Julia Maria Mönig

I. Ethics by Design: Grundlagen und ethische Aspekte

Ethics by Design ist ein Ansatz zur Technikgestaltung, der sicherstellen soll, dass Werte von Beginn an und während des gesamten Lebenszyklus' einer Technologie in deren Gestaltung miteinbezogen werden. Der Begriff wird seit den späten 2010er Jahren im europäischen Kontext insbesondere in Bezug auf die Entwicklung von menschenzentrierten, vertrauenswürdigen Künstlichen Intelligenz (KI)-Anwendungen verwendet. Im deutschsprachigen Diskurs bezieht sich »Ethics by Design« in einem weiteren Sinne auf die Gestaltung von Technologien im Allgemeinen. Begriffsgeschichtlich handelt es sich »vermutlich« (Brey & Dainow, 2023) um eine Verallgemeinerung des Prinzips »*privacy by design*«.

Übergeordnetes Ziel neben der Verhinderung von Verletzungen ethischer Werte und Grundlagen der liberal-demokratischen Ordnung ist in der Europäischen Union (EU) die Idee, dass Ethik nicht, wie ihr oft unterstellt wird, Innovation verhindere, sondern gerade im internationalen Wettbewerb einen »*unique selling point*« und »*key asset*« darstellen könne. KI »*made in Europe*« hätte somit den Vorteil einer Garantie dafür, dass Werte bewahrt und somit das allgemeine Persönlichkeitsrecht und letztlich die Menschenwürde geschützt werden. Dies lässt sich auch diesbezüglich verstehen, dass im globalen Wettbewerb um KI die USA und China aufgrund anderer zugrundeliegender rechtlicher, politischer und wirtschaftlicher Systeme einen uneinholbaren technischen Fortschritt haben.

Die vorliegenden Ansätze sind teilweise vor beziehungsweise während der Arbeit an der europäischen KI-Verordnung formuliert worden. Diese beruht in Bezug auf ethische Grundsätze u. a. auf der Vorarbeit der explizit von der Europäischen Kommission eingesetzten *High Level Expert Group on AI (AI-HLEG)*, auf Statements der *European Group on Ethics in Science and New Technologies (EGE)*,

sowie auf diversen Kommentaren im Rahmen der *Public consultation* und den Anmerkungen aus Parlament und Rat während der verschiedenen Iterationen der Entscheidungsfindung für den finalen Gesetzestext.

1. Ethics by Design – Ursprünge und Grundlagen

Ausgangspunkt der Idee, dass Technik durch ihre Gestaltung zum Guten bzw. auf gesellschaftlich gewünschte Auswirkungen hin beeinflusst werden kann, ist die Beobachtung, dass Technik nie neutral ist. Das sog. Erste Gesetz des Technologiehistorikers Malvin Kranzberg lautet: »Technology is neither good nor bad; nor is it neutral« (Kranzberg, 1986). In Zusammenhang mit der Annahme, dass Technik nie neutral ist, steht auch die Überzeugung, dass Technologien im Allgemeinen und KI-Systeme im Besonderen nie isoliert zu betrachten sind, sondern immer Teil eines soziotechnischen Systems sind (Dignum, 2020, S. 216).

1.1 Privacy / X by Design

Privacy by Design wurde als Konzept in den 1990er Jahren von der ehemaligen Datenschutzbeauftragten (*Information and Privacy Commissioner*) der kanadischen Provinz Ontario, Ann Cavoukian, veröffentlicht. Privacy by Design umfasst sieben Prinzipien: u. a., dass es proaktiv und nicht reaktiv sein solle und »by default«, also standardmäßig von vorneherein gelten solle (Cavoukian, 2011). Die im Jahre 2016 in Kraft getretene europäische Datenschutzgrundverordnung schreibt »data protection by design« und »data protection by default« vor (Verordnung 2016/679). In der deutschsprachigen Fassung werden die Begriffe »Datenschutz durch Technikgestaltung« (Art. 25 DSGVO) bzw. »Datenschutz durch Technik« (Erwägungsgrund 78 DSGVO) und »datenschutzfreundliche Voreinstellungen« (Art. 25 DSGVO) verwendet (Verordnung 2016/679). Eine deutsche Übersetzung des Begriffs »Ethics by Design« könnte also »Ethik durch Technikgestaltung« lauten.

Neben »Privacy by Design« hat sich eine ganze »by Design«-Familie (Nurock et al., 2021) etabliert, auch subsumiert unter dem

Begriff »X-by-design« (AI-HLEG, 2019). Unter anderem umfasst sie »Safety by Design«, »Security by Design« oder auch »Human Rights by Design«.

1.2 Benachbarte Konzepte

Neben Ethics by Design gibt es zahlreiche andere Konzepte, die darauf abzielen, Produkte, Prozesse, Unternehmen und Technologien durch Gestaltungsprozesse und bereits während der Designphase ethisch(er) und orientiert an liberal-demokratischen Werten zu gestalten. Die Idee, dass Ethiker*innen als Designer*innen (»ethicist as designer«, van Wynsberghe & Robbins, 2014) fungieren sollen, Design ethisch ausgerichtet sein solle (»ethically aligned design«, Institute of Electrical and Electronics Engineers [IEEE], 2018), sowie die Forderung, dass technologiebasierte Produkte und Dienste qua Design ethisch (»Ethic*al* by Design«, Nurock et al., 2021) sein sollen [Hervorhebung von mir, JMM], spiegeln nicht nur einen allgemeinen »*ethics turn*« wider, der in den späten 2010er Jahren durch ein erhöhtes Interesse an KI- und Algorithmenethik deutlich wurde (Dignum et al., 2018) und sich in zahlreichen Leitlinien (*guidelines*) und Ethik-Kodizes verschiedenster Provenienz manifestierte,[1] sondern auch einen »*design turn in applied ethics*« (van den Hoven, 2017). Darüber hinaus gibt es Design-Ethik in einem weiteren Sinne, beispielsweise als Berufsethos von Menschen in Designberufen. Des Weiteren gibt es Fälle und Veröffentlichungen, in denen der Begriff »Ethics by Design« nicht für Technologieentwicklung, sondern z. B. in Bezug auf ethische Unternehmens- und Organisationsentwicklung verwendet wird (vgl. z.B. Moore, 2010). Verwandte Konzepte umfassen *Values in Design* (Simon, 2016) und, prominent, *Value-based Design* (vgl. unten). Das *Institut für Digitale Ethik (IDE)* an der Hochschule der Medien Stuttgart definiert Ethics by Design als »beruhend auf den Methoden und Konzepten eines Value Sensitive Design« (IDE, o. J.). Sarah Spiekermann und Till Winkler schlagen

1 Verschiedenste Akteur*innen aus Wirtschaft, Wissenschaft und Gesellschaft, einschließlich großer (Digital-) Unternehmen und NGOs veröffentlichten eigene Ethik-Richtlinien. Corrêa et al. (2023) identifizierten 2023 200 Richtlinien und Empfehlungen für KI-Governance.

die Praxis des »Value-based Engineering« zum Erreichen von Ethics by Design vor (2020).

Wie einige Bezeichnungen bereits vermuten lassen, sind die meisten Ansätze werteorientiert. In der Literatur wird die Art der betrachteten Werte dabei (teilweise) näher beschrieben:[2] moralische, gesellschaftliche und rechtliche Werte (Dignum et al., 2018), sowie »*human values*« in Abgrenzung zu »*functional values*«, die auch von anderen, nicht spezifisch ethischen Ansätzen wie User-centered Design als Grundlage angesetzt werden. Im Values for Design-Ansatz wird davon ausgegangen, dass bestehende Designmethoden Werte von Nutzenden und gesellschaftliche Werte realisieren, wobei es sich im Allgemeinen eben nicht um moralische Werte handelt, in manchen Fällen jedoch moralische Werte beinhaltet sind (Vermaas et al., 2015); deutlich wird dies durch die terminologische Unterscheidung zwischen »user values«, »social values« und »moral values«. Des Weiteren wird davon ausgegangen, dass die Methoden entweder »*designer-driven*« oder »*user-driven*« sein können. Im ersteren Falle reflektieren und definieren die Gestaltenden oder die Auftraggebenden die einfließenden, umzusetzenden und resultierenden Werte. Die Verantwortung für das Design liegt also bei der bzw. dem Designer*in. Es können auch explizit nicht nur Werte der Nutzenden, sondern gesellschaftliche Werte (*social values*) in den Gestaltungsprozess einfließen und diesen leiten. Im letzteren Fall liegt der Schwerpunkt darauf, die Wünsche, aber auch Besorgnisse von Nutzenden zu berücksichtigen. Explizit umgesetzt wird dies mit partizipativen Ansätzen, die – potentielle – Anwender*innen in die Gestaltung einbinden. Die Werte und Prinzipien bedingen sich teilweise gegenseitig.

Als politisches Steuerungsinstrument im Rahmen der Forschungsförderung, das jedoch einen ganzheitlicheren Ansatz verfolgt, wird seit Anfang der 2010er Jahre das Konzept von *Responsible Research and Innovation (RRI)* im europäischen Rahmen eingesetzt (von Schomberg & Hankins, 2019). In dieser Hinsicht stehen ebenfalls die Begriffe ELSI (*Ethical, Legal and Social Implications*) und ELSA (*Ethical, Legal and Social Aspects*), die die ethischen, rechtlichen und

2 Vgl. auch Abschnitt 3. (»Ethics by Design als Teil der Digitalen Ethik und der Wertebegriff«) des zweiten Teils (Umsetzung, Potenziale und Grenzen) des vorliegenden Sachstandsberichts.

sozialwissenschaftlichen Aspekte bzw. Implikationen in Forschungsprojekten adressieren sollen, jedoch teilweise auch weiterhin, parallel zu oder in Kombination mit Ethics by Design verwendet werden.[3]

Kritisiert wurde am ELSI-Ansatz, dass er oft parallel zu den technischen Forschungen laufe und Erkenntnisse u. U. erst am Schluss der Forschungsförderphase vorlägen, so dass diese nicht (mehr) in das Projekt und somit in die Technikgestaltung einfließen könnten (d'Aquin et al., 2018; vgl. auch unten). Um zu vermeiden, dass ELS-Aspekte nur nebenherlaufen, wird ebenfalls der Ansatz der »integrierten Forschung« verfolgt (Gransche & Manzeschke, 2020). In zwei Handbüchern aus den benachbarten Feldern, die jedoch aus den Jahren 2015 und 2019 stammen, wird Ethics by Design (noch) nicht als Begriff erwähnt (van den Hoven et al., 2015; von Schomberg & Hankins, 2019).

Neben Ethics *by* Design ist verschiedentlich auch die Rede von Ethics *in* Design (Datenethikkommission, 2019; Deutscher Ethikrat, 2020) und zusätzlich von Ethics *for* Design (Dignum, 2020). Virginia Dignum definiert die drei Ausprägungen folgendermaßen, wobei entsprechend ihres Artikels von 2018 von Ethics by Design für autonome Systeme ausgegangen wird:

> »Ethics by Design: the technical/algorithmic integration of ethical reasoning capabilities as part of the behavior of artificial autonomous systems.
> Ethics in Design: the regulatory and engineering methods that support the analysis and evaluation of the ethical implications of AI systems as these integrate or replace traditional social structures.
> Ethics for Design: the codes of conduct, standards, and certification processes that ensure the integrity of developers and users as they research, design, construct, employ, and manage artificial intelligent systems.« (Dignum, 2020).[4]

3 So hieß beispielsweise im BMBF-geförderten Projekt KoFFI, das am Institut für Digitale Ethik eines der ersten Projekte war, in denen Ethics by Design umgesetzt wurde, das entsprechende Arbeitspaket ELSI (Erbach et al., 2020).

4 Ethics by Design: die technisch/algorithmische Integration von ethischen Denkfähigkeiten als Teil des Verhaltens künstlicher autonomer Systeme. Ethics in Design: die regulatorischen und technischen Methoden, die die Analyse und Bewertung der ethischen Implikationen von KI-Systemen unterstützen, wenn diese traditionelle soziale Strukturen integrieren oder ersetzen. Ethik für Design: die Verhaltenskodizes, Normen und Zertifizierungsprozesse, die die Integrität von Entwickler*innen und Anwender*innen bei der Erforschung, dem Design,

Im Folgenden werden exemplarisch zwei unterschiedliche, mit Ethics by Design verwandte Konzepte näher beleuchtet. Die Wahl der beiden ist durch ihre Unterschiedlichkeit begründet. Das eine ist seit Jahrzehnten etabliert und in der Forschungscommunity bekannt. Es steht ein breiter Korpus an empirischer und theoretischer Forschungsliteratur zur Verfügung, während das andere ein Konzept ist, das kompakt auf einer Seite dargestellt wird und von zwei einzelnen Designer*innen in der Mitte des letzten Jahrzehnts als Antwort auf den zunehmenden Überwachungskapitalismus (Zuboff, 2019) formuliert wurde. Dies zeigt die Breite der bestehenden Herangehensweisen.

1.2.1 Value Sensitive Design

Value Sensitive Design (VSD) ist ein Ansatz, der in den 1990er Jahren formuliert wurde, mit dem Ziel, Werte in das Design von (Informations-)Technologie einzuschreiben und die Aufmerksamkeit auf die moralische und soziale Dimension von Design zu richten: »*shaping technology with moral imagination*« (Friedman & Hendry, 2019). VSD baut auf Erkenntnissen der Mensch-Maschine-Interaktions-Forschung auf, unterscheidet sich jedoch durch seinen Fokus auf moralische Werte von anderen Ansätzen, die funktionale oder instrumentelle Werte betrachten (wie beispielsweise Benutzerfreundlichkeit). Wie Ethics by Design ist VSD proaktiv, iterativ, in den Forschungs- und Entwicklungsprozess integriert und darauf ausgerichtet, Werte bereits in einer frühen Phase des Designprozesses zu berücksichtigen. Ausgangspunkt ist die Annahme, dass jedes einzelne Design spezifische Features, Chancen und Optionen ermöglicht, während andere nicht zum Zuge kommen. Im Unterschied zu Ethics by Design, das in manchen Interpretationen (z.B. derjenigen der Europäischen Kommission), spezifisch auf Künstliche Intelligenz ausgerichtet ist, ist Value Sensitive Design aufgrund einer aktiven Entscheidung hierzu technologieagnostisch (Friedman & Hendry, 2019, S. 41).

---

der Konstruktion, dem Einsatz und dem Management künstlicher intelligenter Systeme sicherstellen.

Im Rahmen des Value Sensitive Design-Ansatzes wird in drei Phasen versucht, Werte in Technologie zu transferieren.[5] Die erste Phase der konzeptuellen Analyse (*conceptual analysis*) ist informiert durch ethische und moralphilosophische Einsichten, die für das vorliegende Design relevant sind. In der zweiten Phase (*empirical mode of investigation*) werden empirische Daten zur Unterstützung der in Phase eins untersuchten Werte mit einbezogen sowie empirische Daten, die Feedback zur Unterstützung der technischen Untersuchung eines spezifischen Designs zur Verfügung stellen (van den Hoven & Manders-Huits, 2017). In der dritten Phase, der technischen Analyse (*technical analysis*), werden technische Designspezifikationen und Variablen untersucht, die bestimmte Werte im Kontext der zu gestaltenden Technologie fördern oder verhindern könnten. Wissentlich oder unwissentlich könnten Entscheidungen während des Designprozesses die moralischen und politischen Implikationen, die eine Technologie in der Praxis haben könnte, beeinflussen (van den Hoven & Manders-Huits, 2017, S. 331).

Eng verwandt mit Value Sensitive Design ist das Konzept »Values at Play«, das ebenfalls dreischrittig vorgeht und in einer Entdeckungsphase (*discovery phase*) versucht, Werte zu identifizieren. Im Anschluss werden in der Übersetzungsphase (*translation phase*) die zuvor identifizierten Werte in die Architektur und Eigenschaften der Technologie übersetzt, bevor in einer Verifizierungsphase (*verification phase*) überprüft wird, ob die Werte erfolgreich implementiert wurden (van den Hoven & Manders-Huits, 2017). Van den Hoven betonte bereits 2005, dass es sich um eine Ausprägung von »*doing ethics*« handle (van den Hoven & Manders-Huits, 2017). Es wird explizit davon ausgegangen, dass sich menschliche Gestaltung des Designs und die Technologien bzw. das Design reziprok beeinflussen (Friedman & Hendry, 2019). Insofern spiegeln alle Technologien bis zu einem gewissen Grad die menschlichen Werte wider und wirken auf sie ein. Laut Friedman und Hendry (2019) wäre es deshalb keine verantwortliche Position, Werte im Designprozess außen vor zu lassen. Im Gegenteil würden gerade kreative Möglichkeiten für technische Innovation und zur Verbesserung der menschlichen Bedingtheit bereitgestellt, wenn Werte im Design berücksichtigt würden.

5 Siehe auch Abschnitt 4.1. (»Praktische Methode des Value Sensitive Design«) des zweiten Teils (Umsetzung, Potenziale und Grenzen) des vorliegenden Sachstandsberichts.

Werte definieren Friedman et al. als das, was für Menschen in ihrem Leben wichtig ist, mit einem Fokus auf Ethik und Moral (Friedman & Hendry, 2019, S. 45). Hierzu wird unter Berücksichtigung ihrer Verbundenheit (*interconnectedness*) ein breites Set an Werten untersucht. Friedman und Hendry geben die folgende Liste von Werten mit ethischer Bedeutung an, die ihrer Meinung nach oft in »System Design« impliziert sind: Menschliches Wohlergehen / Gemeinwohl, Eigentum und Besitz, Privatsphäre, Unvoreingenommenheit, universelle Nutzbarkeit, Vertrauen, Autonomie, informierte Zustimmung, Verantwortlichkeit, Höflichkeit, Identität, Gelassenheit und ökologische Nachhaltigkeit (*Human welfare, ownership and property, privacy, freedom from bias, universal usabiltiy, trust, autonomy, informed consent, accountability, courtesy, identity, calmness and environmental sustainability* (Friedman & Hendry, 2019, S. 50 f.) und berücksichtigen auch, dass es zu Spannungen, Konflikten und Abwägungen zwischen diesen kommen kann. Hinter der Methode steckt eine bewusste Entscheidung, nicht auf spezifische Werte, Technologien, Bevölkerungsgruppen oder Kontexte zu fokussieren. In ihrem Werk von 2019 listen Friedman und Hendry siebzehn Value-sensitive Design-Methoden auf, u.a. Stakeholder-Analyse, *Value source analysis,* bei welcher zwischen den expliziten Projektwerten, den persönlichen und professionellen Werten der Designer*innen sowie den Werten anderer direkter und indirekter Stakeholder unterschieden wird, und beispielsweise einem teilstrukturierten Interview, bei dem die Werte der Befragten gegenüber einer Technologie herausgearbeitet werden (*value-oriented semi-structured interview)* (Friedman & Hendry, 2019, S. 86 ff.)

1.2.2 The Ethical Design Manifesto

Eine Kurzformel für ethisches Technologiedesign stellt das »Ethical Design Manifesto« (Ind.ie, 2016) dar, das als Antwort auf das vorherrschende Paradigma des Überwachungskapitalismus ausbuchstabiert wurde. Ausgehend von Abraham Maslows Theorie der menschlichen Motivation (Maslow, 1943) entwickelte das Non-Profit-Unternehmen *Ind.ie* (heute aufgegangen in der *Small Technology Foundation*) einen Ansatz zu ethischem Design, der sich auf einen Blick darstellen lässt. Laut Aussage des Autors Aral Balkan wird es bereits in verschiedenen Firmen angewendet, bzw. wird in Designfirmen

sichtbar aufgehängt, um die tägliche Arbeit zu inspirieren. Die Spitze der Pyramide, die sich in drei Teile teilt, gipfelt in »Respekt«. Der oberste Teil bezieht sich auf menschliche Erfahrung (»*human experience*«), welche »*delightful*« sein soll (zu deutsch: entzückend, wunderbar, angenehm). Die mittlere Ebene bezieht sich darauf, dass Technologie, die menschliche Anstrengung respektiere, funktional, praktisch und zuverlässig sein solle. Die Basis der Pyramide bilden Menschenrechte. Interessanterweise werden in diesem Bereich, der auf einem abstrakten Begriff fußt, konkrete technisch umsetzbare und operationalisierbare Eigenschaften angegeben, etwa, dass Technologie, welche Menschenrechte respektiere, dezentralisiert und Ende-zu-Ende-verschlüsselt sein solle, oder Software frei und quelloffen sein solle, etc. Hinterfragen lässt sich die mit der Spitze der Pyramide verbundene Forderung, dass Technik »unsichtbar« und »magisch« sein solle. Während es zutreffend ist, dass sie »einfach funktionieren« sollte, gehen von unsichtbarer Technologie (z.B. im Sinne des *Internet of everything*, wenn alle Alltagsgegenstände mit dem Internet verbunden sind) und die durch »Magie« geschaffene Distanz aufgrund von Unerklärbarkeit, eigene ethische und datenschutzbezogene Problem aus.

Tabelle 1: Ethical Design im Überblick

	Respekt		
Technologie ist	Angenehm	Menschliche Erfahrung	Technologie respektiert
	Funktional, praktisch und zuverlässig	Menschliche Anstrengung	
	Dezentralisiert, privat, offen, interoperationabel, zugänglich, sicher und nachhaltig	Menschenrechte	

2. Das Konzept »Ethics by Design«

Die bloße Tatsache, dass Ethik in einem frühen Entwicklungsstadium mitbedacht wird, ist nicht ausschlaggebend für Ethics by Design. Wie d'Aquin et al. aufzeigen, ist dies ebenso der Fall in »klassischen«

Herangehensweisen einiger Fächer (z.B. Sozialwissenschaften, Medizin, Biologie), in denen eine Ethikkommission darüber entscheidet, ob dem Antrag der Forschenden auf Durchführung ihres Projekts nach Prüfung ethischer Gesichtspunkte stattgegeben wird. Dies adressiere ethische Fragen jedoch nicht proaktiv, wie es Ethics by Design in Anlehnung an den ersten Grundsatz von Privacy by Design unternimmt. Des Weiteren gebe es in diesem Falle lediglich eine binäre Entscheidung, ob die Forschung durchgeführt werden dürfe oder nicht, aber keine ethische Forschung. Selbst diese kann ihre Grenzen haben, wenn sie lediglich begleitend erfolgt, wie in früheren Phasen der ELSI-»Begleitforschung«, wie d'Aquin et al. ebenfalls anhand eines Anwendungsbeispiels für das EU-geförderte REVERIE-Projekt ausführen, bei dem eine gesonderte Aufgabe (»*task*«) darin bestand, die potentiellen ethischen Implikationen zu untersuchen (d'Aquin et al., 2018), deren Erkenntnisse jedoch nicht in das Gesamtprojekt einfließen konnten, da sie erst zeitgleich mit den allgemeinen Projektergebnissen vorlagen. Ein Beispiel für einen gelungenen Fall von Ethics by Design war das vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Projekt Kooperative Fahrer-Fahrzeug-Interaktion (KoFFI), in dem Ethics by Design durch verschiedene Maßnahmen umgesetzt wurde, ein Austausch mit den Projektpartnern erfolgte, empirische Studien durch ethische Fragen ergänzt wurden und beispielsweise die zu Beginn des Projekts definierte ethische Wertematrix überarbeitet wurde (Erbach et al., 2020; Grimm & Mönig, 2020).

Vereinzelt wurden auch explizite Instrumente (»*tools*«) sowie automatisierte Assessments zur Implementierung von Ethics by Design entwickelt (vgl. z.B. Urquhart & Craigon, 2021; Mehlich & Woopen, 2025).

2.1 Die Handreichung »Ethics By Design and Ethics of Use Approaches for Artificial Intelligence«

Der »Ethics by Design for artificial intelligence«-Ansatz in seiner vorliegenden Form wurde maßgeblich in den EU-geförderten For-

schungsprojekten SIENNA[6] und SHERPA[7] erarbeitet (Brey & Dainow, 2023).[8] Die Ergebnisse der Projekte (vgl. Jansen et al., 2021; Brey & Dainow, 2023) beruhen auf früheren Ansätzen (namentlich Dignum et al., 2018 und d'Aquin et al., 2018), welche ihrerseits keine vollständig ausformulierte und anwendbare Methodik lieferten. Dignum et al. (2018) beschäftigen sich jedoch mit der Frage nach der Ethik und Moral autonomer Systeme und nehmen somit den Begriff »Ethics by Design« wörtlich, da diesen Maschinen, wenn es einen Konsens gäbe, sie zu programmieren, moralische Urteilsfähigkeiten *qua Design* einprogrammiert werden müssten.

Für die Literaturauswertung zum Forschungsstand wurden Vorarbeiten aus dem SIENNA-Projekt, die Handreichung zur (Selbst-)Evaluierung von EU-geförderten Forschungsprojekten in der Antragstellung »Horizon Europe Ethics Appraisal Procedure for AI«, das IEEE7000–2021 »Standard Model Process for Addressing Ethical Concerns during System Design« (IEEE, 2021) sowie der organisationale Ethics by Design-Ansatz des World Economic Forum (WEF, 2020) zur verantwortungsvollen Verwendung von Technologie ausgewertet (Jansen et al., 2021). Brey und Dainow (2023) beschreiben die Anwendung, das Framework mit Werten und die zugehörigen Anforderungen (*design requirements*) sowie die Implementierung von Ethics by Design. Das eigentliche Framework wird in einer ausführlichen Handreichung für EU-geförderte Forschungsprojekte, die KI-Anwendungen entwickeln oder einsetzen, dargelegt. Es muss jedoch – im Gegensatz zum o.g. Ethics Self-Assessment, nicht verpflichtend angewendet werden (Europäische Kommission [EC], 2021).

---

6 H2020-Projekt SIENNA: Stakeholder-Informed Ethics for New technologies with high socio-ecoNomic and human rights impAct, Laufzeit 1.10.2017 — 31.3.2021, https://www.sienna-project.eu.

7 H2020-Projekt SHERPA: Shaping the Ethical Dimensions of Smart Information Systems. A European Perspective, Laufzeit 1.5.2018 — 31.10.2021, https://www.project-sherpa.eu/.

8 Aktuell beschäftigt sich das Horizon Europe-Projekt TECHETHOS explizit mit Ethics by Design, wobei nach eigener Aussage auf Ergebnisse der folgenden zuvor geförderten Projekte zurückgegriffen wird: SIENNA und SHERPA, sowie SATORI (Stakeholders Acting Together On the ethical impact assessment of Research and Innovation, Laufzeit 1.1.2014 — 30.9.2017, https://satoriproject.eu) und PANELFIT (Participatory Approaches to a New Ethical and Legal Framework for ICT, Laufzeit 1.11.2018 — 30.4.2022, https://www.panelfit.eu/).

Die Herangehensweise beruht auf sechs ethischen Prinzipien, aus denen sich ethische Anforderungen ergeben. Die Prinzipien haben sich auf europäischer sowie auf internationaler Ebene als Querschnittsanforderungen herausgestellt (EC, 2021, S. 5) und umfassen Respekt vor dem menschlichen Handeln (*respect for human agency*); Privatheitsschutz, Schutz persönlicher Daten und Datengovernance (*privacy, personal data protection and data governance*); Fairness (*fairness*); individuelles, soziales und ökologisches Wohlergehen (*individual, social, and environmental well-being*); Transparenz (*transparency*); Rechenschaftspflicht und Kontrolle / Aufsicht (*accountability and oversight*) (EC, 2021). Den sechs Prinzipien werden nach einer kurzen Definition – teilweise recht präzise – Anforderungen zugeordnet. In Bezug auf den Respekt vor menschlichem Handeln (»*Human Agency*«) solle beispielsweise sichergestellt werden, dass KI-Anwendungen ohne menschliche Aufsicht und die Möglichkeit, Rechtsmittel einzulegen, nicht autonom Entscheidungen über die folgenden Bereiche treffen. Zum einen nicht über fundamentale persönliche Angelegenheiten (die z.B. direkt das Privat- oder Berufsleben beeinflussen etc.), und die normalerweise von Menschen durch freie persönliche Wahl entschieden würden. Zum anderen über grundlegende wirtschaftliche, soziale und politische Fragen, die normalerweise in kollektiven Beratungen entschieden würden, oder den Einzelnen in ähnlicher Weise erheblich betreffen (EC, 2021, S. 6). An dieser Stelle überschneiden sich die ethischen Forderungen mit bereits geltendem Recht, wie z.B. Art. 22 der Datenschutz-Grundverordnung (Verordnung 2016/679) und berühren teilweise Anforderungen, die auch in der im August 2024 in Kraft getretenen europäischen KI-Verordnung enthalten sind (Verordnung 2024/1689).

Im zweiten Teil der Handreichung werden praktische Schritte zur Umsetzung von Ethics by Design erläutert. Ein Fünf-Ebenen-Modell stellt dar wie Grundsätze, ethische Anforderungen, Ethics by Design-Leitlinien, KI-Methologien sowie Tools und Methoden ineinander über gehen. Weiter wird ein generisches Modell für die Entwicklung von KI-Anwendungen vorgestellt. Für jede Phase werden erneut ethische Anforderungen vorgestellt, die in diesem Falle als konkretere Aufgaben (»*tasks*«) formuliert sind. Die sechs Phasen umfassen 1. Spezifikation der Ziele (*specification of objectives*), 2. Spezifikation der Anforderungen (*specification of requirements*), 3. allgemeines Design (*high-level design*), 4. Datenerhebung und Auf-

bereitung (*data collection and preparation*), 5. detailliertes Design und Entwicklung (*detailed design and development*) und 6. Testen und Evaluation (*testing and evaluation*). Laut Autor*innen kann dieses Modell auf bestehende KI-Modelle angewendet werden, wobei die Schritte iterativ sein können. Es sollte eine ethische Risikoanalyse vorgenommen werden, nach Möglichkeit von einer*einem Ethiker*in.[9] Zur Erleichterung der Implementierung von Ethics by Design werden weiter vier Praktiken vorgestellt, denen Anforderungen und Hinweise zugeordnet werden, wie Ethics by Design in der Praxis (eines Forschungsprojektes) umgesetzt werden kann. Es handelt sich um 1. Projektplanung- und management (*Project management*), 2. externe Beschaffung eines KI-Systems (*Acquisition*), 3. Einsatz und Implementierung (*Deployment and implementation*) und 4. Überwachen der ethischen Anforderungen (*Monitoring*). Abschließend wird eine Checkliste zur Verfügung gestellt, in der eine verkürzte Version der ethischen Anforderungen aufgeführt wird, mit Platz für eigene Angaben, wie die Risiken eingedämmt werden können, wenn die jeweilige Frage, z.B. nach dem Ziel, dass End-User*innen Kontrolle gegeben wird, nicht mit »ja« beantwortet wurde.

Der Ansatz von Philip Brey und Brandt Dainow trägt dem Umstand Rechnung, dass es in der *Open Source Community* sowie in der Industrie bereits diverse Tools gibt, die auf verschiedenste Weise Werte im Technologieentwicklungsprozess berücksichtigen und / oder dazu dienen, Rechenschaft über die Arbeit von Programmierenden abzulegen.[10] Dies erfolgt nicht unbedingt, um bestimmte Werte zu stützen oder hochzuhalten. So dienen beispielsweise Dokumentationen über die eigenen Schritte im Programmierprozess u.a. auch der besseren Zusammenarbeit, wenn zu einem späteren Zeit-

9 An anderer Stelle wird jedoch betont, dass in diesem Framework nicht unbedingt Ethiker*innen beteiligt sein müssten, sondern dass im Gegenteil technische bzw. fachliche Expertise notwendig sei, um die Dinge richtig einschätzen zu können. Dies unterscheidet den Ansatz von Brey et al. u.a. von dem vom IDE.

10 Das OECD.AI Policy Observatory führt einen »Catalogue of Tools and Metrics for Trustworthy AI«, d. h. Datenbanken, in denen mit Stand Anfang 2025 919 Tools und 130 Metriken verzeichnet sind. Die Suche nach »by Design« führt aktuell zu zwei Tools https://oecd.ai/en/catalogue/tools?terms=by%20design&page=1 (16.2.2025). Für eine Zuordnung einer Auswahl von Tools zur Erfüllung bestimmter ethischer Prinzipien und Werte siehe Kluge Correa und Mönig (2024).

punkt andere Personen am eigenen Code weiterarbeiten, unterstützen aber auch die Forderung nach dem Grundsatz der Transparenz. Als Beispiel wird »*Datasheets for Datasets*« genannt (Gebru et al., 2021).

2.2 Narrative Ethik by Design

Der Ansatz des Instituts für Digitale Ethik schlägt vor, dass potentielle ethische Konflikte zu einem frühen Zeitpunkt der Technikentwicklung adressiert werden sollen.[11] Grimm und Mönig (2020) betonen, dass der von ihnen entwickelte Fragebogen im Rahmen bestehender Qualitätsmanagementmaßnahmen oder Qualitätskontrollen eingesetzt werden kann. Darüber hinaus liegt ihm ein weites Verständnis von (ethischer) Stakeholder-Beteiligung zugrunde. Da sich Werte verändern, soll der Prozess der Befragung während der Technologieentwicklung (z.B. operationalisiert durch den vorliegenden Fragebogen) iterativ wiederholt werden. Da »Anwendungsfälle« (»*use cases*«) und sogar die Behandlung von unwahrscheinlichen Grenzfällen (sog.»Edge Cases«), die Probleme aufzeigen können, zur »*best practice*« gehören, können Anwendungsbeispiele mit ethischer Dimension ebenfalls während der Produktentwicklung betrachtet und diskutiert werden. Ein Fokus des Instituts für Digitale Ethik liegt diesbezüglich auf einem narrativen Ansatz, der sich zum Einen in Befragungsmethoden äußert (beispielsweise dem Einsatz von narrativen Interviews, vgl. Erbach et al., 2020). Zum Anderen liefern Narrative die Möglichkeit, ethische Reflexion anzustoßen und Werte und Werteverletzungen sichtbar und begreifbar zu machen (vgl. auch Keber, 2021; Hohendanner, 2024). Operationalisiert wurde der Ethics by Design-Ansatz durch das automatisierte Ethik-Assessment-Instrument ELSI-SAT und ELSI-SAT Health and Care.[12]

11 Siehe den zweiten Teil (Umsetzung, Potenziale und Grenzen) des vorliegenden Sachstandsberichts.

12 ELSI-Screening- und Awarenesstool (SAT) https://www.elsi-sat.de/, https://www.elsi-sat-health-and-care.de.

2.3 Das Whitepaper »Towards an Ethics by Design Approach for AI«

Eines der jüngsten unter den derzeit verfügbaren Ethics by Design-Frameworks ist die 2024 von AI4People veröffentlichte Publikation »Towards an Ethics by Design Approach for AI« (AI4People, 2024).[13] In Bezug auf die aktuelle Entwicklung ist dabei interessant, dass der Ansatz nach dem Aufsehen, das die Veröffentlichung von ChatGPT 3.5 im November 2022 hervorrief sowie das – teilweise sogar Fachpublikum erstaunende – Tempo, mit dem in der Folge weitere generative KI-Anwendungen einer breiten Öffentlichkeit zur Verfügung standen, erschienen ist. Die Handreichung der Europäischen Kommission mit dem von Brey et al. formulierten Framework erschien in ihrer ersten Fassung im November 2021, der Entwurf für die europäische KI-Verordnung im April 2021.

Das Whitepaper richtet sich, im Gegensatz zum Fokus auf Institutionen, die Forschungsförderung in Anspruch nehmen möchten, sowohl an öffentliche, namentlich EU-Institutionen, als auch an private Akteur*innen. Ziel ist ein Wettbewerbsvorteil für die Unternehmen sowie die Vermeidung von Folgekosten, wenn ethische Risiken entstanden sind. Die Einhaltung ethischer Grundsätze (»*compliance*«) soll nicht nur die elementaren Rechte von individuellen Personen, sondern auch gesellschaftliche Güter wie die Erhaltung von demokratischen Institutionen und der Gewaltenteilung unterstützen. Der Ansatz von AI4People gliedert sich in fünf Phasen und diesen vorgelagerte „erste Schritte“:

- Erste Schritte: Verstehen des Unternehmenskontextes und Aufbau des Fundaments,
- Phase 1: Verstehen des KI-Systems: Scoping und Spezifikationen,
- Phase 2: Vorläufige Folgenabschätzung (ethisches Impact Assessment),
- Phase 3: Design des vertrauenswürdigen KI-Systems,

13 Überraschenderweise trägt das Whitepaper einen vergleichbaren Titel wie der bereits 2018 erschienene Artikel von d'Aquin et al. (2018), »Towards an ‚Ethics by Design' Methodology for AI Research Projects«. Jedoch wird von den Autor*innen von AI4People betont, dass ihr Ethics by Design-Ansatz offen für Veränderungen und Anpassungen sei, was den Titel (wenn auch nicht die Nähe zum früheren Paper) erklärt.

– Phase 4: Implementierung des vertrauenswürdigen KI-Systemdesigns,
– Phase 5: Überwachung des vertrauenswürdigen KI-Systemdesigns.

Die erste und zweite Phase zählen zum Prozess der Ideenfindung. Die Designphase entspricht bzw. beginnt in diesem Modell erst mit Phase 3. Jede Phase wird kurz beschrieben. Die Hauptaktivitäten und zentralen Ergebnisse werden in einer übersichtlichen Form aufgelistet. Zusammen mit der Konklusion werden Empfehlungen für EU-Institutionen gegeben, wie Ethics by Design weiter gefördert werden kann. Im Anhang finden sich zwölf übersichtliche »*Guidance*«-Dokumente, die jedes für sich betrachtet und verwendet werden können, z.B. mit einer Übersicht über relevante rechtliche Regulierungen; neben der KI-Verordnung und der Datenschutz-Grundverordnung zählen hierzu beispielsweise u. U. auch der *Digital Services Act*.

Tabelle 2: Die 5 Phasen gemäß dem AI4People-Ansatz (Eigene Übersetzung und Anpassung nach AI4People, 2024)

Lebenszyklus-Phasen der Systementwicklung						
Lebenszyklusphase der KI-Entwicklung	Ideenfindung	Systemzulassung (»AI System Approval«)	Design	Entwickeln und Testen	Einsetzen	Überwachen und Einsetzen
EbD Phasen (vorgelagert: »Erste Schritte«)	Phase 1: Verstehen des KI-Systems: Scoping und Spezifikationen Phase 2: Vorläufige Folgenabschätzung (ethisches Impact Assessment)		Phase 3: Design des vertrauenswürdigen KI-Systems	Phase 4: Implementierung des vertrauenswürdigen KI-Systemdesigns	Systemeinführung (»AI System go-live«)	Phase 5: Überwachung des vertrauenswürdigen KI-Systemdesigns

3. Politische Dimension

Der Begriff »Ethics by Design« figuriert in verschiedenen Dokumenten der Europäischen Union. Bereits 2018 schrieb die Europäische Kommission im »Coordinated Plan on Artificial Intelligence«Europa könne »bei der Entwicklung der KI und ihrer Nutzung zum Gemeinwohl, bei der Verfolgung eines auf den Menschen ausgerichteten (»menschzentrierten«) Ansatzes und bei der Förderung der Grundsätze einer integrierten Ethik weltweit führend werden.« (EC, 2018a, S. 9) Interessanterweise wurde an dieser Stelle »ethics-by-design-principles« mit dem Begriff »integrierte Ethik« übersetzt (EC, 2018b, S. 8).

Ethics by Design besitzt also, wie oben bereits deutlich wurde, in seiner derzeitigen Ausprägung explizit eine politische Dimension. Zum einen dient es in Kombination mit dem Ethik-Self-Assessment im Rahmen von *Horizon Europe* als Forschungsförderinstrument. Zum anderen ist eine seiner operationalisierbaren Formen gleichsam durch »Auftragsforschung« entstanden. Während in Standardisierungsgremien laut der deutschen Bundesregierung explizit mehr Stakeholder*innen aus der Wirtschaft vertreten sein sollen, ist auch dies ein politisches Instrument, zumal die europäische KI-Verordnung unter dem »*New Legislative Framework*« explizit mit dem technologischen Fortschritt durch Standardisierung und harmonisierte Normen Schritt zu halten versucht. Auch von der durch die deutsche Bundesregierung eingesetzten Datenethikkommission wurde ethische Technikgestaltung gefordert.[14]

Während verschiedene politische Akteur*innen den Einsatz von Ethics by Design fordern, weist das *Institut für Technikfolgenabschätzung* der Österreichischen Akademie der Wissenschaften (ITA) in einem Bericht für das österreichische Parlament auf die »administrative Bürde« hin, die mit der Umsetzung von X by Design-Konzepten verbunden sein könnte (ITA, 2021).

14 Der deutsche Ethikrat verwendete in seiner Stellungnahme »Mensch und Maschine. Herausforderungen durch Künstliche Intelligenz« zwar den Begriff »Ethics by Design« nicht, allerdings liest sich sein Verweis auf Value Sensitive Design wie eine Definition von Ethics by Design. Des Weiteren wird von ethischem Design sowie von Privacy und Security by Design gesprochen (Deutscher Ethikrat, 2023).

4. Kritik am Konzept »Ethics by Design«

Verschiedene Kritikpunkte sind gegenüber Ethics by Design geltend gemacht worden. Luciano Floridi argumentiert, die Methode sei paternalistisch, weshalb er ihr seinen Ansatz des »Pro-ethical Design« gegenüberstellt (Floridi, 2016). Dem Konzept des »Values in design« wird als Befürchtung entgegengebracht, es beruhe auf der Annahme, dass das System in jedem Fall gebaut werde. Dies sei jedoch weniger hilfreich für die Entscheidung, ob ein System im Falle gravierender (ethischer) Bedenken eventuell gar nicht erst entwickelt werden bzw. wieder zurückgezogen werden sollte (Crawford & Calo, 2016). Mark Coeckelbergh bezieht diesen Einwand auch auf Ethics by Design, das in diesem Falle eine »Barriere« für Ethik darstellen würde (Coeckelbergh, 2020). Außerdem kritisiert er die diesen Ansätzen zugrunde liegende Annahme, dass wir unsere ethischen Werte vollständig artikulieren könnten; was wir jedoch nicht unbedingt immer können (Coeckelbergh, 2020). Im EU-geförderten Projekt SHARESPACE wird »Ethics by Design« mit einem »Good Enough Ethics«-Ansatz (zu Deutsch etwa »hinreichend gute Ethik«) verknüpft, da sich im Laufe des Projekts abzeichnete, dass sich die durch die Ethics by Design-Methodologie sichtbar gewordenen ethischen Probleme (»*issues«)* nicht im Laufe der Projektlaufzeit würden bewältigen lassen. Es handelt sich also um eine Erweiterung des Ansatzes, jedoch mithilfe einer »Verkürzung«. Des Weiteren wird befürchtet, dass Ethics by Design als kosmetische »Buzzwords« im Wettbewerb um Drittmittel im Rahmen der Ausschreibung für die EU-Förderung verwendet werden (könnten) (Precision Drug Repurposing for Europe and the World [REPO4EU], 2024). Die 2018 durch die damalige deutsche Bundesregierung eingesetzte Datenethikkommission betonte, dass Ethics by Design kein »Garant für ethische Produkte und Dienstleistungen« sei und sich Ethik nicht an Technik delegieren ließe. Welche ethischen Prinzipien »wann und wie umgesetzt werden« solle nicht allein Entwickler*innen überlassen werden, »sondern kontextspezifisch und ggf. unter Einbeziehung Betroffener ausgehandelt werden« (Datenethikkommission, 2019, S. 74). In dieser Beziehung kommt hinzu, dass Tools und Metriken nicht unbedingt aus sich selbst heraus verständlich sind, bzw. dass Metriken, die vermeintlich dem selben Ziel dienen bzw.

denselben Wert schützen sollen, nicht unbedingt erreichen, wozu sie eingesetzt werden (vgl. beispielsweise zu(r Wirksamkeit von) Fairness-Metriken Verma & Rubin, 2018). Als problematisch wird es außerdem gesehen, wenn versucht wird, Ethics by Design oder vergleichbare Governancemechanismen im Sinne eines »soft law« zur Vermeidung strengerer, durchsetzbarer Regulierungen zu etablieren. Zudem droht in diesem Zusammenhang u. U. auch eine verstärkte Einflussnahme durch Stakeholder oder Akteure, die nicht unbedingt den (gesamt-)gesellschaftlichen Nutzen im Blick haben. Obwohl mit der *Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD)*, dem *Weltwirtschaftsforum (WEF)* und dem *Institute of Electrical and Electronics Engineers (IEEE)* global agierende Institutionen Ethics by Design führend voranbringen, steht darüber hinaus der Vorwurf im Raum, dass die zu Grunde gelegten Werte eurozentristisch bzw. sog. »westliche« Werte seien.[15]

5. Ausblick: Ausbildung und weitere Aspekte

Wie kann nun also Ethik über die beschriebenen Ansätze hinaus bereits »by Design« in Forschungs- und Entwicklungsprozesse einbezogen werden? Verschieden Akteur*innen haben vorgeschlagen, dass Ethics by Design in die Lehre von technischen und ingenieurswissenschaftlichen Studiengängen implementiert werden soll.[16] Darüber hinaus kann es Workshops, Zertifikatslehrgänge, Online-Kurse und andere Formen der Weiterbildung geben, die das Thema vermitteln.[17] Begleitend zum Einsatz ethischer Tools und Checklisten, muss außerdem die ethische Deliberation weiter geführt werden, u.a., weil sich Werte ändern können, und da rein technische Lösungen (Stichwort »*techno-solutionism* / *technological fix*«) oft nicht ausreichen, um ethische Befürchtungen und unethische Ausgangssituationen, wie z.B. Bias in Datensätzen, zu beheben. Hierbei sollte

15 Spiekermann and Winkler (2020) begegnen diesem Einwand proaktiv mit ihrem Requirement 5b.

16 Siehe hierzu den dritten Teil (Ausbildung sozial-verantwortlicher Ingenieur*innen) des vorliegenden Sachstandsberichts.

17 Für einen Überblick über bestehende Kurse zu Ethics by Design (Stand Oktober 2021) sowie einen Vorschlag für einen Seminarplan vgl. Annex 3 des SIENNA-Deliverables 5.7 Jansen et al. (2021).

auch die Frage gestellt werden, um welche Ethik es sich handelt und welche Werte geschützt werden sollen. Auch sollte es Möglichkeiten geben, Rückfragen an (Ethik-)Expert*innen zu stellen und ethische Bedenken z. B. einem Unternehmen mitteilen zu können. Dabei gilt es zu klären, was diese Expert*innen ausmacht. Positiv ist hervorzuheben, dass ein ganzheitlicher Blick gewagt wird und bestehende Ansätze und Praktiken gewürdigt werden. Da mit der Technologie auch bestimmte Machtstrukturen unser Leben bereits durchdrungen haben, sollte Ethik zukünftig nicht nur »by Design«, sondern auch »by Default« in Technik integriert werden. Dabei dürfen diese Ansätze jedoch nicht dazu führen, dass verbindliche rechtliche Regelungen aufgeweicht werden, da die freiwillige Einhaltung von Regeln auf Grenzen stößt. Insgesamt sollte beachtet werden, dass Werte u. U. nicht eindeutig operationalisiert werden können, und Ethik eine emotionale und unsagbare Komponente besitzt. Rote Linien, ob eine Technologie zum Einsatz kommt oder ggf. aus dem Verkehr genommen wird, sollten als Option in Ethics by Design mitgedacht werden.

Abb. 1: Ethics by Design auf einen Blick

Literaturverzeichnis

AI4People. (2024). *AI4People's Institute Report Towards an Ethics by Design Approach for AI*. https://ai4people.org/wp-content/uploads/2024/06/Towards-an-Ethics-by-Design-Approach-for-AI.pdf

Brey, P., & Dainow, B. (2023). Ethics by design for artificial intelligence. *AI and Ethics, 4*, 1265–1277. https://doi.org/10.1007/s43681-023-00330-4

Cavoukian, A. (2011). *Privacy by Design*. https://www.sfu.ca/~palys/Cavoukian-2011-PrivacyByDesign-7FoundationalPrinciples.pdf

Coeckelbergh, M. (2020). *AI ethics*. The MIT Press. https://doi.org/10.7551/mitpress/12549.001.0001

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns (New York, N.Y.)*, *4*(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature, 538*(7625), 311–313. https://doi.org/10.1038/538311a

d'Aquin, M., Troullinou, P., O'Connor, N. E., Cullen, A., Faller, G., & Holden, L. (2018). Towards an „Ethics by Design" Methodology for AI Research Projects. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES 18*, S. 54–59). Association for Computing Machinery. https://doi.org/10.1145/3278721.3278765

Datenethikkommission. (2019). *Gutachten der Datenethikkommission der Bundesregierung*. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf

Deutscher Ethikrat. (2020). *Robotik für gute Pflege* [Stellungnahme]. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-robotik-fuer-gute-pflege.pdf

Deutscher Ethikrat. (2023). *Mensch und Maschine. Maschine – Herausforderungen durch Künstliche Intelligenz* [Stellungnahme]. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf

Dignum, V. (2020). Responsibility and Artificial Intelligence. In M. D. Dubber, F. Pasquale & S. Das (Hrsg.), *The Oxford Handbook of Ethics of AI*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.12

Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., Génova, G., Haim, G., Kließ, M. S., Lopez-Sanchez, M., Micalizio, R., Pavón, J., Slavkovik, M., Smakman, M., van Steenbergen, M., Tedeschi, S., van der Toree, L., Villata, S., & and de Wildt, T. (2018). In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18,* S. 60–66). Association for Computing Machinery. https://doi.org/10.1145/3278721.3278745

Europäische Kommission (EC). (2018a). *Koordinierter Plan für künstliche Intelligenz*. https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:52018DC0795

Europäische Kommission (EC). (2018b). *Coordinated Plan on Artificial Intelligence*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0795.

Europäische Kommission (EC). (2021). *Ethics By Design and Ethics of Use Approaches for Artificial Intelligence*. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf

Erbach, R., Maurer, S., Meixner, G., Koller, M., Grimm, P., & Mönig, J. M. (2020). KoFFI—The New Driving Experience. In G. Meixner (Hrsg), *Smart Automotive Mobility. Human-Computer Interaction Series* (S. 155–211). Springer. https://doi.org/10.1007/978-3-030-45131-8_3

Floridi, L. (2016). Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration. *Science and Engineering Ethics, 22*(6), 1669–1688. https://doi.org/10.1007/s11948-015-9733-2

Friedman, B., & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press. https://doi.org/10.7551/mitpress/7585.001.0001

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for Datasets. *arXiv:1803.09010*. https://doi.org/10.48550/arXiv.1803.09010

Gransche, B., & Manzeschke, A. (Hrsg.). (2020). *Das geteilte Ganze: Horizonte Integrierter Forschung für künftige Mensch-Technik-Verhältnisse*. Springer VS. https://doi.org/10.1007/978-3-658-26342-3

Grimm, P., & Mönig, J. M. (2020). Ethical Recommendations for Cooperative Driver-Vehicle-Interaction – Guidelines for Highly Automated Driving. In G. Meixner (Hrsg.), *Smart Automotive Mobility. Human-Computer Interaction Series* (S. 213–229). Springer. https://doi.org/10.1007/978-3-030-45131-8_4

High-Level Expert Group on Artificial Intelligence (AI-HLEG). (2019). *Ethics Guidelines for Trustworthy AI*. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

Hohendanner, M. (2024). Design. In P. Grimm, K. E. Trost & O. Zöllner (Hrsg.), *Handbuch Digitale Ethik* (S. 613–623). Nomos.

Ind.ie (2016). *Ethical Design Manifesto*. https://ind.ie/ethical-design/

Institut für Digitale Ethik (IDE). (o. J.). *Ethics by Design*. Zugriff am 24.02.2025. https://www.hdm-stuttgart.de/digitale-ethik/forschung/ethics_by_design

Institute of Electrical and Electronics Engineers (IEEE). / IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2018). *Ethically Aligned Design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2 – For Public Discussion*. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

Institute of Electrical and Electronics Engineers (IEEE). (2021). IEEE Standard Model Process for Addressing Ethical Concerns during System Design. *IEEE Std 7000–2021*. https://doi.org/10.1109/IEEESTD.2021.9536679

Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften (ITA) & Austrian Institute of Technology (AIT). (2021). *Foresight und Technikfolgenabschätzung: Monitoring von Zukunftsthemen für das Österreichische Parlament* [Projektbericht Nr. ITA-AIT-15]. https://epub.oeaw.ac.at/0xc1aa5576_0x003d04ab.pdf

Jansen, P., Henschke, A., Erden, Y., Marchiori, S., Brey, P., & Hoefsloot, M. (2021). *Ethics by Design and Research Ethics for AI*. D5.7 of the H2020-SHERPA project. https://doi.org/10.21253/DMU.16912345.v1

Keber, T. (2021). Digital Ethics by Process? Technical conflicts and policy ethics committees in Europe. *Informatio, 26*(1), 216–229. https://portal.amelica.org/ameli/journal/265/2652175013/2652175013.pdf

Kluge Corrêa, N., & Mönig, J. M. (2024). *Catalog of General Ethical Requirements for AI Certification* [Whitepaper]. Center for Science and Thought. https://doi.org/10.48550/arXiv.2408.12289

Kranzberg, M. (1986). Technology and History: "Kranzberg's Laws". *Technology and Culture, 27*(3), 544–560. https://doi.org/10.2307/3105385

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396. https://doi.org/110.1037/h0054346

Mehlich, J., & Woopen, C. (2025). From applied ethics to innovation practice: an ethics-by-design approach for constructive consideration of ELSI in technological design decisions. *Journal of Responsible Innovation, 12*(1), 2459451. https://doi.org/110.1080/23299460.2025.2459451

Nurock, V., Chatila, R., & Parizeau, M.-H. (2021). What Does "Ethical by Design" Mean? In B. Braunschweig & M. Ghallab (Hrsg.), *Reflections on Artificial Intelligence for Humanity* (S. 171–190). https://doi.org/10.1007/978-3-030-69128-8_11

Moore, S. L. (2010). *Ethics by Design: Strategic Thinking and Planning for Exemplary Performance, Responsible Results, and Societal Accountability*. HRD Press.

Precision Drug Repurposing for Europe and the World (REPO4EU). (2024). *Ethics and privacy-by-design – why these are no buzzwords in REPO4EU*. https://repo4.eu/2024/01/10/ethics-and-privacy-by-design-why-these-are-no-buzzwords-in-repo4eu/

Simon, J. (2016). Values in Design. In J. Heesen (Hrsg.), *Handbuch Medien- und Informationsethik* (S. 35–36). J. B. Metzler. https://doi.org/10.1007/978-3-476-05394-7_49

Spiekermann, S., & Winkler, T. (2020). *Value-based Engineering for Ethics by Design*. https://arxiv.org/abs/2004.13676

Urquhart, L. D., & Craigon, P. J. (2021). The Moral-IT Deck: a tool for ethics by design. *Journal of Responsible Innovation, 8*(1), 94–126. https://doi.org/10.1080/23299460.2021.1880112

van den Hoven, J. (2017). The Design Turn in Applied Ethics. In J. van den Hoven, S. Miller & T. Pogge (Hrsg.), *Designing Ethics* (S. 11–31). Cambridge University Press. https://doi.org/10.1017/9780511844317.002

van den Hoven, J., & Manders-Huits, N. (2017). Value-sensitive Design. In K. Miller & M. Taddeo (Hrsg.), *The Ethics of Information Technologies* (S. 329–332). Routledge. https://doi.org/10.4324/9781003075011-23

van den Hoven, J, Vermaas, P. E., & van de Poel, I. (Hrsg.). (2015). *Handbook of Ethics, Values, and Technological Design*. Springer.

van Wynsberghe, A., & Robbins, S. (2014). Ethicist as Designer: A Pragmatic Approach to Ethics in the Lab. *Science and Engineering Ethics, 20*(4), 947–961. https://doi.org/10.1007/s11948-013-9498-4

Verma, S., & Rubin, J. (2018). Fairness Definitions Explained. *Proceedings of the 2018 ACM/IEEE International Workshop on Software Fairness (FairWare), Sweden*. https://doi.org/10.1145/3194770.3194776

Vermaas, P. E., Hekkert, P., Manders-Huits, N., & Tromp, N. (2015). Design Methods in Design for Values. In J. van den Hoven, P. E. Vermaas & I. van de Poel (Hrsg.), *Handbook of Ethics, Values, and Technological Design* (S. 179–201). Springer. https://doi.org/10.1007/978-94-007-6970-0_10

Verordnung 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG. http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=DE

Verordnung 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz). http://data.europa.eu/eli/reg/2024/1689/oj

von Schomberg, R., & Hankins, J. (Hrsg.) (2019). *International Handbook on Responsible Innovation: A Global Resource*. Edward Elgar Publishing. https://doi.org/10.4337/9781784718862

World Economic Forum (WEF). (2020). *Ethics by Design: An organizational approach to responsible use of technology* [Whitepaper]. http://www3.weforum.org/docs/WEF_Ethics_by_Design_2020.pdf

Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*. Profile Books.