–Whittier*, which implies that a duty to obtain informed consent could be imposed on nurses and nonphysician personnel where they were charged with the task of obtaining the informed consent of the patient: *Anderson v. PIH Health Hospital-Whittier* (Cal. Ct. App., Apr. 8, 2022, No. B308407) [nonpublished opinion].

1207 Under California Business and Professions Code section 2746.54, subdivision (a)(1) and (5), informed consent must be obtained regarding the fact that 'the certified nurse-midwife is not supervised by a physician and surgeon' and that '[t]here are conditions that are outside of the scope of practice of a certified nurse-midwife that will result in a referral for a consultation from, or transfer of care to, a physician and surgeon'.

situations involving AI where the status or expertise of primary caregivers has been lessened or altered.

2. Healthcare institutions

There are different mechanisms for holding an institutional healthcare provider liable for negligence under United States and Californian tort law. This encompasses both a theory of vicarious liability, for the negligent actions of parties suitably associated with an institution, and corporate liability where the corporation owes a direct duty of care to the patient.

Vicariously liability attributes liability to an institution for another's breach of their duty. As such, it does not generate a new obligation, but merely identifies an additional party that a plaintiff can hold responsible for their injury in medical malpractice.[1208] There is ample precedent that a hospital can be liable for the negligence of their agent or their ostensible agent in this manner.[1209] Given that an informed consent action is merely a species of medical malpractice, a healthcare institution may also be vicariously liable for shortcomings related to informed consent.[1210] Yet, as this duty does not provide additional protection, but only reinforces the existent one, it will not be examined further here.

The other basis for liability is provided by the concept of corporate institutional liability. This refers to an institution's 'violation of a duty—as a corporation—owed directly to the patient which resulted in injury'.[1211] Various duties have been imposed on healthcare institutions (specifically hospitals) under this head, whereby a plaintiff need not refer to the specific actions of any agent to establish their claim.[1212] A hospital has, for instance, been held to owe a duty to exercise reasonable care in 'screening the competency of its medical staff to insure the adequacy of medical care rendered to patients at its facility'.[1213] However, the nature of this duty is clearly restricted and, also in light of the previous section's finding that informed

1208 *Ermoian v. Desert Hospital* (2007) 152 Cal.App.4th 475, 501-502.
1209 *Elam v. College Park Hospital* (1982) 132 Cal.App.3d 332, 337; *Ermoian v. Desert Hospital* (2007) 152 Cal.App.4th 475, 502-503.
1210 See the extensive analysis in *Ermoian* on this basis (even though the Court of Appeal ultimately upheld a finding that there had been no negligence): *Ermoian v. Desert Hospital* (2007) 152 Cal.App.4th 475, 514-516.
1211 *Elam v. College Park Hospital* (1982) 132 Cal.App.3d 332, 338, fn. 5.
1212 *Murillo v. Good Samaritan Hospital* (1979) 99 Cal.App.3d 50, 55.
1213 *Elam v. College Park Hospital* (1982) 132 Cal.App.3d 332, 338-347.

consent duties are applicable only to a subset of medical professionals, one must ask whether a hospital can owe such a duty to the patient directly.

Referring back to the analysis in Chapter 2, it must be recalled that such a direct form of liability is expected to assume added relevance to our analysis of AI. Applications of the technology will allow for circumstances where it is only an institution (rather than any individual agent) who can provide the user with information on the technology's use and nature. Specifically, instances were discussed where individuals engage with AI that triage, and partially determine, a clinical process before they enter a specific professional-patient relationship.

Having legally mandated informational duties in these, relatively unorthodox, scenarios is arguably necessary to address AI's unique autonomy challenges (especially its relative independence) and constitutes an important guarantor of patient self-determination. Unfortunately, Californian courts have declined to impose such a duty and have thereby situated themselves well within a wider trend running through the American case law on this matter.

The first step in this analysis, is to reiterate the distinction drawn above between mere duties to warn of products and informational obligations that can secure the patient's informed consent. Although a healthcare provider may owe the former duties,[1214] they are narrowly delineated and are of little assistance in addressing the problems associated with AI. What is of concern to us, is whether a wider, more malleable informational obligation can be imposed on these institutions.

The leading precedent in California on this aspect is *Walker v. Sonora Regional Medical Center*. This case reiterated that a hospital was under a corporate duty to protect the patient from harm, but it declined to hold that this also encompassed an obligation to provide the patient with information (laboratory test results) relevant to her clinical decision.[1215] Admittedly, this case was subject to special circumstances, given that federal regulation and state legislation expressly limited the persons to whom this information could be disclosed.[1216] Nevertheless, in reaching its conclusion, the *Walker* court referred to policy-based argumentation that is in

1214 *Bigler-Engler v. Breg, Inc.* (2017) 7 Cal.App.5th 276, 317-318.

1215 *Walker v. Sonora Regional Medical Center* (2012) 202 Cal.App.4th 948, 959-963. See also: *California Jurisprudence* (Third Edition 2022) Healing Arts and Institutions § 416.

1216 ibid 960-962.

line with the approach of other U.S. states, which have similarly rejected informed consent claims against hospitals and healthcare institutions.[1217] In particular, the court referred to the primary importance of the physician-patient relationship, where disclosure of the relevant information could be presumed, and to the undesirability of the hospital interposing itself into this relationship.[1218]

Furthermore, it stated 'that it is physicians or other licensed medical practitioners, not hospitals as corporate entities, who actually practice medicine' and render advice to the patient on the basis of their individual circumstances.[1219] In other words, the court hinted that one of the bases for imposing informed consent obligations – a relationship of personal dependency with an imbalance of knowledge – was absent.[1220]

Lastly, it appears that the court further characterised the hospital's role as a supplementary, advisory one. It was analogous to the role of a consultant, as in the case of *Mahannah* that was discussed above. [1221]As Gatter has outlined, this rejection of hospital liability for deficiencies in informed consent is almost universal and the kinds of justifications advanced in *Walker* could, by now, be said to represent a traditional policy within U.S. common law.[1222]

At the same time, it may be said that the ruling provided scope for development. The Court of Appeal was keen to limit its findings to the specific circumstances before it.[1223] Moreover, it paid lip service to the

1217 For an overview of these arguments and their widespread acceptance, see: Gatter, 'The Mysterious Survival of the Policy against Informed Consent Liability for Hospitals' (2006) 81(4) Notre Dame Law Review p. 1203.

1218 *Walker v. Sonora Regional Medical Center* (2012) 202 Cal.App.4th 948, 961-963.

1219 ibid 965, fn. 19.

1220 This basis will be discussed further below in relation to the informed consent standard.

1221 *Walker v. Sonora Regional Medical Center* (2012) 202 Cal.App.4th 948, 961, fn. 13.

1222 Note that the exceptions to this *almost* universal approach relate to vicarious liability (which has been addressed above) and a hospital's housing of clinical research, which is inapplicable to the therapeutic use of AI: Gatter, 'The Mysterious Survival of the Policy against Informed Consent Liability for Hospitals' (2006) 81(4) Notre Dame Law Review p. 1203, 1218-1223.

1223 'Whatever may be the scope of a hospital's duty under other circumstances, we must focus our inquiry on the particular facts before us': *Walker v. Sonora Regional Medical Center* (2012) 202 Cal.App.4th 948, 963.

hospital's role in facilitating a patient's autonomy.[1224] If the relevance of the autonomy principle were embraced, then it is arguable that policy factors would favour the imposition of a duty on healthcare institutions where they directly rely on AI. In this eventuality: the hospital's role would become that of primary caregiver and thus cease to be supplementary; the relevant expertise would not be supplied by a physician, but by the selected device; and, accordingly, there would be no danger of interfering with the professional-patient relationship. The *Walker* court's emphasis on policy considerations for the purpose of identifying duties of care ought to facilitate such innovation.[1225]

Consequently, a development of the law in this direction could certainly find support. However, there is no clear evidence that the courts, in California or beyond, have thus far drawn a strong association between the corporate liability of hospitals and patient autonomy. More significantly, the outlined policy justifications for refusing to impose informed consent duties on hospitals have been found wanting for some time – even before one considers the introduction of AI use.[1226] Therefore, judging by the present state of Californian law, and the firm entrenchment of the rule against allowing informed consent claims against hospitals within almost all states, it cannot be assumed that a relevant duty of care would be imposed on healthcare institutions.

3. Summation

Concluding this element of the negligence action, it is possible to focus the analysis going forward on the duties of a restricted group of medical professionals that constitute a patient's primary caregivers or points of contact. Although this group has been argued to be relatively broad, it still leaves one considerable restriction that is unconnected to autonomy. Namely, professionals who utilise AI in a secondary capacity will have no direct obligations to disclose this use. As a consequence, there will be

1224 '[T]he Hospital arguably took the single most effective measure toward achieving the desired result of having Amber receive information and counseling regarding the laboratory test': ibid 966-967.

1225 'The determination that a legal duty is owed in a particular set of circumstances is "only an expression of the sum total of those considerations of policy which lead the law to say that the particular plaintiff is entitled to protection."': ibid 958.

1226 Gatter, 'The Mysterious Survival of the Policy against Informed Consent Liability for Hospitals' (2006) 81(4) Notre Dame Law Review p. 1203, 1232.

situations where a qualitatively similar, or even identical, violation of an individual's autonomy occurs, constituting a breach of duty in the sense outlined below, but the patient will only be informed of one of them.

Institutions may be vicariously liable when the outlined class of individuals constitute their agents or ostensible agents, but they are not envisaged to owe informational duties that are relevant to AI's autonomy challenges at the corporate level. This generates one further potential lacuna in the common law's response to AI's autonomy challenges.

C. Breach

Under American tort law, if a duty of care arises, then a defendant must be found to have breached that duty to be liable.[1227] For the purposes of California, the California Civil Code section 1714, subdivision (a) refers to an individual evincing a 'want of ordinary care or skill in the management of his or her property or person' and case law has interpreted this as a continuation of the common law 'standard of an ordinarily prudent man under normal circumstances'.[1228]

Specifically in the medical malpractice context, in *Huffman v. Lindquist*, the California Supreme Court cited previous case law for the proposition that:

> The "law has never held a physician or surgeon liable for every untoward result which may occur in medical practice" (…) but it "demands only that a physician or surgeon have the degree of learning and skill ordinarily possessed by practitioners of the medical profession in the same locality and that he exercise ordinary care in applying such learning and skill to the treatment of his patient"[1229]

This reference to the learning and skill ordinarily possessed by members of the profession serves to specify the standard of ordinary prudence for the circumstances of clinical practitioners, circumstances which include their specialised education and training.[1230] In other words, the standard of

1227 Dobbs, Hayden and Bublick, *Dobbs' Law of Torts: Practitioner Treatise Series* (Second Edition 2022) § 124.

1228 *Fouch v. Werner* (1929) 99 Cal.App. 557, 564-565.

1229 *Huffman v. Lindquist* (1951) 37 Cal.2d 465, 473.

1230 *Flowers v. Torrance Memorial Hospital Medical Center* (1994) 8 Cal.4th 992, 997-998.

care remains constant, but it is applied in a form that is relevant to their situation.[1231]

For the purposes of establishing a breach of this specified standard, it is understandable why it will usually be necessary to rely on expert testimony.[1232] An exception is normally only appropriate where the alleged breach can be adjudged by reference to the layman's common experience.[1233]

It should further be noted that the judgment cited above states an innovation of American common law, the so-called 'locality rule'. Hereby, the relevant assessment must be restricted to the practice within a particular region.[1234] However, the implications of this rule for the standard of care have been contested in California.[1235] It now appears that the locality of the practitioner does not definitively delimit the practice of the professional community, but constitutes only one relevant factor to be considered in the overall assessment of the standard.[1236]

With this general framework in place, one can turn to the standard that is applicable under the professional's duty to disclose information, which is a

1231 See the further elaboration of the standard in *Meier v. Ross General Hospital*. Here the court approved the rule that 'when a physician chooses one of alternative accepted methods of treatment, with which other physicians disagree, and which is in fact unsuccessful, the jury may not automatically deem him negligent': *Meier v. Ross General Hospital* (1968) 69 Cal.2d 420, 434. See also: *Clemens v. Regents of University of California* (1970) 8 Cal.App.3d 1, 12-13.

1232 *Huffman v. Lindquist* (1951) 37 Cal.2d 465, 474-475.

1233 ibid 474-475; 'Some questions concerning medical negligence require no expertise. Technical knowledge is not requisite to conclude that complications from a simple injection (…), a surgical clamp left in the patient's body (…), or a shoulder injury from an appendectomy (…) indicate negligence. Common sense is enough to make that evaluation': *Franz v. Board of Medical Quality Assurance* (1982) 31 Cal.3d 124, 141.

1234 McDonald, *California Medical Malpractice: Law & Practice* (Revised Edition 2022) § 2:6.

1235 'California decisions have wafted considerably on the rule's meaning and significance. Even now the topic remains in a nether realm of inconsistency, confusion and double-talk': ibid § 2:6.

1236 'It is now well established, however, that the locality of the practitioner is merely one of the "circumstances" to be considered in evaluating the physician's adherence to the standard of care': *California Jurisprudence* (Third Edition 2022) Healing Arts and Institutions § 442. For an overview of the development of the matter see: *Rainer v. Buena Community Memorial Hosp.* (1971) 18 Cal.App.3d 240, 259-260.

form of professional negligence.[1237] It will be seen that, although the courts have been prepared to shift from the above position somewhat, they have done so only to a very limited extent. As a result, Californian negligence law maintains a relatively narrow focus on the defendant's obligations, as defined by their profession and as evidenced by expert testimony.

In the following, I first subject this standard of review to a closer analysis, ascertaining its requirements, particularly with a view to questioning the extent to which it can protect patient autonomy. On this basis I then examine classes of cases that provide analogies to AI's autonomy challenges.

1. The informed consent standard

The terminology of informed consent was established for the first time in *Salgo v. Leland Stanford Jr. University Bd. of Trustees*. Without expressly committing to a negligence analysis, it was said that liability arose with a failure to disclose 'any facts which are necessary to form the basis of an intelligent consent by the patient to the proposed treatment'.[1238] This understanding of the doctrine was to inspire the recognition of informed consent claims within the United States more generally,[1239] but it left unanswered crucial questions regarding the specific categories of information whose disclosure the law would require and how the test, allegedly deriving from the patient's needs, was to be reconciled with negligence law's orthodox orientation towards professional practice.[1240]

i. The meaning of reasonable disclosure

Cobbs v. Grant was the first Californian case to directly confront 'the yardstick to be applied in determining the reasonableness of disclosure'

1237 *Bigler-Engler v. Breg, Inc.* (2017) 7 Cal.App.5th 276, 321-322.

1238 *Salgo v. Leland Stanford Jr. University Bd. of Trustees* (1957) 154 Cal.App.2d 560, 578.

1239 Faden, King and Beauchamp, *A History and Theory of Informed Consent* (1986) 125-129.

1240 The court did mention the element of risk, highlighting 'that in discussing the element of risk a certain amount of discretion must be employed consistent with the full disclosure of facts necessary to an informed consent': *Salgo v. Leland Stanford Jr. University Bd. of Trustees* (1957) 154 Cal.App.2d 560, 578.

and sought to provide the requisite responses.[1241] It remains the governing precedent for the standard of care in Californian informed consent cases.

In a first significant step, the court departed from the majority of U.S. states and the orthodox approach outlined above, by finding that the determination of the information to be disclosed could not be left to medical custom. Leaning into the focus on the patient previously exhibited in *Salgo* and the federal case of *Canterbury v. Spence*,[1242] it was held that vesting '[u]nlimited discretion in the physician is irreconcilable with the basic right of the patient to make the ultimate informed decision regarding the course of treatment to which he knowledgeably consents to be subjected'.[1243] As such, the patient was viewed as the primary decision-maker in the medical encounter: an important prerequisite for establishing the relevance of their practical autonomy.

In giving the opinion of the Supreme Court, Justice Monk went to commendable lengths to outline the nature of this right to informed decision making and to contextualise it for the purposes of the physician-patient relationship. As Chapter 5 argued, many of these statements are consistent with the procedural conception of autonomy that underlies the present analysis.

Not only was it recognised that a patient lacks parity of medical knowledge with their physician, but also that they possess 'an abject dependence upon and trust in [their] physician for information upon which [they rely] during the decisional process'.[1244] From this 'emerges a necessity, and a resultant requirement, for divulgence by the physician to his patient of all information relevant to a meaningful decisional process (…) it is the prerogative of the patient, not the physician, to determine for himself the direction in which he believes his interests lie'.[1245] Conclusively it was stated: '[t]he scope of the physician's communications to the patient, then, must be measured by the patient's need, and that need is whatever information is material to the decision'.[1246] The patient's decision-making process was to be centre stage in the determination of reasonable disclosure.

1241 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243.
1242 *Canterbury v. Spence* (D.C. Cir. 1972) 464 F.2d 772.
1243 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243.
1244 ibid 242.
1245 ibid 242-243.
1246 ibid 245.

ii. The operationalisation of reasonable disclosure

Given these astute observations on the patient's position during the medical encounter and given their recognised interest in the process of decision making, it would initially appear that the disclosure required under *Cobbs* should be capable of addressing some of the challenges posed by clinical ML devices. The patients are the ones who are assessing what is in their interest and they are the ones who place their trust in the medical professional on the basis of their expertise.

However, the realisation of the stated principles through operational legal standards has proven difficult and has tended to be restrictive.[1247] As Weisbard observed in his analysis of the American approach to informed consent (including *Cobbs*): 'the law has fallen short of its own rhetorical promises, to say nothing of its underlying philosophical premises'.[1248] To understand why this is so, one can begin by looking at the specific formula that Justice Mosk provided.

Alongside its more abstract elaborations, *Cobbs* proposed a two-stage test that effectively restricted the scope for a patient-based assessment. This assessment would be relied upon only to establish 'minimal disclosure' levels.[1249] Hereby 'a medical doctor has a duty to disclose to his patient the potential of death or serious harm, and to explain in lay terms the complications that might possibly occur'.[1250] A subset of available alternatives must also be disclosed (variously described as therapeutic,[1251] feasible,[1252] recommended or reasonable)[1253] together with their comparative advantages and disadvantages.[1254] This is the first limb of the legally defined standard that purports to satisfy the materiality test, conveying the information that has been deemed so decisionally important for the patient.

1247 The *Cobbs* court itself recognised that 'scope of the disclosure required of physicians defies simple definition': *Cobbs v. Grant* (1972) 8 Cal.3d 229, 244.

1248 Weisbard, 'Informed Consent: The Law's Uneasy Compromise with Ethical Theory' (1986) 65(4) Nebraska Law Review p. 749, 751.

1249 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 244-245.

1250 ibid 244-245. See also *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1190-1191; *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285, 1301-1302.

1251 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243.

1252 *Contreras v. St. Luke's Hospital* (1978) 78 Cal.App.3d 919, 927-928.

1253 *Scalere v. Stenson* (1989) 211 Cal.App.3d 1446, 1450-1453.

1254 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 242; *Jamison v. Lindsay* (1980) 108 Cal.App.3d 223, 230; *Thor v. Superior Court* (1993) 5 Cal.4th 725, 735.

A second standard, *prima facie* applying to all other classes of information,[1255] was based on the established practice of the relevant medical community: 'a doctor must also reveal to his patient such additional information as a skilled practitioner of good standing would provide under similar circumstances'.[1256] Successfully establishing this aspect of the claim can be dependent upon expert testimony.[1257] Through this element of the test, one can see how the Supreme Court reintroduced the traditional community standard and struck a balance between the doctrinal requirements of negligence and the imperative of the common law principle of patient autonomy.[1258]

It must be added that, regardless of which category one is dealing with, determining the information required is generally a question for the trier of fact (i.e. for the jury)[1259] and the courts have further emphasised how significant the individual circumstances of a plaintiff are to informed consent claims.[1260] As Applebaum and others have stated: '[a]fter a particular kind of problem is litigated, more concrete standards may begin to evolve (…) But within very broad limits the jury is free to write on a clean slate'.[1261] Such an approach limits the relevance of an abstracted legal assessment of

1255 'When a physician recommends one or more courses of treatment, the information that is "material" (and, hence, that must be disclosed in order to obtain the patient's informed consent) falls into two categories—namely, (1) "minimal" disclosures that are always material, and (2) "additional" disclosures that might be material if "skilled practitioner[s] of good standing" would "provide" those disclosures "under similar circumstances."': *Flores v. Liu* (2021) 60 Cal.App.5th 278, 293.

1256 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 244-245.

1257 *Betterton v. Leichtling* (2002) 101 Cal.App.4th 749, 754-755.

1258 A statement from *Arato v. Avedon* is telling in this respect: 'In *Cobbs v. Grant*, we not only anchored much of the doctrine of informed consent in a theory of negligence liability, but also laid down four "postulates" as the foundation on which the physician's duty of disclosure rests': *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1182-1183. For an examination of such a hybrid analysis more generally, see: Appelbaum, Lidz and Meisel, *Informed consent: Legal theory and clinical practice* (Second Edition 2001) 51-52.

1259 *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1184; *Wilson v. Merritt* (2006) 142 Cal.App.4th 1125, 1135.

1260 'The extent of a physician's duty to disclose is directly controlled by the unique situation of each patient': *Brown v. Regents of University of California* (1984) 151 Cal.App.3d 982, 990-991.

1261 Appelbaum, Lidz and Meisel, *Informed consent: Legal theory and clinical practice* (Second Edition 2001) 49.

the forms of disclosure that a patient would require regarding the use of ML technologies.

That being said, a number of judgments have been prepared to determine more concrete legal parameters that help ascertain the disclosable features under the first limb of *Cobbs*. They have done so under an interpretation of the materiality test that represents one of two strands of jurisprudence in the U.S.[1262] Namely, they have found that disclosable, material information is that which would be regarded as significant by a reasonable person in the patient's position.

This parameter can be gleaned from the decision of *Truman v. Thomas*. In considering whether certain disclosure was legally necessary, the court relied on the aforementioned quote from *Cobbs* that 'all information material to the patient's decision' must be given and then went on: '[m]aterial information is that which the physician knows or should know would be regarded as significant by a reasonable person in the patient's position when deciding to accept or reject the recommended medical procedure'.[1263] On the basis of this standard, the California Supreme Court established for the first time in the United States that a serious risk of *not* undergoing a procedure was disclosable.[1264] This figure of the reasonable patient has been drawn upon by subsequent courts to establish the necessity of various other forms of disclosure.[1265]

In *Moore v. Regents of the University of California*, for example, it was determined that a physician must 'disclose personal interests unrelated to the patient's health'.[1266] In so holding, the court expressly referenced that, although the failure to disclose risks was often the focus of medical malpractice actions, the requirement of materiality was capable of covering more diverse situations: 'the concept of informed consent (...) is broad enough to encompass [the situation where the physician has a personal interest]'.[1267] Because a reasonable patient would want to know that there

1262 Weisbard, 'Informed Consent: The Law's Uneasy Compromise with Ethical Theory' (1986) 65(4) Nebraska Law Review p. 749, 759-760; King, 'The Reasonable Patient and the Healer' (2015) 50(2) Wake Forest Law Review p. 343, 350-351.

1263 *Truman v. Thomas* (1980) 27 Cal.3d 285, 291.

1264 ibid 290-295.

1265 *Wilson v. Merritt* (2006) 142 Cal.App.4th 1125, 1133-1134; *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285, 1302-1305; *Jambazian v. Borden* (1994) 25 Cal.App.4th 836, 844-845; *Flores v. Liu* (2021) 60 Cal.App.5th 278, 292-293.

1266 *Moore v. Regents of University of California* (1990) 51 Cal.3d 120, 131-132.

1267 ibid 129.

is a possibility of extraneous interests affecting the physician's judgment – something that is established *inter alia* by reference to previous judicial pronouncements on non-disclosure related issues – such information is material.[1268] One can therefore see that the materiality test was applied in a way that both focussed on analogy (previous categories of disclosure), as well as appealing to an underlying legal determination of the needs of the reasonable patient.

Californian courts have also been prepared – indeed, somewhat more prepared – to identify classes of information that are not material to a medical decision, thereby restricting the scope for permissible (legal) argumentation. *Cobbs* already referred to 'relatively minor risks' in this manner, and highlighted that a patient must not be given '[a] mini-course in medical science'.[1269] Information that is commonly appreciated has also been excluded from the standard of materiality.[1270]

In addition, the seminal California Supreme Court case on this issue, *Arato v. Avedon*, highlighted the judiciary's general scepticism towards recognising specific categories of material information that would have to be disclosed as a matter of law – such a move was described as unwise.[1271] Accordingly, the court rejected a claim that statistical life expectancy data should have been disclosed.[1272] Whether such information was material was a situational judgment for the jury to make and, more generally, the non-therapeutic interests of a patient were not a suitable basis on which to determine the appropriateness of disclosure.[1273]

Lastly, a restriction regarding information that must be disclosed also emerges from the aforementioned nature of the negligence action: that it does not demand perfect behaviour, but only conduct that meets a reasonable standard. In consequence, if a patient is unable to establish that a physician ought to have known a particular piece of information, then

1268 ibid 129-130. The court referred specifically to a case not dealing with informed consent – *Magan Medical Clinic v. California State Bd. of MedicalExaminers* – which stated that: 'a sick patient deserves to be free of any reasonable suspicion that his doctor's judgment is influenced by a profit motive': *Magan Medical Clinic v. California State Bd. of MedicalExaminers* (1967) 249 Cal.App.2d 124, 132.

1269 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 242, 244.

1270 *Truman v. Thomas* (1980) 27 Cal.3d 285, 291.

1271 *Arato v. Avedon* (1993) 5 Cal.4th 1172, 1185.

1272 ibid 1186-1187.

1273 ibid 1186-1189.

they cannot be obligated to disclose it.[1274] It was concluded in Chapter 2 that a reasonable practice involving ML technologies must be predicated on some knowledge of the general characteristics that have been deemed relevant to the patient's exercise of their autonomy. Without a knowledge of AI's general risk-related status, their provision of expertise, their ability to influence the professional's decision making and the kinds of purposes it is suited for, a professional cannot responsibly employ the technology. At the same time, the possession of information about these factors will be a matter of degree. Where relevant in the subsequent analysis, a potential lack of knowledge and corresponding duty will therefore be pointed out.

iii. Summation

The current standard set by Californian common law for a professional's disclosure obligation to a patient has been conceived as a balance between the demands of the negligence doctrine and the imperatives of informed patient consent. These imperatives can be interpreted as largely consistent with the procedural conception of autonomy adopted in this thesis, under which an understanding of AI's challenges has been developed.

Deeming specific information disclosable is not done through recourse to an untethered normative analysis, however. Rather, it also requires reference to the standard of an objective, reasonable patient and, in practice, one must proceed primarily *via* an intentionally cautious development of the established categories of required disclosure.

Consequently, the following should be read as an argument by analogy that explores the relevance of classes of legally recognised information to AI's novel interferences with patient autonomy. In so far as AI-relevant information can be subsumed directly under established categories, this provides the strongest basis for finding a breach of informed consent obligations. Complementing this, an argument can also be constructed by identifying the needs of the prudent person. This will further reflect, albeit more indirectly, the law's consideration of the demands of patient autonomy.

1274 *Townsend v. Turk* (1990) 218 Cal.App.3d 278, 284-285, citing: *Moore v. Preventive Medicine Medical Group, Inc.* (1986) 178 Cal.App.3d 728,739.

2. The risks of medical AI

In Chapter 3 an analysis was provided of how clinical AI can impact the risk profile of an intervention. Two arguments were advanced: an analysis of certain AI that will impact the specific type of risks a patient is exposed to in their treatment and a claim that understanding certain general characteristics of AI will allow the patient to undertake an appropriate risk calculus, without being overloaded with information. The professional would interfere with the positive dimension of the patient's practical autonomy if they did not disclose such information and foster the requisite understanding.

In fitting these factors into the outlined Californian framework, one can begin by noting that the courts have linked inadequate risk disclosure with a diminishment of positive, practical autonomy. As was argued in Chapter 5, *Cobbs* and its progeny themselves have recognised this connection. It was also seen above that they have emphasised its importance by mandating the disclosure of material, serious risks. Yet, as this indicates, informed consent obligations are not breached by just any failure to disclose information about risks. The courts have carved out a relevant subset, the conditions for which will be applied to AI in this section. In line with our perception of AI's challenges, we first consider AI's specific risks and then their more abstract risk-related status.

i. Specific risks

The first requirement is that a given risk must be established by reference to medical expertise.[1275] In and of itself this appears uncontroversial. Without proving the existence of a risk, a patient cannot complain that the risk materialised or that their decision making was impacted by a failure to disclose it. With regard to AI, the existence of specific risks will have to be proven and this can be problematic due to the uncertainty surrounding, and difficulty testing, even approved AI. As referred to above, if a patient is unable to establish that a physician ought to have known of a risk, then they

1275 *Jambazian v. Borden* (1994) 25 Cal.App.4th 836, 849-850; *Betterton v. Leichtling* (2002) 101 Cal.App.4th 749, 756.

also need not disclose it.[1276] This constitutes a real problem for demanding the disclosure of specific AI-risks.

Assuming that the plaintiff provides proof of risk, however, it must then also be shown that the requisite risk information is material. That is, it would be considered significant by the reasonable person in the patient's position. For specific risks, this can be established according to relatively well-defined criteria: it is measured both by the gravity of the harm and the probability of its occurrence. For instance, in *Truman*, the undisclosed risk of cervical cancer was of a very low probability, but the potential harm (death after a failure of early detection) was of the gravest kind.[1277]

Certain AI risks will be significant in this sense. As previous examples demonstrate, AI will be used in critical functions where there is a potential for very serious harm, such as for the identification of a brain haemorrhages. Moreover, initial high values for accuracy (taken as a rough stand-in for the probability of harm eventuating) were also shown to often require downward revision for various reasons. Depending on the surrounding circumstances and the specific uses, there will thus sometimes be a meaningful probability of this harm eventuating. In such cases, AI do not pose minor, low probability risks, but rather material risks.

However, since an ML device will usually impact the existing material risks of a procedure or decision in which it is integrated, one must further enquire whether its contribution is disclosable independently from the information provided about risks more generally. To be a suitable subject of the disclosure doctrine in these circumstances, a material risk must be of a type that distinguishes it from the overall risk assessment.

Morgenroth v. Pacific Medical Center, Inc. illustrates the issue. The court held that the risk of a stroke during a necessary diagnostic procedure was not different to the risks of death and serious harm that had already been disclosed to the patient.[1278] This meant that mentioning only the general and not the specific risk was not a breach of the defendant's obligation, such disclosure met *Cobbs*' materiality test.[1279] In comparison, *Warren v. Schecter* reached the opposite conclusion upon its facts. The doctor did

1276 *Townsend v. Turk* (1990) 218 Cal.App.3d 278, 284-285, citing: *Moore v. Preventive Medicine Medical Group, Inc.* (1986) 178 Cal.App.3d 728, 739.

1277 'Although the probability that Mrs. Truman had cervical cancer was low, Dr. Thomas knew that the potential harm of failing to detect the disease at an early stage was death': *Truman v. Thomas* (1980) 27 Cal.3d 285, 293.

1278 *Morgenroth v. Pacific Medical Center, Inc.* (1976) 54 Cal.App.3d 521, 534.

1279 ibid 534.

advise the patient of serious risks (the eventuation of which called for a second surgery) and of a risk of death from gastric surgery, but not of the risk of severe osteoporosis.[1280] The latter was regarded as a distinct type, warranting separate disclosure. [1281]

This presents a particular problem for medical AI because, unless they themselves represent a separate procedure (an issue that is considered below), AI are unlikely to generate a new kind of serious harm – as the osteoporosis in *Warren* unquestionably did. For example, while the use of the Acumen Hypotension Prediction Index Software may lead to different kinds of errors and varying degrees of uncertainty in the prediction of a hypotensive event, it does not itself generate the risk of such an event or alter its gravity. Under this approach then, the specific risk changes associated with AI that are integrated into a wider process (i.e. the great majority of cases) will not be material.

Nor is it easy to argue for an alteration of this approach from the perspective of procedural autonomy. Focussing on material risks, as well as classifying risks of a similar kind and magnitude together, serve a useful purpose in the fortification of the patient's positive autonomy: it prevents the patient from being overwhelmed by information in their decision making. As was stated in *Cobbs*: 'the patient's interest in information does not extend to a lengthy polysyllabic discourse on all possible complications'.[1282] Listing the specific effects of AI technology on the risks and benefits of a complex procedure would arguably run afoul of these objectives.

Ultimately, the limited disclosure obligation under Californian law reinforces the view expressed in Chapter 3, that requiring information about the particular risks involved in the use of an AI device is of limited utility in addressing the underlying, more nuanced challenges to the patient's decisional and practical autonomy. Such an approach would not even mandate the disclosure of AI use, its separate characteristics or its dangers. Instead, such matters would be obscured under the head of an overall calculus of risks.

1280 *Warren v. Schecter* (1997) 57 Cal.App.4th 1189, 1195-1196.
1281 ibid 1202.
1282 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 244.

ii. Risk-relevant status

A wholly different approach is conceivable if AI/ML itself constitutes a material consideration because of its abstracted relation to the risk analysis. The distinct nature of the technology would be moved into the foreground, without overloading the patient with information, and their decision making would be facilitated. Such an approach is in line with *Cobbs*' central tenets and, as will be explored here, it can be further supported by Californian courts' treatment of the innovative nature of a procedure as a material, risk-related piece of information.

One can begin by noting that the burden of proving a status seems considerably lighter than that of proving the existence of a specific risk. If one focuses on an abstract classification of AI for disclosure purposes, then it is easier for the plaintiff to demonstrate (1) that the technology that was applied to them was of a certain kind, and (2) that this kind is associated with features that are relevant to the patient's risk assessment.

An analogy for this analysis can be found in *Clemens v. Regents of University of California*. Here the plaintiff claimed that they had been subjected to a procedure with an experimental status and that this bore special significance for disclosure purposes.[1283] Given that this was the first claim of its kind, the plaintiff may be forgiven for failing to show that the procedure they underwent was in fact experimental. While not denying the significance of the status, and admitting that the procedure was new, the court held that the testimony and practice of the defendants (who had been employing it for over two years) did not bear out the classification.[1284]

In the subsequent case of *Trantafello v. Medical Center of Tarzana* it was then possible to build upon this limited recognition. The Court of Appeal considered that the plaintiff's evidence had successfully raised a triable issue of fact regarding the claim that the use of a certain substance in their surgery was innovative. [1285] As will be discussed below, both courts appear to have been satisfied that, if the plaintiffs can make good their claim as to the experimental or innovative status of an intervention, this information ought to have been disclosed to the patient.

For AI it is envisioned that a similar approach can be consistently adopted. A patient should be able to prove that a technology utilised in their

1283 *Clemens v. Regents of University of California* (1970) 8 Cal.App.3d 1, 9.
1284 ibid 7.
1285 *Trantafello v. Medical Center of Tarzana* (1986) 182 Cal.App.3d 315, 320.

care possessed the relevant capabilities, most likely by employing a form of ML (as per the definition of Chapter 2). This will require expert evidence. Establishing the materiality of these AI/ML capabilities requires the second, separate evaluation – drawing on their relation to the risk analysis. That the possession of a risk-related status is an indicator of materiality, mandating disclosure, can be derived most straightforwardly from a deeper look at the two examined cases.

In *Clemens* the court was prepared to accept the validity of a jury instruction that 'a physician seeking consent to a "new or experimental" procedure should inform the patient that it is new or experimental when seeking to consent to it'.[1286] This disclosure was deemed necessary, alongside the disclosure of the more specific risk that was in fact disclosed and which eventuated.[1287]

Furthermore, although *Clemens* preceded *Cobbs*, its continued relevance can be gleaned from *Trantafello*, which stated expressly that the professional's 'duty to inform plaintiff of the alleged innovative nature of the treatment [arises] from the general rules stated in *Cobbs v. Grant* (…) and *Clemens v. Regents of University of California*'.[1288] As in *Clemens*, the *Trantafello* court determined that this disclosure was required separately from information regarding the available options and dangers involved.[1289] And, in line with the above analysis, the significance of this information appears to have been deduced primarily from its suspected impact on the probability of causing the patient physical harm (i.e. its risk-related status).[1290]

As such, I do not agree with McDonald, who argues that the innovative nature of a procedure shapes disclosure requirements by heightening existing risk disclosure obligations associated with a procedure.[1291] Both *Clemens* and *Trantafello* indicate that other obligations are not so much

1286 *Clemens v. Regents of University of California* (1970) 8 Cal.App.3d 1, 9.
1287 ibid 8-9.
1288 *Trantafello v. Medical Center of Tarzana* (1986) 182 Cal.App.3d 315, 320, fn. 2.
1289 'Dr. Richland was required to advise plaintiff prior to obtaining plaintiff's consent, of the innovative nature of the operation and the available options and dangers involved': ibid 320.
1290 'The theory of plaintiff's case is that the generally accepted practice in disk surgery is to implant a bone graft for this purpose; that the use of methyl methacrylate was an innovative procedure not generally accepted in the United States because of a high probability it will not properly fuse or heal to the bone and which has a high incidence of pseudo arthrosis': ibid 319.
1291 McDonald, *California Medical Malpractice: Law & Practice* (Revised Edition 2022) § 2:11.

heightened, as an additional category of material information is established. Once a plaintiff proves that a device or procedure with general risk-related qualities was involved in their care, they have a right to this information, even if specific risk disclosure obligations are not applicable or have been fulfilled. This is what the patient's right to self-determination demands, both as outlined by *Cobbs* in the abstract and in these two cases in particular.

Another case that deserves attention in this respect, is *Daum v. Spinecare Medical Group, Inc*. It was already seen in our earlier analysis that the use of an experimental device was not deemed capable of constituting a battery. Yet this did not prevent the Court of Appeal from stating that a breach of statutory informed consent requirements – which were said to represent 'procedural additions to the general common law requirements' – could be found where an individual was not provided with the written consent forms for the use of an investigational device.[1292]

Although an explicit materiality analysis was rendered unnecessary by the statutory background of the claim,[1293] the court made some useful comments during its causation analysis that, as will be seen below, similarly incorporates reference to the reasonable plaintiff.[1294] Namely, it was found that an investigational or experimental device possessed several risk-related characteristics that can be significant to the prudent patient in Mr. Daum's position.[1295] These factors included: a lack of evidence regarding a device, its unapproved regulatory status, the availability of alternatives, and the plaintiff's preference for a conservative approach (literally: 'Mr. Daum claimed he would have been unwilling to be the subject of an experiment with an unproved device (…) he knew from this experience that "the first things that you will build always have some problems with them."').[1296] Note that none of these factors relate directly to specific complications, but

1292 *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285, 1307-1309. Elsewhere the device was referred to as experimental: ibid 1296.

1293 'The disclosure at issue in *Daum* was required by statute and regulation, and in those circumstances we held it was reversible error to require the jury to consider only expert testimony in deciding whether the defendants had complied with their duty of disclosure': *Betterton v. Leichtling* (2002) 101 Cal.App.4th 749, 755.

1294 For an analysis of the two roles of the reasonable patient see: King, 'The Reasonable Patient and the Healer' (2015) 50(2) Wake Forest Law Review p. 343, 349-350.

1295 Generally the device was described as investigational, but experimental was treated as a synonym: *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285, 1296.

1296 ibid 1311-1312.

that they nevertheless raised a question of fact whether a prudent patient would have declined the relevant surgery.[1297] Transferring such insights to the materiality analysis, this finding arguably suggests that a reasonably prudent patient would also consider the investigational status of the device (with all the associated characteristics) material.

The reasoning of the innovative and experimental cases provides several grounds for comparison to AI. In Chapter 3, a covert employment of ML methods was found to be problematic for a patient's theoretical autonomy because: they have a more tenuous connection to the state of medical knowledge, they are subject to an irreducible risk of relying on confounding variables, there are still unsolved difficulties in their scientific validation according to established methods and there is an inherent opacity in their operation that may obscure a more specific assessment of risks. Furthermore, it is true that, unlike the device in *Daum*, it is not suggested that use of an AI automatically constitutes an investigation or experimentation in and of itself (either under the state or federal statutes). Nevertheless, it was commented in Chapter 3 that one objective of online ML functioning will ordinarily be to improve its own performance. In such cases the analogy with experimental or investigational procedures is heightened by the introduction of this research-related purpose.

On these bases it is argued that Californian courts should be prepared to extend the reasoning of *Cobbs*, *Clemens*. *Trantafello* and *Daum* to AI technology and find that the AI/ML status is in itself material information. This would give due deference to existing doctrine, in the form of *Cobbs*' two-stage analysis, while also appropriately recognising the autonomy principle's role in shaping this area of the law. As far as the breach element of the negligence action is concerned, such a limited extension is capable of addressing one of AI's unique autonomy challenges.

3. Alterations to expertise

As argued in Chapter 3, unknown changes from human to AI expertise arguably present a problem for patient autonomy, hindering or altogether preventing them from exercising their theoretical rationality appropriately. This raises the question whether, in certain scenarios, it is the obligation of the physician to affirmatively disclose basic information about their status

1297 ibid 1312.

and identity, about their relative expertise and about their relation to other professionals involved in the patient's care.

Different strands of Californian case law suggest that the non-disclosure of the (lack of) medical expertise of a caregiving professional constitutes a breach of their duty to obtain the patient's informed consent. This section offers an interpretation of these strands that is consistent with the principle of procedural autonomy, providing it with coherence, and argues that it should be affirmed on this basis. It is further maintained that a logical extension to medical AI will require professionals to disclose those situations in which the technology takes over a significant aspect of their expert judgment.

The first authority that is directly in point, is *Davis v. Physician Assistant Bd.* A physician assistant sought relief against a disciplinary action that revoked their licence.[1298] *Inter alia* this decision had been based on the assistant's negligent failures to obtain patients' informed consent. The patients had not been advised that only the assistant – rather than a medical doctor – would be the one performing their procedures.[1299]

Seeking a basis on which to make this novel determination, the Court of Appeal began by highlighting that legally mandated disclosure was not limited to information concerning risks and alternatives.[1300] Echoing *Moore v. Regents of the University of California*, the court then stated that the concept of informed consent is 'broad enough to include information about whether the person who is going to perform a patient's surgery is a doctor or not'.[1301] By applying the reasonable patient test of materiality, the *Davis* court determined that: '[c]learly identifying the practitioner who would perform surgery, making clear whether the person performing the procedure is a physician assistant and not a doctor, and making clear whether or not a physician would be involved at all are matters relevant to informed consent'.[1302]

In this manner, the court expressly noted the relevance of status and, more concretely, the non-doctor status of the person providing primary care. *Davis*' injunction to inform the patient of (any) physician involvement opens the door for argumentation that there is a broader right to know

1298 *Davis v. Physician Assistant Bd.* (2021) 66 Cal.App.5th 227, 231.
1299 ibid 275.
1300 ibid 277.
1301 ibid 278.
1302 ibid 278.

about the actors involved in one's care, their levels of expertise and their relationship to one another.

However, to be contrasted with this finding is *dicta* from *Quintanilla v. Dunkelman*. Here a breach of professionals' informed consent obligations was found, but primarily on the basis that the nature of the patient's choices was not adequately explained to her.[1303] The court did not judge on several issues surrounding the characteristics of the treating surgeon: the plaintiff claimed that she did not know that the surgeons treating her were not gynaecologists;[1304] the court accepted that she was not aware of the identity of the person who performed the procedure, since she believed that it would be the physician with whom she enjoyed a pre-existing relationship;[1305] and the court commented upon a disparity in the expertise between these two individuals. Whereas the person who she believed to be operating was 'a board-certified general surgeon who performs gynecological surgery', the actual surgeon was 'a general surgeon who spent four to six months in residency in gynecology'.[1306]

One way to interpret the court's failure to problematise these aspects, is as an implicit judgment that the non-disclosure of the physician's personal characteristics does not violate informed consent obligations. This finding could be reconciled with *Davis* on the basis that there was no fundamental issue concerning the surgeon's status in *Quintanilla*: the individuals were clearly qualified and licensed to perform the actions that they did.[1307] But this would be an expansive understanding of a limited discussion. A more straightforward explanation can be found in the court's preference for limiting its finding to orthodox grounds, which were readily available. Namely, the aforementioned failure to disclose the nature of the procedure.

In addition, it is illustrated by the facts of *Quintanilla* that it would be anomalous from the perspective of patient autonomy to regard status, but

1303 *Quintanilla v. Dunkelman* (2005) 133 Cal.App.4th 95, 113-115. See also the discussion *infra*.

1304 ibid 101, 107.

1305 ibid 118-119, 102-103.

1306 ibid 101-102.

1307 What was left unclear in *Davis*, however, is in how far the professional status and the professional regulation of the defendant – that certain acts constituted unlicensed, unauthorised, unsupervised practice, etc. – were relevant. That the licensed or unlicensed nature of conduct was not mentioned in the informed consent discussion arguably indicates that this was not understood as a defining characteristic of the material disclosure: *Davis v. Physician Assistant Bd.* (2021) 66 Cal.App.5th 227, 276-279.

not expertise and identity, as actionable non-disclosures. The patient had very clearly placed her trust in one particular individual with whom she had several interactions and whom she understood to possess a requisite degree of skill to carry out a sensitive procedure on her. Surely this was material information from the perspective of a reasonable patient making their medical decision. Under an approach that focuses on procedural autonomy, a patient should simply be required to prove that the human expertise applied to their care was lessened or morphed to such an extent that it would have assumed practical significance in their decision making.

That *Quintanilla* cannot be taken as instructive on the professional's duty to advise a patient of their relative expertise, is also supported by *Moore v. Preventive Medicine Medical Group, Inc.* The Court of Appeal recognised that a patient must be told the risk of being examined only by a non-expert: 'material information included the risk to Moore if he was not examined by the specialist'.[1308] The lesser capability of the referring physician required this risk to be impressed upon the patient at the point of their interaction.[1309] Disparities in the expertise of one's professional is therefore something that can be material to the reasonable patient's decision making in a specific context.[1310]

However, it is true that the applicability of *Moore* has been delineated much more narrowly than suggested by *Davis*. Subsequent case law has affirmed that a duty to advise is only triggered where a less qualified professional actively refers the patient for expert assessment.[1311] A duty to discuss the disparities of expertise and the associated risks will therefore not arise without more. For example, where the physician could recommend an expert opinion, although they are not required to do so by ordinary negligence principles,[1312] but they do not think it necessary on the facts of the case.[1313]

1308 *Moore v. Preventive Medicine Medical Group, Inc.* (1986) 178 Cal.App.3d 728, 738-739.

1309 ibid 738-739.

1310 For a similar interpretation, see: Terrion, 'Informed Choice: Physicians' Duty to Disclose Nonreadily Available Alternatives' (1993) 43(2) Case Western Reserve Law Review p. 491, 511-512.

1311 *Scalere v. Stenson* (1989) 211 Cal.App.3d 1446, 1450.

1312 ibid 1453.

1313 See the next section for an analysis of the requirements for the disclosure of alternatives.

Moore's duty is also construed narrowly in that the difference in expertise must be relevant to a specific procedure and pose a corresponding risk. What matters are not the general characteristics of the professional and the expert, but rather the specific danger that is associated with a failure to follow through on a particular course of action that has been recommended. In *Moore* this course of action was the physician's statement that one could not know a mole's nature for sure, unless the patient saw a specialist and 'got it removed or studied microscopically'.[1314]

Bringing AI-induced changes directly under *Moore* would prove difficult as a result. One aim of leveraging ML technologies is to supplement a less qualified professional's expertise with specialist knowledge. As such, after AI analysis, there will not often be a duty under ordinary negligence to refer a patient to an 'actual' human expert. Nor will it always be easy to frame the information that is relevant to ML-generated expertise as a risk that is associated with the non-acceptance of a procedure. ML assistance may constitute a broader form of assistance, which is not linked to one specific recommendation, as in the case of AI-Pathway Companion Prostate Cancer.

Consequently, the line of case law originating with *Moore v. Preventive Medicine Medical Group, Inc.* is one important judicial recognition of the significance that differing levels of professional expertise can have on the patient's understanding of their care pathway. But it does not provide obligations to disclose such information in its own right.

In light of these conflicting signals, an interpretation of *Davis* that imposes an obligation to disclose the identity of relevant actors in the patient's care, their status and their levels of expertise arguably offers the most coherent interpretation of these different strands of case law. It would vindicate the importance that has been attached to differing levels of professional expertise and it would reflect the significance that a patient is likely to attach to the identity and the capabilities of their carer.

Such an interpretation could then require the disclosure of shifts of expertise precipitated by AI. Recall in this respect an ML device like IDx-DR. Its use allowed a less qualified professional to assume a task that was previously reserved for a more skilled practitioner. *Davis* supports the proposition that a patient has an interest in knowing the expertise-related differences in the actors involved in their care. This is especially true where one actor, such as the physician, is expected to be the central figure in

1314 *Moore v. Preventive Medicine Medical Group, Inc.* (1986) 178 Cal.App.3d 728, 734.

their care. If it is not obvious that an individual with a lesser status and/or expertise is using such a device, then it is arguable that discrepancies in expertise, and the role of the device, constitute material information.

In addition, a patient could strengthen their argument by demonstrating that the discrepancy has the requisite degree of significance. This could emerge for example – as in *Moore* and *Quintanilla* – from the high-stakes nature of the decision. It could also be inferred where the patient can demonstrate that there was a decrease in the overall level of expertise brought to bear on the decision – as in *Davis*. For example, that an AI did not fully substitute the capabilities of the expected human actor.

Things are arguably different where the primary carer's expertise remains unchanged. Where an AI like Mia (Mammography Intelligent Assessment) is involved, whereby the AI merely provides a second opinion to a human radiologist, there would arguably be no disclosable change in the expertise of relevant actors. Any deficiencies in the AI would be compensated for by the involvement of the human professional.

Such a result constitutes a legitimate, plausible development of existing case law. Nevertheless, the incomplete nature of this case law, as well as the conflicting indications that are provided by it, limit the strength of the arguments that can be made in this respect. Requiring the disclosure of shifts in AI-human expertise must therefore be seen as an available argument of uncertain strength under Californian common law.

4. Information concerning the choice of goals

AI's independent pursuit of objectives was argued to be problematic on two accounts in Chapter 3. A very strong challenge emerged from AI that could partially determine an aspect of a clinical decision. A related interference stemmed from AI's ability to impact human decision making through a lesser form of influence, through nudging. Patients are thereby subjected to latent pressures to pursue non-personalised objectives.

The manner in which the Californian informed consent doctrine secures a patient's control over their goals of treatment is by making available to them information concerning: the choices available in their care, including the purposes of recommended options, and by requiring the relevant professionals to disclose any extraneous interest that they may have in their care. Recourse to the autonomy principle is necessary to subsume the requisite information about ML devices under these heads.

i. Understanding choices

Under California's informed consent doctrine, different options must be disclosed to the patient where they constitute material information.[1315] Ordinarily this will involve a choice between courses of action with distinct risk profiles. For example in *Jamison v. Lindsay* the focus was on dangerous therapeutic interventions, for which 'the therapist must inform the patient of the available alternatives and the hazards involved so that the patient is able to give effective consent to the proposed treatment.'.[1316] Similarly, in the aforementioned case of *Quintanilla v. Dunkelman* it was found that a failure to discuss available choices and their dangers, including the surgical procedures actually performed on the patient, constituted a breach of duty.[1317]

ML devices will be subsumable under the existing paradigm in so far as they are deemed to possess risk-related characteristics. Opting for a diagnosis or treatment method involving such AI over another procedure, or over non-treatment, with a different risk-benefit balance, would require a discussion of both options (subject to the caveat discussed below that it is a reasonable alternative). Yet, disclosing the risk-related characteristics of AI does not provide the patient with information regarding AI's objectives or their influence on the non-risk related goals that are pursued in patient care. To consider whether the disclosure of such factors is required, one must analyse them under the wider rationale which informs a professional's duty to advise a patient of the available options.

Namely, allowing a patient to choose amongst alternatives is a recognition of the fact that they are entitled to choose in light of their own objectives. As Applebaum and others have stated in their analysis of U.S. informed consent law: 'the choice among options cannot be made on the basis of objective criteria (…) Informing the patient about alternatives permits patient to make a decision in light of his values, preferences, goals and needs'.[1318] Where a professional makes a choice amongst options for the patient, they exert a very strong influence on the patient's decision making

1315 'A duty of reasonable disclosure of the available choices with respect to proposed therapy': *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243. See also: *Traxler v. Varady* (1993) 12 Cal.App.4th 1321, 1330-1331.

1316 *Jamison v. Lindsay* (1980) 108 Cal.App.3d 223, 230.

1317 *Quintanilla v. Dunkelman* (2005) 133 Cal.App.4th 95, 114-115.

1318 Appelbaum, Lidz and Meisel, *Informed consent: Legal theory and clinical practice* (Second Edition 2001) 59.

– a decision making that is (as it was remarked in *Cobbs*) dependent on being facilitated by the physician's greater expertise. Without the physician's disclosure of alternatives, there is a very real danger that the patient's care is disconnected from the pursuit of their medical and non-medical interests.

Multiple *dicta* testify to the fact that Californian courts have recognised this rationale. In *Mathis v. Morrissey* the court gave the example of a professional disagreement on the optimal treatment for breast cancer.[1319] It was held that, even if the pursuit of either procedure can be supported within the relevant medical community, it was for the patient to make the final, personal decision, so that the professional recommending one procedure had to disclose the other, recognised schools of thought.[1320] In *Wilson v. Merritt* the court was also prepared to recognise the importance that a paraplegic, in selecting among treatment options, would attach to maintaining the use of their arms and shoulders in light of their dependence on them for their mobility.[1321] The danger of disconnecting the patient from the pursuit of their own, individual objectives therefore clearly shapes the information that must be provided in scenarios involving the discussion of alternatives.

From the expositions in Chapter 3 it can be deduced that AI's control over objectives will not normally lead to a similar disconnect. Even if an AI possesses advanced capabilities and nudges the patient towards making a particular kind of choice among their options, *inter alia* by withholding or framing information, the oversight and assistance of a human professional is envisaged to maintain the patient's ability to direct their care. Selecting treatments involving AI will, for the most part, not amount to a pre-determined choice of one alternative (i.e. the AI-favoured one) above another.[1322]

However, in one identified subset of cases an analogous separation between the AI's goals and a patient's personal preferences was identified. This was where an AI is used to determine an aspect of clinical decision making and possesses a relatively wide discretion to select non-personalised objectives. The examples given in this regard were: (1) an AI that may be used as a general-purpose diagnostic tool, which is capable of generating surprising insights into serious underlying conditions (2) an AI that triages the patient according to its own criteria. Such applications of

1319 *Mathis v. Morrissey* (1992) 11 Cal.App.4th 332, 343.
1320 ibid 343-344. Relying on *Cobbs v. Grant* (1972) 8 Cal.3d 229, 243-244.
1321 *Wilson v. Merritt* (2006) 142 Cal.App.4th 1125, 1139.
1322 Although it must also be remembered that this is a matter of degree, partially dependent on the abilities of the intervening human professional.

ML are comparable to situations where a physician acts upon their clinical judgment to select a procedure without discussing the available options with the patient. In both cases, the pursuit of goals by an external entity binds a patient. Facilitating reasoning on these aspects requires the patient to be informed of the types of commitments that they subscribe to through the decision to use AI, commitments which they could avoid without such a subscription. In other words, the choice whether or not to use AI, and the purposes implicit in this, move to the foreground.

To substantiate the obligation to advise a patient of this information, one must turn to the more specific requirements that the courts have imposed on the disclosure of choices. In particular, a breach is generally only identified where a patient was denied a true choice, one between reasonable, medically indicated options.[1323] As stated in *Vandi v. Permanente Medical Group, Inc.*: 'there is no general duty of disclosure with respect to nonrecommended procedures'.[1324]

Beyond this, however, it refers also to availability. A concrete illustration of this requirement can be provided through *Spann v. Irwin Memorial Blood Centers*. Here, the Court of Appeal rejected the plaintiff's claim that a blood centre was liable for failing to disclose the possibility of a donor reduction programme, which was desirable to mitigate the risk of disease transmission. Since such a programme did not physically exist, and there was no duty under ordinary negligence principles to offer it, there could be no duty to disclose it either.[1325]

This line of authority is most relevant to AI's role in triaging patients, because here the patient is arguably denied no choice for aligning the therapeutic procedure with their preferences at all. Where an AI is integrated into a workflow, automatically prioritising images or callers within a given system for instance, another route may not be in existence. Indeed, the pre-existing option (the default) may be a first come, first served basis, which would be equally, if not more, unreceptive to the patient's values. As a result, if a patient is seeking medical assistance at a particular institution or from a particular specialist, then their automatic subjection to AI triaging could be understood as the only available choice.

1323 *Schiff v. Prados* provides a detailed analysis: *Schiff v. Prados* (2001) 92 Cal.App.4th 692, 701-703.

1324 *Vandi v. Permanente Medical Group, Inc.* (1992) 7 Cal.App.4th 1064, 1071.

1325 *Spann v. Irwin Memorial Blood Centers* (1995) 34 Cal.App.4th 644, 658.

This is reinforced by the use of institutional, financial and geographic constraints to shape the legal definition of availability in the U.S. and California.[1326] As discussed in Chapter 3, an ML triage tool's attraction often lies in it being the most accessible and financially sustainable option that can be offered to patients. If an institution or area uses the tool, and has good financial reasons for doing so, then this reinforces a view of the tool as non-optional. In consequence, there will not be a general requirement to make a patient aware of the possible objectives involved in such non-optional AI use.

By contrast, the information associated with AI's ability to diagnose particular conditions is of a different quality. The patient does have a true choice to make, which is most akin to the scenario considered in *Mathis*. There, disclosure of options was demanded to enable patient determination of a care pathway for a very serious decision: the treatment of breast cancer. Where an AI has the capability to diagnose severe conditions, it is expected that the doctor would have knowledge of this purpose and it represents a meaningful choice for the patient whether they wish to run the risk of confronting this information. Declining diagnostic AI use must ordinarily be seen as a feasible, realistic option that can be exercised once the patient is informed of the relevant objectives.[1327]

In the latter scenario there is an even stronger basis for disclosure, because the device's purpose constitutes part of the nature of the recommended and selected procedure, rather than just being a contextualising piece of information concerning choices.[1328] By selecting an AI that can pursue certain objectives, a professional is already making a certain determination. In this regard, the appropriate analogy is to a case like *Quintanilla*, where the choices at issue were the ones actually imposed on the patient by the physician. Discussing the choices involved means discussing the nature of the procedure that is subsequently carried out by subjecting the patient to an ML device. It is not an abstract option that can be limited by reference to the practice of the medical community. In this case, patient autonomy

1326 Terrion, 'Informed Choice' (1993) 43(2) Case Western Reserve Law Review p. 491, 493-496.

1327 *Truman v. Thomas* was premised on such an approach. The patient had to be informed of the purpose of the pap smear, which was capable of detecting cervical cancer, but was then in a position to make an informed refusal: *Truman v. Thomas* (1980) 27 Cal.3d 285, 293-294.

1328 ibid 293-294.

demands that a meaningful indication of AI objectives and capabilities be provided to the patient.

ii. AI's lesser influence on the pursuit of objectives

The last facet of AI use that was considered for its effect on patient autonomy, was its uncertain potential to influence patients to act in accordance with pre-determined objectives. It was noted that, although it is expected that AI assistance will systematically direct patient choice, it is difficult to actually assess this influence in a given medical decision. Particularly as there will normally be a degree of mediation by an expert physician. A range of factors could determine whether such an expert could help the patient overcome AI-induced biases. The most that the informed consent doctrine can do to respond to this issue, is to convey these concerns to the patient and allow them to adapt their decision making as they see fit.

That this is a type of information which is deemed material under the law of informed consent, emerges from the Supreme Court of California's decision in *Moore v. Regents of the University of California*. Through an analysis of the nature, scope and purpose of informed consent obligations the court determined:

> (1) a physician must disclose personal interests unrelated to the patient's health, whether research or economic, that may affect the physician's professional judgment; and (2) a physician's failure to disclose such interests may give rise to a cause of action for performing medical procedures without informed consent or breach of fiduciary duty.[1329]

The materiality of these interests stemmed from the fact that they had a potential, relatively tenuous influence on the decision maker: 'a physician who does have a preexisting research interest might, *consciously or unconsciously*, take that into consideration in recommending the procedure (…) the physician's extraneous motivation *may affect* his judgment and is, thus, material to the patient's consent'.[1330] In a similar vein, California's highest court referenced *Magan Medical Clinic v. Cal. State Bd. of Medical Examiners*, a case where it had been found that 'a sick patient deserves to be free

1329 *Moore v. Regents of University of California* (1990) 51 Cal.3d 120, 129.
1330 ibid 131 (emphasis added).

of any reasonable suspicion that his doctor's judgment is influenced by a profit motive'.[1331]

Consequently, a professional need not in fact deviate from their patient's desires or impose extraneous goals upon them. It suffices if these goals are present and there is a readily comprehensible manner in which they *may* influence professional and patient decision making.

Moore did not require the disclosure of all such influences, though. It focussed on the presence of non-therapeutic motivating factors, those extraneous to the patient's health. Thus, the patient had stated a cause of action for a breach of informed consent specifically where their physician had an undisclosed economic interest and an undisclosed research interest in their care.[1332] These interests were singled out because they were in opposition to the therapeutic goal of benefiting the patient.[1333] This opposition gave rise to the prospect of 'potentially conflicting loyalties', which the court found so odious to the reasonable patient.[1334] Moreover, as the framing of conflicting or opposing interests suggests, it was not necessary to show that the physician did not have any therapeutic purpose for their actions at all.[1335]

In sum, *Moore* allows for an argument that influences capable of systematically causing a physician to prioritise the non-medical over the medical interests of a specific patient must be disclosed under the informed consent doctrine. This is so even where there is no direct indication that there was no therapeutic purpose or that non-therapeutic objectives were determinative.

For the purposes of medical AI, the precedent laid down in *Moore* identifies a kind of breach that could partially protect against its lesser influences on patient decision making. The first condition, that a readily comprehensible influence on the professional's and patient's process of decision making must be present, but not necessarily active, appears to be fulfilled for many uses of AI. As discussed in Chapters 2 and 3, although it is ultimately dependent upon the design of ML systems, it is expected that this technology will have an increased propensity to influence decision makers and nudge them towards a desired kind of decision. A patient

1331 ibid 129-130; *Magan Medical Clinic v. California State Bd. of Medical Examiners* (1967) 249 Cal.App.2d 124, 132.
1332 *Moore v. Regents of University of California* (1990) 51 Cal.3d 120, 132-133.
1333 ibid 130-131.
1334 ibid 130-131.
1335 ibid 133.

should therefore often be in a position to demonstrate that this potential is present.

Regarding the requirement that this influence must be directed towards certain non-therapeutic objectives, it has also been argued that there will be extraneous factors that flow into a device's reasoning. Two in particular were highlighted in Chapters 2 and 3. First, there will almost inevitably be some financial considerations, at least in the sense that an AI must account for local limitations flowing from the availability of resources and from reimbursement conditions. However, it also stands to reasons that one goal of some devices will go beyond this: to make recommendations in a way that saves an institution costs.

Second, for the subset of AI that operate in an online manner, the objective of improving or maintaining their performance is expected to provide one influence directing the AI's functioning. If not research itself, this activity to improve the general experience of other users (the wider population), is straightforwardly comparable and disconnected from benefitting the individual patient. In this sense, AI may be said to introduce the kind of objectionable, non-therapeutic goals that are likely to generate conflicting interests, which shape the doctor's and patient's decision making.

It could be objected that these nudges are less likely to consciously influence the professional than their own extraneous objectives, bypassing their rationality. However, the above reasoning explicitly recognised that finding a conscious reliance on the relevant objectives was not necessary.

Similarly, while it has been argued that AI are likely to maintain an overriding therapeutic objective, it was explicitly stated in *Moore* that the absence of any therapeutic purpose for the relevant actions is not a requirement. So, where ML devices introduce identifiable financial- or improvement-related nudges to clinical decision-making, it will be possible to argue by analogy to *Moore*.

How far such identification will be possible in practice is the only remaining, and yet most substantial, issue. It is an additional hurdle for a patient dealing with an AI that they must prove the existence of certain objectives and demonstrate how they were (abstractly) capable of directing reasoning. Without easy access to the technology or insights into the processes by which a particular decision, or kind of decision, was reached – including information on how human decisionmakers tend to interact with the ML device – the patient will experience difficulties in showing that objectionable non-therapeutic objectives were present in their care and ought to have been disclosed. Likewise, it may be a considerable challenge

to demonstrate that the doctor had or ought to have had specific knowledge of certain non-therapeutic objectives pursued by the AI. A negligence complaint that AI brought extraneous interests to bear on a patient's care is therefore likely to remain a largely theoretical one. The autonomy-based argumentation would not counteract this underlying problem.

5. Summation

The Californian informed consent standard has been shaped by a careful balancing between the demands of the negligence doctrine and the imperative of protecting the patient's autonomy. In the course of this, it has delineated a number of relatively clear categories that serve as focal points for argumentation. Fitting AI into these categories, while still not straightforward, has allowed one to argue that a patient must be informed about: the risk-relevant status of AI; non-obvious substitutions of the human expertise brought to bear on their care, and; the pre-determination of certain serious choices.

D. Causation

For the purposes of causation, a plaintiff must show that the defendant's breach of duty caused their injury.[1336] Hereby it should be remembered that we are concerned with connecting the breach of informed consent obligations with the eventuation of physical injuries – as has been discussed under the damage element above.

Cobbs v. Grant also serves as the seminal authority on applying this element to cases of informed consent. California's Supreme Court was not content with a subjective test, requiring only that the individual patient would have avoided the damage (e.g. refused their consent for the operation).[1337] Instead, conscious of the evidentiary difficulties that flow from the self-serving testimony that plaintiffs may provide after the fact, the court settled on an objective test: 'what would a prudent person in the patient's position have decided if adequately informed of all significant perils'.[1338] A

1336 *Vasquez v. Residential Investments, Inc.* (2004) 118 Cal.App.4th 269, 288.
1337 *Cobbs v. Grant* (1972) 8 Cal.3d 229, 245.
1338 ibid 245.

subsequent addition has made this test yet more favourable to the defendant. Namely, even if the plaintiff can prove that the reasonable person would not have suffered the damage, then it is still open to the defendant to prove that the individual, particular plaintiff would have consented to the procedure and suffered the damage.[1339]

At times the courts have not entirely adhered to this test, adjusting the reasonable person standard in order to zero in on the hypothetical decision of the individual patient or, at least, on the hypothetical decision of a reasonable patient with some of the same commitments as the individual patient.[1340] As anticipated in Chapter 5, these are arguably cases where the judiciary has been confronted with the reflective dimension of autonomy.

In *Hernandez ex rel. Telles-Hernandez v. U.S.* (a federal decision applying California law) the District Court considered the duty of a professional to disclose to a woman giving birth the relative risks and benefits of continuing with a vaginal delivery, which constituted the plaintiff's preferred and chosen option, and those of a caesarean section.[1341] Rather than simply asking what information the reasonable person would have required, the court was prepared to consider 'Mrs. Telles–Hernandez' emphasis on prenatal care and her desire to deliver her baby without the use of medication' as evidence relevant to causation.[1342] The commitments implicit in these past clinical choices were indicative of the decision she would have made, had she been given adequate disclosure.

A similar approach was implicit in the above-analysed case of *Wilson v. Merritt*. The court accommodated the patient's heightened desire for maintaining mobility in its assessment. It further acknowledged that this desire was demonstrated by the moderately successful therapy that the patient had previously undertaken, and which presented no threat to this goal.[1343]

These cases hardly had the ability – nor did they purport – to overrule *Cobbs*. And the reasonable person standard of causation is arguably too entrenched to provide an opportunity for principle-based developments.

1339 *Truman v. Thomas* (1980) 27 Cal.3d 285, 294, fn. 5; *Warren v. Schecter* (1997) 57 Cal.App.4th 118, 1206; *Flores v. Liu* (2021) 60 Cal.App.5th 278, 297-298.

1340 In addition to the cases outlined below, see: *Morgenroth v. Pacific Medical Center, Inc.* (1976) 54 Cal.App.3d 521, 534-535.

1341 *Hernandez ex rel. Telles-Hernandez v. U.S.* (N.D. Cal. 2009) 665 F.Supp.2d 1064, 1078-1079.

1342 ibid 1078-1079.

1343 *Wilson v. Merritt* (2006) 142 Cal.App.4th 1125, 1128-1129, 1139.

However, *Hernandez* and *Wilson* do suggest that the courts are prepared to utilise a limited discretion to specify the reasonable person standard in a way that is more amenable to the reflective dimension of autonomy and, consequently, to the protection of the deeply held commitments of the individual.

For AI's autonomy challenges this means that, on top of establishing the materiality of the outlined information to the decision making of the prudent patient, a plaintiff will have to demonstrate that this information would also have been of such significance to an abstracted individual that they would have altered their decision and avoided the damage. Ordinarily this would mean showing that there are weighty, rational grounds for refusing a procedure involving an AI or for preferring an alternative. If a patient can additionally demonstrate that they have coherent, long held or robust commitments relevant to AI (for instance with regard to an objective that the AI has overruled), then this may lead to an interpretation of the reasonable person standard that accommodates this commitment.

Much of the above discussion, in this chapter and beyond, has been dedicated to demonstrating that there are compelling grounds for a patient to be concerned about ML technology's influence on their decision making. Yet, it is undeniable that this will not always be sufficient to outweigh other demands and become the dominant concern in a high-stakes clinical situation.

The problem is well illustrated by *Spann v. Irwin Memorial Blood Centers*. Here the plaintiff was aware of the relevant risk and argued only that alternatives and other information related to the plaintiff's condition ought to have been disclosed.[1344] Even supposing that it was a breach of duty not to provide this information, however, the court was satisfied that, in light of the life-threatening seriousness of the patient's condition, a reasonable person in their position would not have declined the intervention.[1345] Implicit in such an approach is the dominant importance that the U.S. courts attach to risk disclosure above all other factors.

In light of this, there will be many situations where a concern for AI-specific factors does not weigh particularly heavily with a reasonable person requiring more-or-less acute medical care. From the adduced examples, we may point to circumstances where a professional fails to disclose alterations in human-AI expertise (particularly where there is no loss in the cumulat-

1344 *Spann v. Irwin Memorial Blood Centers* (1995) 34 Cal.App.4th 644, 658.
1345 ibid 657.

ive level of expertise) and to situations where an AI indirectly influences a patient's choices. Would the knowledge that an AI incorporates some objectives to improve its own performance, and systematically nudges a user, really sway a patient to reject its recommendation? The strongest cases for causation could, by contrast, be advanced with regard to risk-related information and with regard to a failure to advise the patient of their pre-determined commitments by subscribing to AI use. Here, the relation to a specific decision is clear and the importance of this kind of information for clinical choices is well-established.

To a certain extent this approach may be justifiable as a rough means of addressing only significant autonomy violations. But it has already been discussed that autonomy is not only impacted by physical injury. Moreover, the objective causation test laid down in *Cobbs* fails to address autonomy interferences that are recognised as particularly severe under a procedural theory; a focus on the reasonable patient will sometimes allow an external perspective to override the individual's deeply held subjective commitments. Therefore, causation provides another element that significantly broadens the class of cases where the patient's decisional autonomy is impaired without remedy.

E. Awarding damages

When we first began the exposition of tort law's role in the protection of a plaintiff's informed consent, it was stated that the primary means by which its obligations would be enforced is through a *post facto* award of damages – albeit such awards were frequently associated with a prophylactic effect. During our analysis of the negligence action, it was further seen that the first major hurdle for any plaintiff to overcome is demonstrating that they had suffered some form of legally cognisable injury or damage. Under Californian precedent, the only realistic category of injury that could ground an action for the analysed types of AI autonomy interferences, was argued to be physical injury. Merging these two strands, this section examines the kinds of damages that a plaintiff can hope to recover if the requirements of negligence have been made out.

Regarding the first aspect, the type of damage to be recovered, one can begin with the premise that the 'remedy in tort is compensatory in nature and damages are generally intended not to punish a negligent defendant

but to restore an injured person as nearly as possible to the position he or she would have been in had the wrong not been done'.[1346] Compensatory damages may be categorised either as economic or non-economic under Californian statute.[1347] Our focus will be on the latter, given our starting premise that the primary damage suffered to ground a successful claim is physical and that any consequent autonomy violations are associated with 'subjective, non-monetary losses including, but not limited to, pain, suffering, inconvenience, mental suffering, emotional distress, loss of society and companionship, loss of consortium, injury to reputation and humiliation'.[1348]

In particular, in awarding damages for these losses, the law seeks to remedy a host of values deriving from the patient's autonomy, including an individual's 'capacity to participate in the physical world' and to use their 'cognitive and expressive faculty' to synchronise their 'individual existence culturally with the lives of a host of others'.[1349] In principle, a patient who suffers physical damage will be able to claim for all of the detriment that flows proximately therefrom.[1350]

In the informed consent context this includes damages for the injuries and losses the plaintiff would have avoided had they been adequately informed. This was established in *Warren v. Schecter*, where the patient suffered both from complications that the professional had disclosed and from one complication that they had, in breach of their duty, failed to inform her about.[1351] The court held that the patient could recover 'not only for the undisclosed complications, but also for the disclosed complications, because she would not have consented to either surgery had the true risk been disclosed, and therefore would not have suffered either category of complications'.[1352]

However, it must also be noted that, in response to a crisis involving a dwindling availability of medical insurance in California, the legislature enacted MICRA.[1353] This statute limits the non-economic damages that an

1346 *Turpin v. Sortini* (1982) 31 Cal.3d 220, 232.
1347 California Civil Code section 1431.2, subdivision (b).
1348 ibid section 1431.2, subdivision (b)(2).
1349 McDonald, *California Medical Malpractice: Law & Practice* (Revised Edition 2022) § 16:7.
1350 California Civil Code section 3333.
1351 *Warren v. Schecter* (1997) 57 Cal.App.4th 1189, 1195.
1352 ibid 1195.
1353 *Reigelsperger v. Siller* (2007) 40 Cal.4th 574, 577-578.

individual can recover against a healthcare provider or healthcare institution based on professional negligence.[1354] Under amendments coming into effect January 1st 2023, the plaintiff is limited to a total recovery of either $1,050,000 for actions not involving wrongful death,[1355] or $1,500,000 for actions that do involve wrongful death.[1356] It has been established that these capitations apply to claims involving damage caused by breach of informed consent, as these actions are properly characterised as forms of professional negligence.[1357]

Given the largely statutory basis for these provisions, there is also not a great deal of influence that the autonomy principle can have. But neither does it appear necessary to demand a change for the purposes of our present analysis. Although some of AI's autonomy violations have been argued to be significant, it is not obvious that they, by themselves, warrant excessive amounts of compensation. Indeed, the relevant caps are more likely to be met by the eventuation of any accompanying physical injury. So long as the system of damages is prepared to recognise and compensate an individual for the autonomy-related implications of their injury, which is arguably the case regardless of MICRA's provisions, then the autonomy principle does not demand more of this element.

III. Conclusion

In conclusion, it has been claimed that California realistically provides only one common law mechanism, through which an argumentation based on the autonomy principle can effectuate a degree of protection against AI's novel challenges: negligence.

This is not to say that this principle has had no impact on the other relevant mechanism, the battery cause of action. It was found that several elements of this tort were in fact loosened out of considerations for patient autonomy. Yet this still had little practical effect, since the case law went

1354 California Civil Code section 3333.2, subdivision (a).
1355 ibid section 3333.2, subdivision (b).
1356 ibid section 3333.2, subdivision (c). Each of these capped sums is further broken down into maximum amounts of $350,000 and $500,000 respectively for healthcare providers, healthcare institutions and unaffiliated defendants – those who are not owners of another specified entity, in a joint venture with such an entity or have a contractual relationship with such an entity: ibid section 3333.2, subdivision (j)(3).
1357 *Bigler-Engler v. Breg, Inc.* (2017) 7 Cal.App.5th 276, 321-322.

on to impose narrowly and rigidly drawn conditions on the validity of the patient's consent. Their consent would only be invalidated on the basis of substantial changes in physical characteristics of the procedure, changes in professional identity or changes that clearly undermined the therapeutic motivation of the professional. Ultimately, these requirements for a valid consent were ill-suited to address any of AI's unique challenges and they were entrenched on the basis of rule-specific considerations. A requisite amendment or extension of these rules through an appeal to principle was not realistically possible.

The negligence action, by contrast, was found to cover some kinds of cases involving AI's challenges. It was significant that the rhetoric underlying the informed consent analysis bore strong similarities to our analysis of procedural autonomy and, more concretely, precedent had already utilised this analysis to identify violations of informed consent obligations that could be developed to address the ML characteristics evaluated in Chapter 3. A patient could argue that various aspects of AI's risk-related status, their alterations of professional experience and their independent pursuit of objectives should be disclosed to them.

Overall, a plaintiff's claim would nevertheless be severely impaired by a relatively stringent adherence to negligence's doctrinal structure. Several elements are likely to preclude an effective protection of a patient's procedural autonomy, including: a limitation of claims to those involving: the eventuation of physical injury; AI use by primary care professionals – excluding institutions and secondary carers, and; a type of undisclosed information that would have been so significant as to cause a reasonable patient to avoid the injury.

This would result in a state of affairs where any protection against AI's interferences in a patient's procedural autonomy is, while not impossible, incidental at best. If a patient does succeed, then the award of compensatory damages under this claim will, additionally, be capped. Yet this is in line with general rules on the recovery of non-economic losses and does not, by itself, appear inadequate for the kinds of violations under discussion.

In comparison, a prophylactic effect – in the sense that professionals will seek to inform patients of AI features to avoid potential future liability – appears doubtful in light of the many well-established hurdles placed in the way of a successful claim and the more tangential relationship of most of the autonomy challenges to physical injury. In spite of the problematic nature of AI's novel features, their hidden and ancillary position will mean

that many patients will never have the opportunity to confront them and direct their thoughts and actions accordingly.

Chapter 8: Assessment of the comparison and of its wider significance

In reviewing the line of argumentation pursued throughout this work one must begin with the deployment of artificial intelligence (AI) or machine learning (ML) in medicine. As has been elaborated, this is a sophisticated novel technology that is rapidly gaining acceptance in the healthcare sphere and holds yet more promise. This should not be forgotten as one focuses on the series of problems that have been identified in relation to patient autonomy.

Chief among these challenges is the inherent opacity of AI decision making and its fraught relationship with established scientific knowledge. This is marked by the difficulty of carrying out established methods of performance evaluation to gauge the technology's functioning. As a result, even established uses of AI/ML exhibit an uncertainty that justifies analogies to forms of innovative or unlicensed treatment: they possess risk-related characteristics.

Another challenge emerged from AI's ability to pursue goal-directed action relatively independently. Without knowledge of these goals and this functioning, the choice to use certain devices could pre-determine a patient's decision, failing to facilitate their decision-making with regard to fundamental aspects of their care. To a lesser degree, ML devices would also be in a position to surreptitiously insert their own objectives into shared processes of decision making. These influences could be conceptualised as biases, nudges or manipulations that hinder a patient's ability to make theoretically rational choices that reflect their own values.

Lastly, the *raison d'être* of many uses of ML technologies was to mimic the capabilities of human experts. It was outlined how this resulted in various forms of cooperation: sometimes substituting a human specialist, sometimes providing them with a second opinion. It was judged that such cooperation could constitute an epistemological challenge for patients, who are normally justified in relying on the pronouncements of human experts whom they trust in their reasoning. This trust was found not to be transposable to machine-rendered outputs.

These normative, bioethical challenges were then argued to be of significance to the two selected legal systems. Both the UK and Californian com-

mon law have operationalised the concept of autonomy as a legal principle and both have appealed to the dimensions that make up our understanding of procedural patient autonomy. It was already noted at this stage, that the Californian courts were more prepared to impose rule-specific, doctrinal limitations on their reasoning with the autonomy principle. One may say that these countervailing concerns were generally given more weight.

Under the influence of this principle, the mechanisms of battery and negligence have been interpreted to anticipate the common law response to some aspects of AI's distinct challenges. This chapter will begin by comparatively evaluating the precise nature of these responses, assessing their adequacy and considering alternative forms of legal protection that may be necessary to overcome inherent shortcomings.

In the course of this, one should be mindful of the multifaceted ways in which the law is interacting with a technology that is precipitating widespread societal changes. There is not one reactive relationship between the two. Legal reasoning plays a proactive role, anticipating and guiding the appropriate responses that can be made to this innovation. An appreciation of this dimension runs throughout this chapter, culminating in a critical revision of the prominent assumptions identified in the law and technology literature in the introduction of this work.

I. The limits of the common law

Having examined the approaches of English and Californian common law, one can begin to discern some of the striking similarities and differences. The state of affairs after the application of the normative principle has been summarised in Table 1 and Table 2 below. These provide a side-by-side comparison of the battery and negligence mechanisms in the UK and California.

Table 1: Application of the battery action to AI in the UK and California

Jurisdiction	Classification	Elements of the claim						
		Contact	Intention	Hostility or unlawfulness	AI-related information required for valid consent			
UK	Outcome of autonomy-based argumentation	Direct, immediate interference	Intentional touching with a view to use AI	No further requirement	That significant outputs may be generated by the AI without the patient's subsequent ability to intervene			
	Circumstances still excluded	Decisions not to treat and non-physical interactions	Unintentional touching	None	AI use *per se*	AI's risk-related status	Shifts in human-AI expertise	Lesser AI influences
California, U.S.	Outcome of autonomy-based argumentation	Direct and indirect contact	Inferred intention to deviate from patient consent	Unconsented to touching	No specific AI-related information			
	Circumstances still excluded	Decisions not to treat	Patient-imposed conditions on care of which the professional is unaware	Cases also excluded by consent element	AI use *per se*	AI's risk-related status	Shifts in human-AI expertise	AI determinations and lesser AI influences

Key of autonomy-based argumentation	
Strong argument for the stated position	No reasonably available argument for a change of the stated position
Available argument of uncertain strength for the stated position	Corollary to an argued position

Table 2: Application of the negligence action to AI in the UK and California

Jurisdiction	Classification	Elements of the claim					
		Damage	Duty	AI-related information required to avoid breach			Causation
UK	**Outcome of autonomy-based argumentation**	Physical harm or significant violation of a patient's autonomy	AI use by a medical professional	AI's risk-related status	That the selection of an AI can pre-determine a choice of alternative	That AI has led to a substantial depreciation of expected expertise	Individual's decision seriously affected by disclosure, or they would have delayed or avoided harm
	Circumstances still excluded	Lesser autonomy interferences	AI use only by institution	Lesser forms of AI influence			Individual's decision not seriously affected, or they would not have delayed or avoided harm
California, U.S.	**Outcome of autonomy-based argumentation**	Physical harm	AI use by a primary caregiving professional	AI's risk-related status	That the selection of an AI can pre-determine a choice of alternative	That AI has led to a significant, non-obvious changes to status or expertise	Reasonable patient would have altered their decision after disclosure
	Circumstances still excluded	Autonomy violations *per se*	AI use only by a secondary caregiver and/or institution	Lesser forms of AI influence			Reasonable patient would not have altered their decision after disclosure

Key of autonomy-based argumentation	
Strong argument for the stated position	No reasonably available argument for a change of the stated position
Available argument of uncertain strength for the stated position	Corollary to an argued position

A. Negligence

The point of departure for our comparative assessment are the informational obligations that were seen to be the most extensive and nuanced – and therefore the most relevant to medical ML devices – in both jurisdictions. These are those imposed under the negligence mechanism. After comparing the nature of mandated disclosure in both jurisdictions (under the breach element) and its relation to AI's challenges, we move on to consider the impact of the wider limitations imposed by the nature of these torts. Throughout, the strength and effect of autonomy-based argumentation is assessed.

1. Informational requirements under the breach element

Under the breach element of negligence quite some overlap has been identified between the two systems. This is not surprising given that the analysis was shaped by comparable autonomy principles. Nevertheless, distinctions already emerged regarding the specific norms or tests that had been elaborated by the leading cases of *Cobbs v. Grant* and *Montgomery v Lanarkshire Health Board* for California and the United Kingdom respectively.[1358]

Beginning with the similarity of the two approaches, both tests for disclosure appealed to the figure of the patient as the primary decision-maker. Both courts were highly critical of medical paternalism, which would have been the consequence of a test catering exclusively to professional custom. Instead, their tests were directed at the needs of the patient and asserted the patient's right to assess relevant information by reference to their own values. This placed an onus on the professional to facilitate decision-making, while not overburdening the patient with information. In short, they combined dimensions of decisional and practical autonomy to arrive at an appropriate definition of reasonable disclosure.

But there were also revealing differences. The Californian rhetoric painted a picture of the patient as much more dependent on their professional, less able to find and evaluate evidence for themselves. This arguably led to a standard of care that remained more deferential to the experts and was developed in an incremental fashion, with a view to the certainty that

1358 *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430; *Cobbs v. Grant* (1972) 8 Cal.3d 229.

the clinical profession required. In line with this, Californian common law relied more heavily on objective criteria such as the reasonable person and, to some extent, the reasonable professional.

The UK, by contrast, paired its recognition of the professional's facilitative role with a more independent view of the patient. The latter was able to find, evaluate and integrate evidence into their broader decision making and the advice of experts was subservient to their consumers' choice. This led to a more open elaboration of the test. Certain categories of information and restrictions upon these were highlighted, but disclosure obligations ultimately depended on the needs of the individual patient.

In this manner, both operationalisations of the informed consent standard reinforced the relevance of the procedural account of patient autonomy, albeit with different emphases. At the same time, and for their own distinct reasons, they indicated that autonomy-based arguments had to be combined with forms of analogical reasoning. In California this was necessitated by the express limitation of disclosure obligations by reference to established categories and the focus on incremental development. In the UK it emerged from the indeterminacy of the normative framework. Without more specific guidance in place, there was a wide discretion in the manner in which autonomy-based disclosure arguments could be framed, but their strength remained uncertain. To establish whether the common law would require a professional to advise a patient of AI-related characteristics the most forceful arguments appealed to both patient autonomy and concrete applications.[1359]

i. Risk-relevant characteristics

A coalescence of principle and analogy was evident in relation to the risk-relevant status of medical AI. The disclosure of specific risks was seen to be the type of disclosure that enjoyed the most support by analogical reasoning in both jurisdictions – and to some extent it could be applied to

1359 As Green and Sales have pointed out the arguments from analogy and of principle are always intertwined: 'a judge may resort to reasoning by principle and/or analogy. That exercise is reciprocal. Reasoning by principle will allow the judge consistently to apply the law across analogous situations; reasoning by analogy will assist the judge in more accurately identifying and articulating the underlying principle': Green and Sales, 'Law, Technology and the Common Law Method in the United Kingdom' [2023](5) Europäische Zeitschrift für Wirtschaftsrecht p. 205, 211.

the novel technology. However, this was not extensive enough to adequately facilitate a patient's assessment of AI's relationship to risk. In other words, the best-supported argument by analogy did not offer adequate protection to the underlying principle of patient autonomy.

This is where the common law's adaptability and flexibility becomes relevant. Information concerning the innovative or unlicensed nature of procedures was not directly covered by the established disclosure norms of either jurisdiction. Nevertheless, with a view to the patient's need, there are lines of case law mandating such disclosure.

The strongest support for this argument was identified in California. Multiple cases on innovative or experimental procedures had affirmed the obligation of the professional to disclose these general characteristics. In the course of this, the court referenced a number of factors also found to be relevant under our conception of patient autonomy, such as: a lack of evidence regarding a device, its unapproved regulatory status, the availability of alternatives, and the plaintiff's preference for a conservative approach.[1360] In sum, by reference to existing case law and particularly through an implicit reliance on the autonomy principle, general risk-related characteristics have been held disclosable in California.

This provides a strong basis for the argument that AI's comparable status must likewise be disclosed. The argument by analogy may still be imperfect, in virtue of the different factual circumstances under consideration, but the argument from principle is made all the stronger by the existence of this line of cases and their reasonings' relation to the outlined principle.

In England the situation is somewhat different. There is no discernible practice of the courts extending risk disclosure obligations to generalised features, such as the innovative nature of the procedure. There is at least one case that has apparently been decided on this basis.[1361] Yet the extent to which it employed reasoning related to the autonomy principle was limited, referring primarily to the uncertainty of a procedure and its lack of validation.

To supplement this shortcoming weaker forms of argumentations were introduced. Statements from other cases could be drawn upon that were

1360 *Daum v. SpineCare Medical Group, Inc.* (1997) 52 Cal.App.4th 1285.

1361 *Mills v Oxford University Hospitals NHS Trust*, [2019] EWHC 936 (QB), (2019) 170 BMLR 100.

either *obiter*,[1362] or remained connected to a narrower form of risk analysis.[1363] Arguments advanced in the UK academic literature on this issue, which strongly supported the disclosure of risk-relevant characteristics, were also adduced. These factors provide indirect support for our line of argumentation and indicate that further extensions to AI characteristics will be arguable.

In summation, in both jurisdictions the general characteristics of AI, capable of impacting a patient's risk assessment, will have to be disclosed when assessed in light of the autonomy principle. In California, there is a forceful argument to this effect. In the UK there is an available argument of uncertain strength.

ii. Goal-directed action by AI

Another category of disclosure that was supported by a combined analysis of existing case law and principle related to the patient's control over choices in their care. In both jurisdictions the disclosure of alternative procedures has been firmly rooted in the need to connect a certain care pathway to the patient's unique values and preferences. This represented a clear and consistent appeal to the reflective dimension of autonomy, which was stated particularly powerfully in *Montgomery* in the UK and, to a lesser extent, in a series of Californian cases.[1364] The evident relevance of this principle provides a strong argument for the disclosure of the danger that an AI device pre-determines certain choices of the patient, by for example making a significant, serious diagnosis. This is reinforced by an analogy to well-established case law that holds it impermissible for a human expert to pre-determine a patient's choice by positively failing to provide them with information.

A possible limitation that must be conceded in both jurisdictions, is that a patient must have an available or reasonable alternative before a duty to advise arises. This was found to exclude at least some cases where the goal-directed action of AI had the potential to disconnect the patient from

1362 *Jones v Taunton and Somerset NHS Foundation Trust* [2019] EWHC 1408 (QB), [2019] 6 WLUK 193.

1363 *Webster v Burton Hospitals NHS Foundation Trust* [2017] EWCA Civ 62, (2017) 154 BMLR 129.

1364 The examples of *Mathis v. Morrissey* (1992) 11 Cal.App.4th 332 and *Wilson v. Merritt* (2006) 142 Cal.App.4th 1125 were given.

the pursuit of their individualised goals. Nonetheless, as it has been argued that a process of autonomous decision-making will only be disturbed if an identifiable decision is undermined, this requirement arguably represents an application of the autonomy principle, rather than a deviation from it. To the extent that the case-law tracks instances of 'identifiable' or 'real' decision-making accurately, and imposes a corresponding obligation to disclose the pre-determination of a relevant choice, there is no basis for censure.

iii. Human-AI expertise

Regarding shifts in human-AI expertise, the analogies that can be drawn in either jurisdiction remain relatively limited, although arguments from principle can be constructed on their basis. One possible interpretation of the English case of *Jones v Royal Devon and Exeter NHS Foundation Trust* was that: substituting the level expertise that the patient expected to be used in their procedure, without proper disclosure, violated their informed consent.[1365] This could be supported by limited *obiter dicta* in other cases and by arguments from principle, guiding the application of the higher-level *Montgomery* test. However, the scope of this argument is potentially quite limited. Disclosure may only be mandated where, as in *Jones*, the patient demonstrated a strong belief in an extraordinarily high level of expertise or, at least, where the shift in expertise was so substantial to be considered material information for the patient.

Similarly, in California the most relevant authority, *Davis v. Physician Assistant Bd*, stands relatively isolated in virtue of the fact that it assessed a disciplinary action and that it emphasised the difference in the professional's status, rather than their expertise *per se*.[1366] As such it did not constitute a direct precedent for the claim that a breach of informed consent followed from a professional's lack of expertise. Such relevance could only be inferred. At a broader level, there were conflicting indicators as to the relevance of a clinician's expertise to a patient's informed consent. It was possible to reference more targeted informed consent actions, concerned

1365 *Jones v Royal Devon and Exeter NHS Foundation Trust* [2015] Lexis Citation 3571.
1366 *Davis v. Physician Assistant Bd.* (2021) 66 Cal.App.5th 227.

with the risks of not seeking more expert advice for a specific procedure.[1367] A reconciliation of these separate strands with *Davis* reinforced the principled argument that such information must be disclosed to facilitate a patient's autonomous decision making.

In sum, case law in both jurisdictions touched on the problem that was highlighted by AI's ability to substitute human expertise. This can support the principled argument that a failure to disclose shifts in expertise, especially substantial shifts towards the application of technological expertise, can amount a breach of the patient's informed consent. However, the existent cases do not provide clear analogical support for this proposition – still less do they resolve the issue of whether a relative lack of expertise of a human specialist *vis-à-vis* an AI would be sufficient to trigger a disclosure obligation. Rather, it may be the case that an alteration in professional status (in California) or the expectation of a particularly high level of expertise (in the UK) is required. This is why there is an available autonomy-based argument of uncertain strength, encompassing at least a subset of cases where the use of AI devices alters human capabilities.

iv. Informational manipulation

The type of disclosure that was the most challenging to subsume under established legal categories, related to AI's ability to manipulate a patient's decision making. This was deemed to be an issue for procedural autonomy in so far as such manipulation did not engage the patient's own reflective reasoning about their therapeutic goals. In other words, where the nudging possessed a surreptitious non-therapeutic objective.

In this respect it is interesting that the Californian Supreme Court case of *Moore v. Regents of the University of California* constructed a disclosure category aimed precisely at situations where human physicians' decision-making had been potentially subject to these same kinds of influences.[1368] An argument by analogy would theoretically support disclosure of AI biases that may affect professional and patient reasoning. However, this was not capable of providing practicable recourse to a patient who was unlikely to be able to prove that the AI had manipulated them in favour of certain non-therapeutic objectives. Nor would it align with negligence's focus on

1367 *Moore v. Preventive Medicine Medical Group, Inc.* (1986) 178 Cal.App.3d 728; *Scalere v. Stenson* (1989) 211 Cal.App.3d 1446.

1368 *Moore v. Regents of University of California* (1990) 51 Cal.3d 120.

the reasonable defendant, whereby the advising professional would have known, or ought to have known, of such objectives. Here the autonomy principle was ultimately of little assistance.

In the UK these practical hurdles would no doubt be present as well. On top of this, its negligence case law had not yet identified the non-disclosure of comparable influences to be breaches of a professional's duty to obtain informed consent. It had only recognised, at a more abstract level, that external pressures of a sufficient severity could undermine the patient's ability to give their (informed) consent.[1369]

In consequence, there is no purchase for principled, autonomy-based argumentation to establish an AI's potential nudges – and the induced biases – as relevant information that must be disclosed. This is a gap that is left even by the present, forward-looking interpretation of the most autonomy-receptive element of the negligence action.

2. Non-informational requirements

The analysis of the breach element has revealed several differences between the two jurisdictions, not least in the strength of available forms of argumentation for the disclosure of different categories of AI-related characteristics. At the same time, a considerable overlap could be identified between them. This can be explained by the close association between rule-based and principled argumentation – whereby the test for the breach of disclosure obligations is receptive to considerations of autonomy – and by the specification of procedural autonomy principles in both jurisdictions.

Things are very different regarding the non-information-related elements of the negligence action: damage, duty and causation. The relevance of the autonomy principle to these elements is less readily constructed, but they serve as important checks on the success of any claim and thus the protection afforded by obligations to disclose information.

The Californian courts have exhibited a particularly strict adherence to rule-specific limitations. The relevant formulation of the damage element continues to require the eventuation of physical harm to the patient, the group of individuals who may have an obligation to obtain a patient's informed consent remains restrictive – primary caregiving professionals – and the causation test focuses on what the reasonable patient, rather

1369 *Thefaut v Johnston* [2017] EWHC 497 (QB), [2017] 3 WLUK 328.

than the individual patient with their specific values, would have done. Stated succinctly, the common law has here been reticent to sacrifice coherence and doctrinal stability in favour of flexibility and adaptation to the demands of the autonomy principle. The scenarios involving AI that have, as a result, been excluded from the imposition of informed consent requirements can be seen in Table 2.

The UK courts have taken a very different approach. Decided cases and the autonomy principle indicate that disclosure obligations concerning AI extend to the patient's entire care team. Beyond this, the significance attached to the realisation of the autonomy principle has, at the very least, generated considerable uncertainty around the nature of the damage and causation elements of the negligence action.

Arguments can reasonably be constructed that the English courts: (1) have recognised a sufficiently serious autonomy violation *per se* as a form of damage and/or (2) have found that if there is a breach of obligation relating to the patient's autonomy, then the orthodox but-for analysis under the causation element (linking breach and damage) is relaxed. Arguably, the patient need now only prove that their autonomy was sufficiently impaired, or that breaches of informational obligations would have delayed or changed a decision that led to the eventuation of physical harm. Moreover, it is beyond doubt that *Montgomery* has led to a causation test that is protective of an individual's reflective autonomy: i.e. their long-held, defensible preferences.

In the UK therefore, the common law has shown itself to be adaptable to shifts in wider society, favouring the stronger protection of the autonomy principle. Potentially this means that a much smaller class of AI uses, which are likewise outlined in Table 2, fall outside the purview of the discussed disclosure obligations. Yet this adaptation has, thus far, come at the cost of considerable uncertainty and incoherence. Although arguments in favour of stronger, more comprehensive autonomy protections are available by reference to the highest legal authority, it appears that they do not carry much force with the lower courts, who are still concerned to maintain doctrinal stability.

Finally, it is intriguing that one remaining non-informational shortcoming in both common law systems, the lack of stringent informational obligations on institutional users of AI, has been partially addressed by a specific statutory intervention in the UK. Namely, through the *UK General Data Protection Regulation* (UK GDPR) and *Data Protection Act* (DPA) 2018. But it was noted that, beyond the disclosure of AI use *per se*, the demands of

this statutory regulation are currently still unclear and in need of principled interpretation.

B. Battery

Our analysis of the battery action presents us, somewhat surprisingly, with a rather different picture than the one outlined above. Both in terms of the informational requirements imposed on valid consent and in terms of the impact of the autonomy principle on non-consent requirements.

1. Informational requirements for valid consent

The informational disclosure obligations that were so receptive to the autonomy principle in negligence, were much less so in battery. Rule-specific factors – encompassing the need to do justice to medical professionals in both jurisdictions and the limitation on damages for negligence, but not battery, actions in California – mandated a more limited obligation to disclose information.

In the UK this more limited duty meant that shifts in human-AI expertise did not have to be disclosed, in spite of potential analogies that could be drawn to established case law. The only feasible argument that could be advanced in favour of disclosure related to certain basic categories of information concerning the nature and broad purpose of a procedure. This was argued to correspond to decisionally necessary true beliefs, which demanded strong protection under the procedural account of autonomy, and to include the AI's potential pre-determination of a significant choice. To this extent, some protection could be afforded to autonomy. Still, even this argument was of a highly uncertain strength given the lack of clarity in this area of the law.

In California by contrast, the narrow delineation of relevant categories precluded any forms of disclosure that could help address AI's more nuanced autonomy challenges. For example, disclosure regarding shifts of expertise. In addition, a narrow focus on the physical nature of a procedure precluded the requirement of valid consent from providing protection for more fundamental, decisionally necessary beliefs. This included beliefs concerning the goal-directed nature of AI functioning.

2. Non-informational requirements

As in negligence, battery's non-informational requirements have the potential to limit the scope of the action severely. In the UK this is exemplified by the condition that there must be a direct interference with the patient's body and, in addition, that there must have been some intention to use an AI with an undisclosed, significant objective before such a direct interference takes place. The courts have given no indication that they would be prepared to abrogate these elements in light of considerations of patient autonomy. Quite the opposite.

By contrast, the rule-specific non-informational considerations in California have been framed fundamentally by reference to the autonomy principle. Violation of patient consent has established the unlawful nature of a procedure, it has allowed the courts to infer the requisite intent on behalf of the professional and it has provided the basis for an argument that indirect contact, crucially including the prescription and administration of medication, falls under the scope of battery.

Given the contrast with the approach taken by the two jurisdictions under negligence, these findings appear incongruous. It is possible that they exhibit a certain trade-off within the battery norm itself. If the common law opts to frame the consent requirement less restrictively, as in the UK, then other limiting factors must correspondingly be emphasised to keep it within acceptable doctrinal limitations. If the consent requirement is framed highly restrictively, as in California, then one can afford to relax other requirements, which depend on the patient first passing the hurdle of invalid consent. In the final analysis therefore, the adaptability of this tort appears to be kept within relatively tight bounds by both common law systems.

C. Conclusion

The above arguments have highlighted some of the strengths, but also some of the limitations of common law tort systems' ability to respond to novel factual circumstances and technological innovations. Flexible, adaptable categories of disclosure have been framed under the negligence claim in both jurisdictions. Indeed, these could be argued to cover a meaningful proportion of AI's novel autonomy challenges – that is, interpreted appropriately in light of the underlying autonomy principle. Nonetheless, the dif-

ferences between the Californian and English approaches have illustrated the issues associated with such flexibility.

If appeals to principle become too overt and untethered from more granular norms and guidance, then there is a real danger that effective protection is jeopardised by the resulting uncertainty. After a change in the standard of care, English law is in a state of flux and development and it has thus far failed to provide such norms. California, by contrast, has been more careful to operate by reference to incremental evolution and objective criteria. Arguably, this strikes a more appropriate balance between coherence and legal certainty on the one side and the protection of the autonomy principle on the other.

A similar argument may be transposed to the way in which both jurisdictions have handled the non-informational requirements of the negligence action. In several regards the UK has been prepared to infuse many elements of the action, which are designed to generate stability, with unorthodox considerations of patient autonomy. Californian law has refused to take this step. But again, one must make a context-sensitive analysis. The ability to protect patient autonomy in light of AI's more nuanced challenges is fundamentally dependent upon a loosening of these broader doctrinal requirements and striking the right balance arguably requires the development of different kinds of doctrinal limitations, ones which are receptive to principle. For example, through defining actionable damage by reference to a sufficiently restrictive, operationalised conception of patient autonomy – as the procedural account has been said to provide. This would require a proactive court to recognise that fundamental shifts in the nature of the negligence action have already taken place.[1370]

An alternative, as some authors have proposed, may be the creation of a *sui generis* tort with distinct elements.[1371] This would have the advantage of maintaining the structure of negligence, while also offering the principled

1370 As Green and Sales remark, the ability to make these kinds of fundamental decisions may be the preserve of the Supreme Court: Green and Sales, 'Law, Technology and the Common Law Method in the United Kingdom' [2023](5) Europäische Zeitschrift für Wirtschaftsrecht p. 205, 209. Recall also that Leggatt LJ noted that the Supreme Court may need to reconsider the area of damage and its relation to patient autonomy: *Duce v Worcestershire Acute Hospitals NHS Trust* [2018] EWCA Civ 1307, (2018) 164 BMLR 1 [92].

1371 Nolan, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European Tort Law p. 259, 376-382; Mulligan, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies p. 55, 71-74.

protection that modern society appears to be demanding. However, currently the courts have given no indication that they would be prepared to take this step.

In the final analysis it must therefore be admitted that neither the Californian nor the English common law system currently strike the outlined balance. The former system emphasises rigidity and objectivity at the expense of the protection of principle. The latter seeks to protect the principle, but has thus far failed to convert this into sufficiently clear and specific norms.

As a result, the common law on informed consent is limited in its ability to respond to the novel, nuanced types of autonomy challenges posed by medical AI. At most the identified disclosure obligations under negligence's breach element could serve a signalling function, falling short of offering substantive protection. To be certain to avoid legal sanctions, medical professionals ought to be careful to address the following AI characteristics (where applicable to the specific device): their risk-relevant status, the substantial, non-obvious shifts they may cause in human-to-AI expertise and the ability of AI to pre-determine certain significant objectives in their relatively independent functioning. If a patient is not informed of these, however, they will probably lack recourse to the courts, unless they incidentally happen to fulfil the outlined additional requirements.

II. Guidance for a bespoke statutory scheme

In this section it is argued that lawmakers and regulators and, more generally actors who design modes of regulation complementing or substituting the common law, can derive useful guidance from the legal operation examined throughout this work. To be exact, it is considered how an approach building on this – a supplementary statutory scheme – could be designed to complement the identified weaknesses of the common law. By conceptualising a desirable legal solution to the autonomy challenges of clinical AI/ML, this sets the scene for the final, underlying analysis of how legal reasoning interacts with the phenomenon of technological innovation.

Of course, by recommending a legislative intervention for a subject that traditionally falls within the common law's purview, the present section taps into a much deeper discussion. Namely, the appropriate roles played by courts and legislatures in a modern society and – as addressed in Chapter

1 – the relative advantages and disadvantages of accommodating social and technological progress through either modality.[1372]

While the preceding argumentation has served to explore primarily the adaptability of the common law, as well as the normative legitimacy of its structure, this section seeks to strike a middle path that capitalises on the strengths of both sources of law. By advocating for a legislative scheme that draws on the common law mode of reasoning, is supplementary to its informed consent framework and contains mechanisms for adaptation, it is hoped that the advantages of the developed approach can be broadly maintained. Simultaneously, the statutory form is not subject to the identified defects of the current system of judge-made law. It offers improvements in terms of specificity, clarity and in terms of monitoring compliance and enforcement. It is these advantages that this section additionally seeks to exploit, proposing a scheme that guarantees an effective protection of patient autonomy in relation to clinical ML devices.

Existing legislative structures will be considered in order to make this case and in order to determine the nature that such a scheme could take in the specific contexts of California and the UK. While legislators are much less constrained than judges by settled legal positions – and are technically at liberty to tailor solutions to the instrumental demands of innovation – it will be seen that they have nevertheless tended to institute incremental changes in this area. In effect, they are also liable to be directed by the existing legal material, both by the prevailing common law position and by the form of existent statutory and regulatory interventions.

In consequence, to inform the recommendations for a bespoke statutory scheme targeted at the resolution of AI's particular autonomy problems, we undertake two analyses in this section. First, we examine the nature of existing legislative informed consent schemes in the two jurisdictions. Second, we determine whether their underlying rationale can be transferred to the regulation of medical ML devices and, if so, what shape an adequate, bespoke solution would take. At this stage one can then draw on the results of the application of our legal mode of reasoning in the common law.

1372 For fundamentally different views on this matter, compare: Friedman, *Law and Society: An Introduction* (1977) 164-165; Calabresi, *A Common Law for the Age of Statutes* (1985) 2.

A. Existing consent statutes

As referred to in Chapter 1, both of the selected jurisdictions have imposed certain statutory regimes for the disclosure of information. For us, the most interesting precedent relates to mandated disclosure of certain pieces of information that may be additional to, or understood to reinforce, the kinds of disclosure mandated at common law.

1. United Kingdom

In the United Kingdom three instances can be cited in this respect. The statutory requirements for: clinical trials of investigational medicinal products; the storage and use of embryos and gametes, and; the removal, storage and use of human tissues.

The first example is provided by the consent requirements for clinical trials of investigational medicinal products. Under the *Medicines for Human Use (Clinical Trials) Regulations* 2004 an individual gives informed consent to participate in a trial only where they are 'informed of the nature, significance, implications and risks of the trial'.[1373] This consent must be evidenced or recorded in writing.[1374] In addition, an investigator must arrange an interview with the subject where they have the 'opportunity to understand the objectives, risks and inconveniences of the trial and the conditions under which it is to be conducted',[1375] they must be provided with a point of contact for the purposes of obtaining further information,[1376] and they are to be informed of their right to withdraw from the trial at any time.[1377]

Overall, in terms of the substantive categories of information that must be provided, the legislation largely replicates the common law. But in so doing it clearly demarcates and emphasises aspects that are more tangential to established forms of disclosure, such as: the objectives of the trial, its conditions and the right to withdraw. Furthermore, it establishes various formal or procedural mechanisms that frame the disclosure: it must be in writing, there must be an interview with the subject and there must be a point of contact for further discussion.

1373 Medicines for Human Use (Clinical Trials) Regulations 2004 schedule 1, part 1, paragraph 3(1)(a).
1374 ibid schedule 1, part 1, paragraph 3(1)(b).
1375 ibid schedule 1, part 3, paragraph 1.
1376 ibid schedule 1, part 3, paragraph 5.
1377 ibid schedule 1, part 3, paragraph 2.

The *Human Fertilisation and Embryology Act* 1990 can be considered apart from this. It requires that, for the purposes of the licence that is a necessary prerequisite for the 'donation, storage and use of gametes and embryos',[1378] certain consent conditions must be complied with.[1379] In particular, a consent to the use of an embryo must specify one of several possible purposes, in terms of treatment, training or research, and a consent to storage must specify the maximum period of storage and what is to be done with the gametes, embryo or admixed embryo if the person is unable to vary or withdraw consent.[1380] The licensing authority may specify further consent requirements in its directions.[1381]

More fundamentally, alongside such specific categories of information there are also general requirements regarding the fact that an individual must receive 'such relevant information as is proper' and that they receive counselling regarding the implications of their decision.[1382] Consent must assume a certain form. Normally this means that it must be in writing and signed.[1383]

In consequence, while there is some overlap with the common law requirements, indicated by references to 'proper' or 'informed' consent, one can see that there is considerable detailed, specific guidance that relates to the innovative nature of the subject matter. Moreover, the statute puts in place mechanisms regarding: the development of more elaborate guidance, the form of consent, and the need for counselling.

The last statute to be examined is the *Human Tissue Act* 2004. This is a wide-ranging statute that primarily governs the consent necessary from relatives for the removal, storage and use of human tissue from deceased persons.[1384] By contrast to the first two schemes, this largely leaves the consent requirements for the removal of material from living persons to

1378 *Jennings v Human Fertilisation and Embryology Authority* [2022] EWHC 1619 (Fam), (2023) 189 BMLR 17 [24].

1379 Human Fertilisation and Embryology Act 1990 (as amended) section 12(1)(c) and schedule 3.

1380 ibid schedule 3, paragraph 2(1)-(2). It also requires them to be informed about the conditions for varying or withdrawing consent: ibid schedule 3, paragraph 4-4A.

1381 ibid schedule 3, paragraph 2(3). Currently see: Human Fertilisation and Embryology Authority, 'Code of Practice' (2021 Ninth Edition) <https://portal.hfea.gov.uk/knowledge-base/read-the-code-of-practice/> accessed 26.3.2023.

1382 Human Fertilisation and Embryology Act 1990 (as amended) section 13(6)-(6A), and schedule 3, paragraph 3(1).

1383 ibid schedule 3, paragraph 1(1).

1384 Human Tissue Act 2004 schedule 1.

the regulation of the common law.[1385] Substantively, it does not elaborate upon vague notions of 'appropriate consent'[1386] or 'qualifying consent'.[1387] However, here too a procedure was created whereby an institution (the Human Tissue Authority) must provide more specific guidance.[1388]

2. California

Beginning with the common ground that can be found in our U.S. case study, one can point to California's statutory regime covering consent to medical experimentation.[1389] This posits stringent disclosure obligations, mandating a research subject to be provided with the experimental subject's bill of rights and information on: 'the nature and purpose of the experiment'; 'an explanation of the procedures to be followed in the medical experiment, and any drug or device to be utilized'; discomforts and risks; benefits to be expected (if applicable); alternative drugs, devices or procedures, and; available forms of medical treatments in case of complications.[1390]

In addition, a subject must also be given the opportunity to ask questions, they must be advised that they can withdraw from the treatment without prejudice, they must be handed a copy of the written consent form and they must make their decision free from 'any element of force, fraud, deceit, duress, coercion, or undue influence'.[1391] In this manner one can see that, alongside aspects of information disclosure, formal safeguards (the written nature of consent and specific rights to ask questions) have been put in place.

More generally, the Californian legislature has been much more proactive than its transatlantic counterpart in its complementation and substitution of common law informed consent requirements. The state imposes

1385 ibid section 45 and schedule 4.
1386 ibid section 3. For an analysis see: McHale, ''Appropriate Consent' and the Use of Human Material for Research Purposes: The Competent Adult' (2006) 1(4) Clinical Ethics p. 195.
1387 Human Tissue Act 2004, schedule 4, part 1.
1388 ibid section 26.
1389 California Health and Safety Code sections 24170ff (the 'Protection of Human Subjects in Medical Experimentation Act').
1390 ibid section 24173 and section 24172, subdivisions (a)-(f).
1391 ibid section 24172, subdivisions (g)-(j).

statutory requirements regarding: the treatment of breast cancer,[1392] the performance of hysterectomies,[1393] sperm or ova removal,[1394] and cosmetic implants.[1395] It is beyond the scope of this work to conduct a detailed analysis of these numerous provisions, but certain shared characteristics can be identified.

In this respect, it is notable that several common law classes of disclosure are replicated in the outlined provisions in various combinations, such as the nature of the procedure, risks, advantages and alternative treatment methods.[1396] At the same time, more specific pieces of information are also required. For example, in the case of breast cancer treatment it includes: available treatment options that are at the clinical trials stage, available telephone numbers for organisations that can provide information to the patient, a discussion of breast reconstruction surgery and statistics on the incidence of breast cancer.[1397] At least one of these categories – the disclosure of statistical information –clearly overrides the default common law position which was laid down by *Arato v. Avedon*, as examined in Chapter 7.[1398]

Statute has also made provision for various kinds of formalities and procedural mechanisms. This includes the development of written summaries by departments, as well as their provision by professionals,[1399] and the necessity of written consent.[1400] Regarding the procedural operationalisation of consent requirements, there are then also mechanisms for the recurrent revision of such information summaries,[1401] and for the imposition of automatic civil penalties for repeated failures of providers to obtain informed consent.[1402] This serves to illustrate the broader institutional instruments available to the legislature in implementing and enforcing its mandates.

1392 ibid section 109275.
1393 ibid sections 1690-1691.
1394 California Business and Professions Code section 2260.
1395 ibid section 2259.
1396 California Health and Safety Code section 109275, subdivision (c); ibid section 1690, subdivision (a); California Business and Professions Code section 2259, subdivision (e).
1397 California Health and Safety Code section 109275, subdivision (c)(3).
1398 *Arato v. Avedon* (1993) 5 Cal.4th 1172.
1399 California Health and Safety Code section 109275, subdivisions (a), (c).
1400 ibid section 1690, subdivision (b).
1401 ibid section 109275, subdivision (c)(2).
1402 California Business and Professions Code section 2260, subdivision (e).

3. Rationale

To examine the intent underlying these provisions, the rationale for taking certain issues of clinical consent outside of the common law, one may look towards: any discernible legislative intent, the context for the provisions' passing and the nature of the legislation itself.

California does not generally record the hearings or reports underlying its legislative process. However, in so far as motivations can be deduced from the context, there are indicators that the common law was deemed to have been operating unsatisfactorily. For example, referring *inter alia* to California and the outlined regulation of consent to breast cancer treatment, Pope has noted: 'in the late 1970s and early 1980s, physicians were not disclosing less invasive treatment options to their breast cancer patients. In response, 14 states enacted statutes that require physicians to present the advantages, disadvantages, and risks of all medically viable alternative therapies'.[1403] In other words, the common law was deemed insufficient to ensure that the medical profession disclosed specific kinds of alternatives.

In spite of the UK's generally greater transparency regarding legislative intent, the relationship to the existing common law was only meaningfully addressed in the passing of the *Human Tissue Act* 2004. A public scandal, regarding the non-consented-to removal of tissue – primarily from deceased children – had led to the realisation that there was a *lacuna* in the pre-existing regime, governed partly by common law and partly by statute.[1404] Referring to this scandal in its Explanatory Note, the Act explicitly stated that 'the current law in this area was not comprehensive, nor as clear and consistent as it might be for professionals or for the families involved'.[1405] In response, it sought to establish the fundamental significance of consent in this area.

The other examples of statutory regulation of informed consent in the UK were not clearly based on the unsatisfactoriness of the *status quo*. The *Medicines for Human Use (Clinical Trials) Regulations* 2004 were not enacted *as* primary legislation, nor were they enacted *under* a targeted,

1403 Pope, 'Certified Patient Decision Aids: Solving Persistent Problems with Informed Consent Law' (2017) 45(1) Journal of Law, Medicine & Ethics p. 12, 17-18.

1404 Jones, *Medical Negligence* (Sixth Edition 2021) para 7-174; McHale, ''Appropriate Consent' and the Use of Human Material for Research Purposes' (2006) 1(4) Clinical Ethics p. 195, 195.

1405 Explanatory Notes to the Human Tissue Act 2004 paragraph 5.

enabling Act of Parliament,[1406] Rather, they were passed as a regulation implementing an EU directive and detailed discussions on informed consent remained limited. The focus was on the desirability of such reforms in general. Likewise, while the discussions underlying the *Human Fertilisation and Embryology Act* 1990 were much more extensive, they focussed on the more controversial aspects of the Bill. In so far as consent was discussed in relation to these two instruments, it was primarily seen as an added, legitimacy-conferring factor for a novel – otherwise morally and politically suspect – scheme.[1407]

On the basis of these examples, one is left with two motivating factors. First, confirming the general phenomenon stated at the outset of this section: identifying a deficiency in the operation and application of the common law regarding consent can be an important prerequisite for a statutory intervention. Second, where a novel kind of statutory regime is introduced to regulate a certain clinical circumstance (such as experimentation or new uses of embryos), a legislative framing of consent can serve to bolster the legitimacy of the resulting framework. It is hard to derive a judgment on the adequacy of the common law's operation *per se* in such situations.

However, one can then derive more specific insights from the nature of the statutory provisions themselves. Rather intuitively, these appear to respond primarily to subject matters that are perceived to pose problems of particular social significance. In the informed consent context this includes above all challenges framed in terms of the value of patient autonomy.

For example, it seems that such interventions are thought most appropriate where there is a heightened danger that patient decision making will be subverted by the external imposition of choices upon them. Similarly, it

1406 The enabling Act was the European Communities Act 1972 section 2(2).

1407 In relation to the Medicines for Human Use (Clinical Trials) Regulations 2004, the need for consent was juxtaposed with safeguards for those without capacity to consent: Joint Committee on the Draft Mental Incapacity Bill, 'Draft Mental Incapacity Bill: Session 2002–03 Volume I' (2003) paragraphs 285-289. In relation to the Human Fertilisation and Embryology Bill 1990, it was stated in the House of Commons for example: 'The Bill is also about freedom of conscience. It does not seek to impose sanctions on those who think that research is wrong. It goes to great lengths to define consent. As Lord Hailsham put it in another place, those who would impose an absolute prohibition should ask themselves, what kind of right, in a free and liberal democracy, do they think that they have to say no to a highly responsible group of people working for the benefit of humanity and subject to the authority of Parliament?': House of Commons Debate 2 April 1990, volume 170, columns 953-953.

appears that the regulated circumstances are frequently ones where important interests of the individual are at stake. These kinds of concerns were succinctly stated by the Californian legislation on experimental treatment:

> The Legislature hereby finds and declares that medical experimentation on human is vital for the benefit of mankind, however such experimentation shall be undertaken with due respect to the preciousness of human life and the right of individuals to determine what is done to their own bodies (…) There is, and will continue to be, a growing need for protection for citizens of the state from unauthorized, needless, hazardous, or negligently performed medical experiments on human beings.[1408]

Quite similar statements can be made in relation to the regulation of the use and storage of embryos, of breast cancer treatment, or the performance of hysterectomies.[1409]

The types of specific categories of disclosure that the statutory mechanisms elaborate, and the manner in which these relate to the existing common law, are also telling. Where such categories are additional to established informed consent precedents, or they clearly override them, this indicates an identified under-inclusiveness of the judge-made law. However, where they merely offer a concretisation of the more general common law standards – which appeared to be in the great majority of cases – it suggests that the common law, while applicable, was deemed too uncertain or conflicted.

In this manner the contribution of the creative process of legal reasoning is recognised. The outputs of this process maintain their relevance even

1408 California Health and Safety Code section 24171.

1409 In relation to breast cancer, see: ibid section 109250: 'Despite intensive campaigns of public education, there is a lack of adequate and accurate information among the public with respect to presently proven methods for the diagnosis, treatment, and cure of cancer. Various persons in this state have represented and continue to represent themselves as possessing medicines, methods, techniques, skills, or devices for the effective diagnosis, treatment, or cure of cancer, whose representations are misleading to the public, with the result that large numbers of the public, relying on the representations, needlessly die of cancer, and substantial amounts of the savings of individuals and families relying on the representations are needlessly wasted. It is, therefore, in the public interest that the public be afforded full and accurate knowledge as to the facilities and methods for the diagnosis, treatment, and cure of cancer available in this state and that to that end there be provided means for testing and investigating the value or lack thereof of alleged cancer remedies, devices, drugs, or compounds, and informing the public of the facts found, and protecting the public from misrepresentation in these matters'.

in relation to (one may presume) much more instrumentally minded legislators who are much less bound by doctrinal limitations. Primarily these outputs are concretised or supplemented so as to allow for a more targeted, effective realisation.

Lastly, it is notable that the examined statutory schemes often outline formal and procedural elements that the common law generally does not, or cannot, lay down. The requirement of written information summaries and consent procedures would be one example. More substantial are the various implementation and enforcement mechanisms that have been discussed: the establishment of regulatory agencies, the issuance of guidance, the need for certain points of contact, automatic civil penalties, etc. The creation and operation of such measures is certainly beyond the capabilities of the common law judge and the adversarial justice system.

B. An informed consent statute for AI

Based on our analysis of AI's challenges, the common law's transformation of these challenges into legally cognisable categories of thought and the legislative solutions already in existence, it will be shown that a supplementary statutory scheme for AI offers one appropriate response to the emergence of this technology.

The starting point here must be the elaborated deficiencies in the common law's ability to meet AI's autonomy challenges. One principal difficulty related not to the fact that disclosure concerning ML devices fell directly outside of established categories, but rather that there was a relationship of uncertain strength between these categories and the specific informational needs generated by AI. Even where strong arguments were identified in principle and by analogy, these remained contestable and subject to general limitations that were, at best, incidental to patient autonomy. A statutory concretisation and 'streamlining' of such requirements would fit well with established precedent and it would draw on the insights of the common law analysis.

Mandating that a patient must be provided with categories of information *via* statute would be a readily available means for resolving the considerable legal uncertainty and the under-inclusiveness of the *status quo*. It should establish, specifically, that the patient be advised of the examined classes of information. This includes: the purpose of the AI use, its general functioning *vis-à-vis* the patient's goals and in relation to scientific know-

ledge, the relation between technology and expertise, its risk-relevant characteristics and, arguably, its potential for influencing decision-making. This is an important autonomy-maintaining baseline. Beyond this, the dynamic, rapidly evolving nature of the technology must be accounted for. More detailed disclosure obligations ought to be subject to mandatory review and revision by a designated authority at regular intervals.

Second, in the AI context the deficiencies of implementing informed consent through the operation of established common law mechanisms has proven a particular area of concern. It has been difficult to target its norms at the right actors, to fulfil the non-consent requirements of its actions, and to ascertain the appropriate amount of compensation that is to be offered in response to violations. At times, particularly regarding the subtle, manipulative influences of ML devices, the disclosure of information *per se* has also appeared to constitute a blunt instrument. The kinds of mechanisms that have been utilised under established legislative schemes would be suited to overcome these hurdles.

Comprehensive informational obligations could be imposed on institutions where they use the technology to engage with patients directly. Compliance with these duties could then be overseen, again, by designated authorities. As one must expect healthcare institutions to be closely involved in the development, selection and deployment of medical AI, this would provide a justifiable basis for the institution-wide regulation of consent requirements. Rather than insisting on the isolation of individual actors, there could be a broader obligation to ensure that healthcare teams using AI have proper consent-taking procedures in place.

Furthermore, one feature of AI is the pervasiveness of its use and its integration with other aspects of clinical care – an aspect that distinguishes it from the established legislated-for areas. It would therefore be of particular significance that related informational obligations are not too demanding. Requiring written consent for the use of ML devices, for example, could jeopardise the integration of the technology into healthcare workflows and the benefits that would accompany this. Practicability in the AI context should mean a sufficiently general summary in circumstances where substantial interferences with autonomy are possible.

As was seen in California and the UK regulation of clinical trials, a suitable means of accommodating these concerns would be mandated discursive engagements with the patient or the provision of wider decision-aids (written summaries). Points of contact where the individual can receive further information, if interested, can also be provided. In these cases,

the patient's decision-making is facilitated and they have the choice to proactively engage with the provided aids.

Conversely, when it comes to the issue of enforcement where information-facilitation procedures have failed, the two jurisdictions have further provided instructive precedents. Given the sometimes minor and potentially frequent infringements of consent requirements through AI deployment, it may be desirable to establish standardised awards, as exist under UK common law or in the statutory regulation of sperm and oval removal in California. As in the latter case, these could then also be targeted specifically at repeat offenders, without depriving patients of their existing civil remedies.

Indeed, as stated in the introduction of this section, it is regarded as desirable that, in so far as is possible, the statutory interventions remain supplementary to the common law, which will provide the broader informed consent framework. Pairing the autonomy principle with the precise norms of informed consent does not offer a perfect response to clinical ML devices in either jurisdiction. Nevertheless, it provides several, potentially adaptable, avenues for argumentation that offer promising starting points. In case of doubt as to the applicability of a specific legislative framework for AI devices, or as to the effect of a bespoke statutory scheme, it should be made clear that the common law remains an applicable, supporting instrument.[1410]

Likewise, the creation of tailored informed consent mechanisms would be aided by embedding them into a wider regulatory framework for clinical ML. This thesis has not purported to deal with many other problems associated with the technology that could inspire such a step, problems relating to medical device regulation, data protection, discrimination and more. However, in so far as AI, especially medical AI, is legislated for regarding one or more of these areas, it would create a real impetus, or even a need, for the legislature to confront the matter of patient consent. Political and legal actors –within, but also far beyond, the UK and the U.S. – have begun to engage with such reforms.[1411] The UK GDPR's explanation requirements

1410 Moses in Marchant, Allenby and Herkert, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (2011) 84-85.

1411 One broader example is provided by the recent statements of the German Ethics Council, *inter alia* on bias and informed consent: Deutscher Ethikrat, 'Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz' (2023) 159-162. For the UK and the U.S. see Chapter 1, Section II.A.

for automated decision-making provide an early example establishing the desirability of this approach in one of our two jurisdictions.

If the legislature moves to provide a regime for the regulation of medical AI, then the arguments presented here consequently gain added force. In this process it should be ensured that patient's decision making, their fundamental interest in autonomy, is adequately facilitated. The creative contribution of common law reasoning offers valuable guidance for these purposes and it is envisioned that it will continue to perform a supporting role.

III. Legal reasoning and technological innovation

In the introduction of this work three prominent assumptions on the interrelationship between law and technological innovation were outlined and contested. These were: that legal reactions to technological change are to be judged by reference to extra-legal, policy-orientated standards; that technological progress is inevitably racing ahead of a largely inert legal system, and; that extra-legal regulatory modalities offer a better chance of realising pertinent instrumental objectives and keeping up with the regulatory needs of society. In light of the developed principle-based methodology and our assessment of its application to the autonomy problems presented by medical AI/ML devices, we are now at a stage to scrutinise these three assumptions on the basis of this concrete example.

A. Understated challenges of non-legal regulation

One argument that emerged from the literature was that direct regulatory interventions, such as altering the architecture of a program's code, provide an available and often preferable alternative to legal norms for the purposes of accomplishing various innovation-related policy goals. This aspect is evaluated first because it was an assumption that our work grappled with in Part I.

In the course of considering the nature of AI devices, and how they would realistically be integrated into their medical context, it became apparent that the functioning of the technology suffered from various shortcomings. These shortcomings included the black box nature of ML models – in particular a subclass of deep neural networks (DNNs) – and

the flawed nature of performance evaluations for the purposes of gauging the functioning of medical AI through standard clinical indicators.

Assessing whether and how these defects could be remedied through the design of the technology itself was an important prerequisite for understanding how a legal solution would function in relation to its factual subject matter. As such, the availability of non-legal alternatives assumed an orthodox position within the framework of the comparison. It was merely supplemented by a more dynamic, forward-looking perspective to anticipate whether technological solutions would be forthcoming in the near future. This matched the more general emphasis on anticipating the development of a technology that is still emerging and which has yet to be comprehensively implemented.

In taking this approach, it was examined whether it was realistic that the design or operation of medical AI/ML technology would be refashioned so as to avoid the various autonomy-related issues. This involved technological fixes that seemingly presented an immediate solution, as well as secondary legal interventions that would frame these fixes. For example, a regulation that would require developers to carry out, catalogue and respond to more suitable performance evaluations – i.e. ones which were tailored to ML – was considered as one possible solution. Similarly, the development of explainable or interpretable AI models was a direct way in which the technological architecture could be designed in order to accommodate the professionals' and, relatedly also, the patients' need for information.

A question that can now be asked is how such technological fixes are to be judged *vis-à-vis* the specific legal recommendations presented here: a combination of common law evolution and the prospect of a bespoke, supplemental statutory disclosure scheme. One can begin answering this question by engaging with the oft-cited benefits of technological solutions, above all the notion that technological fixes emerge with a rapidity that (better) tracks the speed with which corresponding technological challenges emerge.[1412] In the present scenario it was demonstrated that this was simply not the case.

As was noted in Chapters 2 and 3, considerable uncertainty persists concerning the best way(s) to evaluate ML performance, especially in the protean medical context. It is likely that an appropriate scheme has yet to be designed.

1412 See Chapter 1, Section II.C.3. in this regard.

With respect to explainable and interpretable ML methodologies, this flaw is even more strongly pronounced. These technologies are in a state of development, throwing up a whole host of complications that do not take one very much beyond the original problem. This is especially true for the, normatively more desirable, modality of interpretable AI. A rapid adoption and implementation of interpretable methods by designers of medical ML devices appears far-fetched at this moment.

Yet clinical AI's challenges to autonomy are already manifesting themselves. As a result, it is not inapposite to speak of a 'technological lag' or a 'phasing problem' for those seeking to regulate the challenges of innovation through extra-legal solutions. By contrast, the legal system offers an imperfect but normatively defensible approach that can be implemented now. Appropriately conceived, informed consent mechanisms are broadly applicable and will be relevant to many deployments of clinical ML devices.

A further factor advanced in favour of technological solutions was their direct, unmediated effect, which translated into the successful realisation of desired objectives. Within a certain subclass of cases, for instance where individuals are performing tasks exclusively through the operation of a software system, this aspect may hold true.[1413] Yet, in the complex social relationships within which AI medical devices are embedded – cooperating with clinical professionals to facilitate patient decision-making – a design fix does not, and cannot hope to, have this effect.[1414]

Our normative analysis indicates that it remains an open question how improved performance evaluations and forms of interpretability would have to be utilised to further patient autonomy. Not all information is significant to an individual patient, too much will overwhelm them, and it depends on how the professional chooses to convey what they know.

Consequently, it is not enough for the law to frame the necessary code. To achieve desired, direct effects, behaviour guiding norms remain indispensable. In such situations the technological solution is, at most, the indirect facilitator of the more immediate, legal intervention which sets the appropriate bounds of disclosure. Once again one can identify a more complex interrelationship between law and technology, which departs from the

1413 Recall that a growing cyberspace was the primary context for such early claims: Reidenberg, 'Lex Informatica' (1997) 76(3) Texas Law Review p. 553; Lessig, *Code: Version 2.0* (2006); Berman, *Law and Society Approaches to Cyberspace* (2007).

1414 See: Chapter 2, Section IV. Consider especially the uncertain effects that different forms of architectural design ('nudges') had in this context.

narrative of law as a secondary facilitator of a better-suited technological fix.

This leads me to the final point, which relates to the subject matter of the next section. Even granting that technological solutions possess a rapidity and a direct, comprehensive effect that the law does not, it is questionable whether they are also targeted at the right ends and, crucially, how it is ensured that such a correct targeting is maintained. The immediacy of technological action is worth little, after all, if it does not actually engage with the relevant challenge.

The contrast between explainable and interpretable AI provides a good example of this fundamental point. Developers and commentators evidently understood and shared the concern for the preservation of user autonomy, but they perceived the underlying problem of opacity in several different ways. For many, especially those favouring explainability, it was evidently closely associated with the need to bolster user trust, since a lack of this hampered implementation. That the generated explanation was not truly revealing a model's functioning, or even exhibited manipulative tendencies, was not necessarily considered to be in conflict with the goal of autonomy enhancement. That goal could be stated at a sufficiently abstract level and it could be pursued in an uncontested fashion. This illustrates that those focussed on technology policy still do not provide a means for identifying the right ends in concrete situations. In the next section it is seen how the law does provide such guidance.

B. The significance of law's resistance to instrumentalisation

A further trend that was identified within the law and technology literature treated the legal system as an object that could be instrumentalised, in a discretionary manner and relatively comprehensively, in the pursuit of external ends. Most especially this was in the pursuit of policy objectives that were, rather broadly, associated with a desirable or effective regulation of innovation.[1415] Our completed analysis relativises this assumption by appealing to the way in which the law has interacted with, and thereby transformed, a bioethical principle. Corresponding to our division between Part II and Part III of this work, this transformation had two noteworthy facets.

1415 See Chapter 1, Section II.C.1.

Part II illustrated how the law is able to interact with a goal that is external to it and which may be normatively contestable. Yet, to operationalise this in legal reasoning, and to gauge how the legal system would react to its imperatives, one had to do much more than just take over, or put into effect, an external value. That value had to be aligned with the structure and content of the existing system and it had to be transformed by reference to this.

More precisely, our evaluation rejected – by focussing on the function of a legal concept – argumentation that appealed directly to external values. Such an appeal did not account for the structured manner in which legal norms interacted with each other and how they were subject to contestation and revision (the topic of the next section). Nor did it take seriously the contextual limitations that an appeal to any individual objective must be subjected to, whether these stemmed from other principles, rules or consequentialist modes of reasoning.

Only once one had transformed the extra-legal concept into a legal one, could these structural components be satisfactorily captured. In the present case, it was only by transforming the value of autonomy into a legal principle that its operation within the law could be properly accounted for. Any external policymaker seeking to utilise the legal system for their instrumental ends would do well to bear in mind this structural component and the broader, multi-faceted way in which the legal system reacts to the introduction of external norms. A favoured objective will have to be reconciled with the existing commitments that will continue to exert their force to achieve a certain balance of interests.

Admittedly, this aspect is most pertinent to an analysis focussing on the common law mode of reasoning. Nonetheless, it relates to a lesser extent also to other modalities, such as statutory interpretation. In our jurisdictions this application was explored in relation to UK GDPR. Hereby it was evident that the autonomy principle could be invoked to concretise the demands of several open-textured requirements. A similar insight also accompanied the examination of existent informed consent schemes in both California and the UK. One could deduce common autonomy-related rationales and a propensity for incremental development based partly on the existing common law.

Where the law is receptive to a given conception of a value that is to be advanced, such as procedural autonomy, it has been contended that the scope within which that value is utilised is a further important factor. It is a significant consideration that the legal system, while striving to achieve

coherence at a general level, additionally identifies areas or subject matters in which special concerns can be accommodated. This fosters coherence in a distinct field. In Part II it was apparent, especially in the UK, but to a lesser extent also in the American legal system, that patient autonomy was fashioned according to specific considerations that did not apply, for instance, to the autonomy concept at play in disputes of property or contract law. Medical law provided its own reference points, allowing for a more targeted, richer analysis.

On the instrumentalist view this compartmentalisation is an anachronism, hindering the development of a coherent approach to technology regulation.[1416] But again, if these theorists and associated policymakers seek to insert their favoured interests into existing legal frameworks, it is surely indispensable that they account for these fields – whether or not they then seek to move beyond them.

In addition, it is maintained here that, without regarding these structures, one impoverishes the quality of the overall normative analysis. Arguably, one reason why bioethics and law have enjoyed a productive interrelationship, above all regarding the protection of patient autonomy, is that bioethicists have paid close attention to the specific legal mechanisms that have evolved to shape the realisation of this value.[1417] Conversely, it has also been important that the health law field, in building a coherent set of norms, sought to adapt and transform bioethical concerns into legally cognisable forms.[1418] In the course of this interaction a host of interests have been concretised that allow for, perhaps imperfect, but nevertheless significant forms of regulation.

This leads one to the final aspect that was exhibited, above all, in Part III. Considering legal reasoning to be transparently subject to external objectives would not only posit an overly malleable picture of the law – one which is likely to lead to misplaced, ineffective regulatory efforts. More than this, it would distort and jeopardise the substantial analytical resources provided by the outlined operation.

1416 Guihot, 'Coherence in Technology Law' (2019) 11(2) Law, Innovation and Technology p. 311, 319-321. See also the critical perspective of: Sommer, 'Against Cyberlaw' (2000) 15(3) Berkeley Technology Law Journal p. 1145, 1150-1151.

1417 Faden, King and Beauchamp, *A History and Theory of Informed Consent* (1986) 114-150.

1418 Wall in Phillips, Campos and Herring, *Philosophical Foundations of Medical Law* (2019).

Integrating a value such as patient autonomy into the law, examining its functioning and anticipating its impact offer a unique opportunity to develop a rich understanding of that value. It will require the identification of contested objectives, reflection upon preferable resolutions in concrete circumstances and the specification and revision of one's assumptions. Regulators who seek to invoke legal and extra-legal modalities to achieve their goals would lose these insights if the instrumental conception of the law were the only interpretation on offer.

One significant example that was cited in Section I. above and which underscores this point, relates to the differing concretisations of the procedural autonomy concept by UK and Californian courts – specifically regarding the standard of care in informed consent cases. Whereas both jurisdictions accounted for the necessary decisional and practical dimensions of our abstract autonomy concept, they then went on to realise them in diverging ways. The UK paired an independent, proactive perspective of the patient with a professional's supporting role. This generated a legal mechanism that was closely orientated towards the facilitation of their specific subjective needs. By contrast, Californian common law posited a relatively passive understanding of the patient, conceiving of them as broadly dependent on their physician's guidance. This led to an emphasis on objective criteria and on the need for legal certainty regarding the advice that must be provided by the professional. Not only does this generate insights into the possible specifications of procedural autonomy, it also exemplifies how different interests can be accommodated alongside relevant specifications and it further highlights the kinds of consequences that flow from the relevant choice.

Even in those instances where the operation of the common law appeared to entirely discount an otherwise promising normative approach, a productive interaction between abstract evaluation and specific legal argumentation could nevertheless be derived. For instance, the need to 'fit' AI's potential for informational manipulation under a Californian strand of case law that dealt with the non-therapeutic interests of human practitioners may have appeared to ignore the very novelty of the technology: that it was acting independently of human thought and mediation. However, this engagement directed the analysis towards concrete, pertinent considerations that maintained their instructive role for AI/ML devices. It was important to know that there was an existing, well-reasoned determination that an influence on patient decision-making did not have to be proven to have affected human judgment before it could become an appropriate subject

of censure. Further, it was significant that the presence of an overriding therapeutic objective could not justify the non-disclosure of entirely non-therapeutic influences.

Moreover, precisely the remaining limitations of the common law, in terms of monitoring and enforcement, pointed the way towards responses that appropriately targeted AI/ML functioning. The advantages of such a view were noted in this chapter, where it was considered how legislatures, although liberated from many specific constraints, do capitalise upon the guidance derived from the existent legal system's operation. It was noted how they in turn operate according to their own related rationales to offer complementary legal solutions to autonomy-related problems.

Perhaps these insights could be derived in an entirely abstract fashion, *via* a method that simply concentrates on understanding the nature of desirable ends and on anticipating their non-legal realisation. In this manner, as the technological instrumentalists appear to prefer, reliance on any particular system would be eschewed and their argument could remain aloof from doctrinal quirks. Yet, as the comparative lawyer knows, these quirks can constitute functional responses to complex problems. Learning how these things are in fact done, especially in highly dynamic and ever-changing environments, tends to yield unanticipated insights, rich accounts of the problem at hand and various possible methods for addressing them.

C. Law's nuanced normative dynamism

The final argument that was identified in Chapter 1 holds that the law is purely reactive to, and must lag problematically behind, the challenges thrown up by technological progress.[1419] To the extent that this constitutes a purely negative assessment of the law's functioning, this work has already departed from it. The previous section has sought to show one underappreciated positive facet of the law's limited, structured adaptation in response to external demands. Still, it was a further task of this thesis to examine how legal adaptation proceeds – and can proceed – from a methodological standpoint. On the basis of the specific assessment of AI/ML autonomy

1419 Chapter 1, Section II.C.2.

challenges one must ask how the law's responsiveness ought to be evaluated in this doctrinal respect.[1420]

To make this assessment, this work did not purport to operate solely within the confines of existing rules, nor did it restrict itself to a narrow casuistic analysis. Rather, through its normative approach it departed from the state-of-the-art examinations of legal informed consent requirements and medical ML devices.[1421] More broadly, it marked a departure from the law and technology literature that tended to argue either in favour of more liberal or more conservative approaches to rule interpretation – depending primarily on the instrumental demands of particular situations. By detailing the role that legal principles can play in the common law's adjustment to technological progress, a much more nuanced picture has emerged. This demonstrates that it would be misleading to adhere to a bifurcated approach, one that conceives of law's dynamism in the face of innovation as stemming from either a greater or lesser responsiveness to external, instrumentalist demands.

Under the principle-based approach it remains the case that certain avenues for change remain altogether foreclosed – several examples have been marked in Tables 1 and 2. For instance, where there is a well-established line of cases articulating the limited nature of a given class of rules, argumentation based upon the autonomy principle cannot be expected to overcome legal inertia. The narrow requirements pertaining to consent under the Californian battery doctrine provide an illustration of such a circumstance. It is an entirely orthodox proposition to hold that the weight of countervailing rigid and specific norms outweighs appeals to principle in such situations.

A similar conclusion is to be reached in cases where there is a conspicuous absence of authority. Namely, where a certain finding would constitute an altogether untested extension of existing doctrine, lacking even incidental support. Again, taking the generally more restrictive Californian system

1420 Practical difficulties of adaptation – such as: judge-made solutions being dependent on the right kinds of cases being brought forward, or the legislature functioning *via* relatively cumbersome information gathering and decision-making procedures – must be distinguished in this respect. The crucial question for present purposes is whether the law provides the conceptual tools to develop itself in the face of change. For another statement of this distinction, but addressing primarily the institutional component, see: Ard, 'Making Sense of Legal Disruption' (2022) Forward Wisconsin Law Review p. 42.

1421 Recall especially: Cohen, 'Informed Consent and Medical Artificial Intelligence' (2020) 108(6) The Georgetown Law Journal p. 1425.

as the example; arguing that actionable damage under negligence could be extended to autonomy harms *per se*, and that this would be an available adaptation to the pressures of AI's autonomy challenges, would be to misunderstand the limitations of principle-based reasoning. The autonomy principle itself is not capable of generating a cause of action and here there are significant countervailing concerns, albeit they are more abstract. They include above all: the principle of legal certainty and the consequentialist considerations that accompany it. In this situation the distinct considerations raised by medical AI are highly unlikely to call for a reassessment of these significant factors.

While this first class of cases may be described as pursuing a conservative approach, limiting the law's receptivity to desirable change, this would be to miss the point. It is a context-sensitive assessment that must be responsive, first and foremost, to the existent legal material. If this truly represents a strong constraint on reasoning, then arguing for a liberal or conservative mode of argumentation is of little avail.

Conversely, it was seen that there are instances where the emergence of medical ML technologies can be expected to exert strong pressures on the courts to review the state of existing doctrines in light of the demands of the autonomy norm. Under our principle-based analysis the relevant reconfiguration could take the form of novel interpretations of existing standards, of norm creation, of norm change or of exceptions that are made to otherwise unaffected rules. This too presents a more proactive and varied approach than is generally assumed in the wider literature.

Nevertheless, in the preceding examination, clear-cut illustrations of legal dynamism – demarcated in the above tables – also constituted a relatively small subset of the examined material. They emerged most prominently in relation to the interpretation of the breach element of negligence. For example, in both the UK and California a strong argument could be made that a requirement existed to disclose AI's ability to foreclose certain decision-making opportunities – even though this represented a novel phenomenon stemming from the technology's more independent functioning. Here, an extrapolation from the existing doctrinal framework was facilitated by a relatively open standard of care, which invoked autonomy-based reasoning, as well as by more specific rules that had been laid down in the case law. In particular, the courts had, by requiring the disclosure of available and reasonable alternatives, elaborated upon the significance of the reflective dimension of autonomy that also motivated the identification of the pertinent AI problem. It is these structural factors that pushed for a

liberal approach to emerge in this context, not a desire to realise an external objective.

In between these clear-cut opportunities for a dynamic adaptation of the law and those occasions where a rigid adherence to the existing rules was the most likely outcome, there was a third category of case. These were instances where it appeared likely that the introduction of AI/ML devices would require a reconsideration of existing norms, but the strength of available arguments was uncertain.

Under UK law, for example, it was important to consider whether interferences with individual autonomy constituted an actionable form of damage for the purposes of negligence, as this would allow the system to respond coherently to medical AI's challenges. This required an argument for norm generation, or at least adaptation, to be advanced. Although several significant judgments supported such a development, it could not be definitively asserted in light of prominent and persistent criticisms of courts and commentators. The latter were concerned, above all, with the doctrinal limitations of the tort of negligence and with the force of the principles of legal coherence and certainty.

But once more, rather than simply advocating for a liberal, instrumental approach that transcended these doctrinal concerns, a realistic argument in favour of legal dynamism required one to confront them. It was here that the autonomy principle, which had been developed to capture both AI's challenges and to fit within the legal system's specific requirements, became relevant. It was argued to be sufficiently concrete and objective to offer one acceptable conceptualisation of a cognisable form of damage. Therefore, by operating within the restrictions set by the pertinent norms, it was possible to make a structured argument that identified an opportunity for a proactive legal response to medical AI/ML.

This latter aspect – the way in which the dynamic or static nature of a system of norms is determined by the state of development of its more abstract principles – is the final manner in which this work has contributed to a more nuanced understanding of the law's adaptation to technological developments. Those who focus on specific rules appear to imply that the law is particularly prone to generate the 'pacing problem' because it is a cumbersome process to bring about the multifaceted, wide-ranging amendments that can be required by innovations that affect many distinct norms.

However, this does not account for the synergistic effects that a reconceptualisation of broader legal norms can similarly bring about. The poten-

tial for such an approach is evidenced by the impact that the autonomy principle has already had on the British negligence cause of action. Mutually reinforcing developments have spanned across specific norms, which encompass those pertaining to: actionable damage, the standard of care and causation. This demonstrates how the reconceptualisation of the legal system's higher normative commitments can have widespread knock-on effects on the shape of specific rules.

In relation to medical AI this means that, in a system such as the UK where principles pertinent to the regulation of the technology are experiencing a degree of flux, a greater adaptability to the demands of the innovation can be expected. The law is more open to mutually reinforcing adaptations, although knock-on effects on specific rules may still need to be worked out and the challenges posed by the technology may even help clarify the shape that these should take. Things are admittedly quite different in a more settled system, such as California, where a relatively stable balance between specific norms and abstract standards has been established.

Overall, it is to be noted that the law's propensity for development cannot be distilled into a simple formula aiming at the greater or lesser realisation of innovation-related objectives – it is intimately connected with the law's operation and the context-specific restrictions that this imposes. Accordingly, this work has subscribed to a nuanced differentiation that tracks the manner in which abstract and concrete norms interact to create space for creative legal activity. In spite of this context-specific focus, a more general insight can also be gleaned from the examination of medical AI under this lens: the potential for legal dynamism, at least at a doctrinal level, is much more extensive than is often recognised.

IV. Conclusion

Beyond evaluating the results of the common law comparison and outlining the lessons that can be learned from this for the broader regulation of medical ML devices, this final chapter has emphasised how the specific investigation relates to a much more fundamental discussion concerning the relationship between law and technological innovation. It has been demonstrated how the developed methodology provides a differentiated account that recognises the unique and persistent contribution of legal reasoning to this field.

To this end, it is necessary to question three prevalent assumptions concerning: the primacy of instrumental objectives, the relative inertia of the legal modality and an associated preference for extra-legal regulatory instruments. The preceding evaluation reveals that technological instruments may themselves lag behind, and be less effective than, available legal solutions. Moreover, the law continues to provide a multifaceted normative framework that assists in conceptualising the challenges presented by innovation, as well as indicating the nature of adequate solutions to these challenges. Lastly, in considering the law's ability to adapt dynamically one should take a nuanced approach that accounts for the pervasive role played by legal principles. It is the tensions caused by them that shape the creative, proactive role that legal reasoning assumes in accommodating the law to technological progress.

Returning to the specific research question posed in Chapter 1, the foregoing analysis demonstrates that, although the doctrines of informed consent may not themselves provide an adequate level of protection for patient autonomy as artificial intelligence is introduced into medicine, they nevertheless provide an available and inherently adaptable mechanism. The common law is in a position to respond, albeit imperfectly, to a rapidly emerging set of challenges. In addition, the mode of reasoning that characterises its operation points the way towards a coherent, more comprehensive solution. It has been suggested that this should take the form of a supplementary legislative scheme, functioning alongside the more general conditions of the common law and continuing to accommodate the evolving demands of the principle of patient autonomy.

Bibliography

I. Literature

Abràmoff Michael D., Lavin Philip T., Birch Michele and others, 'Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices' (2018) 1 NPJ Digital Medicine, 1–8.

Ackerman Bruce A., 'The Storrs Lectures: Discovering the Constitution' (1984) 93(6) The Yale Law Journal, 1013–1072.

Afnan Michael A. M., Liu Yanhe, Conitzer Vincent and others, 'Interpretable, Not Black-Box, Artificial Intelligence Should Be Used for Embryo Selection' [2021](4) Human Reproduction Open, 1–8.

Alberdi Eugenio, Povykalo Andrey, Strigini Lorenzo and others, 'Effects of Incorrect Computer-Aided Detection (Cad) Output on Human Decision-Making in Mammography' (2004) 11(8) Academic Radiology, 909–918.

Allenby Braden R., 'Governance and Technology Systems: The Challenge of Emerging Technologies' in Marchant, Allenby and Herkert, The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem (Springer Netherlands 2011).

Alon-Barkat Saar and Busuioc Madalina, 'Human-AI Interactions in Public Sector Decision-Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice' (2023) 33(1) Journal of Public Administration Research and Theory, 153–169.

Alpaydin Ethem, Machine Learning (Revised Edition, The MIT Press 2021).

Amirthalingam Kumaralingam, 'Causation and the Gist of Negligence' (2005) 64(1) The Cambridge Law Journal, 32–35.

Andorno R., 'The Right Not to Know: An Autonomy Based Approach' (2004) 30(5) Journal of Medical Ethics, 435-9.

Angus Derek C., 'Randomized Clinical Trials of Artificial Intelligence' [2020](11) The Journal of the American Medical Association, 1043–1045.

Appelbaum Paul S., Lidz Charles W. and Meisel Alan, Informed consent: Legal theory and clinical practice (Second Edition, Oxford University Press 2001).

Arabian Armand, 'Informed Consent: From the Ambivalence of Arato to the Thunder of Thor' (1994) 10(3) Issues in Law & Medicine, 261–298.

Ard B. J., 'Making Sense of Legal Disruption' (2022) Forward Wisconsin Law Review, 42–63.

Armitage Mark, Charlesworth John and Percy Rodney A. Charlesworth & Percy on Negligence (Fifteenth Edition, Sweet & Maxwell 2022).

Arvind T. T. and McMahon Aisling M., 'Responsiveness and the Role of Rights in Medical Law: Lessons from Montgomery' (2020) 28(3) Medical Law Review, 445–477.

Askland Andrew, 'Introduction: Why Law and Ethics Need to Keep Pace with Emerging Technologies' in Marchant, Allenby and Herkert, The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem (Springer Netherlands 2011).

Austin Louise, 'Correia, Diamond and the Chester Exception: Vindicating Patient Autonomy?' (2021) 29(3) Medical Law Review, 547–561.

Austin Louise, 'Grimstone v Epsom and St Helier University Hospitals NHS Trust: (It's Not) Hip to Be Square' (2018) 26(4) Medical Law Review, 665–674.

Azevedo Cunha Mario V. de, Andrade Norberto N. G. de and Lixinski Lucas and others (eds), New Technologies and Human Rights: Challenges to Regulation (Ashgate Publishing 2013).

Balganesh Shyamkrishna and Parchomovsky Gideon, 'Structure and Value in the Common Law' (2015) 163(5) University of Pennsylvania Law Review, 1241–1310.

Bankowski Zenon, MacCormick Neil and Marshall Geoffrey, 'Precedent in the United Kingdom' in MacCormick and Summers, Interpreting Precedents: A Comparative Study (Routledge 2016).

Barker Kit, Fairweather Karen and Grantham Ross (eds), Private Law in the 21st Century (Hart Publishing 2017).

Baron Jonathan, Rationality and Intelligence (Cambridge University Press 2005).

Bauer Nadja, Ickstadt Katja and Lübke Karsten and others (eds), Applications in Statistical Computing (Springer International Publishing 2019).

Beauchamp Tom L., 'The Failure of Theories of Personhood' (1999) 9(4) Kennedy Institute of Ethics journal, 309–324.

Beauchamp Tom L. and Childress James F., Principles of Biomedical Ethics (Fifth Edition, Oxford University Press 2001).

Beck James M. and Azari Elizabeth D., 'FDA, Off-Label Use, and Informed Consent: Debunking Myths and Misconceptions' (1998) 53(1) Food and Drug Law Journal, 71–104.

Beck Thorsten, Demirgüç-Kunt Asli and Levine Ross, 'Law and Finance: Why Does Legal Origin Matter?' (2003) 31(4) Journal of Comparative Economics, 653–675.

Becker Ulrich, 'Sozialrecht und Sozialrechtswissenschaft' (2010) 65(4) Zeitschrift für öffentliches Recht, 607–652.

Bell John and Ibbetson David, European Legal Development: The Case of Tort (Cambridge University Press 2012).

Benjamens Stan, Dhunnoo Pranavsingh and Meskó Bertalan, 'The State of Artificial Intelligence-Based FDA-Approved Medical Devices and Algorithms: An Online Database' (2020) 3 NPJ Digital Medicine.

Bennett Casey C. and Doub Thomas W., 'Expert Systems in Mental Health Care: AI Applications in Decision-Making and Consultation' in Luxton, Artificial Intelligence in Behavioral and Mental Health Care (Elsevier Reference Monographs 2016).

Berkeley Istvan S. N., 'The Curious Case of Connectionism' (2019) 2(1) Open Philosophy, 190–205.

Berlin Isaiah and Harris Ian, Liberty (Second Edition, Oxford University Press 2017).

Berman Paul S. (ed), Law and Society Approaches to Cyberspace (Ashgate Publishing 2007).

Binns Reuben, 'Algorithmic Accountability and Public Reason' (2018) 31(4) Philosophy & Technology, 543–556.

Bjerring Jens C. and Busch Jacob, 'Artificial Intelligence and Patient-Centered Decision-Making' (2021) 34(2) Philosophy & Technology, 349–371.

Bloche M. G., 'The Invention of Health Law' (2003) 91(1) California Law Review, 247–322.

Bloustein Edward J., 'Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser' in Schoeman, Philosophical Dimensions of Privacy (Cambridge University Press 2009).

Blumenthal-Barby Jennifer S. and Naik Aanand D., 'In Defense of Nudge-Autonomy Compatibility' (2015) 15(10) The American Journal of Bioethics, 45–47.

Boden Margaret A., 'GOFAI' in Frankish and Ramsey, The Cambridge Handbook of Artificial Intelligence (Cambridge University Press 2014).

Bohr Adam and Memarzadeh Kaveh (eds), Artificial Intelligence in Healthcare (Academic Press 2020).

Brajer Nathan, Cozzi Brian, Gao Michael and others, 'Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission' (2020) 3(2) JAMA Network Open, 1-14.

Braude Hillel D., 'Skilled Know-How, Virtuosity, and Expertise in Clinical Practice' in Schramme and Edwards, Handbook of the Philosophy of Medicine (Springer Netherlands 2017).

Brazier Margaret and Cave Emma, Medicine, Patients and the Law (Sixth Edition, Manchester University Press 2016).

Brazier Margaret and Lobjoit Mary, 'Fiduciary Relationship: An Ethical Approach and a Legal Concept?' in Erin and Bennett, HIV and AIDS: Testing, Screening, and Confidentiality (Oxford University Press 2001).

Brazier Margaret, 'Patient Autonomy and Consent to Treatment: The Role of the Law,' (1987) 7(2) Legal Studies, 169–193.

Brownsword Roger, 'An Interest in Human Dignity as the Basis for Genomic Torts' (2003) 42(3) Washburn Law Journal, 413–488.

Brownsword Roger, 'Law Disrupted, Law Re-Imagined, Law Re-Invented' (2019) 1 Technology and Regulation, 10–30.

Brownsword Roger, Law 3.0: Rules, Regulation, and Technology (Routledge 2021).

Brownsword Roger, Rights, Regulation, and the Technological Revolution (Oxford University Press 2008).

Brownsword Roger and Somsen Han, 'Law, Innovation and Technology: Fast Forward to 2021' (2021) 13(1) Law, Innovation and Technology, 1–28.

Brownsword Roger and Yeung Karen (eds), Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes (Hart Publishing 2008).

Brownsword Roger, Scotford Eloise and Yeung Karen (eds), The Oxford Handbook of Law, Regulation and Technology (Oxford University Press 2016).

Buckner Cameron, 'Deep learning: A Philosophical Introduction' (2019) 14(10) Philosophy Compass, 1–19.

Bumgarner Joseph M., Lambert Cameron T., Hussein Ayman A. and others, 'Smartwatch Algorithm for Automated Detection of Atrial Fibrillation' (2018) 71(21) Journal of the American College of Cardiology, 2381–2388.

Bunnik Eline M., Jong Antina de, Nijsingh Niels and others, 'The New Genetics and Informed Consent: Differentiating Choice to Preserve Autonomy' (2013) 27(6) Bioethics, 348–355.

Burrell Jenna, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) Big Data & Society, 1–12.

Byrne Peter (ed), Rights and Wrongs in Medicine (Oxford University Press 1986).

Calabresi Guido, A Common Law for the Age of Statutes (Harvard University Press 1985).

California Jurisprudence (Third Edition, Bancroft-Whitney 2022).

Cane Peter, 'Rights in Private Law' in Nolan and Robertson, Rights and Private Law (Hart Publishing 2012).

Carvalho Diogo V., Pereira Eduardo M. and Cardoso Jaime S., 'Machine Learning Interpretability: A Survey on Methods and Metrics' (2019) 8(8) Electronics, 832.

Chan Tracey E., 'Legal and Regulatory Responses to Innovative Treatment' (2013) 21(1) Medical Law Review, 92–130.

Chang Anthony, 'The Role of Artificial Intelligence in Digital Health' in Wulfovich and Meyers, Digital Health Entrepreneurship (Springer Cham 2020).

Chico Victoria, Genomic Torts: The English Tort Regime and Novel Grievances (Routledge 2010).

Chillag Nancy A., 'Negligent Infliction of Emotional Distress as an Independent Cause of Action in California: Do Defendants Face Unlimited Liability' (1982) 22(1) Santa Clara Law Review, 181–210.

Christman John, 'Autonomy and Personal History' (1991) 21(1) Canadian Journal of Philosophy, 1–24.

Clark T. and Nolan D., 'A Critique of Chester v Afshar' (2014) 34(4) Oxford Journal of Legal Studies, 659–692.

Cockburn Tina and Fay Michael, 'Consent to Innovative Treatment' (2019) 11(1) Law, Innovation and Technology, 34–54.

Cockfield Arthur J., 'Towards a Law and Technology Theory' (2003) 30(3) Manitoba Law Journal, 383–416.

Cockfield Arthur and Pridmore Jason, 'A Synthetic Theory of Law and Technology' (2007) 8(2) Minnesota Journal of Law, Science & Technology, 475–513.

Coggon John, 'Varied and Principled Understandings of Autonomy in English Law: Justifiable Inconsistency or Blinkered Moralism?' (2007) 15(3) Health Care Analysis: Journal of Health Philosophy and Policy, 235–255.

Cohen Felix S., 'Transcendental Nonsense and the Functional Approach' (1935) 35(6) Columbia Law Review, 809–849.

Cohen I. G., 'Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?' (2020) 108(6) The Georgetown Law Journal, 1425–1470.

Cohen I. G., Amarasingham Ruben, Shah Anand and others, 'The Legal and Ethical Concerns That Arise From Using Complex Predictive Analytics in Health Care' (2014) 33(7) Health Affairs (Project Hope), 1139–1147.

Coyle Casey A., 'Gonzales v. Carhart: Justice Kennedy at the Intersection of Life Interests, Medical Practice and Government Regulations Comment' (2008) 27(2) Temple Journal of Science, Technology & Environmental Law, 291–314.

Crootof Rebecca and Ard B. J., 'Structuring Techlaw' (2021) 34(2) Harvard Journal of Law & Technology, 347–418.

Cross Rupert, Precedent in English law (Third Edition, Clarendon Press 1979).

Curchoe Carol L., Flores-Saiffe Farias Adolfo, Mendizabal-Ruiz Gerardo and others, 'Evaluating Predictive Models in Reproductive Medicine' (2020) 114(5) Fertility and Sterility, 921–926.

Currie David P., 'Positive and Negative Constitutional Rights' (1986) 53(3) The University of Chicago Law Review, 864–890.

Custers Bart and Heijne Anne-Sophie, 'The Right of Access in Automated Decision-Making: The Scope of Article 15(1)(h) GDPR in Theory and Practice' (2022) 46 Computer Law & Security Review, 105727.

Daly Erin, 'Reconsidering Abortion Law: Liberty, Equality, and the New Rhetoric of Planned Parenthood v. Casey' (1995) 45(1) American University Law Review, 77–150.

Davenport Thomas H. and Glaser John P., 'Factors Governing the Adoption of Artificial Intelligence in Healthcare Providers' (2022) 1(1) Discover Health Systems.

Davies Simon J., Vistisen Simon T., Jian Zhongping and others, 'Ability of an Arterial Waveform Analysis-Derived Hypotension Prediction Index to Predict Future Hypotensive Events in Surgical Patients' (2020) 130(2) Anesthesia and Analgesia, 352–359.

Deakin Simon, 'Organisational Torts: Vicarious Liability Versus Non-Delegable Duty' (2018) 77(1) The Cambridge Law Journal, 15–18.

Debrabander Jasper and Mertes Heidi, 'Watson, Autonomy and Value Flexibility: Revisiting the Debate' [2021] Journal of Medical Ethics, 1043–1047.

Deo Rahul C., 'Machine Learning in Medicine' (2015) 132(20) Circulation: Cardiovascular Quality and Outcomes, 1920–1930.

Di Nucci Ezio, 'Should We Be Afraid of Medical AI?' (2019) 45(8) Journal of Medical Ethics, 556–558.

Dillon John J., DeSimone Christopher V., Sapir Yehu and others, 'Noninvasive Potassium Determination Using a Mathematically Processed ECG: Proof of Concept for a Novel "Blood-Less, Blood Test"' (2015) 48(1) Journal of Electrocardiology, 12–18.

Dobbs Dan B., Hayden Paul T. and Bublick Ellen M., Dobbs' Law of Torts: Practitioner Treatise Series (Second Edition, Thomson West 2022).

Donnelly Mary, Healthcare Decision-Making and the Law: Autonomy, Capacity and the Limits of Liberalism (Cambridge University Press 2010).

Donovan Mary, 'Is the Injury Requirement Obsolete in a Claim for Fear of Future Consequences' (1993) 41(5) UCLA Law Review, 1337–1396.

Dror Yehezkel, 'Law and Social Change 1958-1959' (1959) 33(4) Tulane Law Review, 787–802.

Dube Simant, An Intuitive Exploration of Artificial Intelligence: Theory and Applications of Deep Learning (Springer International Publishing 2021).

Dunn Michael, Fulford K. W. M., Herring Jonathan and others, 'Between the Reasonable and the Particular: Deflating Autonomy in the Legal Regulation of Informed Consent to Medical Treatment' (2019) 27(2) Health Care Analysis: Journal of Health Philosophy and Policy, 110–127.

Duxbury Neil, 'The Law of the Land' (2015) 78(1) The Modern Law Review, 26–54.

Dworkin Gerald, 'The Nature of Autonomy', The Theory and Practice of Autonomy (Cambridge University Press 2012).

Dworkin Ronald, Taking Rights Seriously (Duckworth 1987).

Eisenberg Melvin A., The Nature of the Common Law (Harvard University Press 1988).

Ekstrom Laura W., 'A Coherence Theory of Autonomy' (1993) 53(3) Philosophy and Phenomenological Research, 599–616.

Epstein Richard A., 'The Static Conception of the Common Law: Legal and Economic Perspectives' (1980) 9(2) The Journal of Legal Studies, 253–275.

Erb Randall J., 'Introduction to Backpropagation Neural Network Computation' (1993) 10(2) Pharmaceutical Research, 165–170.

Erin Charles A. and Bennett Rebecca (eds), HIV and AIDS: Testing, Screening, and Confidentiality (Oxford University Press 2001).

Ezra David B., 'Smoker Battery: An Antidote to Second-Hand Smoke' (1989) 63(4) Southern California Law Review, 1061–1112.

Faden Ruth R., King Nancy M. P. and Beauchamp Tom L., A History and Theory of Informed Consent (Oxford University Press 1986).

Farber Daniel A. and Frickey Philip P., 'In the Shadow of the Legislature: The Common Law in the Age of the New Public Law' (1991) 89(4) Michigan Law Review, 875–906.

Feng Tan K., 'Failure of Medical Advice: Trespass or Negligence?' (1987) 7(2) Legal Studies, 149–168.

Fischer John M. and Ravizza Mark, Responsibility and Control: A Theory of Moral Responsibility (Cambridge University Press 2000).

Foster Charles, Choosing Life, Choosing Death: The Tyranny of Autonomy in Medical Ethics and Law (Hart Publishing 2009).

Frankfurt Harry G., 'Freedom of the Will and the Concept of a Person' (1971) 68(1) The Journal of Philosophy, 5–20.

Frankish Keith and Ramsey William M. (eds), The Cambridge Handbook of Artificial Intelligence (Cambridge University Press 2014).

Freeman Karoline, Geppert Julia, Stinton Chris and others, 'Use of Artificial Intelligence for Image Analysis in Breast Cancer Screening Programmes: Systematic Review of Test Accuracy' (2021) 374 BMJ (Clinical Research Edition), 1-15.

Friedman Lawrence M., Law and Society: An Introduction (Prentice-Hall 1977).

Friedman Lawrence M. and Ladinsky Jack, 'Social Change and the Law of Industrial Accidents' (1967) 67(1) Columbia Law Review, 50–82.

Froese Tom and Ziemke Tom, 'Enactive Artificial Intelligence: Investigating the Systemic Organization of Life and Mind' (2009) 173(3-4) Artificial Intelligence, 466–500.

Fuller Lon L., 'Means and Ends' in Winston, The Principles of Social Order: Selected Essays of Lon L. Fuller (Revised Edition. Hart Publishing 2001).

Funer Florian, 'Accuracy and Interpretability: Struggling with the Epistemic Foundations of Machine Learning-Generated Medical Information and Their Practical Implications for the Doctor-Patient Relationship' (2022) 35(1) Philosophy & Technology.

Funer Florian, 'The Deception of Certainty: How Non-Interpretable Machine Learning Outcomes Challenge the Epistemic Authority of Physicians' (2022) 25(2) Medicine, Health Care and Philosophy, 167–178.

Gaille Marie and Horn Ruth, 'Solidarity and Autonomy: Two Conflicting Values in English and French Health Care and Bioethics Debates?' (2016) 37(6) Theoretical Medicine and Bioethics, 441–446.

Gardner John, Torts and Other Wrongs (Oxford University Press 2019).

Garg Amit X. Adhikari Neill K. J. McDonald Heather and others, 'Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review' (2005) 293(10) The Journal of the American Medical Association, 1223–1238.

Gatter Robert, 'The Mysterious Survival of the Policy against Informed Consent Liability for Hospitals' (2006) 81(4) Notre Dame Law Review, 1203–1274.

Gerke Sara, Minssen Timo and Cohen Glenn, 'Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare' in Bohr and Memarzadeh, Artificial Intelligence in Healthcare (Academic Press 2020).

Gilmore Grant, 'Legal Realism: Its Cause and Cure' (1961) 70(7) The Yale Law Journal, 1037–1048.

Giuffrida Luisa A., 'Moore v. Regents of the University of California: Doctor, Tell Me Moore' (1991) 23(1) Pacific Law Journal, 267–314.

Goddard Kate, Roudsari Abdul and Wyatt Jeremy C., 'Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators' (2012) 19(1) Journal of the American Medical Informatics Association, 121–127.

Goebelsmann Christina L., 'Putting Ethics and Traditional Legal Principles Back into California Tort Law: Barring Wrongful-Birth Liability in Preimplantation Genetic Testing Cases' (2010) 43(2) Loyola of Los Angeles Law Review, 667–692.

Goertzel Ben and Pennachin Cassio (eds), Artificial General Intelligence (Springer 2007).

Gold Stephanie S., 'An Equality Approach to Wrongful Birth Statutes' (1996) 65(3) Fordham Law Review, 1005–1041.

Goldberg John C. and Zipursky Benjamin, 'Torts as Wrongs' (2010) 88(5) Texas Law Review, 917–986.

Goldberg Richard (ed), Medicinal Product Liability and Regulation (Hart Publishing 2013).

Goodhart Arthur L., 'Case Law in England and America' (1930) 15(2) Cornell Law Review, 173–193.

Goodman Bryce and Flaxman Seth, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' (2017) 38(3) AI Magazine, 50–57.

Goodwin Jean, 'Accounting for the Appeal to the Authority of Experts' (2011) 25(3) Argumentation, 285–296.

Grattet Ryken, 'Sociological Perspectives on Legal Change: The Role of the Legal Field in the Transformation of the Common Law of Industrial Accidents' (1997) 21(3) Social Science History, 359–397.

Green Sarah and Sales Philip, 'Law, Technology and the Common Law Method in the United Kingdom' [2023](5) Europäische Zeitschrift für Wirtschaftsrecht, 205–214.

Grubb Andrew, 'Battery and Administration of Anaesthetic: Davis v. Barking, Havering and Brentwood Health Authority' (1993) 1(3) Medical Law Review, 389–391.

Grubb Andrew, 'Failed Sterilisation: Duty to Provide Adequate Warning' (1995) 3(3) Medical Law Review, 297–299.

Grubb Andrew and Pearl David S., Blood Testing, AIDS, and DNA Profiling: Law and Policy (Jordan Publishing 1990).

Grzybowski Andrzej and Brona Piotr, 'Analysis and Comparison of Two Artificial Intelligence Diabetic Retinopathy Screening Algorithms in a Pilot Study: IDx-DR and Retinalyze' (2021) 10(11) Journal of Clinical Medicine, 1–8.

Guidotti Riccardo, Monreale Anna, Ruggieri Salvatore and others, 'A Survey of Methods for Explaining Black Box Models' (2019) 51(5) ACM Computing Surveys, 1–42.

Guihot Michael, 'Coherence in Technology Law' (2019) 11(2) Law, Innovation and Technology, 311–342.

Günther Christian M., 'Legal vs. Extra-Legal Responses to Public Health Emergencies' (2022) 29(1) European Journal of Health Law, 131–149.

Gutwirth Serge, Hert Paul de and Sutter Laurent de, 'The Trouble with Technology Regulation: Why Lessig's 'Optimal Mix' Will Not Work' in Brownsword and Yeung, Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes (Hart Publishing 2008).

Hage Jaap C. and Pfordten Dietmar von der (eds), Concepts in Law (Springer Netherlands 2009).

Hansen Pelle G. and Jespersen Andreas M., 'Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy' (2013) 4(1) European Journal of Risk Regulation, 3–28.

Haqq Luke I., 'The Impact of Roe on Prenatal Tort Litigation: On the Public Policy of Unexpected Children' (2020) 13(1) Journal of Tort Law, 81–160.

Harris Cailin, 'Statutory Prohibitions on Wrongful Birth Claims & Their Dangerous Effects on Parents' (2014) 34(2) Boston College Journal of Law & Social Justice, 365–396.

Harris Neville (ed), Social Security Law (Oxford University Press 2000).

Harris Neville, 'The Welfare State, Social Security, and Social Citizenship Rights' in Harris, Social Security Law (Oxford University Press 2000).

Harrison Kevin and Boyd Tony, The Changing Constitution (Edinburgh University Press 2006).

Hatib Feras, Jian Zhongping, Buddi Sai and others, 'Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis' (2018) 129(4) Anesthesiology, 663–674.

Heidenreich Colleen W., 'Clarifying California's Approach to Claims of Negligent Infliction of Emotional Distress' (1995) 30(1) University of San Francisco Law Review, 277–312.

Herring Jonathan, ''Choosing Life, Choosing Death, The Tyranny of Autonomy in Medical Ethics and Law, by Charles Foster'' (2010) 30(2) Legal Studies, 330–333.

Herring Jonathan, Medical Law and Ethics (Fourth Edition, Oxford University Press 2012).

Herring Jonathan and Foster Charles, '"Please Don't Tell Me": The Right Not to Know' (2012) 21(1) Cambridge Quarterly of Healthcare Ethics, 20–29.

Herring Jonathan and Wall Jesse, 'The Nature and Significance of the Right to Bodily Integrity' (2017) 76(3) The Cambridge Law Journal, 566–588.

Hervey Matt and Lavy Matthew (eds), The Law of Artificial Intelligence (Sweet & Maxwell 2021).

Heywood Rob and Miola José, 'The Changing Face of Pre-operative Medical Disclosure: Placing the Patient at the Heart of the Matter' (2017) 133((Apr)) Law Quarterly Review, 296–321.

Hildebrandt Mireille, Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology (Edward Elgar Publishing 2015).

Hinkle Rachael K. and Nelson Michael J., 'The Transmission of Legal Precedent among State Supreme Courts in the Twenty-First Century' (2016) 16(4) State Politics & Policy Quarterly, 391–410.

Hoeren Thomas and Maurice Niehoff, 'Artificial Intelligence in Medical Diagnoses and the Right to Explanation' (2018) 4(3) European Data Protection Law Review, 308–319.

Hoffmann-Riem Wolfgang, Innovation und Recht - Recht und Innovation: Recht im Ensemble seiner Kontexte (Mohr Siebeck 2016).

Hohmann Hanns, 'The Nature of the Common Law and the Comparative Study of Legal Reasoning' (1990) 38(1) The American Journal of Comparative Law, 143–170.

Holley Kerrie and Becker Siupo, AI-First Healthcare: AI Applications in the Business and Clinical Management of Health (O'Reilly 2021).

Holmes Jr. Oliver W., The Common Law (Little, Brown and Company 1881).

Hornung Gerrit, Grundrechtsinnovationen (Mohr Siebeck 2015).

Hosseini Mohammad-Parsa, Lu Senbao, Kamaraj Kavin and others, 'Deep Learning Architectures' in Pedrycz and Chen, Deep Learning: Concepts and Architectures (Springer International Publishing 2020).

Hostiuc Sorin (ed), Clinical Ethics at the Crossroads of Genetic and Reproductive Technologies (Elsevier 2018).

Hostiuc Sorin, 'Predictive Genetic Testing in Multifactorial Disorders' in Hostiuc, Clinical Ethics at the Crossroads of Genetic and Reproductive Technologies (Elsevier 2018).

Humphreys Paul, 'The Philosophical Novelty of Computer Simulation Methods' (2009) 169(3) Synthese, 615–626.

Hyun Insoo, 'Authentic Values and Individual Autonomy' (2001) 35(2) The Journal of Value Inquiry, 195–208.

Igual Laura and Seguí Santi (eds), Introduction to Data Science (Springer International Publishing 2017).

Igual Laura and Seguí Santi, 'Unsupervised Learning' in Igual and Seguí, Introduction to Data Science (Springer International Publishing 2017).

Iheukwumere Emmanuel O., 'Doctor, Are You Experienced? The Relevance of Disclosure of Physician Experience to a Valid Informed Consent' (2002) 18(2) The Journal of Contemporary Health Law and Policy, 373–419.

Jackson Emily, ''Informed Consent' to Medical Treatment and the Impotence of Tort' in McLean, First Do No Harm (Routledge 2016).

Johns Margaret Z., 'Informed Consent: Requiring Doctors to Disclose Off-Label Prescriptions and Conflicts of Interest' (2007) 58(5) Hastings Law Journal, 967–1024.

Johnson Richard and Yi Zhu Yuan (eds), Sceptical Perspectives on the Changing Constitution of the United Kingdom (Hart Publishing 2023).

Jolliffe I. T., Principal Component Analysis (Second Edition, Springer New York 2002).

Jones Meg L., 'Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw' [2018](2) University of Illinois Journal of Law, Technology & Policy, 249–284.

Jones Michael, Medical Negligence (Sixth Edition, Sweet & Maxwell 2021).

Kähler Lorenz, 'Norm, Code, Digitalisat' in Kuhli and Rostalski, Normentheorie im digitalen Zeitalter (Nomos 2023).

Kaminski Margot E., 'Technological "Disruption" of the Law's Imagined Scene: Some Lessons from Lex Informatica' (2021) 36(3) Berkeley Technology Law Journal, 883–914.

Kazzazi Fawz, 'The Automation of Doctors and Machines: A Classification for AI in Medicine (ADAM framework)' (2021) 8(2) Future Healthcare Journal, 257–262.

Kearl Kurtis J., 'Turpin v. Sortini: Recognizing the Unsupportable Cause of Action for Wrongful Life' (1983) 71(4) California Law Review, 1278–1297.

Keating Rebecca and Wright Laura, 'AI and Professional Liability' in Hervey and Lavy, The Law of Artificial Intelligence (Sweet & Maxwell 2021).

Kelley Patrick J., 'Wrongful Life, Wrongful Birth, and Justice in Tort Law' (1979) Fall(4) Washington University Law Quarterly, 919–964.

Kellmeyer Philipp, 'Ethical Issues in the Application of Machine Learning to Brain Disorders' in Mechelli and Vieira, Machine Learning: Methods and Applications to Brain Disorders (Academic Press 2019).

Kelly Christopher J., Karthikesalingam Alan, Suleyman Mustafa and others, 'Key Challenges for Delivering Clinical Impact with Artificial Intelligence' (2019) 17(1) BMC Medicine, 1–9.

Kennedy Ian, 'The Doctor-Patient Relationship' in Byrne, Rights and Wrongs in Medicine (Oxford University Press 1986).

Kennedy Ian and Grubb Andrew, 'Testing for HIV Infection: The Legal Framework' (1989) 86(7) Law Society Gazette, 30-35.

Keown John, 'The Ashes of Aids and the Phoenix of Informed Consent' (1989) 52(6) The Modern Law Review, 790–800.

Keren-Paz Tsachi, 'Compensating Injury to Autonomy in English Negligence Law: Inconsistent Recognition' (2018) 26(4) Medical Law Review, 585–609.

Keren-Paz Tsachi, 'Compensating Injury to Autonomy: A Conceptual and Normative Analysis' in Barker, Fairweather and Grantham, Private Law in the 21st Century (Hart Publishing 2017).

Keren-Paz Tsachi, 'Gender Injustice in Compensating Injury to Autonomy in English and Singaporean Negligence Law' (2019) 27(1) Feminist Legal Studies, 33–55.

Kiener Maximilian, 'Artificial Intelligence in Medicine and the Disclosure of Risks' (2020) 36(3) AI & Society, 705–713.

Kim Dohyun, You Sungmin, So Soonwon and others, 'A Data-Driven Artificial Intelligence Model for Remote Triage in the Prehospital Environment' (2018) 13(10) PloS One.

King Nancy M. P., 'The Reasonable Patient and the Healer' (2015) 50(2) Wake Forest Law Review, 343–362.

Kirby Michael, 'New Frontier: Regulating Technology by Law and 'Code'' in Brownsword and Yeung, Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes (Hart Publishing 2008).

Kirchhoffer David G. and Richards Bernadette J. (eds), Beyond Autonomy (Cambridge University Press 2019).

Krishnan Maya, 'Against Interpretability: A Critical Examination of the Interpretability Problem in Machine Learning' (2020) 33(3) Philosophy & Technology, 487–502.

Kudina Olya and Boer Bas de, 'Co-Designing Diagnosis: Towards a Responsible Integration of Machine Learning Decision-Support Systems in Medical Diagnostics' (2021) 27(3) Journal of Evaluation in Clinical Practice, 529–536.

Kuhli Milan and Rostalski Frauke (eds), Normentheorie im digitalen Zeitalter (Nomos 2023).

Laing Judith M. and McHale Jean V. (eds), Principles of Medical Law (Fourth Edition, Oxford University Press 2017).

Landes William M. and Posner Richard A., The Economic Structure of Tort Law (Harvard University Press 1987).

Lawson Craig, 'The Family Affinities of Common-Law and Civil-Law Legal Systems' (1982) 6(1) Hastings International and Comparative Law Review, 85–132.

Lebovitz Sarah, Levina Natalia and Lifshitz-Assaf Hila, 'Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What' (2021) 45(3) MIS Quarterly, 1501–1526.

Lessig Lawrence, 'The Constitution of Code: Limitations on Choice-Based Critiques of Cyberspace Regulation' (1997) 5(2) CommLaw Conspectus: Journal of Communications Law and Policy, 181–192.

Lessig Lawrence, Code: Version 2.0 (Basic Books 2006).

Lewens Tim (ed), Risk: Philosophical Perspectives (Routledge 2007).

Li Aiguo, Walling Jennifer, Ahn Susie and others, 'Unsupervised Analysis of Transcriptomic Profiles Reveals Six Glioma Subtypes' (2009) 69(5) Cancer research, 2091–2099.

Liu Hin-Yan, Maas Matthijs, Danaher John and others, 'Artificial Intelligence and Legal Disruption: A New Model for Analysis' (2020) 12(2) Law, Innovation and Technology, 205–258.

López-Rubio Ezequiel, 'Computational Functionalism for the Deep Learning Era' (2018) 28(4) Minds & Machines, 667–688.

Luxton David D. (ed), Artificial Intelligence in Behavioral and Mental Health Care (Elsevier Reference Monographs 2016).

Lyell David and Coiera Enrico, 'Automation Bias and Verification Complexity: A Systematic Review' (2017) 24(2) Journal of the American Medical Informatics Association, 423–431.

MacCallum Gerald C., 'Negative and Positive Freedom' (1967) 76(3) The Philosophical Review, 312–334.

MacCormick Neil, Legal Reasoning and Legal Theory (Oxford University Press 1994).

MacCormick Neil and Summers Robert S. (eds), Interpreting Precedents: A Comparative Study (Routledge 2016).

Maclean Alasdair R., 'Consent, Sectionalisation and the Concept of a Medical Procedure' (2002) 28(4) Journal of Medical Ethics, 249–254.

Maclean Alasdair R., 'The Doctrine of Informed Consent: Does It Exist and Has It Crossed the Atlantic?' (2004) 24(3) Legal Studies, 386–413.

Maclean Alasdair R., Autonomy, Informed Consent and Medical Law: A Relational Challenge (Cambridge University Press 2009).

Mahase Elisabeth, 'Birmingham Trust and Babylon Health Discuss Pre-A&E Triage App' (2019) 365(l2354) BMJ (Clinical Research Edition).

Mandel Gregory N., 'Legal Evolution in Response to Technological Change' in Brownsword, Scotford and Yeung, The Oxford Handbook of Law, Regulation and Technology (Oxford University Press 2016).

Marchant Gary E., 'The Growing Gap Between Emerging Technologies and the Law' in Marchant, Allenby and Herkert, The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem (Springer Netherlands 2011).

Marchant Gary E., Allenby Braden R. and Herkert Joseph R. (eds), The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem (Springer Netherlands 2011).

Mayer-Schonberger Viktor, 'Demystifying Lessig' [2008](4) Wisconsin Law Review, 713–746.

McDonald George, California Medical Malpractice: Law & Practice (Revised Edition, Thomson West 2022).

McDougall Rosalind J., 'Computer Knows Best?: The Need for Value-Flexibility in Medical AI' [2019](45) Journal of Medical Ethics, 156–160.

McDougall Rosalind J., 'No We Shouldn't Be Afraid of Medical AI; It Involves Risks and Opportunities' (2019) 45(8) Journal of Medical Ethics, 559.

McGregor Harvey, Edelman James, Colton Simon and others, McGregor on Damages (Twenty-First Edition, Sweet & Maxwell 2021).

McHale Jean V., ''Appropriate Consent' and the Use of Human Material for Research Purposes: The Competent Adult' (2006) 1(4) Clinical Ethics, 195–199.

McHale Jean V., 'Consent to Treatment: The Competent Patient' in Laing and McHale, Principles of Medical Law (Fourth Edition. Oxford University Press 2017).

McLean Sheila A. M. (ed), First Do No Harm (Routledge 2016).

Meagher Dan, 'Is There a Common Law Right to Freedom of Speech?' (2019) 43(1) Melbourne University Law Review, 269–302.

Mechelli Andrea and Vieira Sandra (eds), Machine Learning: Methods and Applications to Brain Disorders (Academic Press 2019).

Meisel Alan, 'A Dignitary Tort as a Bridge between the Idea of Informed Consent and the Law of Informed Consent' (1988) 16(3-4) Law, Medicine and Health Care, 210–218.

Meisel Alan, 'The Right to Die: A Case Study in American Lawmaking' (1996) 3(1) European Journal of Health Law, 49–74.

Mele Alfred R., 'Motivated Irrationality' in Mele and Rawling, The Oxford Handbook of Rationality (Oxford University Press 2004).

Mele Alfred R. and Rawling Piers (eds), The Oxford Handbook of Rationality (Oxford University Press 2004).

Michaels Ralf, 'The Functional Method of Comparative Law' in Reimann and Zimmermann, The Oxford Handbook of Comparative Law (Second Edition. Oxford University Press 2019).

Michelucci Umberto, Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks (Apress 2018).

Miguel Beriain Iñigo de, 'Should We Have a Right to Refuse Diagnostics and Treatment Planning by Artificial Intelligence?' (2020) 23(2) Medicine, Health Care, and Philosophy, 247–252.

Minsky Marvin, The Society of Mind (First Edition, Simon & Schuster Paperbacks 1988).

Minssen Timo, Gerke Sara, Aboy Mateo and others, 'Regulatory Responses to Medical Machine Learning' (2020) 7(1) Journal of Law and the Biosciences, 1–18.

Molnar Christoph, Interpretable Machine Learning: A Guide for Making Black Box Models Interpretable (Lulu 2019).

Montanez Savannah R., 'Pregnant and Scared: How NIFLA v. Becerra Avoids Protecting Women's Reproductive Autonomy' (2019) 56(3) San Diego Law Review, 829–852.

Montgomery Jonathan, 'Law and the Demoralisation of Medicine' (2006) 26(2) Legal Studies, 185–210.

Moore Nancy J., 'Intent and Consent in the Tort of Battery: Confusion and Controversy' (2012) 61(6) American University Law Review, 1585–1656.

Morgan Jonathan, 'Torts and Technology' in Brownsword, Scotford and Yeung, The Oxford Handbook of Law, Regulation and Technology (Oxford University Press 2016).

Morik Katharina, 'A Note on Artificial Intelligence and Statistics' in Bauer and others, Applications in Statistical Computing (Springer International Publishing 2019).

Moses Lyria B., 'Adapting the Law to Technological Change: A Comparison of Common Law and Legislation Courts and Parliament' (2003) 26(2) UNSW Law Journal, 394–417.

Moses Lyria B., 'Sui Generis Rules' in Marchant, Allenby and Herkert, The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem (Springer Netherlands 2011).

Moses Lyria B., 'The Legal Landscape Following Technological Change: Paths to Adaptation' (2007) 27(5) Bulletin of Science, Technology & Society, 408–416.

Moses Lyria B., 'Why Have a Theory of Law and Technological Change?' (2007) 8(2) Minnesota Journal of Law, Science & Technology, 589–606.

Mosier Kathleen L., Skitka Linda J., Heers Susan and others, 'Automation Bias: Decision Making and Performance in High-Tech Cockpits' (1998) 8(1) The International Journal of Aviation Psychology, 47–63.

Mourby Miranda, Ó Cathaoir Katharina and Collin Catherine B., 'Transparency of Machine-Learning in Healthcare: The GDPR & European Health Law' (2021) 43 Computer Law & Security Review.

Muehlematter Urs J., Daniore Paola and Vokinger Kerstin N., 'Approval of Artificial Intelligence and Machine Learning-Based Medical Devices in the USA and Europe (2015–20): A Comparative Analysis' (2021) 3(3) The Lancet Digital Health, 195–203.

Mueller Andy, 'The Knowledge Norm of Apt Practical Reasoning' (2021) 199(1-2) Synthese, 5395–5414.

Mulligan Andrea, 'A Vindicatory Approach to Tortious Liability for Mistakes in Assisted Human Reproduction' (2020) 40(1) Legal Studies, 55–76.

Murphy Thérèse (ed), New Technologies and Human Rights (Oxford University Press 2009).

Naeem Muddasar, Rizvi Syed T. H. and Coronato Antonio, 'A Gentle Introduction to Reinforcement Learning and its Application in Different Fields' (2020) 8 IEEE Access, 209320–209344.

Narla Akhila, Kuprel Brett, Sarin Kavita and others, 'Automated Classification of Skin Lesions: From Pixels to Practice' (2018) 138(10) The Journal of Investigative Dermatology, 2108–2110.

Ngo Brandon, Nguyen Diep and van Sonnenberg Eric, 'The Cases for and against Artificial Intelligence in the Medical School Curriculum' (2022) 4(5) Radiology: Artificial intelligence, 1–8.

Nilsson Nils J., The Quest for Artificial Intelligence (Cambridge University Press 2009).

Nolan Donal, 'Damage in the English Law of Negligence' (2013) 4(3) Journal of European Tort Law, 259–281.

Nolan Donal, 'Negligence and Autonomy' [2022](2), 356–383.

Nolan Donal, 'New Forms of Damage in Negligence' (2007) 70(1) The Modern Law Review, 59–88.

Nolan Donal, 'Varying the Standard of Care in Negligence' (2013) 72(3) The Cambridge Law Journal, 651–688.

Nolan Donal and Robertson Andrew (eds), Rights and Private Law (Hart Publishing 2012).

Nolan Paul, 'Artificial Intelligence in Medicine - Is Too Much Transparency a Good Thing?' [2023] The Medico-Legal Journal, Onlinefirst.

O'Neill Jennifer, 'Lessons From the Vaginal Mesh Scandal: Enhancing the Patient-Centric Approach to Informed Consent for Medical Device Implantation' (2021) 37(1) International Journal of Technology Assessment in Health Care, e53, 1-5.

Orfali Kristina, 'A Journey Through Global Bioethics' (2019) 16(3) Journal of Bioethical Inquiry, 305–308.

Ormerod David C. and Laird Karl, Smith, Hogan, & Ormerod's Criminal Law (Fifteenth Edition, Oxford University Press 2018).

Papathanasiou Jason, Zaraté Pascale and Freire de Sousa Jorge (eds), EURO Working Group on DSS: A Tour of the DSS Developments Over the Last 30 Years (Springer International Publishing 2021).

Parasuraman Raja and Manzey Dietrich H., 'Complacency and Bias in Human Use of Automation: An Attentional Integration' (2010) 52(3) Human Factors, 381–410.

Parfit Derek, On What Matters: Volume One (Oxford University Press 2011).

Parikh Harsh, Hoffman Kentaro, Sun Haoqi and others, 'Why Interpretable Causal Inference is Important for High-Stakes Decision Making for Critically Ill Patients and How To Do It' (2022) Preprint, 1–31.

Parikh Ravi B. and Helmchen Lorens A., 'Paying For Artificial Intelligence in Medicine' [2022](5) NPJ Digital Medicine.

Pattinson Shaun D., Medical Law and Ethics (Sixth Edition, Sweet & Maxwell 2020).

Paulo Norbert, The Confluence of Philosophy and Law in Applied Ethics (Palgrave Macmillan UK 2016).

Pedrycz Witold and Chen Shyi-Ming (eds), Deep Learning: Concepts and Architectures (Springer International Publishing 2020).

Pennachin Cassio and Goertzel Ben, 'Contemporary Approaches to Artificial General Intelligence' in Goertzel and Pennachin, Artificial General Intelligence (Springer 2007).

Perry Stephen, 'Risk, Harm, Interests and Rights' in Lewens, Risk: Philosophical Perspectives (Routledge 2007).

Phillips Andelka M., Campos Thana C. de and Herring Jonathan (eds), Philosophical Foundations of Medical Law (Oxford University Press 2019).

Ploug Thomas and Holm Søren, 'Doctors, Patients, and Nudging in the Clinical Context--Four Views on Nudging and Informed Consent' (2015) 15(10) The American Journal of Bioethics, 28–38.

Ploug Thomas and Holm Søren, 'The Four Dimensions of Contestable AI Diagnostics - A Patient-Centric Approach to Explainable AI' (2020) 107 Artificial Intelligence in Medicine.

Ploug Thomas and Holm Søren, 'The Right to Refuse Diagnostics and Treatment Planning by Artificial Intelligence' (2020) 23(1) Medicine, Health Care, and Philosophy, 107–114.

Pope Thaddeus M., 'Certified Patient Decision Aids: Solving Persistent Problems with Informed Consent Law' (2017) 45(1) Journal of Law, Medicine & Ethics, 12–40.

Poscher Ralf, 'The Hand of Midas: When Concepts Turn Legal, or Deflating the Hart-Dworkin Debate' in Hage and Pfordten, Concepts in Law (Springer Netherlands 2009).

Posner Richard A., Law and Legal Theory in England and America (Clarendon Press 2003).

Post David G., 'Against "Against Cyberanarchy"' (2002) 17(4) Berkeley Technology Law Journal, 1365–1387.

Povyakalo Andrey A., Alberdi Eugenio, Strigini Lorenzo and others, 'How to Discriminate Between Computer-Aided and Computer-Hindered Decisions: A Case Study in Mammography' (2013) 33(1) Medical Decision Making, 98–107.

Prainsack Barbara and Buyx Alena, 'Thinking Ethical and Regulatory Frameworks in Medicine From the Perspective of Solidarity on Both Sides of the Atlantic' (2016) 37(6) Theoretical Medicine and Bioethics, 489–501.

Priaulx Nicolette, 'Joy to the World! A (Healthy) Child Is Born! Reconceptualizing Harm in Wrongful Conception' (2004) 13(1) Social & Legal Studies, 5–26.

Priaulx Nicolette, The Harm Paradox: Tort Law and the Unwanted Child in an Era of Choice (Routledge 2014).

Price David, 'Remodelling the Regulation of Postmodern Innovation in Medicine' (2005) 1(2) International Journal of Law in Context, 121–141.

Price II W. N., 'Artificial Intelligence in Health Care: Applications and Legal Implications.' (2017) 14(1) The SciTech Lawyer, 10–13.

Price II W. N., 'Distributed Governance of Medical AI' (2022) 25(1) SMU Science & Technology Law Review, 3–22.

Price II W. N., 'Medical AI and Contextual Bias' (2019) 33(1) Harvard Journal of Law and Technology, 65–116.

Price II W. N., Sachs Rachel E. and Eisenberg Rebecca S., 'New Innovation Models in Medical AI' (2022) (2022) 99(4) Washington University Law Review, 1121– 1173.

Price Monroe E., 'The Newness of New Technology' (2001) 22(5-6) Cardozo Law Review, 1885–1913.

Pugh Jonathan, Autonomy, Rationality, and Contemporary Bioethics (Oxford University Press 2020).

Purshouse Craig, 'Autonomy, Affinity, and the Assessment of Damages: ACB v Thomson Medical Pte Ltd [2017] SGCA 20 and Shaw v Kovak [2017] EWCA Civ 1028' (2018) 26(4) Medical Law Review, 675–692.

Purshouse Craig, 'How Should Autonomy Be Defined in Medical Negligence Cases?' (2015) 10(4) Clinical Ethics, 107–114.

Purshouse Craig, 'Judicial Reasoning and the Concept of Damage: Rethinking Medical Negligence Cases' (2015) 15(2-3) Medical Law International, 155–181.

Purshouse Craig, 'Liability for Lost Autonomy in Negligence: Undermining the Coherence of Tort Law?' (2015) 22(3) Torts Law Journal, 226–249.

Purshouse Craig, 'Should Lost Autonomy be Recognised as Actionable Damage in Medical Negligence Cases?' (Thesis for the degree of PhD in Bioethics and Medical Jurisprudence, University of Manchester 2015).

Purshouse Craig, 'The Impatient Patient and the Unreceptive Receptionist: Darnley v Croydon Health Services NHS Trust [2018] UKSC 50' (2019) 27(2) Medical Law Review, 318–329.

Raz Joseph, 'Legal Principles and the Limits of Law' (1972) 81(5) Yale Law Journal, 823–854.

Reidenberg Joel R., 'Governing Networks and Rule-Making in Cyberspace' (1996) 45(3) Emory Law Journal, 911–930.

Reidenberg Joel R., 'Lex Informatica: The Formulation of Information Policy Rules through Technology' (1997) 76(3) Texas Law Review, 553–594.

Reimann Mathias and Zimmermann Reinhard (eds), The Oxford Handbook of Comparative Law (Second Edition, Oxford University Press 2019).

Rhodes Rosamond, Francis Leslie P. and Silvers Anita (eds), The Blackwell Guide to Medical Ethics (John Wiley & Sons 2008).

Richards Bernadette J., 'Autonomy and the Law: Widely Used Poorly Defined' in Kirchhoffer and Richards, Beyond Autonomy (Cambridge University Press 2019).

Rid Annette. and Wendler David, 'Risk-Benefit Assessment in Medical Research – Critical Review and Open Questions' (2010) 9(3-4) Law, Probability and Risk, 151–177.

Robertson Andrew and Tang Hang W. (eds), The Goals of Private Law (Hart Publishing 2009).

Robertson Andrew, 'Constraints on Policy-Based Reasoning in Private Law' in Robertson and Tang, The Goals of Private Law (Hart Publishing 2009).

Rubel Alan, Castro Clinton and Pham Adam, Algorithms and Autonomy (Cambridge University Press 2021).

Rudin Cynthia, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) Nature Machine Intelligence, 206–215.

Rudin Cynthia and Ustun Berk, 'Optimized Scoring Systems: Toward Trust in Machine Learning for Healthcare and Criminal Justice' (2018) 48(5) Interfaces, 449–466.

Rudin Cynthia, Chen Chaofan, Chen Zhi and others, 'Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges' (2022) 16 Statistics Surveys, 1–85.

Rumelhart David E., Hinton Geoffrey E. and McClelland James L., 'A General Framework for Parallel Distributed Processing' in Rumelhart, James L. McClelland and PDP Research Group, Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations (1999).

Rumelhart David E., James L. McClelland and PDP Research Group (eds), Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations (1999).

Russell Stuart J., 'Rationality and Intelligence' (1997) 94(1-2) Artificial Intelligence, 57–77.

Rustad Michael L. and Koenig Thomas H., 'Cybertorts and Legal Lag: An Empirical Analysis' (2003) 13(1) Southern California Interdisciplinary Law Journal, 77–140.

Ryan Mark, 'In AI We Trust: Ethics, Artificial Intelligence, and Reliability' (2020) 26(5) Science and Engineering Ethics, 2749–2767.

Saghai Yashar, 'Salvaging the Concept of Nudge' (2013) 39(8) Journal of Medical Ethics, 487–493.

Saporta Adriel, Gui Xiaotong, Agrawal Ashwin and others, 'Benchmarking Saliency Methods for Chest X-Ray Interpretation' [2022](4) Nature Machine Intelligence, 867–878.

Sartor Giovanni, 'Human Rights in the Information Society: Utopias, Dystopias and Human Values' in Azevedo Cunha and others, New Technologies and Human Rights: Challenges to Regulation (Ashgate Publishing 2013).

Savulescu Julian, 'Autonomy, the Good Life, and Controversial Choices' in Rhodes, Francis and Silvers, The Blackwell Guide to Medical Ethics (John Wiley & Sons 2008).

Savulescu Julian, 'Rational Desires and the Limitation of Life-Sustaining Treatment' (1994) 8(3) Bioethics, 191–222.

Scarpazza Cristina, Baecker Lea, Vieira Sandra and others, 'Applications of Machine Learning to Brain Disorders' in Mechelli and Vieira, Machine Learning: Methods and Applications to Brain Disorders (Academic Press 2019).

Schmidhuber Jürgen, 'Deep Learning in Neural Networks: An Overview' (2015) 61 Neural Networks, 85–117.

Schoeman Ferdinand D. (ed), Philosophical Dimensions of Privacy (Cambridge University Press 2009).

Schönberger Daniel, 'Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications' (2019) 27(2) International Journal of Law and Information Technology, 171–203.

Schramme Thomas and Edwards Steven (eds), Handbook of the Philosophy of Medicine (Springer Netherlands 2017).

Schultz Marjorie M., 'From Informed Consent to Patient Choice: A New Protected Interest' (1985) 95(2) The Yale Law Journal, 219–299.

Seabourne Gwen, 'The Role of the Tort of Battery in Medical Law' (1995) 24(3) Anglo-American Law Review, 265–298.

Sejnowski Terrence J., The Deep Learning Revolution (MIT Press 2018).

Selbst Andrew D. and Powles Julia, 'Meaningful Information and the Right to Explanation' (2017) 7(4) International Data Privacy Law, 233–242.

Sharma Nisha, Ng Annie Y., James Jonathan J. and others, 'Large-Scale Evaluation of an AI System as an Independent Reader for Double Reading in Breast Cancer Screening' (2021) Pre-Print, 1–13.

Sharpe Virginia A. and Faden Alan I., Medical Harm: Historical, Conceptual, and Ethical Dimensions of Iatrogenic Illness (Cambridge University Press 2001).

Shell Richard G., 'Contracts in the Modern Supreme Court' (1993) 81(2) California Law Review, 431–530.

Sherrard Michael and Gatt Ian, 'Human Immunodeficiency Virus (HIV) Antibody Testing: Guidance from an Opinion Provided for the British Medical Association' (1987) 295(6603) British Medical Journal, 911–912.

Shortreed Susan M., Laber Eric, Lizotte Daniel J. and others, 'Informing Sequential Clinical Decision-Making Through Reinforcement Learning: An Empirical Study' (2011) 84(1-2) Machine Learning, 109–136.

Siegel Reva B. and Greenhouse Linda, Before Roe v. Wade: Voices that Shaped the Abortion Debate Before the Supreme Court's Ruling (Kaplan 2010).

Simons Kenneth W., 'A Restatement (Third) of Intentional Torts' (2006) 48 Arizona Law Review, 1061–1102.

Smith Stephen A., 'Duties, Liabilities, and Damages' (2012) 125(7) Harvard Law Review, 1727–1756.

Sommer Joseph H., 'Against Cyberlaw' (2000) 15(3) Berkeley Technology Law Journal, 1145–1232.

Somsen Han, 'Regulating Human Genetics in a Neo-Eugenic Era' in Murphy, New Technologies and Human Rights (Oxford University Press 2009).

Spindelman Marc, 'Are the Similarities between a Woman's Right to Choose an Abortion and the Alleged Right to Assisted Suicide Really Compelling' (1996) 29(3) University of Michigan Journal of Law Reform, 775–856.

Stevens Laura M., Mortazavi Bobak J., Deo Rahul C. and others, 'Recommendations for Reporting Machine Learning Analyses in Clinical Research' (2020) 13(10) Circulation: Cardiovascular Quality and Outcomes, 782-793.

Stoljar Natalie, 'Informed Consent and Relational Conceptions of Autonomy' (2011) 36(4) The Journal of Medicine and Philosophy, 375–384.

Strauß Stefan, 'Deep Automation Bias: How to Tackle a Wicked Problem of AI?' (2021) 5(2) Big Data and Cognitive Computing, 1–14.

Sujan Mark, Furniss Dominic, Grundy Kath and others, 'Human Factors Challenges for the Safe Use of Artificial Intelligence in Patient Care' (2019) 26(1) BMJ Health & Care Informatics, 1–5.

Sun Ron, 'Connectionism and Neural Networks' in Frankish and Ramsey, The Cambridge Handbook of Artificial Intelligence (Cambridge University Press 2014).

Sunstein Cass R., After the Rights Revolution: Reconceiving the Regulatory State (Harvard University Press 1993).

Sunstein Cass R., Legal Reasoning and Political Conflict (Oxford University Press 1998).

Sutherland Lauren, 'Montgomery: Myths, Misconceptions and Misunderstanding' [2019](3) Journal of Personal Injury Law, 157–167.

Sutherland Lauren, 'The Law Finally Reflects Good Professional Practice: Montgomery v Lanarkshire Health Board' [2015](123) Reparation Bulletin, 4–8.

Sutton Richard S. and Barto Andrew, Reinforcement Learning: An Introduction (Second Edition, MIT Press 2018).

Syrett Keith, 'Institutional Liability' in Laing and McHale, Principles of Medical Law (Fourth Edition. Oxford University Press 2017).

Szolovits Peter, Artificial Intelligence in Medicine (Westview Press 1982).

Terrion Halle F., 'Informed Choice: Physicians' Duty to Disclose Nonreadily Available Alternatives' (1993) 43(2) Case Western Reserve Law Review, 491–524.

Thaler Richard H. and Sunstein Cass R., Nudge (The Final Edition, Allen Lane 2021).

Theodoridis Sergios, Machine Learning: A Bayesian and Optimization Perspective (Second Edition, Elsevier 2020).

Thomasian Nicole M., Eickhoff Carsten and Adashi Eli Y., 'Advancing Health Equity with Artificial Intelligence' (2021) 42(4) Journal of Public Health Policy, 602–611.

Throckmorton Archibald H., 'Damages for Fright' (1923) 57(6) American Law Review, 828–853.

Todd Stephen, 'Common Law Protection for Injury to a Person's Reproductive Autonomy' (2019) 135 Law Quarterly Review, 635–659.

Topol Eric J., Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again (Basic Books 2019).

Topol Eric J., 'High-Performance Medicine: The Convergence of Human and Artificial Intelligence' (2019) 25(1) Nature Medicine, 44–56.

Tranter Kieran, 'The Law and Technology Enterprise: Uncovering the Template to Legal Scholarship on Technology' (2011) 3(1) Law, Innovation and Technology, 31–83.

Tribe Laurence H., 'Technology Assessment and the Fourth Discontinuity: The Limits of Instrumental Rationality' (1973) 46(3) Southern California Law Review, 617–660.

Tsoukias Alexis, 'Social Responsibility of Algorithms: An Overview' in Papathanasiou, Zaraté and Freire de Sousa, EURO Working Group on DSS: A Tour of the DSS Developments Over the Last 30 Years (Springer International Publishing 2021).

Turner Jacob, Robot Rules: Regulating Artificial Intelligence (Springer International Publishing 2019).

Turton Gemma, 'Informed Consent to Medical Treatment Post-Montgomery: Causation and Coincidence' (2019) 27(1) Medical Law Review, 108–134.

Twerski Aarib D. and Cohen Neil B., 'Informed Decision Making and the Law of Torts: The Myth of Justiciable Causation' [1988](3) University of Illinois Law Review, 607–666.

Vansweevelt Thierry and Glover-Thomas Nicola (eds), Informed Consent and Health: A Global Analysis (Edward Elgar Publishing 2020).

Varuhas J. N., 'The Concept of 'Vindication' in the Law of Torts: Rights, Interests and Damages' (2014) 34(2) Oxford Journal of Legal Studies, 253–293.

Veatch Robert M., 'Doctor Does Not Know Best: Why in the New Century Physicians Must Stop Trying to Benefit Patients' (2000) 25(6) The Journal of Medicine and Philosophy, 701–721.

Vieira Sandra, Pinaya Walter H. L. and Mechelli Andrea, 'Introduction to Machine Learning' in Mechelli and Vieira, Machine Learning: Methods and Applications to Brain Disorders (Academic Press 2019).

Vieira Sandra, Pinaya Walter H. L. and Mechelli Andrea, 'Main Concepts in Machine Learning' in Mechelli and Vieira, Machine Learning: Methods and Applications to Brain Disorders (Academic Press 2019).

Vieira Sandra, Pinaya Walter H. L., Garcia-Dias Rafael and others, 'Deep Neural Networks' in Mechelli and Vieira, Machine Learning: Methods and Applications to Brain Disorders (Academic Press 2019).

Voter Andrew F., Meram Ece, Garrett John W. and others, 'Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Intracranial Hemorrhage' (2021) 18(8) Journal of the American College of Radiology, 1143–1152.

Wachter Sandra, Mittelstadt Brent and Floridi Luciano, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) International Data Privacy Law, 76–99.

Wachter Sandra, Mittelstadt Brent, Russell and others, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2017) 31(1) Harvard Journal of Law & Technology, 841–887.

Wagemans Jean H. M., 'The Assessment of Argumentation from Expert Opinion' (2011) 25(3) Argumentation, 329–339.

Wall Jesse, 'How the Philosophy Gets In' in Phillips, Campos and Herring, Philosophical Foundations of Medical Law (Oxford University Press 2019).

Walton Douglas N., Appeal to Expert Opinion: Arguments from Authority (Pennsylvania State University Press 1997).

Wang Pei, 'On Defining Artificial Intelligence' (2019) 10(2) Journal of Artificial General Intelligence, 1–37.

Wang Pei, 'What Do You Mean by "AI"?' in Wang, Goertzel and Franklin, Artificial General Intelligence 2008: Proceedings of the First AGI Conference (IOS Press 2008).

Wang Pei, Goertzel Ben and Franklin Stan (eds), Artificial General Intelligence 2008: Proceedings of the First AGI Conference (IOS Press 2008).

Weisbard Alan J., 'Informed Consent: The Law's Uneasy Compromise with Ethical Theory' (1986) 65(4) Nebraska Law Review, 749–767.

Westbrook Johanna I., Coiera Enrico W. and Gosling A. S., 'Do Online Information Retrieval Systems Help Experienced Clinicians Answer Clinical Questions?' (2005) 12(3) Journal of the American Medical Informatics Association, 315–321.

Wicks Elizabeth, Human Rights and Healthcare (Hart Publishing 2007).

Winner Langdon, 'Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology' (1993) 18(3) Science, Technology, & Human Values, 362–378.

Winston Kenneth I. (ed), The Principles of Social Order: Selected Essays of Lon L. Fuller (Revised Edition, Hart Publishing 2001).

Witkin Bernard E., Summary of California Law (Eleventh Edition, Thomas Reuters 2022).

Witzleb Normann and Caroll Robyn, 'The Role of Vindication in Torts Damages' (2009) 17(1) Tort Law Review, 16–44.

Wolf Susan M., 'Shifting Paradigms in Bioethics and Health Law: The Rise of a New Pragmatism' (1994) 20(4) American Journal of Law & Medicine, 395–415.

Wolf Susan, Freedom Within Reason (Oxford University Press 1993).

Wood Elena A., Ange Brittany L. and Miller D. D., 'Are We Ready to Integrate Artificial Intelligence Literacy into Medical School Curriculum: Students and Faculty Survey' (2021) 8 Journal of Medical Education and Curricular Development, 1-5.

Wulfovich Sharon and Meyers Arlen (eds), Digital Health Entrepreneurship (Springer Cham 2020).

Yang Xi, Bian Jiang, Hogan William R. and others, 'Clinical Concept Extraction Using Transformers' (2020) 27(12) Journal of the American Medical Informatics Association, 1935–1942.

Yu Kun-Hsing, Beam Andrew L. and Kohane Isaac S., 'Artificial Intelligence in Healthcare' (2018) 2(10) Nature Biomedical Engineering, 719–731.

Zacher Hans F., 'Das soziale Staatsziel' in Zacher, Abhandlungen zum Sozialrecht (Müller Juristischer Verlag 1993).

Zacher Hans F., Maydell Bernd von and Eichenhofer Eberhard (eds), Abhandlungen zum Sozialrecht (Müller Juristischer Verlag 1993).

Zednik Carlos, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34(2) Philosophy & Technology, 265–288.

Zweigert Konrad and Kötz Hein, Einführung in die Rechtsvergleichung: Auf dem Gebiete des Privatrechts (Third Edition, Mohr Siebeck 1996).

II. Material

'FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency' (2021).

'Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining' (2015).

'The 2021 International Symposium on Networks, Computers and Communications: October 31-November 2, 2021, Dubai, UAE' (2021).

AI Healthcare Coalition, 'AI Healthcare Coalition Appreciates CMS' Efforts to Support Access to Innovative AI Services' (2021) <https://ai-coalition.org/news/ai-healthcare-coalition-appreciates-cms-efforts-to-support-access-to-innovative-ai-services> accessed 26.3.2023.

Annette Markham and others (eds), 'AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society' (2020).

Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2018).

Bhatt Umang, Xiang Alice, Sharma Shubham and others, 'Explainable Machine Learning in Deployment' (Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27.01.2020-30.01.2020).

Caruana Rich, Lou Yin, Gehrke Johannes and others, 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission' (Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney NSW Australia, 10.8.2015-13.8.2015).

Chang Anthony, 'Common Misconceptions and Future Directions for AI in Medicine: A Physician-Data Scientist Perspective' in Riaño, Wilk and Teije, Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings (2019).

Chaudhury Hassan, 'AI in Health in the United Kingdom: An Overview for SME's and Research Institutes on the Trends, Challenges and Opportunities for AI Applications in the British Healthcare Sector' (Market Report, 2021) <https://www.rvo.nl/sites/default/files/2021/06/AI-in-Health-UK-market-report_0.pdf> accessed 26.3.2023.

Coghlan Andy, 'Could HIV tests land doctors in court?' (19.1.1994) <https://www.newscientist.com/article/mg14119100-600-could-hiv-tests-land-doctors-in-court/> accessed 9.3.2021.

Deutscher Ethikrat, 'Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz' (Stellungnahme, 2023).

Di Nucci Ezio, Jensen Rasmus T. and Tupasela Aaro, 'Ethics of Medical AI: The Case of Watson for Oncology' (English version of the paper — Kunstig Intelligens og Medicinsk Etik: Tilfaeldet Watson for Oncology, to be printed in the volume 8 Cases i Medicinsk Etik 5.12.2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3432317> accessed 5.4.2020.

Digital Diagnostics, 'IDx-DR' <https://www.digitaldiagnostics.com/products/eye-disease/idx-dr/> accessed 7.3.2022.

Edwards Lifesciences, 'Edwards' Acumen Hypotension Prediction Index Launches In The U.S.' (18.3.2022) <https://www.edwards.com/ns20180319> accessed 7.3.2022.

Fibricheck, 'What is FibriCheck and how does it work?' <https://www.fibricheck.com/what-is-fibricheck-and-how-does-it-work/> accessed 7.3.2022.

Franke Ulrike, 'Artificial Intelligence Diplomacy: Artificial Intelligence Governance as a New European Union External Policy Tool' (Study Requested by the AIDA Committee, 2021) <https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2021)662926> accessed 26.3.2023.

General Medical Council, 'Consent: Patients and Doctors Making Decisions Together' (London 2008).

General Medical Council, 'Decision Making and Consent' (London 2020).

GOV.UK, 'Regulating medical devices in the UK: What you need to do to place a medical device on the Great Britain, Northern Ireland and European Union (EU) markets' (Guidance 1.1.2022) <https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk> accessed 7.3.2022.

Grgic-Hlaca Nina, Redmiles Elissa M., Gummadi Krishna P. and others, 'Human Perceptions of Fairness in Algorithmic Decision Making' (WWW '18: Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23.04.2018-27.4.2018).

Hamon Ronan, Junklewitz Henrik, Malgieri Gianclaudio and others, 'Impossible Explanations?: Beyond Explainable AI in the GDPR From a COVID-19 Use Case Scenario' (FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, 03.03.2021 - 10.03.2021).

Human Fertilisation and Embryology Authority, 'Code of Practice' (2021 Ninth Edition) <https://portal.hfea.gov.uk/knowledge-base/read-the-code-of-practice/> accessed 26.3.2023.

IBM, '5725-W51 IBM Watson for Oncology: Sales Manual' (2020) <https://www.ibm.com/common/ssi/cgi-bin/ssialias?appname=skmwww&htmlfid=897%2FENUS5725-W51&infotype=DD&subtype=SM&mhsrc=ibmsearch_a&mhq=IBM%20WATSON%20ONcology> accessed 18.3.2023.

IDx LLC, 'Fully Automated Diagnostic Device Receives CE Certification; IDx LLC Planning For Rollout Across Europe' (6.5.2013) <https://www.prnewswire.com/news-releases/fully-automated-diagnostic-device-receives-ce-certification-idx-llc-planning-for-rollout-across-europe-206263101.html> accessed 7.3.2022.

Information Commissioner's Office, 'What are 'controllers' and 'processors'?' (17.10.2022) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/controllers-and-processors/what-are-controllers-and-processors/#1> accessed 18.3.2023.

International World Wide Web Conference Committee (ed), 'WWW '18: Proceedings of the 2018 World Wide Web Conference' (2018).

Johns Hopkins Technology Ventures, 'Digital Health Startup That Assists Emergency Department Decision Making Acquired' (2022) accessed 17.3.2023.

Joint Committee on the Draft Mental Incapacity Bill, 'Draft Mental Incapacity Bill: Session 2002–03 Volume I' (2003).

Lakkaraju Himabindu and Bastani Osbert, '"How Do I Fool You?": Manipulating User Trust via Misleading Black Box Explanations' (AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 07.02.2020-09.02.2020).

Laugel Thibault, Lesot Marie-Jeanne, Marsala Christophe and others, 'The Dangers of Post-Hoc Interpretability: Unjustified Counterfactual Explanations' (Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 8.10.2019-8.16.2019).

Law Commission, 'Liability for Psychiatric Illness' (Law Commission Consultation Paper No 137, 1995).

Mann Brian, 'Health Care Software Firm Fined $145M In Opioid Scheme With Drug Companies' (1.2.2020) <https://www.npr.org/2020/02/01/801832788/healthcare-software-firm-fined-145m-in-opioid-scheme-with-drug-companies?t=1615393792393> accessed 10.3.2021.

Matheny Michael, Israni Sonoo T., Ahmed Mahnoor and others, 'Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril' (NAM Special Publication 2019) <https://nam.edu/artificial-intelligence-special-publication/> accessed 5.4.2020.

Marchiori Chiara, Dykeman Douglas, Girardi Ivan and others, 'Artificial Intelligence Decision Support for Medical Triage' [2020] AMIA Annual Symposium Proceedings, 793–802.

MaxQ AI, Ltd, 'MaxQ-AI Receives CE Mark Approval for Accipio™Ix Intracranial Hemorrhage Artificial Intelligence Software Platform' (22.5.2018) <https://www.prnewswire.com/news-releases/maxq-ai-receives-ce-mark-approval-for-accipioix-intracranial-hemorrhage-artificial-intelligence-software-platform-300652488.html> accessed 7.3.2022.

MaxQ Artificial Intelligence, 'ACCIPIO®—Solution Architecture and Design: A White Paper' <https://www.maxq.ai/resources> accessed 7.3.2022.

Mireille Hildebrandt and others (eds), 'FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency: January 27-30, 2020, Barcelona, Spain' (2020).

NHS Transformation Directorate, 'The NHS AI Lab' <https://transform.england.nhs.uk/ai-lab/> accessed 26.3.2023.

NHSX, 'NCCID case study: Setting standards for testing Artificial Intelligence' (21.2.2022) <https://www.nhsx.nhs.uk/ai-lab/explore-all-resources/develop-ai/nccid-case-study-setting-standards-for-testing-artificial-intelligence/> accessed 6.3.2022.

Nix Mike, Onisiforou George and Painter Annabelle, 'Understanding Healthcare Workers' Confidence in AI' (Report 1 of 2 2022) <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai> accessed 11.11.2022.

Nuffield Council on Bioethics, 'Bioethics Briefing Note: Artificial Intelligence (AI) in Healthcare and Research' (2018) <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf> accessed 17.6.2022.

Okay Feyza Y., Yildirim Mustafa and Ozdemir Suat, 'Interpretable Machine Learning: A Case Study of Healthcare' (2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 10.31.2021-11.2.2021).

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, 'Deciding to Forego Life-Sustaining Treatment: A report on the Ethical, Medical and Legal Issues in Treatment Decisions' (Washington, DC 1983).

Riaño David, Wilk Szymon and Teije Annette ten (eds), Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings (2019).

Rohaidi Nurfilzah, 'IBM's Watson Detected Rare Leukemia In Just 10 Minutes' (16.8.2016) <https://www.asianscientist.com/2016/08/topnews/ibm-watson-rare-leukemia-university-tokyo-artificial-intelligence/> accessed 4.9.2022.

Ross Casey and Swetlitz Ike, 'IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close' (5.9.2017) <https://www.statnews.com/2017/09/05/watson-ibm-cancer/> accessed 28.3.2023.

Sarit Kraus (ed), 'Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence' (2019).

Schemmer Max, Kühl Niklas, Benz Carina and others, 'On the Influence of Explainable AI on Automation Bias: Research in Progress' (19.4.2022) <https://arxiv.org/pdf/2204.08859> accessed 6.6.2022.

Selanikio Joel, 'A Closer Look at FDA's AI Medical Device Approvals' (12.10.2022) <https://www.futurehealth.live/blog/2022/10/10/closer-look-at-fda-ai-approvals> accessed 19.3.2023.

Siemens Healthineers, 'AI-Pathway Companion Prostate Cancer from Siemens Healthineers approved for use in Europe as medical device' (3.3.2020) <https://www.siemens-healthineers.com/fr-be/press-room/press-releases/pr-aipathwaycomp-ce.html> accessed 7.3.2022.

Smart Andrew, James Larry, Hutchinson Ben and others, 'Why Reliabilism Is Not Enough: Epistemic and Moral Justification in Machine Learning' (AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, USA, 07.02.2020-09.02.2020).

Stember Joseph and Shalu Hrithwik, 'Deep Reinforcement Learning With Automated Label Extraction From Clinical Reports Accurately Classifies 3D MRI Brain Volumes' (17.6.2021) <https://arxiv.org/pdf/2106.09812> accessed 6.3.2022.

Tayo Benjamin O., 'Simplicity vs Complexity in Machine Learning — Finding the Right Balance' (11.11.2019) <https://towardsdatascience.com/simplicity-vs-complexity-in-machine-learning-finding-the-right-balance-c9000d1726fb> accessed 6.3.2022.

U.S. Food & Drug Administration, 'Accipiolx 510(k) Summary' (K182177 26.20.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K182177> accessed 7.3.2022.

U.S. Food & Drug Administration, 'Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices' (5.10.2022) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices> accessed 19.3.2023.

U.S. Food & Drug Administration, 'Artificial Intelligence and Machine Learning in Software as a Medical Device' (2021) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> accessed 6.3.2022.

U.S. Food & Drug Administration, 'De Novo Classification Request for Acumen Hypotension Prediction Index Feature Software' (De Novo Summary (DEN160044) 16.3.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN160044> accessed 7.3.2022.

U.S. Food & Drug Administration, 'De Novo Classification Request for IDx-DR' (De Novo Summary (DEN180001) 11.4.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?ID=DEN180001> accessed 7.3.2022.

U.S. Food & Drug Administration, 'FibriCheck 510(k) Summary' (K173872 28.9.2018) <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K173872> accessed 7.3.2022.

United Lincolnshire Hospitals NHS Trust, 'ULHT trialling artificial intelligence software to support breast cancer screening' (16.8.2019) <https://www.ulh.nhs.uk/news/ulht-trialling-artificial-intelligence-software-to-support-breast-cancer-screening/> accessed 7.3.2022.

World Economic Forum, 'The 'AI divide' between the Global North and Global South' (16.1.2023) <https://www.weforum.org/agenda/2023/01/davos23-ai-divide-global-north-global-south/> accessed 26.3.2023.

Zhang Daniel, Maslej Nestor and Brynjolfsson Erik and others, 'Artificial Intelligence Index Report 2022' (2022) <https://aiindex.stanford.edu/report/> accessed 26.3.2023.