4 Kritische Überprüfung der Begriffe

4.1 Klärung von Begriffen

Ein weiterer Aspekt der begrifflichen Klarheit aus ethischer Sicht sind die Begriffe, die man im Diskurs über digitale Transformation und künstliche Intelligenz verwendet. Warum muss man diese Begriffe klären? «Ich weiß nicht recht, ob gerade die Informatik oder ihre Teildisziplin Künstliche Intelligenz eine so große Vorliebe für Euphemismen hat. Wir sprechen so spektakulär und so bereitwillig von Computersystemen, die verstehen, die sehen, die entscheiden, die urteilen (...), ohne dass wir selbst unsere eigene Oberflächlichkeit und unermessliche Naivität in Bezug auf diese Begriffe erkennen. Und indem wir so sprechen, betäuben wir unsere Fähigkeit, (...) uns ihres Endnutzens bewusst zu werden (...) Man kann diesem Zustand nicht entkommen, ohne sich immer wieder zu fragen: 'Was tue ich eigentlich?' 'Was ist die endgültige Anwendung und Verwendung der Produkte meiner Arbeit?' und schließlich: 'Bin ich zufrieden oder schäme ich mich, zu dieser Verwendung beigetragen zu haben?'»[301]

Darüber hinaus wird man bei der Frage nach der Definition von «Roboter», «künstlicher Intelligenz» und «digitaler Transformation» (ein Oberbegriff, der Automatisierung, Maschinisierung, Maschinisierung, Robotisierung, Digitalisierung umfasst) auf deren begriffliche Unschärfe[302] aufmerksam, die es zu überwinden gilt.

4.2 Roboter

Ein «Roboter» kann definiert werden als «eine automatisch gesteuerte, programmierbare, bewegliche Mehrzweckmaschine mit mehreren Freiheitsgraden, die entweder ortsfest oder mobil für den Einsatz in industriellen Automatisierungsanwendungen sein kann»[303]. Der Begriff «Roboter» stammt aus einem Theaterstück mit dem Titel «W.U.R. Werstands

301 Weizenbaum 1987: 45.
302 Vgl. Ohly 2019a: 20.
303 Europäische Norm 1992.

universal Robots», das von Karel Čapek geschrieben und 1920 veröffentlicht wurde.[304] Der Titel deckt sich mit dem Namen eines Unternehmens, das humanoide Maschinensklaven herstellt, die den Menschen von der Arbeit befreien sollen, stattdessen aber versuchen, den Menschen zu vernichten. Aus ethischer Sicht ist die Zuschreibung von «mehreren Freiheitsgraden» aufgrund der obigen Ausführungen in Kapitel 3 Kann ethisches Urteilsvermögen an Technologien delegiert werden? problematisch.

Weitere Attribute, die Robotern zugeschrieben werden, sind ein «Körper», die Fähigkeit, über Sensoren «auf vielfältige Weise mit seiner räumlich nahen Umgebung in Kontakt zu treten», die «Auswirkungen auf die räumliche Umgebung» und ein Computer.[305]

Ein «intelligenter Roboter» kann verstanden werden als «eine Maschine, die in der Lage ist, Informationen aus ihrer Umgebung zu extrahieren und das Wissen über ihre Arbeit zu nutzen, um sich sicher in einer sinnvollen und zielgerichteten Weise zu bewegen»[306], während ein «autonomer Roboter» «ohne menschliches Eingreifen arbeiten kann und mit Hilfe verkörperter künstlicher Intelligenz in seiner Umgebung agieren und leben kann»[307]. Aus ethischer Sicht ist die Zuschreibung von «Autonomie» auf der Grundlage der obigen Ausführungen in Kapitel 3 Kann ethisches Urteilsvermögen an Technologien delegiert werden? problematisch.

4.3 Künstliche Intelligenz

Künstliche Intelligenz kann definiert werden als «Maschinen, die in der Lage sind, auf menschenähnliche Weise zu 'denken' und höhere intellektuelle Fähigkeiten und berufliche Fertigkeiten zu besitzen, einschließlich der Fähigkeit, sich selbst aus ihren eigenen Fehlern zu korrigieren»[308] oder als «die Wissenschaft und Technologie von Maschinen mit Fähigkeiten, die nach dem Standard der menschlichen Intelligenz als intelligent gelten»[309]. Der Begriff «künstlich» in «künstliche Intelligenz» unterstreicht,

304 Vgl. Čapek 2017.
305 Ohly 2019a: 21-22.
306 Arkin 1998: 2.
307 Tzafestas 2016: 37.
308 Tzafestas 2016: 22-25.
309 Jansen et al. 2018: 5.

dass die «Intelligenz durch technische Mittel dargestellt oder simuliert wird»[310].

Diese beiden Definitionen von künstlicher Intelligenz sind anthropozentrisch. Künstliche Intelligenz kann auch unabhängig von der menschlichen Intelligenz definiert werden als das Bestreben, «Computer dazu zu bringen, die Art von Dingen zu tun, die der Verstand tun kann»[311].

1950 warf die Veröffentlichung von Alan Turing über Rechenmaschinen und Intelligenz die Frage auf, ob Maschinen denken können. Ein einfacher heuristischer Test – der so genannte «Turing-Test» – sollte helfen, diese Hypothese zu überprüfen: Könnte ein Computer auf der Grundlage von getippten oder weitergeleiteten Nachrichten ein Gespräch führen und Fragen so beantworten, dass ein misstrauischer Mensch den Computer für einen Menschen halten würde?[312] John Searle stellt dieses Argument in Frage, indem er mit dem «Chinese Room Thought Experiment» den Unterschied zwischen Denken und dem Vortäuschen von Denkfähigkeit hervorhebt: Ausgehend von der Eingabe chinesischer Schriftzeichen kann eine künstliche Intelligenz eine Ausgabe in Form anderer chinesischer Schriftzeichen erzeugen, indem sie den Anweisungen eines Computerprogramms folgt, wie die chinesischen Schriftzeichen zu kombinieren sind. Wenn die Ausgabe einen menschlichen Chinesisch-Sprechenden so überzeugt, dass dieser glaubt, sie stamme von einem menschlichen Chinesisch-Sprechenden, wäre der «Turing-Test» zwar bestanden, aber würde dies wirklich bedeuten, dass die künstliche Intelligenz Chinesisch versteht? Oder zeigt es nur die Fähigkeit der künstlichen Intelligenz an, die Fähigkeit zu simulieren, Chinesisch zu verstehen? Würde man einen Menschen in einen geschlossenen Raum setzen und ihm Eingaben und Anweisungen geben, wie er diese Eingaben zu einer Ausgabe verarbeiten soll, die den Eindruck erweckt, dass er Chinesisch versteht, könnte der Mensch zwar vortäuschen, Chinesisch zu verstehen, aber er würde es nicht wirklich verstehen.[313]

310 Coeckelbergh 2020: 203.
311 Boden 2016: 1.
312 Vgl. Turing 2009.
313 Einer der wichtigsten Punkte, die man aus dem Gedankenexperiment des «Chinese Room Argument» ziehen kann, ist, dass Syntax weder gleichwertig noch ausreichend für Semantik ist. Das Argument des «Chinese Room» zeigt, dass symbolische Manipulation, so überzeugend sie auch erscheinen mag, aufgrund der entscheidenden Rolle, die die Semantik beim Verstehen spielt, weder gleichwertig noch ausreichend für das Verstehen ist.

Ebenfalls 1950 schlug Claude Shannon die Entwicklung einer Maschine vor, der man das Schachspielen[314] mit roher Gewalt oder durch Auswertung einer kleinen Menge strategischer Züge des Gegners beibringen könnte.[315]

Künstliche Intelligenz ist auch die Bezeichnung für ein wissenschaftliches Gebiet, das sich um das Verständnis von Intelligenz bemüht und mit Informatik, Algorithmen und Logik verbunden ist. Ausgangspunkt für die Forschung auf dem Gebiet der künstlichen Intelligenz war die Überzeugung, dass «jeder Aspekt des Lernens oder jedes Merkmal der Intelligenz im Prinzip so genau beschrieben werden kann, dass eine Maschine dazu gebracht werden kann, es zu simulieren»[316].

Die künstliche Intelligenz als wissenschaftlicher Bereich unterscheidet sich von der «Kognitionswissenschaft (cognitive science)», die sich ebenfalls mit der Erforschung allgemeiner Prinzipien der Intelligenz befasst,[317] aber mit Psychologie, Biologie, Neurobiologie usw. verbunden ist,[318] und der «embodied cognitive science», die sich auf die Mechanismen konzentriert, die intelligentem Verhalten zugrunde liegen, und die von der Überzeugung ausgeht, dass «das individuelle Gehirn nicht der einzige Ort des kognitionswissenschaftlichen Interesses sein sollte. Kognition ist kein Phänomen, das erfolgreich untersucht werden kann, wenn man die Rolle des Körpers, der Welt und des Handelns ausklammert»[319].

Künstliche Intelligenz kann definiert werden als «die Lehre von den Berechnungen, die es ermöglichen, wahrzunehmen, zu denken und zu handeln»[320]. Eine weitere Dimension wird durch die folgende Prägung der künstlichen Intelligenz hinzugefügt:[321] «Ein KI-System besteht aus drei Hauptelementen: Sensoren, Betriebslogik und Aktoren. Die Sensoren sammeln Rohdaten aus der Umgebung, während die Aktoren den Zustand der Umgebung verändern. Die entscheidende Stärke eines KI-Systems liegt in seiner operativen Logik. Für eine gegebene Reihe von Zielen und auf der Grundlage von Eingabedaten von Sensoren liefert die Betriebslogik Ausgaben für die Aktoren. Diese haben die Form von Empfehlungen, Vorhersagen oder Entscheidungen, die den Zustand der Umwelt beeinflussen

314 Vgl. Shannon 1950.
315 Vgl. Universität von Washington 2006.
316 McCarthy et al. 2006: 12.
317 Pfeifer / Scheier 1999: 5.
318 Vgl. Pfeifer / Scheier 1999: 5-6.
319 Clark 1999: 350.
320 Winston 1992: 5.
321 Vgl. Russel / Norvig 2009; Gringsjord / Govindarajulu 2018.

können.[322] Die folgende Definition ermöglicht ein umfassendes Verständnis: «KI-System: Ein KI-System ist ein maschinenbasiertes System, das für eine gegebene Reihe von vom Menschen definierten Zielen Vorhersagen, Empfehlungen oder Entscheidungen treffen kann, die reale oder virtuelle Umgebungen beeinflussen. KI-Systeme sind so konzipiert, dass sie mit unterschiedlichem Grad an Autonomie arbeiten.

- Lebenszyklus von KI-Systemen: Der Lebenszyklus von KI-Systemen umfasst folgende Phasen: i) ‚Entwurf, Daten und Modelle', eine kontextabhängige Abfolge, die Planung und Entwurf, Datenerfassung und -verarbeitung sowie Modellbildung umfasst; ii) ‚Verifizierung und Validierung'; iii) ‚Einsatz'; und iv) ‚Betrieb und Überwachung'. Diese Phasen laufen oft iterativ ab und sind nicht unbedingt aufeinander folgend. Die Entscheidung, ein KI-System aus dem Betrieb zu nehmen, kann zu jedem beliebigen Zeitpunkt während der Betriebs- und Überwachungsphase getroffen werden.
- KI-Wissen: KI-Wissen bezieht sich auf die Fähigkeiten und Ressourcen wie Daten, Code, Algorithmen, Modelle, Forschung, Know-how, Schulungsprogramme, Governance, Prozesse und bewährte Verfahren, die erforderlich sind, um den Lebenszyklus von KI-Systemen zu verstehen und daran teilzunehmen.
- KI-Akteure: KI-Akteure sind diejenigen, die eine aktive Rolle im Lebenszyklus des KI-Systems spielen, einschließlich Organisationen und Einzelpersonen, die KI einsetzen oder betreiben.»[323]

KI-Systeme haben einen Lebenszyklus mit den folgenden Phasen:

- «1) Entwurf, Daten und Modellierung umfassen mehrere Aktivitäten, deren Reihenfolge bei verschiedenen KI-Systemen unterschiedlich sein kann:
  - Die Planung und Gestaltung des KI-Systems umfasst die Formulierung des Konzepts und der Ziele des Systems, der zugrundeliegenden Annahmen, des Kontexts und der Anforderungen sowie möglicherweise den Bau eines Prototyps.
  - Die Datenerfassung und -verarbeitung umfasst das Sammeln und Korrigieren von Daten, die Prüfung auf Vollständigkeit und Qualität sowie die Dokumentation der Metadaten und Merkmale des Datensatzes. Zu den Metadaten eines Datensatzes gehören Informationen darüber,

322 OECD 2019a.
323 OECD 2019c: I.

wie ein Datensatz erstellt wurde, wie er zusammengesetzt ist, welche Verwendungszwecke er hat und wie er im Laufe der Zeit gepflegt wurde.
  - Die Modellbildung und -interpretation umfasst die Erstellung oder Auswahl von Modellen oder Algorithmen, deren Kalibrierung und/oder Training und Interpretation.
- 2. Die Verifizierung und Validierung umfassen die Ausführung und Abstimmung von Modellen mit Tests zur Bewertung der Leistung in verschiedenen Dimensionen und Bereichen.
- 3. Die Einführung in die Live-Produktion umfasst Pilotprojekte, die Überprüfung der Kompatibilität mit bisherigen Systemen, die Einhaltung gesetzlicher Vorschriften, das Management organisatorischer Veränderungen und die Bewertung der Benutzer:innenfreundlichkeit.
- 4. Der Betrieb und die Überwachung eines KI-Systems umfassen den Betrieb des KI-Systems und die kontinuierliche Bewertung seiner Empfehlungen und (beabsichtigten und unbeabsichtigten) Auswirkungen im Hinblick auf die Ziele und ethischen Erwägungen. In dieser Phase werden Probleme identifiziert und Anpassungen vorgenommen, indem auf andere Phasen zurückgegriffen wird oder ein KI-System gegebenenfalls aus der Produktion genommen wird.[324]

Heutzutage wird die künstliche Intelligenz in «schwache künstliche Intelligenz» und «starke künstliche Intelligenz» unterteilt.[325] Während eine «schwache KI» entwickelt wird, um eine bestimmte, begrenzte Aufgabe zu erfüllen, will eine «starke KI» der menschlichen Intelligenz ähnlich sein oder sie sogar übertreffen. Eine andere Definition bezeichnet «Artificial Narrow Intelligence» als «KI, die sich auf einen Bereich spezialisiert. Es gibt eine KI, die den Schachweltmeister im Schach schlagen kann, aber das ist das Einzige, was sie kann»[326]. «Künstliche allgemeine Intelligenz (...) oder KI auf menschlichem Niveau (...) bezieht sich auf einen Computer, der in allen Bereichen so intelligent ist wie ein Mensch – eine Maschine, die jede intellektuelle Aufgabe ausführen kann, die auch ein Mensch kann.»[327] (Darüber hinaus gibt es eine «Superintelligenz», die oben in Kapitel 3 Kann ethisches Urteilsvermögen an Technologien delegiert werden? kurz

324 OECD 2019a.
325 Vgl. UNESCO 2018.
326 Urban 2015.
327 Urban 2015.

angesprochen wurde und weiter unten in Unterkapitel 7.7 Transhumanismus diskutiert wird.) Ob und wann eine «Künstliche Allgemeine Intelligenz» jemals erreicht wird, wird diskutiert, ist aber umstritten.[328]

4.4 Datenbasierte Systeme

Aus ethischer Sicht ist zu beachten, dass künstliche Intelligenz oft nicht für sich allein steht, sondern ihr Potenzial in Kombination mit anderen Technologien entfaltet,[329] und dass künstliche Intelligenz «immer auch sozial und menschlich ist: Bei KI geht es nicht nur um Technologie, sondern auch darum, was Menschen damit machen, wie sie sie nutzen, wie sie sie wahrnehmen und erleben und wie sie sie in ein größeres sozio-technisches Umfeld einbetten.»[330]

Aus ethischer Sicht wird der oben genannte Ansatzpunkt kritisiert: «Intelligenz ist nicht darauf beschränkt, ein bestimmtes kognitives Problem zu lösen, sondern es kommt darauf an, *wie* das geschieht.»[331] Angesichts des Charakters der künstlichen Intelligenz sind aus ethischer Sicht Zweifel angebracht, ob der Begriff überhaupt angemessen ist, da die künstliche Intelligenz zwar die menschliche Intelligenz zu imitieren versucht, dies aber auf einen bestimmten Bereich der Intelligenz (z.B. bestimmte kognitive Fähigkeiten) beschränkt ist.[332] Darüber hinaus ist davon auszugehen, dass künstliche Intelligenz der menschlichen Intelligenz allenfalls in bestimmten Bereichen der Intelligenz ähnlich, aber niemals gleich werden kann. Wie oben in Kapitel 3 Kann ethisches Urteilsvermögen an Technologien delegiert werden? ausgeführt, ist beispielsweise die moralische Fähigkeit einer der Bereiche menschlicher Intelligenz, die künstliche Intelligenz nicht erreichen kann.

Schließlich ist die Kritik an der «künstlichen Intelligenz» auf konzeptioneller Ebene ethisch relevant, was sich z.B. an der Verwendung des Begriffs «vertrauenswürdige künstliche Intelligenz» zeigen lässt: «Der zugrundeliegende Leitgedanke einer ‚vertrauenswürdigen KI' ist zunächst schon mal begrifflicher Unsinn. Maschinen sind nicht vertrauenswürdig,

328 Vgl. Brooks 2017.
329 Vgl. Coeckelbergh 2020: 78-81.
330 Coeckelbergh 2020: 80.
331 Misselhorn 2018: 17.
332 Vgl. Dreyfus 1972: 29; Dreyfus / Dreyfus 1986.

nur Menschen können vertrauenswürdig sein – oder eben auch nicht. Wenn ein nicht vertrauenswürdiger Konzern oder eine nicht vertrauenswürdige Regierung sich unethisch verhält und in Zukunft eine gute, robuste KI-Technologie besitzt, dann kann er oder sie sich noch besser unethisch verhalten. Die Geschichte von der *Trustworthy AI* ist eine von der Industrie erdachte Marketing-Narrative, eine Gute-Nacht-Geschichte für die Kund:innen von morgen. In Wirklichkeit geht es darum, Zukunftsmärkte zu entwickeln und Ethikdebatten als elegante öffentliche Dekoration für eine groß angelegte Investitionsstrategie zu benutzen.»[333]

Es bestand die Versuchung, es in diesem Buch bei dieser Kritik zu belassen, keine Definition, nur eine Arbeitsdefinition oder nur ein Modell vorzulegen. Nun soll aber doch ein konkreter Vorschlag in Form einer Definition gemacht werden – im Dienste der begrifflichen Substanz, der Konkretheit, der Bestimmtheit und der konzeptionellen Schärfe. Der Begriff «datenbasierte Systeme» wäre angemessener als «künstliche Intelligenz», weil dieser Begriff das beschreibt, was «künstliche Intelligenz» eigentlich ausmacht: Erzeugung, Sammlung und Auswertung von Daten; datenbasierte Wahrnehmung (sensorisch, sprachlich); datenbasierte Vorhersagen; datenbasierte Entscheidungen. Darüber hinaus hilft der Begriff «datenbasierte Systeme» auch, die Hauptstärke und die Hauptschwäche der gegenwärtigen technologischen Errungenschaft in diesem Bereich hervorzuheben. Die Beherrschung einer enormen Datenmenge ist die wichtigste Stärke datenbasierter Systeme.

Der Hinweis auf ihre Kerneigenschaft – nämlich auf Daten zu beruhen und sich in all ihren Prozessen, ihrer eigenen Entwicklung und ihren Handlungen (genauer gesagt ihren Reaktionen auf Daten) ausschließlich auf Daten zu stützen – lüftet den Schleier der unangemessenen Zuschreibung des Mythos der «Intelligenz», der wesentliche Probleme und Herausforderungen datenbasierter Systeme verdeckt, und ermöglicht mehr Genauigkeit, Angemessenheit und Präzision bei der kritischen Reflexion datenbasierter Systeme. Die Unnachvollziehbarkeit, Unvorhersehbarkeit und Unerklärbarkeit der algorithmischen Prozesse, die zu datenbasierten Auswertungen, datenbasierten Vorhersagen und datenbasierten Entscheidungen führen («Black-Box-Problem»[334]), ihre große Anfälligkeit für systemische Fehler, ihre ausgeprägte Anfälligkeit für die Verwechslung von

333 Metzinger 2019.

334 Vgl. Knight 2017a; Bathäe 2018; Weinberger 2018; Knight 2017b; Castelvecchi 2016.

Kausalität und Korrelation (z. B. ein hoher Eis-Konsum von Kindern in einem Sommermonat und eine hohe Anzahl von Kindern, die aufgrund von mehr Mobilität in den Ferien im selben Sommermonat in Autounfälle verwickelt sind, korrelieren zwar miteinander, es besteht jedoch kein kausaler Zusammenhang zwischen den beiden Statistiken, was bedeutet, dass der Eis-Konsum keine Autounfälle verursacht)[335], und die hohe Wahrscheinlichkeit, dass voreingenommene und diskriminierende Daten zu voreingenommenen und diskriminierenden datenbasierten Bewertungen, datenbasierten und diskriminierenden Vorhersagen und datenbasierten und diskriminierenden Entscheidungen führen, stellen die größten Nachteile dar.[336] «Algorithmen sind Meinungen, in Codes eingebettet. Sie sind nicht objektiv.»[337] Sie sind nicht neutral. Sie dienen bestimmten Zielen und Zwecken. Sie müssen beherrscht werden.[338]

Diese oben genannten Hauptnachteile werden durch ethische Herausforderungen der Algorithmen als Teil datenbasierter Systeme begründet oder verstärkt:[339] epistemische Bedenken: «nicht schlüssige Beweise»[340] («wahrscheinliches, aber zwangsläufig unsicheres Wissen»), «undurchschaubare Beweise»[341] («Verbindung zwischen den Daten und der Schlussfolgerung» nicht zugänglich), «fehlgeleitete Beweise»[342] («Schlussfolgerungen können nur so zuverlässig (aber auch so neutral) sein wie die Daten»); normative Bedenken: «ungerechte Ergebnisse»[343] (diskriminierende Wirkung), «transformative Effekte»[344] («Algorithmen können die Art und Weise beeinflussen, wie wir die Welt konzeptualisieren»), «Nachvollziehbarkeit»[345] («Algorithmen sind Software-Artefakte, die bei der Datenverarbeitung eingesetzt werden, und bringen als solche die ethischen Herausforderungen mit sich, die mit dem Design und der Verfügbarkeit neuer Technologien sowie mit der Manipulation großer Mengen personenbezogener und anderer Daten verbunden sind. Dies bedeutet, dass Schäden, die

335 Vgl. Iversen / Gergen 1997: 317-318.
336 Vgl. UNESCO 2019; Fortmann-Roe 2012; Bartoletti 2018; Counts 2018; Europarat 2018a; Europarat 2018b; Wildhaber / Lohmann / Kasper 2019; Coeckelberg 2020.
337 Demuth 2018: 16.
338 Lohmann 2018.
339 Vgl. Mittelstadt et al. 2016: 4-5.
340 Vgl. Mittelstadt et al. 2016: 5.
341 Vgl. Mittelstadt et al. 2016: 6-7.
342 Vgl. Mittelstadt et al. 2016: 7-8.
343 Vgl. Mittelstadt et al. 2016: 8-9.
344 Vgl. Mittelstadt et al. 2016: 9-10.
345 Vgl. Mittelstadt et al. 2016: 10-12.

durch algorithmische Aktivitäten verursacht werden, schwer zu beheben sind (d. h. den Schaden zu erkennen und seine Ursache zu finden), aber auch, dass es selten einfach ist, zu bestimmen, wer für den verursachten Schaden verantwortlich gemacht werden sollte.»[346])

Leider können diese ethischen Herausforderungen nur bis zu einem gewissen Grad durch die Verbesserung der Algorithmen und ihrer Anwendungen bewältigt werden. Natürlich wäre es daher von erheblichem Nutzen, wenn Daten keine Entscheidungen treffen dürften, sondern lediglich als Informationsquelle dienen könnten.

Darüber hinaus wäre es ein wesentlicher positiver Unterschied, wenn datenbasierte Systeme nicht so konstruiert würden, dass das «Black-Box-Problem» entsteht und man dann versucht, die «Black-Box»[347] zu öffnen, sondern von Anfang an mit einem Ansatz gestaltet würden, der nicht nur erklärbar ist («KI, die dem Menschen ihre Handlungen, Entscheidungen oder Empfehlungen erklären oder hinreichend Auskunft darüber geben kann, wie sie zu ihren Ergebnissen gekommen ist»[348]), sondern auch interpretierbar ist,[349] d.h den Menschen in die Lage versetzt, angesichts von Daten eine informierte und durchdachte Position einzunehmen, weil er nicht nur die Datenverarbeitung versteht, die zu diesen Ergebnissen führt, sondern auch in der Lage ist, die Datenverarbeitung zu bewerten und einzuschätzen, was auch in seine Sichtweise integriert ist. Der Mensch sollte nur das entscheiden und tun, was er versteht. Letztlich ist die Forderung nach einem technologischen Ansatz, der das «Black-Box-Problem» vermeidet, eine ethische Grundforderung,[350] kein hoher ethischer Anspruch. Es wird lediglich gefordert, dass Technologien so konzipiert, entwickelt und produziert werden, dass aus ethischer Sicht das Problem vermieden wird, dass man nicht garantieren kann, dass negative Risiken, Auswirkungen oder Folgen vermieden werden können, weil man einfach nicht weiß, was in der «Blackbox» passiert. Dazu gehört auch die Akzeptanz möglicher negativer Auswirkungen auf die technologische Leistungsfähigkeit.[351] Um dies einfach mit einer Analogie zu verdeutlichen: Etwas Ähnliches wie ein «Black-Box-Ansatz» wäre es, in der pharmazeutischen Industrie ein Medikament zu produzieren, von dem man nicht weiß, was es enthält, son-

346 Mittelstadt et al. 2016: 4-5.
347 Vgl. Samek et al. 2017.
348 Coeckelberg 2020: 204.
349 Vgl. Rudin 2019.
350 Vgl. Göbel et al. 2018.
351 Vgl. Seseri 2018.

dern nur hofft, dass es die gewünschten positiven Wirkungen hervorruft, und – falls nicht – ein anderes Medikament zu entwickeln, falls negative Wirkungen auftreten sollten. Niemand würde in diesem Fall akzeptieren, dass man nur die Daumen drückt und hofft, dass nichts Negatives passiert.

Das «Black-Box-Problem» besteht nicht nur in der technologischen Komplexität und der Zweckmäßigkeit,[352] sondern auch in einer Lücke zwischen Mathematik und Ethik, auf die weiter unten in Kapitel 5 Die Komplexität der Ethik eingegangen wird. Selbst wenn alles erklärbar[353] und interpretierbar ist, kann der Prozess, wie man eine Brücke zwischen Mathematik und Ethik schlägt und wie man ethische Prinzipien und Normen in Algorithmen umwandelt, undurchsichtig bleiben.[354] Darüber hinaus umfasst das «Black-Box-Problem» auch implizit verbleibende normative Vorstellungen innerhalb der Datenverarbeitung, wie z. B. Vorurteile, diskriminierende Muster. Weltanschauliche, kulturelle oder soziale Normen provozieren auch eine imperialistische Form der normativen Universalität. Schließlich besteht das «Black-Box-Problem» auch in der Unklarheit über ethisch relevante und ethisch irrelevante Aspekte – was einen Teil der Komplexität der Ethik darstellt, die weiter unten in Kapitel 5 Die Komplexität der Ethik erläutert wird. Aus ethischer Sicht müssen alle diese Aspekte des «Black-Box-Problems» überwunden werden, um die Herausforderung der Intransparenz zu bewältigen.

Dennoch muss man berücksichtigen, dass dies aus ethischer Sicht nicht die Lösung, sondern ein Teil der Lösung ist. «Der gesamte konzeptionelle Raum der ethischen Herausforderungen, die sich aus der Verwendung von Algorithmen ergeben, kann jedoch nicht auf Probleme reduziert werden, die mit leicht zu identifizierenden erkenntnistheoretischen und ethischen Unzulänglichkeiten zusammenhängen.»[355] Die erstgenannten ethischen Herausforderungen und das letztgenannte grundlegende ethische Problem von Algorithmen machen deutlich, dass es unangemessen ist, «die Ethik algorithmischer Assemblagen mit der Transparenz von algorithmischem Code gleichzusetzen»[356].

352 Vgl. Rudin 2019.
353 Vgl. Müller-Dott 2019; Horizon 2020 Commission Expert Group 2020.
354 Ich bin Justus Piater und Aaron Butler dankbar, dass wir diese Idee unabhängig voneinander entwickelt und auf der Jahrestagung des «Innsbrucker Kreises von Moraltheolog:innen und Sozialethiker:innen» vom 2. bis 4. Januar 2020 miteinander diskutiert haben.
355 Mittelstadt et al. 2016: 15.
356 Ananny 2016: 109.

Darüber hinaus kommen die Anforderungen an die Menschen, die datenbasierte Systeme nutzen, ins Spiel. Menschen, die datenbasierte Systeme nutzen, datenbasierte Entscheidungen treffen usw. müssen in die Lage versetzt werden, die Verarbeitung von Daten und die datenbasierten Ergebnisse kritisch zu reflektieren, Daten mit kritischem Denken zu nutzen und datenbasierten Systemen oder Daten gegebenenfalls zu widersprechen.

4.5 Digitale Transformation

Dieser konzeptionelle Vorschlag der «datenbasierten Systeme» hat Auswirkungen auf andere Begriffe, z. B. auf die «digitale Transformation». Der Begriff «digitale Transformation» steht für verschiedene technologiebasierte Veränderungen wie «Digitalisierung», «Automatisierung», «Robotisierung», «Maschinisierung», «Mechanisierung», den «Einsatz datenbasierter Systeme» und den «Umgang mit datenbasierte Supersystemen». Es handelt sich um einen Oberbegriff, der die Begriffe Automatisierung, Maschinisierung, Mechanisierung, Robotisierung, Digitalisierung, «Nutzung datenbasierter Systeme» und «Umgang mit datenbasierten Supersystemen» umfasst. Dabei ist zu berücksichtigen, dass Objekte, Phänomene und Realitäten als «digitale Transformation» bezeichnet werden, obwohl sie selbst nicht digital sind, sondern Produkte oder Ergebnisse digitaler Prozesse sind.[357] «Digital» bedeutet «Aufzeichnung oder Speicherung von Informationen als eine Reihe der Zahlen 1 und 0, um zu zeigen, dass ein Signal vorhanden ist oder nicht, unter Verwendung von oder im Zusammenhang mit digitalen Signalen und Computertechnologie, Darstellung von Informationen in Form eines elektronischen Bildes, unter Verwendung eines Systems, das von einem Computer und anderen elektronischen Geräten verwendet werden kann, in dem Informationen in elektronischer Form als eine Reihe der Zahlen 1 und 0 gesendet und empfangen werden, im Zusammenhang mit Computertechnologie, insbesondere dem Internet, Darstellung von Informationen als ganze Zahlen und nicht in einer anderen Form wie einem Bild, einer Grafik usw.»[358]

Aus ethischer Sicht ist vor allem der Unterschied zwischen der aktuellen technologiebasierten «digitalen Transformation» und früheren technologiebasierten Wandelepochen von besonderem Interesse für das Verständ-

357 Vgl. Ohly 2019a: 25-29.
358 Cambridge Dictionary n.d.

nis der digitalen Transformation. Seine Analyse erfolgt weiter unten im Unterkapitel 7.18 Reduktion der bezahlten Arbeitsplätze – unter anderem mit dem folgenden Aspekt: Im Unterschied zu früheren technologiebasierten Epochen des Wandels[359] geht es bei der «digitalen Transformation» u.a. nicht um Arbeitserleichterungen für Menschen, sondern um den Ersatz von Menschen durch datenbasierte Systeme in der Wertschöpfungskette (z.B. zielen automatische Kassen im Supermarkt nicht darauf ab, die berufliche Tätigkeit einer Kassiererin zu erleichtern, wie es beim Bauern der Fall war, als der Pflug durch einen Traktor ersetzt wurde). Die Kernkonsequenz der digitalen Transformation und des Einsatzes datenbasierter Systeme kennzeichnen diesen technologiebasierten Wandel: *Immer weniger Menschen werden direkt an einer effizienteren und effektiveren Wertschöpfungskette teilnehmen und teilhaben.*[360] Aus ethischer Sicht bedeutet dies, dass aus makroökonomischer Perspektive die Herausforderung nicht in der Menge der zur Verfügung stehenden Mittel liegt, sondern in der Gestaltung eines gerechten Gesellschafts- und Wirtschaftssystems,[361] *u.a.* im Hinblick auf die Verteilung der finanziellen Mittel und der Rechte und Möglichkeiten der Teilhabe und Partizipation, im Hinblick auf die Chancengleichheit für alle, im Hinblick auf die Gewährleistung des Überlebens und eines menschenwürdigen Lebens für alle und die Wahrung des friedlichen Zusammenlebens.

Schließlich kann die zunehmende Interaktion zwischen Mensch und Maschine (z.B. die massive Auswirkung der hohen Präsenz sozialer Medien im Alltag auf soziales Verhalten, soziale Kompetenz und persönliche Interaktion) und deren Eingriffe in soziale und persönliche Prozesse und Beziehungen durch die Veränderung der Lebensbedingungen zu einer Technologisierung und Robotisierung von Ideen, Begriffen, Vorstellungen und Konzepten, des Menschenbildes, der menschlichen Denkweise und einer Infragestellung der Menschenwürde führen. Wenn beispielsweise das Wort «Freiheit» so weit reduziert wird, dass das, was man «Freiheit» nennt, nur noch in dem besteht, was in Parametern beschreibbar, in Algorithmen programmierbar und für technologische Systeme wahrnehmbar ist, dann wird der Reichtum, die Vielfalt, die Präzision und die Tiefe des menschlichen Denkens, der Reflexion und der Sprache dramatisch nivelliert, normalisiert und standardisiert.

359 Vgl. Hessler 2016.
360 Vgl. Kirchschläger 2017a; Kirchschläger 2016b.
361 Vgl. Kirchschläger 2013c.

4.6 Maschinenethik – Roboterethik

Am Ende dieser kritischen Überprüfung von Begriffen sollen Begriffe von dieser Überprüfung nicht ausgeschlossen werden, welche dieses Buch disziplinär verorten – wie oben in Kapitel 1 Einleitung angegeben. Die «Maschinenethik» zielt darauf ab, ein System der Ethik zu entwickeln, das dann in eine Maschine[362] eingesetzt werden kann, um das Gedeihen der Menschen und des Planeten Erde zu fördern, zu unterstützen, zu maximieren, zu erhalten und zu schützen. Dieses Ziel gewinnt angesichts des technologischen Fortschritts in Form von selbstlernenden Maschinen, automatisierten Maschinen (z. B. selbstfahrende Fahrzeuge), datenbasierten Systemen und Super-Datenbasierten Systemen an Bedeutung. «Maschinenethik» diskutiert u.a. die Möglichkeit der Implementierung von Ethik in Maschinen,[363] die materielle Definition der Ethik, die in Maschinen implementiert werden soll,[364] und die Art und Weise, wie diese Implementierung verfolgt werden könnte.[365]

Die «Roboterethik»[366] ist mit der Entwicklung, der Produktion und dem Einsatz von Robotern befasst und wendet sich an Menschen, die mit Robotik zu tun haben (z. B. Designer:innen, Hersteller:innen und Benutzer:innen von Robotern). Ein anderes Verständnis – auch als «Ethik in der Robotik» bezeichnet – bringt die oben genannten Aspekte der Maschinenethik und der Roboterethik zusammen: «Erstens können wir darüber nachdenken, wie Menschen durch oder mit Robotern ethisch handeln können. In diesem Fall sind die Menschen die ethischen Akteur:innen. Darüber hinaus könnten wir praktisch darüber nachdenken, wie man Roboter so gestaltet, dass sie ethisch handeln, oder theoretisch darüber, ob Roboter wirklich ethische Akteure sein können. Hier sind die Roboter die fraglichen ethischen Subjekte. Schließlich gibt es mehrere Möglichkeiten, die ethischen Beziehungen zwischen Menschen und Robotern zu gestalten: Ist es ethisch, künstliche moralische Agenten zu schaffen? Ist es unethisch, hochentwickelte Roboter nicht mit ethischen Argumentationsfähigkeiten auszustatten? Ist es ethisch vertretbar, Roboter-Soldaten, -Polizeibeamte oder -Krankenschwestern zu schaffen? Wie sollten Roboter Menschen be-

362 Vgl. Moor 2006; Moor 1995.
363 Vgl. Allen / Wallach 2014; Floridi 2011; Johnson 2011.
364 Vgl. Mackworth 2011; McLaren 2011; Pereira 2016.
365 Vgl. Segal 2017; Wallach et al. 2008.
366 Vgl. Veruggio / Operto 2008: 1504.

handeln, und wie sollten Menschen Roboter behandeln? Sollten Roboter Rechte haben?»[367]

367 Asaro 2006: 10.