



Editorial	<i>Joachim Wiemeyer (Bochum)</i> Künstliche Intelligenz als technologisches Großprojekt Zu diesem Heft	2
Schwerpunktthema	<i>Anna Maria Riedl (Bonn)</i> Herausforderung Technik	3
	<i>Bernhard Irrgang (Dresden)</i> Datenzentrierte Biologie und KI Paradigmenwandel für Wissenschaft und Technologie	5
	<i>Armin Grunwald (Karlsruhe)</i> Digitalisierung und Künstliche Intelligenz Technikfolgenabschätzung trifft Anthropologie	12
	<i>Katharina Klöcker (Bochum)</i> Erlösung von der Last der Entscheidung? Die Algorithmisierung der Gesellschaft in theologisch-ethischer Reflexion	19
	<i>John Wyatt (London)</i> What does it mean to be human in a world of technology and AI?	26
Arts & ethics	<i>Catrin Misselhorn (Göttingen)</i> Moral für Maschinen? Grundsätze der Maschinenethik	35
	<i>Birgit Feike (Dortmund)</i> „Sei Dein Gott“	28
Interview	<i>Oliver Bendel (Zürich)</i> „Maschinen können keine moralische Verantwortung tragen.“	41
Buchbesprechungen	Kirche neu erfinden	47
	Migrationsethik	48
	Gender in der Theologie	49
	Lehrbuch der Christlichen Sozialethik	51
Der Überblick	Summaries	54
	Résumés	55
	Bisherige Schwerpunktthemen und Vorschau	56
Impressum		U2



Joachim Wiemeyer

In der Bibel ist der Kultur-
auftrag in den Anfangser-
zählungen des Buches Gene-
sis 1,28 angelegt. Viele his-
torische Zeugnisse zeigen,
dass Menschen seit Jahrtau-
senden sowohl Schutz gegen
Naturgefahren entwickelt als
auch durch Beherrschung der
Natur deren Früchte für Men-
schen nutzbarer gemacht
haben. Die Erzählung vom

Turmbau zu Babel (Gen 11,1–9) demon-
striert aber auch eine menschliche Hy-
bris, sich technologische Projekte vor-
zunehmen, die scheitern und in einem
Desaster enden können. Während die
Erfindung und Verbreitung neuer Tech-
nologien bis zum Beginn der industri-
ellen Revolution sehr langsam ablief,
führte die systematische naturwissen-
schaftlich-technische Forschung in en-
ger Verbindung mit der Anwendung in
der Wirtschaft zunächst in westlichen
Industrieländern zu einem anhalten-
den Wirtschaftswachstum und grund-
legenden Veränderungen der Lebens-
weisen (z. B. der Individualmobilität,
Elektrizität). Wenn man sich große Zie-
le (z. B. Mondlandung) setzte, und dafür
große finanzielle Mittel und menschli-
che Intelligenz aufwand, konnten die-
se erreicht werden.

Allerdings sind diese technologi-
schen Möglichkeiten 50 Jahre nach der
ersten Mondlandung nicht weiter ge-
nutzt worden. In der Gegenwart stellt
die Künstliche Intelligenz ein solches
technologisches Großprojekt dar, das in
vielfältigen Kontexten der Wirtschaft,
des privaten Lebens und der zwischen-
menschlichen Kommunikation im Ver-
kehr, der Medizin sowie in der Kriegs-
führung das gesellschaftliche Leben re-
volutionieren soll.

Künstliche Intelligenz als technologisches Großprojekt

Angesichts des gegenwärtigen
Hypes um Digitalisierung und Künst-
liche Intelligenz verwundert es aber,
dass die Arbeitsproduktivität – trotz et-
wa des wachsenden Einsatzes von Ro-
botern – seit Jahren kaum noch steigt
und so langsam wächst wie seit dem
19. Jahrhundert nicht mehr. Selbst bei
einer stagnierenden Wirtschaft entste-
hen noch neue Arbeitsplätze. Die Irrita-
tion über den angeblich „beschleunig-
ten technischen Fortschritt“ und die sta-
gnierende Arbeitsproduktivität wird als
„Produktivitätsparadox“ wissenschaft-
lich diskutiert, ohne dass bisher über-
zeugende Erklärungen gefunden wur-
den. Manche vollmundigen Ankündi-
gungen von „Technikfreaks“ (etwa von
Tesla-Chef Elon Musk), in welchem
Jahr selbstfahrende Autos sich im nor-
malen Verkehr bewegen werden, sind
weit verfehlt worden. In der Arbeitswelt,
dem zentralen Lebensbereich der Men-
schen, entsprechen die Veränderungen
bisher den üblichen Lernprozessen wäh-
rend eines Arbeitslebens bzw. sind Ver-
änderungen im normalen Generations-
wechsel. Im privaten Konsum entschei-
den die Menschen selbst darüber, ob,
wann und wie sie neue technologische
Möglichkeiten nutzen wollen. Zwischen
technologischen Erfindungen und ihrer
Umsetzung in unternehmerischen Pro-

duktionsprozessen sowie in marktfäh-
igen Produkten, ihrer gesellschaftlichen
Zulassung, ihrer Akzeptanz bei breiten
Bevölkerungsgruppen etc. bestehen im-
mer erhebliche Diskrepanzen und län-
gere zeitliche Abläufe.

Diese ermöglichen einen sozialet-
hischen Diskurs über die technologische
Entwicklung, zu dem dieses Heft ei-
nen Beitrag leisten will. Manche Zu-
kunftsvisionen sind so weitreichend,
dass selbstlernende Roboter ein eige-
nes Bewusstsein entwickeln könnten.
Dies würde das menschliche Selbstver-
ständnis in Frage stellen. Den aus die-
sen Herausforderungen erwachsenden
grundlegenden theologisch-ethischen
(Katharina Klöcker), anthropologisch-
theologischen (John Wyatt) und ma-
schinenethischen (Catrin Misselhorn)
Fragen gehen die Beiträge dieses Heft-
es nach, eingeleitet durch die Hinfüh-
rung von Anna-Maria Riedl. Zusätzlich
wird auch auf die politisch-ethische
Frage des Diskurses solcher Probleme
im Kontext einer Technologiefolgen-
abschätzung (Armin Grunwald) ein-
gegangen. Zudem erläutert Bernhard
Irrgang den Paradigmenwechsel in der
Forschung durch KI anhand der Bio-
logie bzw. der Lebenswissenschaften.



Herausforderung Technik



Anna Maria Riedl


„[...] grundsätzlich waren Maschinen nicht selbstbewegend, nicht selbstentworfen, nicht autonom. Sie konnten den Traum des Menschen nicht erfüllen, nur nachäffen. Eine Maschine war kein Mensch, keine Urheberin ihrer selbst, nur eine Karikatur dieses reproduktiven Traums [...]. Schon der Gedanke, dass es anders sein könnte, wäre paranoid gewesen. Heute sind wir nicht mehr so sicher. Die Maschinen des späten 20. Jahrhunderts haben die Differenz von natürlich und künstlich, Körper und Geist, selbstgelenkter und au-Bengesteuerter Entwicklung sowie viele andere Unterscheidungen, die Organismen und Maschinen zu trennen vermochten, höchst zweideutig werden lassen. Unsere Maschinen erscheinen auf verwirrende Weise quicklebendig – wir selbst dagegen beängstigt träge.“

Donna Haraway formulierte diesen Gedanken in ihrem Manifest für Cyborgs bereits in den 1980er Jahren. Sie nimmt damit vorweg, was uns aktuell in den Heilversprechen oder Warnungen vor KI (Künstliche Intelligenz) begegnet, die die momentane Diskussion um die Digitalisierung prägen und den Vorgängerhype um das Web 2.0 und social media abgelöst haben. Die Ausgabe 3/2022 von Amosinternational greift mit dem Thema „Künstliche Intelligenz“ diese Anfragen auf. Im Mittelpunkt des Heftes steht die anthropologische Frage nach dem Humanen angesichts technischer und technologischer Entwicklungen und nach ethischer Orientierung, wenn bislang

als sicher geglaubte Grenzen zwischen Natur und Technik zu verschwimmen scheinen. Denn, so hat Markus Zimmermann bereits 2016 in seinem Artikel „Grenzverschiebungen“ festgehalten: *„Die Natur des Menschen wird heute im Rahmen bio- oder gentechnischer [und technologischer] Eingriffe selbst zum Gegenstand der Veränderung und kann schon von daher nicht zur Orientierung dienen. Unter den Leitbildern der Optimierung oder Verbesserung werden vermeintlich natürlich vorgegebene Grenzen überschritten, so dass die Natur als Gegenüber, als ethische Orientierungsbasis und als Ausweis eines menschlichen Sonderstatus entgleitet.“* (In: Bogner, Daniel/Mügge, Cornelia: Natur des Menschen. Freiburg, 177–178). Nicht nur der Rückzug auf eine (mythisch) überhöhte Natur ist verstellt, sondern auch der Bezug auf eine Wesensnatur des Menschen als Orientierung fällt damit aus. Woran also sich orientieren, wenn Leben keine „stabile ontologische und normative Referenz“ mehr ist, sondern „in einem Maße gestaltbar geworden [...], das jede Vorstellung einer vom menschlichen Handeln unberührte Natur überholt erscheinen lässt“ (Lemke: Eine Analytik der Biopolitik 2008, 79)?

Diese Fragen stellen sich auch, wenn die starke KI (die Mensch-Maschine oder der Cyborg) eine Science-Fiction-Vision ist (und vermutlich noch eine ganze Weile bleiben wird). Nicht nur starre Grenzziehungen zwischen uns – den scheinbar alleinig autonom handelnden menschlichen Sub-

jekten – und den Maschinen als Objekten unserer Handlungen sind angesichts von KI schwieriger geworden. Schwache KI ist in unseren Lebenszusammenhängen längst präsent. Es ist ein Markt entstanden, der die Entwicklung konkret vorantreibt. Ein Beispiel dafür sind die hohe Investition in Biotechnologien bei Google oder die mittlerweile etablierten Institute, Vereine und Think Tanks, die offensiv für transhumanistische Zukunftsvisionen werben. Untersucht man dort vertretene Positionen genauer, zeigt sich,

 Transhumanistische Visionen haben einen Wegbereiter in der bereits stattfindenden Digitalisierung und im Human Enhancement

dass transhumanistische Utopien tief im naturwissenschaftlich-technischen Denken verwurzelt sind. Deutlich wird, die transhumanistischen Visionen haben offensichtlich einen Wegbereiter in der bereits stattfindenden Digitalisierung und im Human Enhancement. Das wirft die Frage auf, welche Handlungsspielräume uns in diesen Entwicklungen bleiben und wie ein theologisch-ethischer Beitrag in diesen zunächst so fachfremd erscheinenden Debatten aussehen kann.



Um das auszuloten, gilt es als erstes in den Blick zu nehmen, mit welchen Veränderungen wir tatsächlich konfrontiert sind. *Bernhard Irrgang* zeichnet zunächst nach, dass wir es nicht nur mit einer neuartigen Verbindung von Biologie und Technologie (z. B. in der synthetischen Biologie) zu tun haben, sondern dass der Paradigmenwechsel wesentlich weiter reicht und sich auf alle Lebens- und Wissenschaftsbereiche auswirkt. Zwei Entwicklungen sind dafür ursächlich: Einerseits die Fülle an vorhandenen und verwendbaren Daten (algorithmischer Turn), die aber nicht automatisch mehr Wissen bedeutet, sondern von allen Wissenschaften eine Interpretation dieser Daten, ihrer Genese und Aussagekraft verlangt. Andererseits lässt sich ein Wandel im Denkstil beobachten hin zu immer größerer Vernetzung und Komplexität. Gerade in der Verbindung von Technologie und Biologie verschwimmen nicht nur die Grenzen der Fächer und die von Natur und Technik, sondern Technik selbst wird vom Objekt, das der Ingenieur baut, zu einem wachsenden Organismus. Dieser Paradigmenwechsel verändert die Möglichkeiten vorauszusagen, wohin sich eine solche Technik entwickeln und welche Folgen sie zeigen wird. Die klassische Technikfolgenabschätzung (TA) gerät damit zunehmend in Schwierigkeiten. Zu vielfältig sind die Entwicklungsmöglichkeiten und zu groß ist die Offenheit der Zukunft. *Armin Grunwald* entwirft das Modell einer hermeneutischen TA, die die aktuell verbreiteten Aussagen über technologische Entwicklungen (Visionen, Versprechen, Horror szenarien) als Lernorte versteht. Indem wir analysieren und transparent machen, was Menschen sich jetzt erhoffen und befürchten, bekommen wir nicht nur einen Einblick in aktuelle Mensch-Technik-Verhältnisse, sondern können darüber auch die Zukunft der Technik beeinflussen, für die jetzt in unserer Gegenwart die Weichen gestellt werden. *Katharina Klöcker* greift diesen Ansatz auf und fragt, wie wir uns aus einer theologisch-ethischen Perspektive gegen die

scheinbare Ausweglosigkeit der Technik stellen können (also die Idee, dass technologische Entwicklungen wie ein D-Zug über uns hinwegrasen und uns nur die Anpassung bleibt). Sie hinterfragt dafür die Zukunftsversprechen, mit denen die Algorithmisierung wirbt (mehr Daten = weniger Leid) kritisch, um Handlungsspielräume zu eröffnen. Einen Schritt weiter geht *John Wyatt*, wenn er dafür wirbt, das reiche christliche Erbe, den Glauben an Menschwerdung und Auferstehung Jesu Christi dafür zu nutzen, die menschliche Verkörperung und Einzigartigkeit in einer Welt intelligenter Maschinen zu würdigen, zu respektieren und zu schützen. Dies versteht er nicht als Plädoyer gegen KI, sondern als Indienstnahme der hochentwickelten Technologien, um die Menschen zu befähigen, sich als menschliche Personen zu entfalten. Zu einer ganz ähnlichen Bewertung gelangt auch *Catrin Misselhorn*, die in ihrem Beitrag zur Maschinenethik die Chancen und Grenzen einer Artificial Morality auslotet, also der Möglichkeiten von Maschinen moralisch zu handeln. Sie entwirft dafür zentrale Grundsätze, die alle eine Orientierung am Menschen in den Mittelpunkt rücken.



Die neuen Technologien sind Orte des Lebens geworden, die einfordern, soziologisch, philosophisch, theologisch und ethisch neu überdacht zu werden

Trotz der unterschiedlichen Zugänge zeichnen die Beiträge so doch gemeinsam den fundamentalen Wandel nach, vor den wir uns gestellt sehen. Deutlich wird, dass der Mensch sich nicht mehr ohne Bezug zur Technik beschreiben kann. Die neuen Technologien sind nicht mehr nur Hilfsmittel, sondern Orte des Lebens geworden, die einfordern, soziologisch, philosophisch, theologisch und ethisch neu überdacht zu werden. Dass es zunächst Naturwissenschaften und Technikforschung sind, die diese Entwicklungen

ZUR AUTORIN

Anna Maria Riedl, geb. 1982, Dr. theol., Juniorprofessorin für Christliche Sozialethik an der Kath.-Theolog. Fakultät der Universität Bonn. Forschungsschwerpunkte: sozial- und moralphilosophische Grundlagenfragen Theologischer Ethik; Politische Ethik, Technikethik und Nachhaltigkeit (KI, Transhumanismus). Aktuelle Publikationen:

- 2022: Technik, in: Christliche Sozialethik. Grundlagen – Kontexte – Themen. Ein Lehr- und Studienbuch, hg. v. Heimbach-Steins, Marianne/Becka, Michelle u. a., Regensburg, 280–299;
- 2021: „Biopolitik als Anthropopolitik“. Theologische Ethik vor der Herausforderung des Transhumanismus, in: Gott, Gaia und eine neue Gesellschaft? Theologie anders denken mit Bruno Latour, hg. v. Bauer, Christian/Bogner, Daniel/Schübler, Michael, Bielefeld.

vorantreiben und als Expertinnen des Diskurses erscheinen, macht die theologisch-ethische Auseinandersetzung keinesfalls überflüssig, ganz im Gegenteil. Hier liegt nicht nur eine Aufgabe, sondern auch eine Chance, noch einmal grundsätzlich zu fragen, welchen Beitrag theologisch-ethische Orientierung leisten möchte und kann. Am Ende sind es vermutlich nicht die Scheingefechte um die Frage, ob das selbstfahrende Auto lieber die alte Dame oder die junge Mutter überfahren soll, die der theologisch-ethischen Reflektion Relevanz verleihen, sondern die ganz großen und grundlegenden Fragen, die wir nicht verstummen lassen dürfen: Wer wollen wir sein als Menschen und in welcher Gesellschaft wollen wir leben? Das gilt umso mehr da, wo Technik uns scheinbar Erlösung von schwierigen Entscheidungen verspricht.

Datenzentrierte Biologie und KI



Paradigmenwandel für Wissenschaft und Technologie

Im Zusammenspiel von KI und Biotechnologie zeigten sich in den letzten 40 Jahren umfangreiche Wandlungsprozesse im Forschungsdesign von Wissenschaft und Technologie – über die Biologie hinaus. Dabei handelt es sich um eine Entwicklung, die der Autor mit dem Begriff „Techno-research“ beschreibt. Dieser Strukturwandel beruht auf einem neuen Forschungs-Paradigma, dem „algorithmic turn“, der allerdings nicht – wie oft behauptet – zum Ende wissenschaftlicher Theorie führt, sondern auf einem neuen Denkstil beruht. Dieses Paradigma entsteht im Übergang von der klassischen Systemtheorie (Kybernetik) zur Synergetik (Theorie dynamischer komplexer Systeme). Die neue Orientierung an Komplexität und die Art, mit ihr datenbezogen umzugehen, verändert Praxis und Theorie von Wissenschaft und Technologie im 21. Jahrhundert.



Bernhard Irrgang

Vorbemerkung: Wie entstand Künstliche Intelligenz?

Ich beginne mit einem kurzen Überblick technologisch-ökonomischer Entwicklungspfade, die zur Genese von KI führten. KI entstand während des Zweiten Weltkriegs durch die Entschlüsselung des geheimen deutschen U-Boot-Codes Enigma durch Alan Turing und eines Programms zur Verfolgung von Raketen bzw. Flugzeugen, um sie abschießen zu können, entwickelt durch Norbert Wiener. Für beide Probleme waren extrem komplexe Berechnungen erforderlich, und zwar von Informationen/Daten, zum einen sprachlichen Ursprungs, zum anderen physikalischer Art (Flugbahnen). Eine weitere technologische Hilfe bot eine Rechenmaschine, deren Architektur John von Neumann als Computer (zunächst als Großrechner) entwarf. Das menschliche Rechnen als Vorbild für Rationalität (Irrgang 1992) konnte nun in eine Maschine ausgelagert werden. Computer hießen vor Erfindung des technischen Gerätes einfach Rechenknechte (und das sind sie auch). Künstliche Intelligenz ist daher ein Name für Datenverarbeitung, die auf sehr verschiedenen Ebenen stattfinden kann. Ein weiterer

entscheidender Entwicklungspfad war die Vernetzbarkeit von Rechnern, die eine ungeheure Ausweitung der Zusammenführung von Datenverarbeitung, Big Data, Plattformökonomie und datenzentrierter Wissenschaft ermöglichte. In der Philosophie des 20. Jahrhunderts war die Sprachforschung die fundamentale philosophische Leitdisziplin. Man nannte dies „linguistic turn“. An seine Stelle trat in den letzten beiden Jahrzehnten ein „algorithmic turn“.

Netzwerkstrukturen als Vorläufer des Internets lassen sich in zwei Bereiche (Irrgang 2022b, 316–383) aufteilen:

(1) KI, „algorithmic turn“ und datenzentrierte Wissenschaft entwickelten sich im Umfeld von Hippies, Technofreaks und Steward Brands Computernetzwerk *Whole Earth Catalog* (ökologisches Welt-Entwicklungs-Modell ab 1968) wie aus dem ARPANET¹ als Kinder von Wissenschaft, Technologie und dem Design der ersten IBM-Com-

puter. Diese Ansätze werden im *Open-Access-Modell* des Internet, welches vor allem der wissenschaftlichen Forschung verpflichtet bleibt, weitergeführt. Es begründet die Freiheit des kreativ-generativen, hypermodern-technologischen Designs des Internets, bleibt gemeinwohlorientiert und könnte in vielen anderen Bereichen der Gesellschaft in öffentlicher Verantwortung betrieben werden (Irrgang 2022b, 10–31).

(2) Der zweite Bereich ist die vor 30 Jahren erfundene GAFAM²-Welt der Plattformökonomien durch neoliberal-turbokapitalistische Verfechter des Traums des *Silicon Valley*. Sie umfasst vor allem Werbung und Verkauf, Multimedia-Anwendungen, Soziale Netzwerke und bezieht sich auf Lifestyleausformungen. Diese stehen im besonderen Bezug zur Eventökonomie und zum Überwachungskapitalismus. Digitalisierung und Plattformökonomie haben die Umweltkrise und den Klimawandel entscheidend verschärft (Zuboff 2018; Irrgang 2022b, 32–78).

¹ ARPANET (Advanced Research Projects Agency Network) war ein Computer-Netzwerk und wurde ab 1968 von einer kleinen Forschergruppe unter der Leitung des Massachusetts Institute of Technology und des US-Verteidigungsministeriums entwickelt.

² GAFAM ist ein Akronym für die größten IT-Unternehmen der Welt: Google, Amazon, Meta (Facebook), Apple, Microsoft.



Einerseits werden Disruptionen, die durch die Plattformökonomien im gesellschaftlichen Bereich in den letzten 30 Jahren entstanden sind, in der Öffentlichkeit bereits ausgiebig diskutiert. Die Menschheit hat andererseits in den letzten 70 bis 30 Jahren im „algorithmischen Turn“ und dem Computer in menschlich-mentaler Einbettung einen methodischen Zauberstab (also ein technologisches Instrument) Open Access und im Kontext von datenzentrierter Wissenschaft erhalten, um in harter Arbeit dem Wissen darüber näherzukommen, „was die Welt im Innersten zusammenhält“. KI entstand und prägte sich in Konvergenz mit Biologie und Biotechnologie aus; ein Prozess, an dem ich seit 1985, damals als Forschungs-Assistent am Institut für Moralthologie und Lehrbeauftragter für ethische Fra-

gen am Genzentrum der LMU München in Martinsried, selbst beteiligt war. Aufgrund dieser Eigenschaften wurde ich zum Nachfolger von Franz Böckle der Arbeitsgruppe ELSA (Ethical, Legal, Social Aspects) des europäischen Humangenom-Projektes in Brüssel, wel-



KI entstand und prägte sich in Konvergenz mit Biologie und Biotechnologie aus

ches gewissermaßen paradigmatisch für die forschungsethische Grundausrichtung mit der Einrichtung von Begleitforschung geworden ist, wenn es um Fragen der gesellschaftlichen und kulturell-ethischen Einbettung von Technik und Wissenschaft geht.

Daher „surfen“ innovative Menschen beim Erlernen des Umgangs mit neuen Technologien unterschiedliche Wellen der Unsicherheit (Irrgang 2008; Irrgang 2010; Irrgang 2022b). Sie erwerben implizites Wissen der reflektierten Praktiker:innen und Tüftler:innen (Irrgang 2022a). Ich habe diese Ansätze in den letzten 25 Jahren unter dem Namen „Technikhermeneutik“ (Irrgang 1996) zu präzisieren versucht. Das Konzept technologisch-ökonomischer Entwicklungspfade ist eine ihrer Ausprägungen (Irrgang 2016). Ein Beispiel wäre der Computer. Technikhermeneutik untersucht technische, gesellschaftliche, ökonomische oder kulturelle Einbettungsfaktoren seiner Entwicklung.

Maschinelles Lernen ist eine Methode für die Verarbeitung großer Datenmengen, z.B. bei der Modellierung einzelner Gene in einer Population. Ich habe eine solche Berechnung von Hand in einem Forschungsseminar der Würzburger Verhaltensbiologie im Sommersemester 1976 selbst verfolgen dürfen. Es war kein ausbaufähiges Modell für die Weiterentwicklung der Evolutionsbiologie, selbst wenn man eine größere Gruppe von Rechenkünstlern beschäftigte. Seit Mitte der 1980er Jahre (Einführung des PCs) bildeten sich eine neue Art von Wissenschaft aus: Sie umfasst Humangenetik und prädiktive Medizin, eine neue Evolutionstheorie auf der Basis des Selbstorganisationsparadigmas, ersetzte artüberschreitenden Gentransfer durch Genom-Editierung⁴ und überführte synthetische Biologie⁵ in „Transformations-Biotechnologie“ (Irrgang 2018). In diesen Prozessen manifestiert sich ein Paradigmenwechsel, der unter dem Schlagwort „Technoresearch“ gefasst und im folgenden ersten Kapitel zunächst überblicksartig beschrieben wird. Die Transformation von Forschung in Technoresearch ist nicht auf

Zum Einstieg: KI in Form von maschinellem Lernen als Blackbox und Grundlage für Technoresearch

Aus wissenschaftsphilosophischer Sicht fanden um die Mitte der 1980er Jahre zwei wissenschaftlich-technologische Revolutionen statt: In den Life Sciences das Humangenomprojekt und die neue Evolutionstheorie, im Bereich der KI der Übergang von der sog. starken regelbasierten und menschliche Experten kopierenden KI zu einer Ingenieur-KI der bewusst artifiziellen KI im Rahmen evolutiver Robotik und eines maschinellen Lernen (siehe Irrgang 2022c): Maschinelles Lernen („Deep Learning“³, neuronale Netze) erzeugt je nach Aufgabe aus verschiedenen digital aufbereiteten Datenmengen mithilfe unterschiedlicher Arten von Algorithmen Entscheidungsbäume, die Schlussfolgerungen aus ihnen ziehen, indem sie Daten gemäß den einem Softwareprogramm zugrunde gelegten Methoden in korrigierenden Feedbackschleifen einer Irrtumskorrektur unterwerfen. Man kann je nach Aufgabe zusätzliche Daten eingeben oder nicht. Der einzelne Rechenweg der von dieser Art von Software erzeugten Ergebnisse kann gemäß

dem alten Wissenschaftsideal theoretisch nicht überprüft werden (z.B. durch nachrechnen). Dies ist aber auch eine völlig neue Art von Mathematik, die ohne zugrundeliegende Computertechnologie nicht betrieben werden kann.

Das durch maschinelles Lernen entstehende Problem wird unter dem Thema ‚Explainability‘ (Erklärbarkeit) der KI diskutiert. Allerdings ist fehlende wissenschaftliche Erklärbarkeit technischer Artefakte bzw. Technologien (man nennt das auch „black-box“) weder für das Design noch den Umgang mit und den Gebrauch von neuen Technologien ein Hindernisgrund, sondern die Regel. Denn Technologien orientieren sich nicht an wissenschaftlicher Wahrheit, sondern an ihren Erfolgen im Alltag, vor allem in Bereichen mit innovativen Potenzialen.

³ Deep Learning bezeichnet einen Teilbereich Künstlicher Intelligenz und eine Methode maschinellen Lernens, welche sich auf künstliche neuronale Netze und große Datenmengen fokussiert. In der Praxis wird Deep Learning dafür genutzt, Bilder zu erkennen, Texte zu verstehen und bei Entscheidungen zu helfen.

⁴ Genom-Editierung ist ein Sammelbegriff für molekularbiologische Techniken zur zielgerichteten Veränderung von DNA, einschließlich des Erbguts von Pflanzen, Tieren und Menschen.

Biologie beschränkt, sondern Indikator für fundamentalere Wandlungsprozesse im Verhältnis von Technologie und Wissenschaft. Zwei zentrale Kennzeichen des Transformationsprozesses werden nachfolgend noch einmal eigens charakterisiert. Im zweiten Kapitel geht es um die Genese und Problematisierung von Datenverarbeitung. Diese Entwicklung, die auch unter Stichworten wie „Big Data“, „algorithmic turn“, Bioinformatik wie datenzentrierte Biologie eingeordnet werden kann, bedeutet nicht das Ende der theoretischen Arbeit in der Wissenschaft (weil allein die gesammelten Daten nicht für sich sprechen), sondern erfordert eine stark forschungsorientierte und technologisch-innovative Wissenschaft, die in der Lage ist, diese

1. Ein neues Paradigma entsteht: Technoresearch

Digitalisierung und Künstliche Intelligenz werden zu neuen Schlüsselfaktoren der hypermodernen Technologiezivilisation, insbesondere in ihrer Konvergenz mit der Transformations-Biotechnologie⁶ und zu einem der zentralen technologisch-ökonomischen Mega-Entwicklungspfade im Übergang vom Industriezeitalter in eine technologische Hypermoderne (vgl. Irrgang 2022b). Seit Mitte der 1970er Jahre konstituierte sich im Umfeld von Biotechnologie, Biomedizin und den Überlebenswissenschaften eine neue Forschungspraxis, die das traditionelle Bild der Natur- und Ingenieurwissenschaften in erheblichem Maße transformierte. Sie erzeugt im Labor, in der Feldforschung, in den Agrarwissenschaften und in der biotechnologischen Züchtungspraxis Ergebnisse, die auf einer schrittweisen Ausweitung der Laborforschung in die Feldforschung beruhen. Bisher hat sich das im Labor erworbene Umgangswissen impliziter Art (methodisch reflektierte Erfahrung) mit dem im Umgang mit in der Natur erworbenen Wissen als kompatibel erwiesen. Noch in den 1970er und 1980er Jahren für grundsätzlich gehaltene Un-

Daten richtig interpretieren und einordnen zu können. Das dritte Kapitel greift den mit Technoresearch verbundenen beschriebenen fundamentalen Wechsel im Denkstil auf – weg von der klassischen Ingenieurstechnik (Kybernetik) hin zu einem immer vernetzteren Denken (Synergetik) und einer Orientierung an Komplexität. Besagter Wechsel kann Leitbildcharakter für die Modellierung weiterer technologisch-ökonomischer Entwicklungspfade komplexer gesellschaftlicher, ökologischer, kultureller, klimarelevanter und für das Überleben der Menschheit dringend erforderlicher Prozesse gewinnen. (Konkretisierungen zu diesem Themenfeld bietet auf knapp 400 Seiten Irrgang 2022b)

terschiede wie zwischen künstlichem Labor und natürlicher Natur oder Verbotsversuchen von artenüberschreitendem Gentransfer wurden hypothetisch (Irrgang 2018).

Wissenschaft ist durch Theoriewandel und das Festhalten von Ergebnissen charakterisiert, Forschung jedoch ist (hoffentlich) innovative Praxis. Forschungshandeln geschieht über das implizite Wissen: die technischen Kompetenzen der Forschenden (das Tüftler-Paradigma: Irrgang 2022a). Dabei ist Technologie eine Mischung aus tatsächlichem technischem *Know-how* und technisch instrumentellen Fertigkeiten sowie Methodologie-Kenntnissen bis hin zu einem Verfahrens-*Know-how*. Letzteres ist nicht bloßes Faktenwissen, sondern schließt ein ak-

tives Verfügen-Können über bestimmte Kompetenzen, methodisch zu arbeiten, ein. Es geht darum, etwas verstehend machen zu können. Die Gentechnik bis zur Jahrhundertwende verwendete die klassischen Methoden des Gentransfers vor der Genom-Editierung (Irrgang 2003) und entsprach dem Leitbild der traditionellen Ingenieurwissenschaften. Die klassische Bioingenieurkunst versprach dabei zwar die Synthese von Biologie und Technik, ihr Leitbild aber war nicht die Biologie, sondern die (mechanische) Ingenieurkunst: Es handelte sich um ein Konstruktionsideal. Seit gut zehn Jahren ändert sich das Leitbild für Bio-Design: Der Übergang von Gentechnik und klassischer synthetischer Biologie zur Transformations-Biotechnologie wurde vor allem durch Genom-Editierung ermöglicht und wäre ohne datenzentrierte Formen von Wissenschaft nicht durchführbar gewesen (Irrgang 2018; Irrgang 2022a, 129–159).



Der Übergang von Gentechnik und klassischer synthetischer Biologie zur Transformations-Biotechnologie wurde vor allem durch Genom-Editierung in Verbindung mit datenzentrierter Wissenschaft ermöglicht

Noch einmal zusammengefasst: Der theoretische Hintergrund für ein Verständnis des Lebendigen änderte sich durch die Verschränkung von Epigenetik⁷ und einer neuen erweiterten Evolutionstheorie seit etwa 25 Jahren. Beide Forschungsfelder hätten ihre Er-

⁵Die synthetische Biologie ist ein Fachgebiet im Grenzbereich von Molekularbiologie, organischer Chemie, Ingenieurwissenschaften, Nanobiotechnologie und Informationstechnik. Dabei werden im Labor biologische Systeme erzeugt, die so in der Natur nicht vorkommen.

⁶Transformations-Biologie beschreibt die Weiterentwicklung der synthetischen Biologie, z. B. durch neue Verfahren, wie genome editing, in die DNA einzugreifen (Neugebauer 2019).

⁷Epigenetik erforscht, welche Faktoren die Aktivität eines Gens und damit die Entwicklung der Zelle zeitweilig festlegen. Sie unter sucht die Änderungen der Genfunktion, die nicht auf Veränderungen der Sequenz der DNA, etwa durch Mutation oder Rekonstruktion, beruhen und dennoch an Tochterzellen weitergegeben werden.

gebnisse nicht erzielen können ohne den „algorithmic turn“, „Deep Learning“ und neuronale Netze sowie in der Biologie durch Formen datenzentrierter Systembiologie, Bioinformatik und Transformations-Biotechnologie. Epigenetik und evolutionäre Entwicklungsbiologie haben bei ihrer Theoriebildung den Übergang von der Kybernetik zur Synergetik, der klassischen Bioingenieurskunst zur datenzentrierten Systembiologie, vollziehen müssen, um erfolgreich zu sein (Lange 2020; Irrgang 2022b; Irrgang 2022c). Dieser Paradigmenwechsel ist zwar keineswegs abgeschlossen, vollzog sich zunächst eher parallel, dann aber mit zunehmender Konvergenz ab den späten 1980er Jahren sowohl in der KI als auch in der entstehenden Transformations-Biotechnologie. In der KI fand eine Transformation von der regelbasierten symbolischen (starken) und theorieorientierten KI zu Formen schwacher Ingenieurs-KI statt, die sich in evolutionären Algorithmen, Deep Learning und neuronalen Netzen manifestierte (Irrgang 2022a, 86–128; Irrgang 2021, Irrgang 2022c). Sie orientierte sich nicht mehr an formaler Logik und Konstruktion, Mechanismen und Automatismen, sondern am Forschungs-Design evolutionärer Entwicklungsbiologie, epigenetischer Expression von Genen und neuronalen Netzen im menschlichen

Gehirn, also komplexen dynamischen biologischen Systemen, die sich mit dem algorithmic turn in einer völlig neuen Art und Weise erfassen ließen und zur Transformation von Wissenschaft in Technoresearch führte.

Dieser Prozess ist allerdings nicht nur auf die Biologie beschränkt, denn wir sind tatsächlich Zeugen einer weit aus größeren interdisziplinären Welle der Natur- und Technowissenschaften als das, was in einer akademischen Disziplin, der Biologie, allein abläuft. Die Konvergenz von KI und Biologie zeigt gemäß meiner Interpretation einen Paradigmenwechsel, der die Struktur der Ingenieurwissenschaft (und mit ihr der Wissenschaft) insgesamt transformiert. Natur-, Human- und Technikwissenschaften (Technoscience als Experimentalkunst und Laboratoriumsforschung) werden zu Formen von Technoresearch auf der Basis von daten- und algorithmenzentrierter Forschung (Irrgang 2022a; Irrgang 2022b). Vergleichbare Konvergenzen wie zwischen KI und Biologie finden wir auch in Klimaforschung, Biomedizin, Humanwissenschaften, Digital humanities, Gesellschaftswissenschaften und Ökonomie. Konvergenz und Transformation meint gegenseitige interdisziplinäre Durchdringung, nicht Ersatz einer durch eine andere Form von Wissenschaft.

umsuchen im interdisziplinären Zusammenhang und als Fortsetzung der Laboratoriums-Forschung in der Virtuellen Realität (Irrgang 2022a). Wissenschaftliche Relevanz wird durch aufgezeichnete Signale und deren Auswertung oft auch automatisiert erzeugt. Der Umgang mit den Rechnern musste allerdings erst erlernt werden. In den Messinstrumenten und Rechnern platzierte elektromagnetische Pick-ups und Kontaktmikrofone ermöglichen statt der menschlichen eine „technisch teilnehmende Beobachtung“ (Fischer u. a. 2020, 9–13)



Technoresearch verlagert Forschung in den Raum des Digitalen

In den letzten drei Dekaden sind Online-Datenbanken, digitale Visualisierungstechniken und automatisierte Datenanalysen bevorzugte Schlüssel-Werkzeuge bei der anwachsenden Vermehrung der Größe und Universalität wissenschaftlich relevanter Information geworden. Sie haben Methoden und die Art und Weise, wie wissenschaftliche Forschung betrieben und theoretisch reflektiert wird, revolutioniert. Ein neues, durch Daten vorangetriebenes Forschungsparadigma ist entstanden, gemäß dem aus Daten Schlussfolgerungen ohne vorausgehende Hypothesenerstellung gezogen werden. Dabei liegt die eigentlich revolutionäre These nicht im Daten-Konzept, sondern besteht in einem neuen Umgang mit Daten. Dieser betrifft nicht nur die Wissenschaft, sondern auch den gesellschaftlichen und politischen Umgang mit wissenschaftlicher Forschung. Eine Konvergenz der digitalen Technologien für die Produktion, Verbreitung und Analyse von Daten mit neuen regulatorischen Regimen für die Resultate wissenschaftlicher Forschung zeichnet sich ab und verbindet sich mit der Frage nach ihrem legitimen

2. Datenzentrierte Wissenschaften und ihre epistemologische Verortung


Technoresearch ist eine Form technologisierter Forschung, die sich in den letzten Jahrzehnten immer mehr durchgesetzt hat. Zugrunde liegt dieser Konzeption ein Modell von Daten, welche als einfache und basale Repräsentanten der untersuchten Gegenstandsbereiche (also insbesondere der Natur oder des Menschen in seiner kulturellen Einbettung) angesehen werden, aus denen durch bottom-up-Verfahren, insbesondere technologischer Art, Wissen erzeugt werden kann. Daten gel-

ten als Grundlage von Fakten, insbesondere wissenschaftliche Daten. Technoresearch ist gekennzeichnet durch die Dematerialisierung von Forschung, durch ihre Verlagerung in den Raum des Digitalen und auf Handlungen, die sich am Rechner vollziehen. Dabei gilt Forschen im digitalen Raum als Bricolage⁸, als Probieren und Her-

⁸ Bricolage steht für ein Verhalten, bei dem der Akteur (Bricoleur) mit den zur Verfügung stehenden Ressourcen Probleme löst, statt sich besondere, speziell für das Problem entworfene Mittel zu beschaffen.



Gebrauch. Dies impliziert epistemologische Herausforderungen, die in den Prozess der Datensammlung involviert sind, z. B. Klassifikation, Interpretation und Vervielfältigung von begrifflichen, materialen und sozialen Strukturen, in welche epistemologische Prozesse eingebettet sind. Eine solche Konzeption wendet sich aber gegen die These vom Ende der Theorie, die mit der These von Big Data verbunden wird, und votiert für einen theorie-zentrierten Weg der Datenermittlung und der die Forschung unterstützenden Institutionen sowohl in methodischer wie in finanzieller Weise (Leonelli 2016, 1–3).

 Online-Datenbanken haben die Art und Weise, wie wissenschaftliche Forschung betrieben wird, revolutioniert

Viele Wissenschaftler sind der Meinung, dass mit dem Beginn des 21. Jahrhunderts eine Zeit des epochalen Wandels in der Art und Weise, wie Wissenschaft und Forschung betrieben wird, angebrochen ist. Für diese Art der Forschung ist eine neue Epistemologie erforderlich (Leonelli 2016, 9). Von der Öffentlichkeit nicht so beachtet wie die ökonomischen Plattformen ist eine digitale Infrastruktur auch für Wissenschaft und Forschung entwickelt worden, in besonderem Maße in den Bereichen der Biologie im Anschluss an das Humangenom-Projekt.⁹ Diese digitale Infrastruktur soll ihre wissenschaftlichen Nutzer dazu befähigen, ihre Forschung ordnungsgemäß durchzuführen. Damit bleibt ein gewisses Spannungsverhältnis zwischen den Versprechungen solcher Infrastrukturen und der realen Einlösung bestehen (Leonelli 2016, 15).


Datenbanken biologischer Art wurden immer dringlicher, seit nach dem Zweiten Weltkrieg die Biologie ihre molekulare Dimension entdeckte (Leonelli 2016, 17). Für diesen Ansatz waren Modellorganismen und ihre Untersuchung bzw. Konstruktion wegwei-

send (Leonelli 2016, 20). Datenbanken stellten allerdings Fragen nach der zugrunde gelegten Art der Kategorienbildung und der Zuschreibung relevanter Ebenen der Beschreibung und Einordnung (Klassifikation). Hierzu waren Interpretationsprozesse erforderlich (Leonelli 2016, 26–28), die von speziell ausgebildeten Datenbank-Betreuern vorgenommen werden mussten (Leonelli 2016, 31–35). Diese sollten Datensuch-Prozesse zur Forschungsunterstützung in methodisch geeigneter Weise aufbereiten und zur Verfügung stellen. Art und Weise der Datenermittlung und -aufarbeitung sind entscheidende Merkmale der Epistemologie einer datenzentrierten Biologie, welche Kennzeichen einer relativen Neuartigkeit des historischen Phänomens und der Eigentümlichkeit der Art des wissenschaftlichen Forschens ist. Sie ist verbunden mit dem, was man gemeinhin unter Big Data versteht (Leonelli 2016, 39).

Der weithin herrschende Naturalismus und Positivismus in der Bestimmung der Begriffe von Information, Daten und Fakten und der Übergang von Daten zu Wissensstrukturen ist für die sich entwickelnden Strukturen von Technoresearch aus epistemologischer Perspektive weitgehend ungeklärt. Häufig ist die technische Natur wissenschaftlicher Daten nicht ausrei-

3. Synergetik – ein neues Paradigma für Wissen

Hinter dem Zusammenwachsen von Biotechnologie und KI steht als neuer Denkstil die allgemeine Theorie dynamischer Systeme (Synergetik). In dieser wird die zeitliche Veränderung der untersuchten Zustände durch Gleichungen beschrieben. Komplexe dynamische Systeme gehorchen fachübergreifend in Physik, Chemie, Biologie und Ökologie denselben oder ähnlichen mathematischen Gesetzen. Da-

 Wissen ist nicht identisch mit Daten, Fakten und Information

chend bewusst. Daten für spezifische Forschungsanliegen und zur Bearbeitung von Projekten müssen sowohl für die methodische Durchführung wie für die epistemologische Absicherung sorgfältig durchdacht sein, denn das gehört zu den Bedingungen für die Transformation von Daten in wissenschaftliches Wissen. Daten sind genauso designt wie die Methoden, durch die sie erzeugt werden. Sie sind materiale Artefakte (Leonelli 2016, 69). Wissen ist nicht identisch mit Daten, Fakten und Information, sondern ergibt sich aus den Zusammenhängen von Einsichten, von Kenntnissen und von kontextuell eingebettetem Wissen, welches durch Theorie bestimmt ist. Leonelli weist dazu auf die epistemische Rolle der Klassifikation angesichts der Variabilität der Realität hin (Leonelli 2016, 127). Von einem Ende der Theorie kann also keinesfalls die Rede sein, sondern nur von einer anderen Art von Theorie, die an Forschung orientiert ist. Daher ist das Framing für Big-Data-Analysen ein epistemologisch wichtiger Punkt (Leonelli 2016, 137), hermeneutische Philosophie nennt das Horizontbildung.

her liefert diese Theorie einen Schlüssel zur Komplexität der Welt. Die Grundidee ist immer dieselbe: Erst komplexe Wechselwirkungen von vielen Elementen führen zur Erzeugung neuer Eigenschaften im Gesamtsystem, die nicht auf einzelne Elemente zurückführbar sind. So ist ein einzelnes Wassermolekül nicht feucht, aber eine Flüssigkeit durch die Wechselwirkung vieler solcher Elemente ist es. Diese Emer-

⁹Das Humangenom-Projekt war ein internationales Forschungsprojekt, mit dem Ziel, das menschliche Genom vollständig zu entschlüsseln.

genz oder Selbstorganisation von vielen neuen Systemeigenschaften wird in der Systembiologie berechenbar und in Computermodellen modellierbar. In diesem Sinne ist die Systembiologie der Schlüssel zur Komplexität des Lebens. Manche Systemeigenschaften sind der jeweiligen Systemumgebung angepasst und setzen sich durch, andere zerfallen und werden ausgesondert. In einer komplexen und sich ständig verändernden Umwelt können nicht alle Eventualitäten in einem Programm berücksichtigt werden. Eine komplexe körperliche Selbstorganisation funktioniert weitgehend ohne bewusste Zentralsteuerung. Das Prinzip, das dahintersteht, ist Learning by doing (Mainzer 2020, 18 f.).

Um den neuen theoretischen Ansatz zu begreifen, müssen wir den Schritt von der linearen zur nichtlinearen Dynamik von Systemen gehen. Lineare Gleichungen sind leicht zu lösen. Nichtlineare Gleichungen in geo-

metrischen Kurven darzustellen, erlaubt nicht immer eine beliebig genaue Berechenbarkeit, selbst mit unseren besten Computern nicht. Ein Beispiel sind die Mehrkörper-Probleme der Himmelsmechanik, bei denen mehr als zwei Himmelskörper durch Gravitation aufeinander wirken. Sie erzeugen Rückkoppelungen, die nichtlinearen Bewegungsgleichungen der Planeten gleichen (Mainzer 2020, 52 f.). Noch komplexer erweisen sich die Probleme im Bereich des Lebens. Schon Zellen sind als komplexe dynamische Netzwerke zu verstehen (Mainzer 2020, 63 f.). Sensorische Transkriptionsnetzwerke und Entwicklungsnetzwerke sind nur zwei Beispiele. Erstere sind für die Wahrnehmung und Reaktion einer Zelle zuständig, während Letztere sich die Entwicklung einer Zelle zu organischem Gewebe bzw. zu einem Organismus organisiert. Biologische und technische Schaltpläne haben also ei-

ne gewisse Ähnlichkeit (Mainzer 2020, 75–80).


Das Zusammenwachsen der beiden Disziplinen ermöglichen mathematische Modelle komplexer zellulärer Netzwerke (Mainzer 2020, 81). Systembiologie rekonstruiert komplexe Schaltpläne und eröffnet einen breiten Pfad zu einer erneuerten Biomedizin. Dazu bedarf es nicht nur großer Datenbanken in den einzelnen medizinischen Disziplinen, sondern fächerübergreifender Computermodelle und klinischer Kooperationen (Mainzer 2020, 101). Die Bioinformatik ist mit Biologie, Informatik, Mathematik, Statistik und Ingenieurwissenschaften verbunden. Im Zentrum stehen Algorithmen und Software, mit denen aus molekularen Datenmassen komplexe Strukturen und Funktionen von z. B. Proteinen erschlossen werden sollen. Der Schlüssel dazu sind molekulare Codes (Mainzer 2020, 107). Zwischen Bioinformatik und KI wird eine

LITERATUR

- Fischer, Philipp u. a. (2020): Datennaturen. Ein Gespräch zwischen Biologie, Kunst, Wissenschaftstheorie und -geschichte, Zürich.
- Irrgang, B. (1992): Die Maschinisierung des Subjektes und die rationale Konstruktion der Gesellschaft. Künstliche Intelligenz als Mäeutik eines neuen Bildes vom Menschen und der Art seines Zusammenlebens?, in: Schmidt, J. (Hg.): Denken und denken lassen. Künstliche Intelligenz. Möglichkeiten, Folgen, Herausforderungen, Neuwied/Berlin 1992, 115–154.
- Irrgang, B. (1996): Von der Technologiefolgenabschätzung zur Technologiegestaltung. Plädoyer für eine Technikermeneutik, in: Jahrbuch für Christliche Sozialwissenschaften 37, 51–66.
- Irrgang, B. (2003): Von der Mendelgenetik zur synthetischen Biologie. Epistemologie der Laboratoriumspraxis Biotechnologie, Dresden.
- Irrgang, B. (2008): Philosophie der Technik, Darmstadt.
- Irrgang, B. (2010): Von der technischen Konstruktion zum technologischen Design. Philosophische Versuche zur Theorie der Ingenieurspraxis, Münster.
- Irrgang, B. (2016): Technologische Entwicklungspfade: Innovation und Folgelasten. Macht und Ohnmacht angewandter Ethik bei der Einbettung nutzerfreundlicher Technologie, Würzburg.
- Irrgang, B. (2018): Art. Gentechnik, in: Goerres Gesellschaft/Herder Verlag (Hgg): Staatslexikon. Band 8, Freiburg, 1126–1138.
- Irrgang, B. (2021): Neuere Literatur zu theoretisch-epistemologischen Fragestellungen der künstlichen Intelligenz, in: Philosophischer Literaturanzeiger 74/3/2021, 263–283.
- Irrgang, B. (2022a): Vom impliziten technischen Wissen zum technologischen Design von KI und Biotechnologie. Cyberphilosophie des Design-Konzeptes der Tüftler und reflektierten Praktiker, Würzburg.
- Irrgang, B. (2022b): Wege zu ökosozial-liberalen hypermodernen Technologiezivilisationen. Networking, KI, Kybernetik und Synergetik – die Rolle von Wissenschaft und Technologie, Würzburg.
- Irrgang, B. (2022c): (Bio)Philosophie, eine neue Evolutionstheorie und die Konvergenz von Biotechnologie und KI, in: Philosophischer Literaturanzeiger 75/2/2022, 178–211.
- Lange, A. (2020): Evolutionstheorie im Wandel. Ist Darwin überholt? Berlin.
- Leonelli, S. (2016): Data-Centric-Biology. A Philosophical Study, Chicago.
- Neugebauer, R. (2019) (Hgg.): Biologische Transformation, Berlin.
- Mainzer, K. (2020): Leben als Maschine: Wie entschlüsseln wir den Corona-Kode? Von der Systembiologie und Bioinformatik zur Robotik und künstlicher Intelligenz, EA 2010, Paderborn.
- Zuboff, S. (2018): Das Zeitalter des Überwachungskapitalismus, Frankfurt/New York.



sehr tief liegende Äquivalenz evolutionärer, mathematischer und technischer Verfahren deutlich, die sich darin manifestiert, dass mit mathematischen Symbolen und durch Computermodelle molekular und zellulär verschlüsselte effektive Verfahren der Natur repräsentiert werden können. Damit eröffnet uns die Erfindung des Computers eine grundlegende und einzigartige, in ihrer Tragweite zunächst nicht absehbare technologische Innovation mit wissenschaftlicher Brisanz: Alle effektiven

 Zwischen Bioinformatik und KI wird eine sehr tief liegende Äquivalenz evolutionärer, mathematischer und technischer Verfahren deutlich

dynamischen, also synergetischen Prozesse sind möglicherweise im Computer modellierbar. Das wäre der Kern einer vereinten Theorie komplexer Netzwerke. Die symbolischen Codes im Computer wären nur unsere menschlich-mentale Art der Informationsverarbeitung, welche atomare, molekulare, zelluläre und evolutionäre Prozesse repräsentieren (Mainzer 2020, 306 f.). Da Computerprogramme von Menschen ersonnen sind und für Menschen verständlich sein müssen, werden sie mit Symbolen von Programmiersprachen dargestellt. Das ist aber nur eine spezielle Kodierung von Information in technischen

Fazit

Mit der Transformation der Lebenswissenschaften (Irrgang 2022a; Irrgang 2022c) haben wir ein Modell und einen Gegenentwurf zu der auf monopolistischen Plattformökonomien beruhenden neoliberal globalisierten und digitalisierten Internet-Industriegesellschaft, die ökologische Probleme in den letz-

ZUM AUTOR

Bernhard Irrgang, geb. 1953, Prof. Dr. Dr., Lehramtsassessor (emeritiert 2019); Promotion in Philosophie zur Skepsis in der Wissenschaftsphilosophie der Aufklärung 1982; Promotion in Theologie 1991 in Würzburg/München über Anthropozentrik und christliche Umweltethik; Habilitation für Philosophie in Bamberg 1996 zur Forschungsethik der Biotechnologie; *Lehre*: fakultätsübergreifend und international Technikphilosophie interdisziplinär und transkulturell sowie angewandte Ethik an der TU Dresden, zuvor in Würzburg und Braunschweig Philosophie, in München und Siegen Moraltheologie, christl. Sozialethik; Dozent VHS München für westliche und altindische Philosophie 87/97; Lehrbeauftragter Genzentrum München 87/93 für Bioethik. Zwischen 1995 und 2009 internationale Lehre zur Gestaltung technologisch-ökonomischer Entwicklungspfade in Südamerika, Süd- und Ostasien, Ostafrika und den USA, Gastprofessur Indien, IIT Chennai 2003; *Politikberatung*: EU 88/89 (Human-genetik ELSA), 2000/07 (Privatheit im Internet, FIDIS); Deutsche Bischofskonferenz 2001/10 (Nachhaltigkeit); BMBF 2012/13 (Industrie 4.0), 2015 (Internetinstitut), 2018/21 (KI und virtuelle Realität).

Heute Ausbau und Weiterführung der Dresdner Technikphilosophie in unabhängiger Forschung und mit wissenschaftlichen Publikationen als Gründungsmitglied und Ehrenpräsident von NetPhilTech e.V. (mit Doktorandinnen/Postdoc-Team) im Umfeld der TU Dresden in digitaler Vernetzung mit Instituten für Technikphilosophie an der Universität Florenz, für Informatik an der TU Kaiserslautern und für Luft- und Raumfahrt an der TU Berlin im Übergangsfeld von Evolutionsbiologie, Bioinformatik, Hirnforschung, KI und Robotik. Diskussion über eine neue Philosophie in der technologisierten Gesellschaft in Blog-Beiträgen organisiert von schweizerischen und österreichischen Universitäten und in Podcasts mit dem Chaos Computer Club. Homepage: <http://www.bernhard-irrgang.eu/>

Systemen. In biologisch-intelligenten Systemen ist eine solche Zwischen-Repräsentation mit Sprachen wohl nicht notwendig, da Informationen durch molekulare und zelluläre Wechselwirkungen verschlüsselt und verstanden werden (Mainzer 2020, 308 f.).

ten 30 Jahren exponentiell verschärft hat. Welche neuen Möglichkeiten in der datenzentrierten Wissenschaft stecken, zeigt die Verminderung der Dauer von Impfstoffentwicklung und -testung von ca. 15 auf unter einem Jahr. Aber technologisch-wissenschaftliche Entwicklung (Technoresearch) ist nur das ei-

ne – die gesellschaftliche Reaktion und Einbettung das andere, von den Querdenkern bis hin zur Strategie Chinas bei der Pandemiebekämpfung. Die enge Verzahnung der Entwicklung von KI, Biologie und Biomedizin in den letzten 50 Jahren macht mir persönlich allerdings Hoffnung, dass die Menschheit ein weiteres Mal einen Ausweg aus einer verzweifelt scheinenden Lage finden wird (Irrgang 2022b, 316–383). Plakatativ gesprochen: Klassische kybernetische Strukturen münden ins Politische umgesetzt in autoritäre Herrschaftsstrukturen, synergetische Selbstorganisation in verschiedene Formen von Demokratie. Kennengelernt habe ich übrigens Synergetik als Student in der Schöpfungstheologie von Alexandre Ganoczy 1974/75.

Digitalisierung und Künstliche Intelligenz

Technikfolgenabschätzung trifft Anthropologie



Die Technikfolgenabschätzung (*technology assessment; TA*) erarbeitet Orientierung zum technischen Fortschritt durch Zukunftsaussagen und ihre Bewertung. Digitalisierung und Künstliche Intelligenz erschweren Zukunftsaussagen angesichts des maschinellen Lernens und neuer Mensch/Technik-Verhältnisse. Selbst der gängige Ansatz der TA, mit Szenarien zu arbeiten statt Vorhersagen zu machen, funktioniert häufig nicht mehr. Zu offen und unwägbar ist die Zukunft. In diesem Artikel stellt der Autor dar, wie die TA mit dieser Situation umgeht. In einer hermeneutischen Perspektive nutzt sie verbreitete Zukunftsaussagen zu Digitalisierung und KI, um über die Gegenwart zu lernen, insbesondere über die sich in den Projektionen der KI zeigenden neuen Mensch/Technik-Verhältnisse. Diese gilt es zu explizieren, um sie der transparenten Debatte zugänglich zu machen und die Ergebnisse für Technikgestaltung in diesem Feld zu nutzen.



Armin Grunwald

Was ist Technikfolgenabschätzung (TA)?

Die TA ist vor etwa fünfzig Jahren in den USA als wissenschaftliche Politikberatung des Kongresses zu Fragen des technischen Fortschritts entstanden. Von Beginn an wurde sie als interdisziplinäres Feld für vorausschauende Folgenforschung und deren Reflexion verstanden und entwickelt. Denn Technikfolgen haben unterschiedliche Aspekte, etwa ökologische, soziale, kulturelle, ethische, rechtliche und ökonomische, so dass ihre Erforschung nicht an Disziplingrenzen enden darf. Das entstehende Folgenwissen soll als ein „Wissen zum Handeln“ Entscheidungsträger, zivilgesellschaftliche Akteure und Bürger, aber auch Ingenieure in die Lage versetzen, besser reflektierte Entscheidungen zur Gestaltung und Nutzung des technischen Fortschritts zu treffen. Dies geschieht teils als Auftragsforschung, teils im Rahmen thematisch unabhängiger Forschung.

Die zentrale Motivation für das Entstehen der TA, ihre Verbreitung und ihre thematische Diversifizierung war

und ist, dass im technischen Fortschritt nicht nur die gewünschten Effekte und Innovationen, sondern immer wieder unerwartete und teilweise gravierende Technikfolgen auftreten. Statt davon überrascht zu werden und die Folgen reparieren oder kompensieren zu müssen, insofern dies überhaupt noch möglich wäre, soll frühzeitige Analyse ermöglichen, mit ihnen besser umgehen zu können (Grunwald 2022). Entsprechend soll die TA den vorausschauenden und umfassenden Blick auf die Folgen des wissenschaftlich-technischen Fortschritts systematisch stärken und entsprechende Orientierung in gesellschaftliche Meinungsbildung und politische Entscheidung einbringen. Es geht darum, die naivfortschrittsoptimistische *wait-and-see*-Haltung zu überwinden, nach der im Wesentlichen positive Folgen zu erwarten sind und eventuelle unerwartet eintretende, negative Technikfolgen im Nachhinein leicht repariert werden könnten. Eine solche Haltung ist für

Entwicklung und Einsatz vieler moderner Technologien weder politisch, ökonomisch noch ethisch verantwortbar (Jonas 1979).

Damit die TA den Anspruch der Vorausschau einlösen kann, benötigt sie belastbares Zukunftswissen über die möglichen, plausiblen oder wahrscheinlichen Folgen neuer Technik. Das Wissen über diese Folgen, z. B. in Form von Energieszenarien oder der Vorausschau zukünftiger Einsatzfelder der Künstlichen Intelligenz (KI) in unterschiedlichen Anwendungsfeldern soll Entscheidungsprozesse zur Forschungsförderung, Innovationspolitik und Regulierung, aber auch für öffentliche Technikdebatten durch Wissen und Bewertung bereichern. Mögliche Technikfolgen werden analysiert und auf Erwünschtheit oder Zumutbarkeit bewertet, so etwa in Klimapolitik, Energiepolitik oder Sicherheitspolitik. Im konsequentialistischen Handlungsschema der TA (Abb. 1) sind Bewertung und Abwägung von Technikfolgen handlungsleitend für zu treffende Entscheidungen, beispielsweise für



die politische und rechtliche Ausgestaltung des Wegs zu Mobilitätssystemen mit autonomen, also KI-gesteuerten Autos.

Zukunftsbezüge der Künstlichen Intelligenz

Üblicherweise wird Technik und damit auch KI als Mittel für menschliche Zwecke angesehen. Dieses Bild ist nicht falsch; so werden auch Algorithmen für bestimmte Aufgaben programmiert und an Referenzdaten trainiert, damit sie diese Aufgaben möglichst gut erfüllen können. Allerdings ist es unvollständig. Es kommt zu weiteren, nicht gewollten und unerwarteten, oft dann auch unerwünschten und problematischen Folgen, die nicht im Blickfeld der Technikentwickler:innen, Ingenieur:innen und Informatiker:innen lagen wie etwa unerwartete Nutzungen neuer Technologien, Missbrauch und Zweckentfremdung oder Risiken für Gesellschaft und Umwelt.

Der Einsatz Künstlicher Intelligenz kann unerwartete, problematische Folgen nach sich ziehen

Klimawandel und Biodiversitätsverlust, die Abhängigkeit von Öl- und Gasimporten und die Gefährdung der demokratischen Diskussionskultur durch Auswüchse der Internetkommunikation und Manipulation sind typische Beispiele für *nicht intendierte* Folgen des technischen Wandels (Grunwald 2022). Dies bietet Anlass, zwei Zukunftsbezüge von Technik allgemein und KI-unterstützter Technik im Besonderen zu unterscheiden: (1) *ex ante* und (2) *ex post*.

(1) KI wird – wie andere Technik auch – entwickelt, um mit den Produkten Ziele und Zwecke zu erreichen. Sie soll bestimmte Funktionen und Leistungsmerkmale im Innovationsprozess realisieren, z.B. in der Mustererkennung für die Diagnostik medizinischer

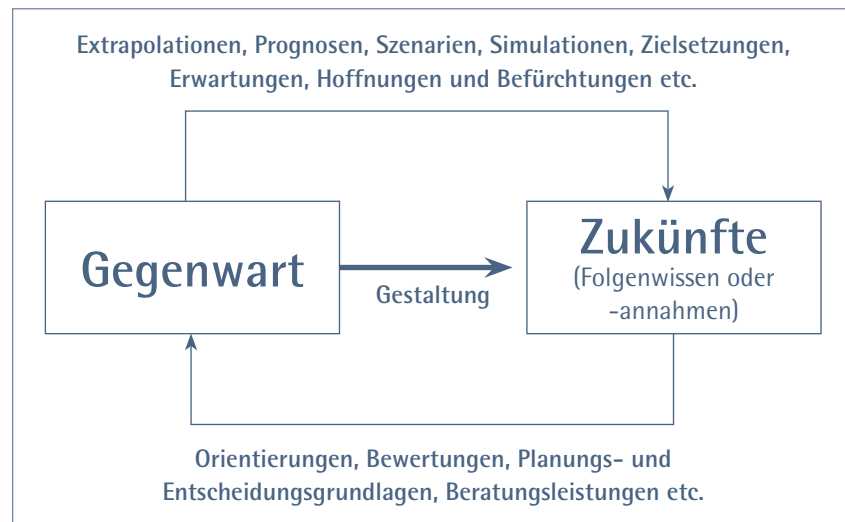


Abb. 1: Das konsequentialistische Schema der Orientierung durch Zukunftsanalyse (modifiziert nach Grunwald 2022)

Sachverhalte, für Sicherheitsbehörden zur Früherkennung von Terrorgefahren oder in der Optimierung internationaler Logistikketten. Der Bedarf zukünftiger Nutzer und die Nachfrage auf zukünftigen Märkten werden erhoben, um Anhaltspunkte für die Richtung der noch zu entwickelnden KI zu gewinnen. KI-Systeme und Produkte werden auf die Zukunft hin entworfen: Ihre Leistungen sollen in einer ökonomischen, politischen und gesellschaftlichen Zukunftswelt Möglichkeiten der Nutzung eröffnen, auf Nachfrage treffen und Akzeptanz am Markt finden. Der Zukunftsbezug *ex ante* ist planend.

(2) Wenn KI-Systeme in diesem Modus entwickelt und in die Anwendung gelangt sind, werden sie üblicherweise zunächst zu den intendierten Zwecken genutzt. Dann jedoch können sie auf die Nutzer zurückwirken und Gewohnheiten, Lebensstile, ökonomische Verhältnisse, soziale Zusammenhänge oder kulturelle Aspekte beeinflussen. Einmal vorhandene Technik verändert die Welt, weil sie neue Zwecksetzungen, Geschäftsmodelle, Lebensstile etc. ermöglicht. Beispiele sind die Geschichte des Telefons und die noch kurze aber beeindruckende Entwicklung der *social media* im Internet mit ihren weltweiten Folgen für Kommunikations- und Vernetzungsgewohnhei-

ten. Der Zukunftsbezug *ex post* meint, dass neue Technologien, etwa KI-gestützte Dienstleistungen, in der Nutzung eine unabsehbare Menge an Folgen jenseits der ursprünglichen Zwecksetzungen mit sich bringen, zu denen auch die bereits erwähnten nicht intendierten und oft problematischen Folgen und Risiken gehören.

Auch noch so positive Absichten und Ziele *ex ante* können nicht verhindern, dass *ex post* erhebliche negative und nicht intendierte Folgen entstehen können. Diese Situation ist zentral für die Verbreitung der TA (Grunwald 2022) – ohne den Zukunftsbezug *ex post* würde sie kaum benötigt. Sie soll in dieser Situation verantwortliches Handeln ermöglichen und unterstützen, indem sie bereits während der technischen Entwicklung mögliche nicht intendierte und unerwünschte Folgen erforscht, bewertet und reflektiert. Diese für jede neue Technik schwierige Situation verschärft sich angesichts der KI.

Denn *erstens* ist für KI-Systeme wie etwa autonome Roboter ihre Fähigkeit zu lernen charakteristisch. Zum ersten Mal in der Technikgeschichte wurden lernfähige Objekte erzeugt und in die Welt entlassen, z.B. als Roboter, die sich in ihrer Tätigkeit weiterentwickeln können. Technik gewinnt da-

durch, wenngleich bislang nur in rudimentärer Form, eine bislang Menschen vorbehaltene Fähigkeit: das systematische Lernen aus Erfahrungen, möglicherweise eine menscheitsgeschichtliche Zäsur. Da die Anlässe, aus denen im laufenden Betrieb etwa eines selbstfahrenden Autos gelernt wird, nicht vorhersehbar sind, können sich die Eigenschaften des Systems durch dieses Lernen in einer nicht vorherzusehenden Weise ändern. Diese Eigenschaft setzt der Vorhersage des Erreichens der Zwecke (Zukunftsbezug *ex ante*) und der Einschätzung möglicher nicht beabsichtigter Folgen (Zukunftsbezug *ex post*) engere Grenzen als bei traditioneller, sich nicht selbst verändernder Technik.

Zweitens verändern sich durch autonom werdende KI-gestützte Technik Subjekt-Objekt-Verhältnisse zwischen Mensch und Technik. Im traditionellen Bild gestalten und nutzen menschliche Subjekte die technischen Objekte. Die Verteilung von Subjekt- und Objektrolle ist eindeutig. Mit KI kommt es jedoch zu anderen Konstellationen. KI-Systeme sind nicht mehr nur eine Menge von einzelnen Objekten wie Instrumente, Apparate und Maschinen, die zum Subjekt Mensch in einem klaren Objektverhältnis stehen. Stattdessen werden Men-

schen in vernetzten Systemen zu Bestandteilen von Konstellationen, in denen sie gelegentlich die Subjekt-, teils aber auch die Objektrolle innehaben. Wenn Menschen beispielsweise in autonomen Fahrzeugen mitfahren, nehmen sie die Rolle von Frachtgut ein, das sich selbst aufgibt. Der Mensch bestimmt zwar das Ziel der Fahrt, hat aber auf die Fahrt selbst keinen Einfluss mehr, sondern wird zum Objekt der Entscheidungen des Bordcomputers. Ähnliches geschieht, wenn Entscheidungskompetenz an digitale Systeme abgetreten wird wie etwa in den so genannten *automated decision-making* (ADM)-Systemen.



Wenn Menschen in autonomen Fahrzeugen mitfahren, nehmen sie die Rolle von Frachtgut ein, das sich selbst aufgibt

Die KI ist daher nicht einfach ein weiteres Technikfeld als neuer Untersuchungsgegenstand der TA wie andere auch, sondern etwas Neuartiges: neuartig in Bezug auf das Verhalten über die Zeit wie auch in Bezug auf die veränderte Rolle von Menschen mit anschließenden anthropologischen Fragen.

Klimawandel beispielsweise ist nicht eine Folge davon, dass es Autos und fossil betriebene Kraftwerke gibt, sondern rührt daher, dass diese Autos und Kraftwerke entsprechend dem gesellschaftlichen Bedarf, also gemäß Lebensstilen und Konsummustern massenhaft eingesetzt werden. Der Zukunftsbezug *ex post* (s.o.) umfasst das Handeln und Verhalten der Menschen. Eine Vorhersage von Technikfolgen müsste also die Vorhersage des menschlichen Handelns auf der Nutzerseite mit einbeziehen. Der Optimismus der 1970er Jahre, dass mit genügend Forschungsaufwand und besseren

Methoden derartige Vorhersagen präzise möglich werden könnten, ist jedoch rasch verschwunden. Und wenn auch aktuell im Feld von KI und Big Data erneut Hoffnungen auf weitgehende Vorhersagbarkeit menschlichen Verhaltens geweckt werden, dürften diese ebenso an der Kreativität und Unvorhersehbarkeit vieler menschlicher und gesellschaftlicher Entwicklungen scheitern, die hinter dem Zukunftsbezug *ex post* von Technik stehen.

Die mangelnde Vorhersagbarkeit ist nichts weiter als die Kehrseite der Offenheit der Zukunft und ihrer Gestaltbarkeit. Klare Vorhersagen sind nur in determinierten Systemen möglich. In der Tat wurde in der frühen TA gemeinsam mit der Ausrichtung auf Prognosen ein *Technikdeterminismus* angenommen, nach dem die Technikentwicklung eigenen Gesetzen folge und gar nicht oder nur begrenzt gestaltet werden könne. Mensch und Gesellschaft müssten sich an Technikfolgen anpassen, wofür ihre möglichst genaue Vorhersage wichtig und in diesem Ansatz auch möglich sei. Dieser Determinismus ist jedoch weder begründbar noch wünschenswert und blendet Gestaltungspotentiale aus. Vor allem haben Versuche präziser Vorhersagen angesichts überraschender gesellschaftlicher Entwicklungen oder neuer politischer Konstellationen zumeist nicht funktioniert. Eigenschaften der KI wie die Möglichkeit des Lernens (s.o.) dürften die Vorhersagbarkeit weiter einschränken. Die ab den 1980er Jahren erfolgte Hinwendung der TA zur Gestaltungsorientierung und ihre parallele Abkehr vom Versuch der Vorhersage dürfte daher in Zeiten der KI noch bestärkt werden.

Auch in der wenigstens teilweisen Gestaltungs Offenheit der Zukunft ist das konsequentialistische Schema (Abb. 1) anwendbar. Allerdings muss es so modifiziert werden, dass unterschiedliche zukünftige Entwicklungen und Technikfolgen gleichermaßen in den Blick genommen werden können. Hierzu dienen *Szenarien*. Mit alterna-

Vorhersagen und Szenarien

Die TA wurde in ihren Anfängen in den 1970er Jahren mit hohen Erwartungen an die mehr oder weniger genaue Vorhersagbarkeit von Technikfolgen konfrontiert. Allerdings sind Technikfolgen nicht einfach Folgen der Technik, die sich mit Mitteln der Ingenieurwissenschaften erforschen und dann exakt vorhersagen lassen. Denn sie ergeben sich nicht nur aus den technischen Objekten und Systemen, sondern bilden sich im Zusammenwirken der technischen Eigenschaften mit ihrer Nutzung in Innovationen, Wertschöpfungsketten, Geschäftsmodellen, am Arbeitsplatz und im Freizeitverhalten etc. Der



tiven Szenarien kann der Raum möglicher oder plausibler zukünftiger Entwicklungen strukturiert werden. Im Gegensatz zu Trendextrapolationen (bei denen gegenwärtige Trends in die Zukunft fortgeschrieben werden) oder anderen prognostischen Verfahren wird nicht der Eindruck *einer* Zukunft erweckt, die man vorhersehen könne, sondern es wird – abhängig von verschiedenen angenommenen Randbedingungen – mit mehreren Zukünften gearbeitet. Szenarien machen nur Sinn im Plural angesichts der Offenheit der Zukunft. Dadurch ist ihnen der Gestaltungs- und Entscheidungsbezug inhärent: Je nach heutigen oder zukünftigen Entscheidungen wird man auf andere Szenarien und damit in andere Zukünfte geführt. Szenarien lehren das Denken in (zumindest teilweise) offenen Gestaltungsräumen. In der Technikfolgenabschätzung müssen sie relevant (bedeutsam), kohärent, plausibel, anschlussfähig und transparent

sein, um ihre orientierende Funktion zu erfüllen (Grunwald 2022).

Szenarien haben sich als Ansatz zur Konzeptualisierung einer gestaltungs-offenen Zukunft zum hauptsächlichen Instrument der TA im Umgang mit Zukunft entwickelt. Auch in anderen strategischen Kontexten haben sie ganz erhebliche Bedeutung gewonnen, so z.B. in der Nachhaltigkeitsforschung und im Unternehmensbereich. Die TA hat dadurch gelernt, mit der Offenheit der Zukunft und der dadurch implizierten Nicht-Vorhersagbarkeit konstruktiv umzugehen und in dieser Offenheit Orientierungsmöglichkeiten zu entwickeln. Dies erfolgt z.B. durch die Bestimmung von so genannten robusten Handlungsstrategien. Hierbei handelt es sich um Maßnahmen bzw. Maßnahmenbündel, die in jedem der betrachteten Szenarien erwünschte, wenn auch im Detail oft unterschiedliche Effekte erwarten lassen.

Zukunftserzählungen als Aussagen über die Gegenwart

Freilich sind auch Szenarien an Voraussetzungen gebunden, die nicht immer erfüllt sind. Die großen Fragen der KI, ob z.B. durch Lernen und Emergenz so etwas wie maschinelles Bewusstsein entstehen kann, ob KI-Systeme fähig werden können, selbst Intentionen zu entwickeln, oder auch nur, ob die weitgehende Übernahme von kaum oder gar nicht transparenten KI-Systemen zu nicht mehr umkehrbaren Abhängigkeiten von diesen Systemen führt, dürften von diesem Typ sein. Sie sind aber nicht die ersten Herausforderungen der TA, bei denen der Ansatz der Szenarien nicht mehr weiterhilft.

Etwa ab dem Jahr 2000 wurde die TA mit den heute so genannten NEST (*new and emerging sciences and technologies*) konfrontiert, beginnend mit der Nanotechnologie (Nordmann et al. 2006). In den frühen Jahren der Zukunftsdebatten zu diesem damals visionären Technikfeld schwankten die

Zukunftsansagen zwischen Paradieserwartungen und Befürchtungen eines Untergangs der Menschheit, und zwar ohne Möglichkeit, diese nach Plausibilität oder Wahrscheinlichkeit zu unterscheiden. Damit waren nicht nur Vorhersagen, sondern auch Szenarien illusorisch. Der Raum des zukünftig Möglichen war so groß und unbestimmt, dass die nachvollziehbare Eingrenzung auf einige plausible Szenarien und die Aussonderung anderer als unplausibel nicht durchführbar war. Parallel wurde die anthropologische Frage nach dem Menschen aufgeworfen, vor allem nach seiner Zukunft. Paradigmatisch dafür standen früh der Essay „Why the future doesn't need us“ (Joy 2000) und Habermas' Überlegungen zur Zukunft der Natur des Menschen (2001). Damit wurden wohl zum ersten Mal Orientierungserwartungen an die TA mit anthropologischen Fragen gekoppelt, die seit-

dem die TA und die benachbarten Felder der Ethik begleiten.



Eine Koppelung der TA mit anthropologischen Fragen gibt es erst seit Beginn der 2000er Jahre

Die TA hat auf diese neue Konstellation mit einer *hermeneutischen* Wende reagiert (Grunwald/Hubig 2018). Ihr zentrales Element ist eine umgekehrte Perspektive auf Zukunftsaussagen. Üblicherweise werden diese als Vorhersagen über zukünftige Ereignisse oder Entwicklungen betrachtet. Dies ist jedoch nur sinnvoll, wenn die Zukunftsaussagen eine gewisse Evidenz haben, wenn also aus Erkenntnissen, Wissen und Daten von Gegenwart und Vergangenheit belastbare Extrapolationen für die Zukunft abgeleitet werden können. Dieses übliche Modell versagt jedoch immer dann, wenn in visionären Feldern Erwartungen oder auch Befürchtungen in die Welt gesetzt werden, die nicht durch Daten und Regelmäßigkeiten der Vergangenheit abgedeckt sind. Stattdessen werden Erwartungen in die Welt gesetzt, dass durch wissenschaftlich-technische Innovation etwas qualitativ Neues in die Welt kommt, das zu qualitativen Änderungen oder sogar disruptiven Brüchen führen kann. In genau diesem Sinn klaffen in den typischen NEST-Debatten wie zur Nanotechnologie vor etwa zwanzig Jahren, zur Synthetischen Biologie vor etwa zehn Jahren und gegenwärtig zu Digitalisierung, Robotik und KI die geäußerten Erwartungen wie Befürchtungen so weit auseinander, dass es keinen Sinn macht, sie als Aussagen über plausible oder wahrscheinliche Entwicklungen in der Zukunft zu interpretieren. Sie scheinen vielmehr spekulativ und beliebig zu sein. Ob und wie aus derartigen Zukunftsbildern Orientierung gewonnen werden kann, erscheint auf den ersten Blick fraglich.

Einen wichtigen Hinweis gibt eine sehr grundsätzliche Beobachtung. Be-



reits von Augustinus wurde im Jahre 397 formuliert (Bekenntnisse XI, 20):

Das ist nun wohl klar und einleuchtend, daß weder das Zukünftige noch das Vergangene ist. Eigentlich kann man gar nicht sagen: Es gibt drei Zeiten, die Vergangenheit, Gegenwart und Zukunft ... In unserem Geiste sind sie wohl in dieser Dreizahl vorhanden, anderswo aber nehme ich sie nicht wahr. Gegenwärtig ist hinsichtlich des Vergangenen die Erinnerung, gegenwärtig hinsichtlich der Gegenwart die Anschauung und gegenwärtig hinsichtlich der Zukunft die Erwartung.

Danach existiert nur die Gegenwart als Zeit, in der Menschen Erfahrungen machen, lernen, entscheiden und handeln können. Vergangenheit und Zukunft sind andere Formen, sie existieren nur in der *Vergegenwärtigung* durch Menschen. In Bezug auf die Vergangenheit besteht diese Vergegenwärtigung in der Erinnerung, in Bezug auf die Zukunft als Erwartung. Während die Erinnerung sich auf Relikte aus der Vergangenheit stützen kann wie Texte, Bilder oder alte Mauern, gibt es nichts Entsprechendes aus der Zukunft. In der Sprache der Digitalisierung: Es sind keine Daten aus der Zukunft verfügbar. Zukunftsaussagen stellen daher sämtlich aus einer Gegenwart heraus entstandene *Erwartungen* dar, die nicht empirisch geprüft werden können. Auch wissenschaftliche Zukunftsaussagen können nur mit *gegenwärtigem* Wissen und *gegenwärtigen* Einschätzungen, Werten, Vorurteilen, Annahmen, Interessen und Wünschen erstellt werden. Diese ‚Zutaten‘ führen auf Bilder von möglichen Zukünften, unterschiedliche Zutaten führen auf unterschiedliche Zukunftsbilder. In den Zukunftsaussagen steckt daher viel Gegenwart – oder konsequenter formuliert: Zukunftsaussagen führen bestimmte Einschätzungen aus der Gegenwart zusammen und spitzen sie auf unterschiedliche Erzählungen zu.



Zukunftsaussagen führen bestimmte Einschätzungen aus der Gegenwart zusammen und spitzen sie auf unterschiedliche Erzählungen zu

Daher kann aus Zukunftserzählungen etwas über die jeweilige Gegenwart gelernt werden. Dies ist die hermeneutisch umgekehrte Erkenntnisrichtung in Bezug auf visionäre Zukunftsaussagen: Statt ihnen etwas über mögliche Entwicklungen der nächsten Jahrzehnte zu entnehmen, werden sie daraufhin befragt, was sie über die Gegenwart aussagen. Der hermeneutische Blick ist komplementär zum konsequentialistischen Schema (Abb. 1). Während letzteres mit der Evidenz für eine bestimmte *Erwartbarkeit* des Eintreffens der betrachteten Zukunftsaussagen steht und fällt, spielt die Erwartbarkeit in der hermeneutischen Perspektive keine Rolle.

Zukunftserzählungen generell wie auch die hier speziell interessierenden visionären Aussagen, wie sie Digitalisierung und KI prägen (Grunwald 2019), werden nicht *entdeckt* wie eine Erkenntnis, sondern *gemacht*. Sie haben Autor:innen und werden erzeugt und kommuniziert von individuellen Menschen, Gruppen, Institutionen und Organisationen. All dies erfolgt in einer bestimmten Gegenwart. Gegenwärtige Wissensbestände, aber auch Zeit- und Problemdiagnosen, Werte, Weltanschauungen, Wunschbilder für zukünftige Welten, gegenwärtige Interessen und Präferenzen, Annahmen über unbeeinflussbare Randbedingungen, Zeitgeistvorstellungen, konkrete Interessen sowie sicher vieles mehr wird dabei verarbeitet. In hermeneutischer Perspektive ist interessant, was diese Zukünfte über aktuelle gesellschaftliche Befindlichkeiten, unterschwellige Sorgen, implizite Hoffnungen und Befürchtungen, verborgene geistesgeschichtliche Traditionen oder kulturelle Zusammenhänge, aber ggf. auch über die Werte, Interessen und Weltanschauungen ihrer Autoren

aussagen (Grunwald/Hubig 2018). In der hermeneutischen Analyse der visionären Erzählungen zur KI geht es darum, ihre Ursachen und Quellen, ihre Hintergründe und Motivationen aufzudecken, sie zu hinterfragen, ihrer Bedeutung nachzuspüren und mögliche Alternativen zu explorieren. Spezifisch für die Technikfolgenabschätzung ist die Frage, warum gerade bestimmte Visionen zur KI verbreitet werden und andere nicht, einschließlich der Frage nach möglichen Alternativen zu den dominanten Erzählungen. Erst im Feld möglicher alternativer Wege (Grunwald 2022) in eine Zukunft mit immer mehr KI-Systemen kann bewusste Gestaltung ansetzen. Und darum muss es letztlich gehen: Die Zukunftserzählungen zur KI bergen, so die Vermutung, Wissen und Einschätzungen, deren Aufdeckung und Erforschung einer transparenteren demokratischen Debatte und der bewussten Gestaltung der KI-gestützten weiteren Digitalisierung zugutekommen können.

Die hermeneutische Perspektive nimmt die Verfahrensweisen und die verwendeten ‚Zutaten‘ (s.o.) in den Blick, mit denen die Zukunftserzählungen von ihren Autor:innen erzeugt werden. Hierzu gehören im Kontext der KI zum Beispiel die Vorstellungen über zukünftige Mensch/Roboter-Kooperationen, etwa in der industriellen Produktion (Industrie 4.0). Hier wird immer wieder davon gesprochen, dass Menschen und Roboter „auf Augenhöhe“ kooperieren sollen. Im technischen Design werden Entwicklungen in die Wege geleitet, die zukünftige Rollen von Menschen betreffen. Dies sind keine Vorhersagen, dass es zu dieser „Augenhöhe“ kommen wird, sondern Planungen im Zukunftsbezug *ex post*. Hermeneutische Analyse hat hier zu hinterfragen, was die „Augenhöhe“ konkret bedeutet, welche anthropologischen Fragen betroffen sind und wie die neuen Kooperationsformen ethisch zu beurteilen sind. Auch wenn Roboter als künstliche Gefährten zur Betreuung einsamer Menschen eingesetzt werden

sollen, stellen sich ähnliche Fragen, die unabhängig vom konsequentialistischen Schema bearbeitet und beantwortet werden können.

Die hermeneutische Perspektive betrifft auch datenbasierte, quantitative und mathematisch strukturierte Modelle, die zur Generierung von Szenarien oder Simulationen der Zukunft eingesetzt werden. Diese bestehen keineswegs nur aus vermeintlich objektiven Erkenntnissen und Daten, sondern enthalten Erzählungen, die im Akt des Modellierens handlungsleitend sind und die späteren Ergebnisse der Szenarien oder Simulationen beeinflussen können. Um aus einzelnen Bestandteilen ein Modell zu bauen, bedarf es verbindender Erzählungen, denn Daten und empirisches Wissen fügen sich nicht wie von selbst zu einem Modell. Diese Erzählungen können sich z. B. in der Wahl eines bestimmten anthropologischen Ansatzes verbergen wie der wirtschaftswissenschaftlichen Erzählung vom *homo oeconomicus* oder dem nachhaltigkeitsorientierten *homo sustinens* (Siebenhüner 2000). Je nachdem, welches Menschenbild dem Modell zugrunde gelegt wird, dürften Szenarien und Simulationen auf andere

Zukünfte führen. Andere Erzählungen können volkswirtschaftliche Einschätzungen zur zukünftigen Ausprägung des Wachstumsprinzips der kapitalistischen Wirtschaft oder auch dominante Einschätzungen zu technischen Sachfragen sein, etwa was die grundsätzliche Überlegenheit einer bestimmten Technologie in der Ermöglichung nachhaltiger Mobilität sein.

In der Beantwortung der hermeneutischen Fragen muss das interdisziplinäre Spektrum der TA erweitert werden. Sprachwissenschaften, hermeneutische



Das interdisziplinäre Spektrum der TA muss um weitere Disziplinen wie Sprach- oder Kulturwissenschaften erweitert werden

Ansätze in Philosophie und Geisteswissenschaften, Kulturwissenschaften und auch die Hermeneutik in der Kunst – insofern z. B. Technikzukünfte mit künstlerischen Mitteln erzeugt und kommuniziert werden – müssen an Bord der TA kommen. Vom prognoseorientierten Ansatz ihrer frühen Zeit hat sie sich damit sehr weit entfernt.

Technikfolgenabschätzung trifft Anthropologie

Die Fähigkeit des Lernens und die Möglichkeit, KI-Systemen Entscheidungshoheit zu überlassen, führen auf neue Fragen nach dem Verhältnis von Mensch und Technik. Einige haben sich bereits oben im Text angedeutet, wenn nicht gar aufgedrängt. Weil die KI mit Attributen wie lernend, intelligent, autonom etc. verbunden wird, stellen sich Fragen nach ihrem Verhältnis zum Menschen. Diese Fragen beziehen sich natürlich auf die Zukunft, etwa wenn die Entwicklung des Arbeitsmarkts angesichts der immer größeren KI- und Roboter-Konkurrenz für menschliche Arbeit in den Blick genommen wird. Diese Zukunft entsteht jedoch nicht von selbst, sondern hängt

von den Entscheidungen ab, die auf dem Weg dorthin getroffen werden. In diesen Entscheidungen spielen Zukunftsvorstellungen eine Rolle, die hermeneutisch analysiert und anthropologisch hinterfragt werden können, ohne in einen Vorhersagemodus zu verfallen. Hier ist ein Ansatzpunkt für demokratische und ethisch reflektierte Gestaltung, denn die Art und Weise, wie heute zukünftige Mensch/Technik-Verhältnisse in den Zukunftserzählungen konzipiert werden, etwa unter dem erwähnten Stichwort der Augenhöhe, beeinflusst, wie sie sich wahrscheinlich entwickeln werden.

Wissenschaft und Technik werfen bekanntlich vielfältige ethische

Fragen auf (Grunwald 2022). Gerade zu digitalen Technologien und KI erschöpfen sich die Herausforderungen darin jedoch nicht. Denn hinter den ethischen Fragen steht die Frage nach dem Menschen, seinem Wesen, seiner Rolle auf dem Planeten und im Kosmos, seiner Zukunft. Bereits Immanuel Kant ließ seine drei berühmten Fragen „Was kann ich wissen? Was soll ich tun? Worauf darf ich hoffen?“ in eine vierte münden: „Was ist der Mensch?“. Sie lässt sich im Spiegel der KI zuspitzen (Grunwald 2021): Wer bist du, Mensch, angesichts von Robotik, Künstlicher Intelligenz und technischen Unsterblichkeitsphantasien? Wer willst Du sein in einer zusehends hoch technisierten Welt? Wie willst Du Freiheit, Verantwortung und Kreativität leben? Was wird aus Dir in einem sich weiter beschleunigenden technischen Fortschritt?

Die Antworten sind vielfältig. Auf der einen Seite stehen Allmachtsfantasien vom Menschen als Weltenherrscher bis hin zur Hybris im Blick auf die Erfolge moderner Naturwissenschaft und Technik. Auf der anderen Seite steht der unterlegene Mensch angesichts immer leistungsfähiger werdender digitaler und zusehends autonomer Technik, denen gegenüber Menschen zusehends unvollkommen sind, so dass ihnen in der Zukunft nur ein Dasein als ihren eigenen Geschöpfen unterlegene Spezies bleibt (Grunwald 2019). Zumindest zwei Stoßrichtungen für die TA deuten sich hier an: (1) die TA kann aktuelle wissenschaftlich-technisch motivierte Zukunftsaussagen und ihre weltanschaulichen Überhöhungen in Bezug auf anthropologische Diagnosen oder Setzungen analysieren. Beispielsweise kann sie Narrative der Digitalvisionäre kritisch untersuchen, ob sie Menschen als datenverarbeitende Maschinen deuten und daran die Gestaltung der Schnittstellen von KI und Mensch ausrichten. Auf diese Weise können die Narrative dem Dialog und damit der Kritik zugänglich gemacht werden; (2) die



TA kann ebenfalls moderne Erzählungen vom Menschen untersuchen, beispielsweise die des *homo deus* (Harari 2018) oder des *homo sustinens* (Siebenhüner 2000). Auch hier würde es um die Aufdeckung inhärenter anthropologischer Prämissen gehen.

Auf diese Weise hat eine interessante Wandlung der TA stattgefunden, jedenfalls im Feld der KI. War die TA ursprünglich entwickelt worden, um den technischen Fortschritt nicht dem planenden Denken in Zukunftsbezügen *ex ante* allein zu überlassen, sondern den Zukunftsbezug *ex post* möglichst frühzeitig und begleitend zur Technikentwicklung immer mitlaufen zu lassen, ist die hermeneutische Perspektive in gewisser Weise eine Rückbesinnung auf den Zukunftsbezug *ex ante* – allerdings in kritischer Absicht als Aufdeckung und Hinterfragung der z. B. anthropologischen Prämissen in den Projektionen zukünftiger Mensch/Technik-Verhältnisse.

Fazit

Die TA war zu ihrer Entstehungszeit in den 1970er Jahren mit der Erwartung

der Vorhersage von Technikfolgen konfrontiert. Seit Jahrzehnten jedoch ist diese letztlich deterministische Perspektive zugunsten eines gestaltungsorientierten Ansatzes überwunden.

Damit verbunden ist die Anerkennung der Offenheit der Zukunft. Erforschung und Bewertung von Technikfolgen in der TA sind auf Politik- und Gesellschaftsberatung zur Gestaltung des technischen Fortschritts ausgerichtet. Das Arbeiten mit Szenarien zukünftiger Entwicklung entspricht diesem Ansatz.

Digitalisierung und Künstliche Intelligenz verschärfen die Anforderungen, überhaupt noch belastbares Zukunftswissen erzeugen zu können. Insbesondere die Fähigkeit der KI zu lernen, stellt eine Herausforderung dar. Im Fall technikvisionärer Erzählungen funktioniert das konsequentialistische Schema der TA nicht mehr.

Es verbleibt, die kursierenden Zukunftserzählungen zur KI als Texte zu verstehen, die von Autor:innen erstellt und von Akteur:innen kommuniziert und teils kontrovers diskutiert werden. Hermeneutisch geht es darum, die Hintergründe der Entstehung und Kommunikation dieser Erzählungen zu deuten, um daraus für die Gegenwart zu lernen.

ZUM AUTOR

Armin Grunwald, geb. 1960, Leiter des Instituts für Technikfolgenabschätzung und Systemanalyse (ITAS) und Professor für Technikethik und Technikphilosophie am Karlsruher Institut für Technologie (KIT). Arbeitsgebiete: Theorie und Methodik der Technikfolgenabschätzung, Digitalisierung, nachhaltige Entwicklung. Aktuelle Bücher: *Der unterlegene Mensch: Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern*, München 2019; *Wer bist du, Mensch? Transformationen menschlicher Selbstverständnisse im wissenschaftlich-technischen Fortschritt*, Freiburg 2021; *Technikfolgenabschätzung. Eine Einführung*, Baden-Baden 2022³.

Digitalisierung und insbesondere KI werfen jenseits von Technikfolgenabschätzung und Ethik auch anthropologische Fragen auf. In der hermeneutischen Ausrichtung der TA gilt es, diese zu untersuchen und zu reflektieren, um proaktive Technikgestaltung zuträglicher Mensch/Technik-Verhältnisse für die Zukunft zu fördern.

LITERATUR

- Grunwald, Armin (2019): *Der unterlegene Mensch: Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern*, München.
- Grunwald, Armin (Hg.) (2021): *Wer bist du, Mensch? Transformationen menschlicher Selbstverständnisse im wissenschaftlich-technischen Fortschritt*, Freiburg.
- Grunwald, Armin (2022³): *Technikfolgenabschätzung. Eine Einführung*, Baden-Baden.
- Grunwald, Armin/Hubig, Christoph (2018): *Technikhermeneutik: ein kritischer Austausch zwischen Armin Grunwald und Christoph Hubig*. *Jahrbuch Technikphilosophie 2018*, Baden-Baden, 321–352.
- Habermas, Jürgen (2001): *Die Zukunft der menschlichen Natur*, Frankfurt.
- Harari, Yuval (2018²): *Homo Deus. Eine Geschichte von morgen*, München.
- Jonas, Hans (1979): *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*, Frankfurt.
- Joy, Bill (2000): *Why the Future Does not Need Us*. *Wired Magazine*, April 2000, 238–263.
- Nordmann, Alfred/Schummer, Joachim/Schwarz, Astrid (Hg.) (2006): *Nanotechnologien im Kontext*, St. Augustin.
- Siebenhüner, Bernd (2000): *Homo sustinens – towards a new conception of humans for the science of sustainability*. *Ecological Economics* 32, 15–25.

Erlösung von der Last der Entscheidung?



Die Algorithmisierung der Gesellschaft in theologisch-ethischer Reflexion

Der Text setzt sich mit der These eines möglicherweise zu erwartenden unaufhaltsamen Transfers der Entscheidungsmacht vom Menschen an die Maschine auseinander, wie sie sich in derzeit kursierenden Zukunftserzählungen zur Algorithmisierung der Gesellschaft findet. Vor dem Hintergrund von Überlegungen, wie solche Zukunftserzählungen für das ethische Nachdenken hermeneutisch erschlossen werden können, reflektiert der Text, welche Konsequenzen die Abgabe der Entscheidungsmacht an künstliche Intelligenzen mit sich brächte. Die These, dass Algorithmische Entscheidungssysteme einen Fortschritt hin zum Besseren darstellen, da sie den überforderten Menschen entlasten und damit auch sein Leid mindern, wird aus einer theologisch-ethischen Perspektive problematisiert.

In seinem Essay *Die Zukunft des Menschen – Das Versprechen der Künstlichen Intelligenz* beschwört der Medienwissenschaftler Roberto Simanowski das Bild einer Welt, in der Algorithmen das Steuer, Künstliche Intelligenz die Macht übernommen haben. Der Autor schaut im Jahr 2020 aus einer imaginierten Zukunft zurück auf das gerade begonnene Jahrzehnt und malt aus, wie tiefgreifend Algorithmen das Leben des Menschen verändert haben werden. Am Ende dieses Jahrzehnts, so prognostiziert es der Autor, werden wir den automatisierten Entscheidungssystemen nicht nur die Suche nach Lebenspartner:innen komplett übertragen haben, sondern die maschinellen Entscheidungssysteme werden uns ein passanter auch zu besseren und gesünderen Menschen gemacht haben. Beispielsweise werden wir uns dank entsprechender Algorithmen durch *Nudging* in Form technischer Impulskontrolle bzw. -manipulation von Lastern und Süchten, wie etwa dem Rauchen, befreit haben. Nicht zuletzt werden wir auch unseren ökologischen Fußabdruck dank der Algorithmen so verkleinert haben, dass


die drohende Klimakatastrophe abgewendet werden kann. Dazu werden Algorithmen deshalb in der Lage sein, weil sie uns dank des zunehmenden Einsatzes von Smartspeakern (wie zum Beispiel *Alexa* oder *Siri*) besser kennen werden als wir uns selbst. Simanowski folgend befinden wir uns inmitten eines unaufhaltsamen *Souveränitätstransfers*, einer schleichenden Übergabe der Entscheidungsmacht vom Menschen an die Maschine. Unumkehrbar sei diese Entwicklung deshalb, weil sich der omnipräsente Einsatz von Algorithmen bereits jetzt als hocheffizient erweise. Einer solchen „Intelligenz, die effektiver ist als man selbst, überlässt man gern das Denken und Entscheiden: beim Autofahren wie beim Klimaretten.“ (Simanowski, 2020)

Simanowskis Vision einer algorithmisierten Welt steht stellvertretend für viele derzeit kursierende Vorstellungen einer sich technologisch immer rasanter entwickelnden Gesellschaft. Mal euphorisch, mal apokalyptisch intoniert wird in Aussicht gestellt, dass unterschiedliche Formen Künstlicher Intelligenzen in den kommenden Jahr-



Katharina Klöcker

zehnten den Menschen in fast allen Bereichen entlasten, überflügeln oder übertrumpfen und letztlich vielleicht sogar überflüssig machen werden. Die Vorstellung der Substitution des Menschen durch Maschinen ist dabei nicht neu; sie ist ein Menschheitsthema, das bereits in der Antike, vor allem aber in der Zeit der beginnenden Industrialisierung thematisiert wurde (vgl. Müller/

 Die Vorstellung der Substitution des Menschen durch Maschinen ist ein Menschheitsthema, das bereits in der Antike thematisiert wurde

Liggieri 2019, 3–14). Mit Martina Heßler lassen sich drei Dimensionen dieser Verhältnisbestimmung unterscheiden: „die Entlastung und Befreiung, die Bedrohung und das Überflüssigwerden der Menschen sowie die Vernichtung der Menschheit und ihre Ersetzung als Gattungswesen durch Maschinen“ (Heßler 2020, 163). Alle drei Dimensionen lassen sich auch im Zeitalter der Digitalisierung identifizieren.



Immer vernehmbarer werden derzeit vor allem apokalyptische Prognosen über Technologien, die den Menschen bedrohen oder gar vernichten werden. So warnen mittlerweile sogar so genannte „Technologievisionäre wie Elon Musk von Tesla Motors, Bill Gates von Microsoft und Apple-Mitbegründer Steve Wozniak [...] vor Superintelligenz als einer ernststen Gefahr für die Menschheit, vielleicht bedrohlicher als Atombomben“¹. Nicht selten werden dabei aus beobachteten Entwicklungen

Prognosen abgeleitet, ohne dass deren Vorläufigkeit und der ihnen inhärente Unsicherheitsvorbehalt ausreichend kommuniziert würden. So folgert beispielsweise Simanowski aus der Tatsache, dass sich viele Autofahrer:innen blindlings auf die Anweisungen des Navigationsgeräts verließen, dass es zulässig sei, dieses Vertrauen in den Algorithmus auch auf andere Lebensbereiche wie etwa die Partner:innensuche zu übertragen.

digitalen Diktatur wiederfinden werden, darüber gehen die Prognosen weit auseinander. Adäquat verstanden und für die ethische Reflexion fruchtbar gemacht werden können solche und andere Zukunftsvisionen vor allem dann, wenn man sie als „gegenwärtige Konstruktionen“ (Grunwald, 2009, 26) begreift, die Austragungsorte gesellschaftlicher Konflikte darstellen. Weil sich in ihnen die eine Gesellschaft durchziehenden Konfliktlinien widerspiegeln, sind diese unterschiedlichen und zum Teil einander diametral entgegen gesetzten Imaginationen technologischer Zukünfte genauer in den Blick zu nehmen (vgl. Grunwald 2007, 57). Zukunftsprognosen und -visionen werden damit für jede ethische Reflexion im Zeitalter immenser technologischer Transformationen zu herausfordernden Wahrnehmungs- und Denk-Orten. In ihnen werden normative Annahmen und Wertvorstellungen sichtbar, die Rückschlüsse auf Ängste und Hoffnungen und das herrschende Selbstverständnis des Menschen in der Gegenwart erlauben.

Sie nicht als solche wahrzunehmen, würde bedeuten, die epistemische Kraft dieser Vorstellungen hinsichtlich ihrer Orientierungsfunktion für die Gegenwart zu verkennen (vgl. Horn 2014, 25 ff.). Ethik als Orientierungswissenschaft kommt deshalb neben der notwendigen Reflexion von Mitteln, Zielen und Folgen neuer Entwicklungen die Aufgabe zu, Vorstellungen zukünftiger technologischer Entwicklungen nicht nur „als antizipierende Rede über Zukünftiges mit anschließender folgenethischer Reflexion“ zu deuten, sondern auch als „als *Ausdrucksformen der Gegenwart*“ (Grunwald 2020, 79) ernst zu nehmen und begreifbar zu machen. Für die ethische Reflexion bedeutet das: Zukunftsvisionen sind nicht nur als solche Ausdrucksformen zu verstehen, sondern auch als auslösende Momente für moralische Bewertungen,


Aufgaben einer Ethik der Digitalisierung

Welche Aufgaben kommen nun aber vor dem Hintergrund dieser Entwicklungen auf eine Ethik im Zeitalter der Digitalisierung zu? Idealerweise wird von einer solchen Ethik wohl erwartet, dass sie eine Art hocheffizienten und untrüglichen Sinnesapparat zur Verfügung stellt, mit dessen Hilfe man sich im immer unübersichtlicher werdenden Gelände der zunehmenden Errungenschaften Künstlicher Intelligenz zu orientieren vermag. Ethik sollte, dieser Sichtweise folgend, zwar nicht ka-

jeweiligen technischen Entwicklung intensiver Reflexion. Doch, und dies ist Ausgangspunkt der folgenden Überlegungen, so wichtig solche kontextbezogenen Abwägungen und Folgenabschätzungen in normativer Hinsicht zweifellos sind, eine Ethik der Digitalisierung kann nicht allein auf ihre beratende und handlungsorientierende Funktion reduziert werden. Vielmehr muss sie sich darüber hinaus in sehr grundlegender Weise mit einer großen Anzahl von metaethischen Fragen befassen. Ein Aspekt, der erst seit Kurzem stärker berücksichtigt wird und hier exemplarisch mit Blick auf Simanowskis Zukunftserzählung besondere Berücksichtigung findet, ist die Frage, wie die Ethik überhaupt – in hermeneutischer Perspektive – mit Zukunftsvisionen, die gerade in der Digitalisierungsdebatte eine so bedeutende Rolle spielen, umgehen sollte.

Zukunftsvisionen als Konstruktionen

Ob die digitalisierte Gesellschaft paradisiische oder doch eher apokalyptische Züge annehmen wird, ob wir uns dank des Algorithmen-Einsatzes entspannt zurücklehnen oder uns in einer

 Ethik sollte den Blick dafür schulen, wo das Befreiungspotential von Maschinen in Bedrohung umschlägt

tegorische Handlungsurteile fällen, aber doch den Blick dafür schulen, wo das Befreiungspotential von Maschinen in Bedrohung umschlägt, wo aus einer Entlastung des Menschen seine Entmachtung zu werden droht und wo der technologische Fortschritt nicht lebenserhaltende oder -verbessernde, sondern destruktive, inhumane, vielleicht sogar die gesamte Menschheit bedrohende Züge anzunehmen beginnt.

Um diesen Erwartungen normativer Orientierung gerecht werden zu können, unterzieht eine Ethik der Digitalisierung neben den Mitteln und Zielen vor allem auch die Folgen der

¹ www.spektrum.de/news/wie-algorithmen-und-big-data-unsere-zukunft-bestimmen/1375933

die diese technologischen Entwicklungen begleiten. Damit tritt die ethische Relevanz dieser hermeneutischen Perspektivierung deutlich hervor, denn zu konstatieren ist: „In modernen Gesellschaften werden Entscheidungen weniger mit Bezug auf Traditionen, sondern viel stärker durch Zukunftsüberlegungen begründet.“ (Grunwald 2021, 195)



Für die Theologie sind technologische Zukunftsvorstellungen auch als *Zeichen der Zeit* wahrzunehmen, die es im *Licht des Evangeliums* zu deuten gilt

Für die Theologie sind technologische Zukunftsvorstellungen dann konsequenterweise auch als *Zeichen der Zeit* wahrzunehmen, die es im *Licht des Evangeliums* zu deuten gilt (GS 4). Eine sich als theologisch verstehende Ethik, die Abschied genommen hat von dem Irrglauben, ihr christliches Profil durch platte Vernunft- oder Wissenschaftskepsis schärfen zu können, wird sich vor diesem Horizont in die Digitalisierungsdebatte einbringen. Sie wird etwa damit ringen müssen, ob sie wirklich

Algorithmic turn

Simanowski entwirft die Dystopie einer sich immer weiter beschleunigenden Algorithmisierung der Gesellschaft. Dieser auch als *algorithmic turn* bezeichnete Prozess kennzeichnet den zunehmenden Einsatz von Algorithmen, mit denen sich große Datenmengen filtern, auswerten oder auf die Erkennung bestimmter Muster hin untersuchen lassen. Unterschieden wird zwischen regelbasierten Systemen Künstlicher Intelligenz, in denen der Mensch die Vorgaben für den Algorithmus vollständig definiert einerseits und lernenden Systemen, in denen die menschliche Konfiguration lediglich die Ausgangslage für weitere selbstständige Ableitungen durch den

Erhellendes – Kritisches wie Konstruktives – zur Frage nach den anthropologischen Implikationen der kursierenden technologischen Zukunftsvisionen beitragen kann. Sie wird Position beziehen, wie ein verantwortlicher Umgang mit der fortschreitenden Digitalisierung der Gesellschaft im Horizont des christlichen Glaubens gestaltet werden kann. Und sie wird bei alledem auch noch gute Gründe dafür finden müssen, warum man christlichen Positionen insbesondere im ethischen Kontext überhaupt (noch) Gehör schenken sollte, in einer Zeit, in der vom moralischen Kapital der Kirche aufgrund ihres desastösen Umgangs mit dem in ihrem Verantwortungsbereich geschehenen sexuellen und spirituellen tausendfachen Missbrauch von Kindern, Jugendlichen und Erwachsenen kaum noch etwas übrig geblieben zu sein scheint.

Im Folgenden soll die eingangs erwähnte Zukunftserzählung Simanowskis, der eine gesellschaftliche Debatte über die in unserer Gesellschaft maßgeblichen Akzeptanzgrenzen im Kontext der Digitalisierung anregen möchte (vgl. Simanowski 2021²), aus theologisch-ethischer Perspektive analysiert werden.

Algorithmus darstellt andererseits (vgl. Enquete-Kommission Künstliche Intelligenz 2020, 51 f.). Wichtig zu betonen ist, dass es bislang keine Systeme Künstlicher Intelligenz gibt, die auch nur annähernd dazu in der Lage sind, Informationen aus verschiedenen Kontexten in größere Zusammenhänge zu stellen und Lösungsstrategien für Pro-

bleme zu entwickeln: „Dies gilt insbesondere bei der Interaktion mit der physischen Welt, die selbst bei für einen Menschen vergleichsweise einfachen Tätigkeiten ein komplexes Zusammenspiel verschiedener Sinneswahrnehmungen in Kombination mit einer komplexen Motorik und Agilität erfordert.“ (Enquete-Kommission Künstliche Intelligenz 2020, 53) Umstritten ist, ob Künstliche Intelligenz überhaupt jemals diese Fähigkeiten menschlicher Intelligenz wird aus- oder nachbilden können; es gilt außerdem „als sehr unwahrscheinlich, dass eine solche Entwicklung kurz oder mittelfristig bevorsteht“ (ebd.).

Allerdings werden aufgrund der stetigen Generierung weiterer Datenmengen Algorithmen immer unentbehrlicher, so genannte ADM-Systeme (*algorithmic decision making systems*) werden vor allem dort eingesetzt, wo Entscheidungen getroffen werden müssen.² Das führt gegenwärtig auch zu einer Verlagerung der Debatte über Künstliche Intelligenz: „Die Frage, ob es angemessen ist maschinelle Lernsysteme als lernend oder intelligent zu beschreiben, wird tendenziell durch die Frage abgelöst, ob sie zu gerechten Entscheidungen führen.“ (Kaminski 2020, 493) In der Praxis wirken sich automatisierte Entscheidungsprozesse auf das Leben weiter Teile der Bevölkerung aus, ohne dass dies den Betroffenen immer bewusst wäre. Sie kommen in nahezu allen Bereichen des Lebens zum Einsatz, so zum Beispiel bei der Zuteilung knapper Gesundheitsleistungen, bei der Vergabe von Krediten oder bei der Auswahl von Bewerber:innen bei der Besetzung von offenen Stellen. Zwar stehen algorithmische Entschei-

² „Algorithmische Entscheidungsfindung bezeichnet den Gesamtprozess von der Datenerfassung über die Datenanalyse bis hin zur Deutung und Interpretation der Ergebnisse und der Ableitung einer Entscheidung oder einer Entscheidungsempfehlung aus den Ergebnissen. Der Entscheidungsprozess ist festgelegt und kann beliebig viele Datenpunkte verarbeiten. Der Algorithmus fungiert dabei als (teil-)automatisiertes Hilfsmittel oder als ausschließlicher Entscheider. Algorithmische Entscheidungsfindung umfasst also mehrere Schritte und Unterprozesse, angefangen von der Entwicklung der Algorithmen selbst, über die Modellierung bis zur Interpretation des Outputs.“ www.bertelsmann-stiftung.de/fileadmin/files/BS/Publikationen/GrauePublikationen/Teilhaba_ausgerechnet.pdf

dungen immer wieder unter dem Verdacht, Diskriminierung zu fördern und für Manipulationen anfällig zu sein, man spricht hier vom so genannten *algorithm bias*³, jedoch werden Algo-

Frei von Entscheidungen

Dieser Gewinn an Effizienz durch Künstliche Intelligenz ist dann auch das ausschlaggebende Moment, warum der Mensch, so Simanowski, einem *Teufelspakt* zustimme, der darin bestehe, nichts mehr entscheiden zu *können* und auch nichts mehr entscheiden zu *wollen*. Der Mensch werde „der Künstlichen Intelligenz, dem neuen Gott, fortan die Erkenntnis von Gut und Böse, falsch und richtig überlassen sowie jede daraus folgende Entscheidung. Der Mensch, ausgedient als Zwischenwirt der Vernunft und entledigt aller weiteren Mühe, diese einzusetzen, wird heimkehren ins Paradies, wo er sein Leben noch nicht selber leben musste“ (Simanowski, 2020). Die Algorithmen – immer vorausgesetzt, es wird eine solche starke Künstliche Intelligenz überhaupt geben – stellen ein Leben in Aussicht, das frei ist von Entscheidungen, einem Zustand noch vor dem Sündenfall vergleichbar. Das Essen von der verbotenen, aber verlockenden Frucht vom Baum der Erkenntnis von Gut und Böse war einmal. Reizvoller erscheint es, auf die Entscheidung zu verzichten und damit zugleich auch der Gefahr zu entgehen, sich zu irren und die Konsequenzen einer möglicherweise falschen Entscheidung verantworten zu müssen. Zudem wird auch die Schuldfrage gleich mit ausgelagert. Algorithmen in Form einer starken Künstlichen Intelligenz versprechen in diesen und ähnlichen Zukunftserzählungen die Erlösung von dem, was das Menschsein riskant, fehleranfällig und letztlich schuldhaft machen könnte.

Die Imagination der Selbstentmachtung des Menschen hinsichtlich seiner Fähigkeit, zwischen Gut und Böse zu

rithmen trotz dieser mit ihrem Einsatz verbundenen Gefahren und Risiken in vielen Bereichen aufgrund ihrer Effizienz und ihrer Kostenersparnis immer häufiger eingesetzt.

entscheiden, stellt für Simanowski in bewusster Überzeichnung die Signatur einer algorithmisierten Gesellschaft dar (vgl. Simanowski 2021², 125). Ähnliche Visionen von der Befreiung von Entscheidungszwängen durch immer intelligenter werdende Maschinen finden sich auch andernorts; es handelt sich um einen verbreiteten Topos vieler mit der Digitalisierung verbundener Zukunftserzählungen. So schrieb



Algorithmen in Form einer starken Künstlichen Intelligenz versprechen in vielen Zukunftserzählungen die Erlösung von dem, was das Menschsein riskant, fehleranfällig und schuldhaft macht

etwa Ray Kurzweil bereits vor mehr als 20 Jahren in seinem Buch *Homo sapiens*: „Wenn die Gesellschaft und die Probleme, denen sie gegenübersteht, immer komplexer und die Maschinen immer intelligenter werden, lassen sich die Menschen von den Maschinen immer mehr Entscheidungen abnehmen, und zwar aus dem einfachen Grund, weil die von Maschinen getroffenen Entscheidungen zu besseren Resultaten führen als ihre eigenen.“ (Kurzweil 1999, 281 f.) Simanowskis Zukunftsvisionen lassen sich auch durch Yuval N. Hararis Überlegungen in seinem Buch *Homo Deus* flankieren, denen zufolge der Mensch nicht von aufbegehrenden Algorithmen verklärt wird,

sondern sich ihnen überzeugt von der Güte ihrer Entscheidungen anvertrauen wird (vgl. Harari 2017, 451).

Der Mensch der Spätmoderne wird in diesen und ähnlichen Visionen, die zwischen Utopie und Dystopie changieren, als ein entscheidungsmüdes Wesen, das jede Wahl eher als Qual empfindet, portraitiert. Simanowski verweist auf die wachsende Komplexität der Welt, die den Leidensdruck auf Menschen erhöhe: Weil jede Entscheidung in zunehmender Unübersichtlichkeit das Risiko in sich berge, sich als falsch zu erweisen, werde der Verzicht auf moralische Autonomie als Akt der Befreiung verstanden. Das entscheidende Moment ist dabei für Simanowski Folgendes: Im Vergleich zur Künstlichen Intelligenz erlebe der Mensch seine Entscheidungskompetenz zunehmend als defizitär, denn er sei im Gegensatz zu den algorithmisierten Entscheidungssystemen gar nicht mehr eigens in der Lage, die wachsenden Datenmengen zu analysieren, zu standardisieren und auf dieser Grundlage möglichst objektive Entscheidungen zu fällen. So übertrage der Mensch den Maschinen die Souveränität.

Simanowski bedient sich für seine dystopische Beschreibung dezidiert religiöser Metaphorik. Die Maschinen ermöglichen die Rückkehr ins Paradies. Der Garten Eden im Zeitalter der Digitalisierung verheiße das Leben in einer ein für alle Mal entscheidungsfreien Zone. Die Mühsal der Entscheidung, als die sich die verbotene Frucht vom Baum der Erkenntnis entpuppt hat, werde (wieder) einem höheren Wesen, einem Gott namens Algorithmus überlassen: „[I]ndem sich der Mensch dem KI-Regime unterstellt, verzichtet er in seinem triumphalsten Augenblick auf den Ungehorsam, der ihn einst dazu führte, Gottes Verbot, vom Baum der Erkenntnis zu essen, zu übertreten.“ (Simanowski 2021², 102)

³ „Für die Güte und die Nachvollziehbarkeit algorithmischer Entscheidungen sind die Verfügbarkeit, aber auch die Qualität, Integrität und Transparenz von Trainingsdaten und Datenmerkmalen, die der Entscheidung zugrunde liegen, von entscheidender Bedeutung.“ (Enquete-Kommission Künstliche Intelligenz 2020, 201)



Leidensminderung als moralisches Qualitätssiegel

Die beschriebenen Zukunftsvisionen zeigen: Die Zukunftserzählung von der totalen Algorithmisierung der Gesellschaft stellt nichts weniger als die Rückabwicklung der Emanzipationsgeschichte des Menschen in moralischer Hinsicht in Aussicht. Dabei handelt es sich um eine Vorstellung, die jede Ethik auf den Plan ruft, auch die theologische, verbinden sich mit einer solchen Dystopie doch anthropologische und moralphilosophische Fragen von großer Tragweite, die – wie eingangs entfaltet – als *gegenwärtige Konstruktionen* wahrzunehmen sind. Um jedoch diese und ähnliche Prognosen als Gegenwartskonstruktionen erschließen zu können, ist es aus ethischer Perspektive notwendig, die Wahrnehmung für die Motive zu schärfen, die einen solchen in Aussicht gestellten und scheinbar alternativlosen freiwilligen Souveränitätstransfer vom Mensch auf die Maschinen rechtfertigen sollen.

In Simanowskis Vision lassen sich zwei miteinander verbundene Versprechen auf die Zukunft hin identifizieren. Zum einen stellt die Algorithmisierung der Gesellschaft eine Effizienzsteigerung bei Entscheidungen in Aussicht. Algorithmen gelten als Werkzeuge, mit denen die zunehmende Datenflut beherrschbar bleibt. Sie ermöglichen es aber nicht nur, mit der offenbar unausweichlichen Datafizierung Schritt zu halten, sondern sie generieren vor allem Entscheidungen, zu denen der Mensch nicht mehr in der Lage sein könnte. Mit der Steigerung der Entscheidungsquantität und – folgt man Harari (vgl. Harari 2017, 460 ff.) – auch der Entscheidungsqualität geht zu-

gleich eine grundlegende Entlastung des Menschen einher. Da der Mensch sich mit einer immer komplexer und unüberschaubarer werdenden, ihn auch in moralischer Hinsicht überfordernden Welt konfrontiert sieht, tragen algorithmisierte Entscheidungen letztlich zur Leidensminderung, die seit jeher als „moralischer Kern des Fortschritts“ (Rendtorff 1997, 262) gilt, bei.

Effizienzsteigerung und Leidensminderung spielen also eine zentrale Rolle für die Legitimierung der Algorithmisierung von Entscheidungsfindungen. Sie gelten auch in anderen Kontexten als Legitimationsgründe für die moralische Akzeptanz von Fortschrittstechnologien. Allen voran ist das moralische Qualitätssiegel schlechthin für Fortschritt als einem „Fortschreiten zum Besseren“ (Kosselleck 2006, 168) die Überwindung oder zumindest Minderung von Leiden. Zweifellos stellt jeder technologische Fortschritt, der darauf abzielt, ein hohes und förderungswürdiges Gut dar. Eine reflektierte Wertschätzung des Fortschritts ist sich jedoch auch der Ambivalenzen bewusst, die jede neue technologische Entwicklung mit sich bringt: „Mit jeder ‚Bewegung zum Besseren‘ nimmt auch die Möglichkeit zu, das im Fortschritt Erstrebte zu verfehlen. Oder seine Möglichkeiten zu missbrauchen.“ (Rendtorff 1997, 261) Das führt zu der Notwendigkeit, immer auch die Kehrseite der in Aussicht gestellten Verbesserungen bei der moralischen Bewertung der jeweiligen Fortschrittstechnologie zu berücksichtigen.

Theologisch-ethische Denkanstöße

Doch wie sieht die Gewinn-Verlust-Rechnung aus, die Simanowski in seiner Zukunftsvision aufstellt? Ihr Clou besteht darin, den Preis, den der Mensch für die totale Algorithmisie-

rung zu zahlen hat, die Aufgabe seiner Entscheidungsfreiheit, nicht als Verlust, sondern als Gewinn auszuweisen. Darin sieht er den Knackpunkt der Algorithmisierung im Hinblick auf die

Frage nach dem Wert der Entscheidung. Im Anschluss daran lässt sich nun eine der meines Erachtens drängendsten Fragen des Digitalzeitalters folgendermaßen formulieren: Verteidigen wir die Mühsal der Entscheidung als Akt der Freiheit? Oder verstehen wir es als Akt der Befreiung, uns der Mühsal der Entscheidung mit Hilfe der Algorithmen zu entledigen – mit allen Konsequenzen, die eine solche Selbstentmächtigung des Menschen als moralisches Wesen mit sich brächte? Diese Frage wird die Gesellschaft in den kommenden Jahren intensiv beschäftigen, die Gründe dafür dürften deutlich geworden sein. Aus einer theologisch-ethischen Perspektive können für diese Debatte folgende Denkanstöße formuliert werden:

1. Dass in derzeit einschlägigen Zukunftsvisionen im Kontext der Debatte um Künstliche Intelligenz Maschinen autorisiert werden sollen, Entscheidungen zu treffen, schärft die Wahrnehmung dafür, dass die Freiheit und Notwendigkeit der Entscheidung in einer immer komplexer werdenden Welt von vielen Menschen vermehrt als Überforderung wahrgenommen wird. Jede, auch die sich als theologisch verstehende Ethik hat sich von daher selbstkritisch zu fragen, ob sie ihrer Aufgabe genügend nachkommt, in der zunehmenden Unübersichtlichkeit Orientierung zu ermöglichen und moralische Akteure zu stärken und zu befähigen, ihrer Verantwortung gerecht zu werden – was nicht bedeutet, Algorithmen nicht klug und effizient in klar definierten Entscheidungsprozessen kontrolliert einzusetzen.

2. Überlässt der Mensch den Maschinen tendenziell immer mehr Entscheidungskompetenz, dann kommt es zu einer folgenschweren Verantwortungsdiffusion. Es öffnet sich ein Raum der Intransparenz von Verantwortlichkeiten, der zu einem Ort ethischer Dauerreflexion werden muss. Denn die Frage, wer die Verantwortung trägt, sie im Schadensfall übernimmt und dafür sorgt, dass zukünftige Schäden abgewendet werden, ist



Algorithmen erhöhen die Gefahr, dass die Frage nach der Verantwortungsübernahme ins Leere läuft

nicht mehr eindeutig zu beantworten. Algorithmen erhöhen also die Gefahr, dass die Frage nach der Verantwortungsübernahme ins Leere läuft. Aber nicht nur Schadensbegrenzung wird unmöglich. Der Verlust eines moralisch verantwortlichen Adressaten und seine Substitution durch eine lernende Künstliche Intelligenz hätte auch zur Konsequenz, dass die Kriterien, was ein Fortschritt hin zum Besseren sein soll, nicht mehr Gegenstand diskursiver Auseinandersetzung sein würde, sondern Teil von technischen Prozeduren, die aus menschlicher Perspektive gar nicht mehr nachvollzogen werden können. Bei alledem sollte nicht aus dem Blick geraten, dass derzeit zum Einsatz kommenden, schwachen Künstlichen

Intelligenzen immer Entscheidungen von Menschen zugrunde liegen, denen dann auch die Verantwortung für etwaige Diskriminierungen zugeschrieben werden kann (vgl. Orwat 2019).

3. Auch wenn gerade die christliche Ethik ihrem Selbstverständnis nach völlig zu Recht Anwältin der Leiden ist und sein muss, wird sie sich im Hinblick auf die Fortschrittsthematik vor allzu reflexartigen Überlegungen schützen müssen. Denn auch das hat die Analyse der Zukunftserzählung

Leidensminderung darf nicht undifferenziert ein Gütesiegel für die moralische Qualifizierung von Fortschritt sein

von der Algorithmisierung der Gesellschaft gezeigt: Leidensminderung darf nicht undifferenziert ein Gütesiegel für die moralische Qualifizierung von

Fortschritt sein. Vielmehr muss gefragt werden, ob es sich dabei um ein Leiden handelt, das nach eingehender Güter- und Übelabwägung um jeden Preis beseitigt werden soll. Eruiert werden muss auch, ob es nicht andere Möglichkeiten gibt, das Ziel der Leidensminderung zu erreichen. Das Prinzip der Verhältnismäßigkeit wird dabei eine zentrale Rolle spielen. Schließlich kann die Leidensbeseitigung im Hinblick auf Entscheidungszwänge auf lange Sicht hin sehr viel mehr Leiden produzieren, wenn man auch die Konsequenzen mit in den Blick nimmt, die aus der soeben beschriebenen Verantwortungsdiffusion erwachsen können. Wer also vorgibt, Leiden durch den Einsatz bestimmter Technologien beseitigen zu wollen, der wird nicht darauf verzichten können und dürfen, zu fragen, ob dieser Vorsatz nicht auf lange Sicht hin sein eigentliches Ziel konter-

LITERATUR

- Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale (2020), Deutscher Bundestag, 19. Wahlperiode, Drucksache 19/23700, auch abrufbar unter: <https://dserver.bundestag.de/btd/19/237/1923700.pdf> (Stand 28.04.2022).
- Grunwald, Armin: Digitalisierung und Künstliche Intelligenz. Hoffnung auf bessere Prognosen, in: Hock, K./Stengel, F./van Oorschot, J. (Hg.): Zukunfts-Sichten zwischen Prognose und Divination, Berliner Theologische Zeitschrift 38 (2021) 195–214.
- Grunwald, Armin: Ethik und Technik, in: Heßler, M./Liggieri, K. (Hg.) (2020): Technikanthropologie. Handbuch für Wissenschaft und Studium, Baden-Baden, 69–82.
- Grunwald, Armin: Umstrittene Zukünfte und rationale Abwägung. Prospektives Folgenwissen in der Technikfolgenabschätzung, in: Technikfolgenabschätzung – Theorie und Praxis 16 (2007) 54–63.
- Grunwald, Armin: Wovon ist die Zukunftsforschung eine Wissenschaft?, in: Popp, R./Schüll, E. (Hg.) (2009): Zukunftsforschung und Zukunftsgestaltung. Beiträge aus Wissenschaft und Praxis, Berlin, 25–35.
- Harari, Yuval Noah (2017): Homo Deus. Eine Geschichte von morgen, München.
- Heßler, Martina: Ersetzung, in: Dies./Liggieri, Kevin (Hg.) (2020): Technikanthropologie, a. a. O., 263–269.
- Horn, Eva (2014): Zukunft als Katastrophe, Frankfurt.
- Kaminski, Andreas: Entscheiden, in: Heßler, M./Liggieri, K. (Hg.) (2020): Technikanthropologie, a. a. O., 493–501.
- Koselleck, Reinhart (2006): Begriffsgeschichten. Studien zur Semantik und Pragmatik der politischen und sozialen Sprache, Frankfurt.
- Kurzweil, Ray (1999): Homo sapiens. Leben im 21. Jahrhundert – was bleibt vom Menschen? Köln.
- Müller, Oliver/Liggieri, Kevin: Mensch-Maschine-Interaktion seit der Antike: Imaginationsräume, Narrationen und Selbstverständnisdiskurse, in: Dies. (Hg.) (2019): Mensch-Maschine-Interaktion, Stuttgart.
- Orwat, Carsten (2019): Diskriminierungsrisiken durch Verwendung von Algorithmen.
- Rendtorff, Trutz: Ethik im Haus der Wissenschaft, in: Münchener theologische Zeitschrift 48 (1997) 255–264.
- Simanowski, Roberto: Die Zukunft des Menschen. Das Versprechen der Künstlichen Intelligenz, 13.12.2020, online abrufbar: www.deutschlandfunk.de/die-zukunft-des-menschen-das-versprechen-der-kuenstlichen-100.html (Stand: 28.04.2022).
- Simanowski, Roberto (2021²): Todesalgorithmus. Das Dilemma der künstlichen Intelligenz, Wien.

kariert, also letztlich sogar mehr Leid produziert als er eliminiert.

4. Im Kontext der Algorithmisierung ist die vorgebliche Leidensminderung, also die Befreiung vom Druck der Entscheidung durch maschinengestützte Entscheidungssysteme, der Anfang vom Ende des Menschen als Wesen, dem Autonomie und Freiheit zukommen, und das dazu in der Lage ist, Gut und Böse und damit auch Gott vom Teufel zu unterscheiden. Die jüdisch-christliche Deutung der moralischen Emanzipation des Menschen zeigt sehr präzise, wo die Gefahrenlinie verläuft. Den Garten Eden deutet Simanowski um in einen Sehnsuchtsort, einen Ort, an dem in alle Ewigkeit Entscheidungsfreiheit herrscht. Aber wäre das Zurück in so eine Art von Paradies, in dem der Mensch – wie einst – nicht teilhätte an der Erkenntnis von Gut und Böse, nicht fatalerweise ein Ort, den er gar nicht mehr als schlechthin guten Ort identifizieren könnte? Der Garten Eden, den uns die jüdisch-christliche Tradition vor Augen stellt, ist für die christliche Ethik auch deshalb so in-

ZUR AUTORIN

Katharina Klöcker, geb. 1972, Dr. theol., Professorin für Theologische Ethik an der Katholisch-Theologischen Fakultät der Ruhr-Universität Bochum. Forschungsschwerpunkte: Fundamentelethische Fragestellungen; Bio- und Medizinethik; Missbrauch in Fürsorgekontexten; Theologisch-ethische Vulnerabilitätsforschung; Sicherheits- und Präventionsethik (Terrorismusbekämpfung); Ethik der Digitalisierung.

Veröffentlichungen zum Thema: Klöcker, K.: Autoritäre Algorithmen, in: Herder Korrespondenz 7/2021, 37–39; Klöcker, K.: Prämissen und Paradigmen der Prädiktion. Theologisch-ethische Marginalien zur Digitalisierung der Medizin, in: Fritz, A./Mandry, C./Proft, I./Schuster, J. (Hg.): Digitalisierung im Gesundheitswesen. Anthropologische und ethische Herausforderungen der Mensch-Maschine-Interaktion (Jahrbuch für Moralthologie 5) Freiburg 2021, 138–157; Klöcker, K.: Zur ethischen Diskussion um Enhancement. Eine kritische Anmerkung zum Transhumanismus aus theologisch-ethischer Perspektive, in: Göcke, Benedikt/Meier-Hamidi, Frank (Hg.): Designobjekt Mensch. Die Agenda des Transhumanismus auf dem Prüfstand, Freiburg 2018, 309–338.

teressant, weil in diesem Garten die Sehnsucht danach, zu wissen, was Gut und Böse ist, und damit die Beanspruchung moralischer Autonomie, offenbar so stark wird, dass sie zu einem Akt des Ungehorsams gegenüber der göttlichen Autorität führt. Für diesen Akt der Selbstermächtigung, das Erringen moralischer Autonomie, zahlt der Mensch einen hohen Preis: Er tauscht

die Komfortzone der Unkenntnis von Gut und Böse gegen einen, zumindest im Vergleich zum Paradies, unwirtlichen Ort ein. Dort wird ihm Verantwortung übertragen und die Möglichkeit, schuldig zu werden. Doch verbindet sich mit diesem Ort auch die Verheißung an den Menschen, dass er die Welt als Mitschöpfer gestalten kann.



What does it mean to be human in a world of technology and AI?



Wir leben in einer Welt, die von scheinbar intelligenten Maschinen beherrscht wird, und wir neigen dazu, unser eigenes Menschsein durch die verzerrende Brille von Technologie und KI zu betrachten. Digitale Technologien führen dazu, dass wir uns auf körperlose, abstrakte und quantifizierbare Informationen konzentrieren und dazu neigen, unsere menschliche Verkörperung zu entwerten. Der nach Gottes Ebenbild geschaffene Mensch ist wesentlich mit Gott verbunden und von Gott mit einer einzigartigen Berufung ausgestattet. Die Menschwerdung und Auferstehung Christi bieten eine reiche Grundlage, um die menschliche Verkörperung und Einzigartigkeit in einer Welt intelligenter Maschinen zu feiern, zu respektieren und zu schützen. Die Herausforderung, vor der wir stehen, besteht darin, innovative und kreative Wege zu finden, um hochentwickelte Technologien der Künstlichen Intelligenz nicht dazu zu nutzen, uns zu erniedrigen, zu entstellen oder zu schmälern, sondern uns zu befähigen, uns als menschliche Personen zu entfalten, die einzigartig geschaffen sind, um Gottes Charakter und Wesen widerzuspiegeln.



John Wyatt

Exploring human identity

What does it mean to be human? It sounds like a question in an undergraduate philosophy examination rather than anything to do with the real world. However, the question seems to have new urgency and relevance in the confusing and divided culture in which we find ourselves. Moreover, there is little doubt that the question about human identity is of renewed importance because of the growing power and potential of artificial intelligence technology. In a world that is going to be more and more dominated by intelligent machines, is there anything distinctive about human beings? If machines can increasingly take on most of the roles that human beings have filled, what are human beings for?

Of course, human beings have always reflected about their own identity and meaning within the vast cosmos. But whereas in the past the deep questions

around human identity seemed merely of speculative and academic interest, now these questions have emerged into popular culture. Paperbacks like *Sapiens* and *Homo Deus* from the best-selling historian Yuval Harari have taken on cult status, selling tens of millions of copies worldwide. In 2017 over 200 million people watched online as an AI program called AlphaGo competed with Lee Sodol, a human world champion player of the ancient game of Go, with the event framed as an ultimate test as to whether “AI has become more intelligent than human beings”.

What does intelligence mean?

There is a common narrative amongst AI technology enthusiasts that goes something like this: ‘Everything human beings have ever achieved is a

product of intelligence. But for the first time in human history our intelligence can be replicated in information processing machines which are constantly increasing in power and complexity. Therefore machines are going to achieve everything that human beings can do’.

Max Tegmark, an influential physicist who is President of the Future of Life Institute, says: “[T]he conventional wisdom among AI researchers is that intelligence is ultimately about information and computation, not about flesh, blood or carbon atoms. This means that there’s no fundamental reason why machines can’t one day be at least as intelligent as us... Everything we love about civilization is a product of intelligence, so amplifying our human intelligence with artificial intelligence has the potential of helping civilization flourish like never be-




fore – as long as we manage to keep the technology beneficial.” (Tegmark, 2018)

And Demis Hassabis, co-founder of the company Google DeepMind, is on record as stating that his goal is to “solve intelligence and then use that to solve

everything else” (Hassabis 2016). We will return to the claim that machines can replicate human intelligence, but first we look at some of the mechanisms by which technology is changing the way we understand ourselves.


psyche was at risk of excessive internal forces leading to potentially destructive consequences (!)

 In the age of digital technologies, humans are considered information-processing machines

We understand our humanity through the lens of technology

Human beings have always tried to understand themselves by comparison with the leading technologies of the time. In the opening pages of his *Treatise on Man*, published in 1662, René Descartes wrote:

“I make the supposition that the body is nothing else but a statue or earthen machine, that God has willed to form entire ... We see clocks, artificial fountains, mills, and other similar machines, which, being only made by men, nevertheless do not lack the force to move themselves in several diverse means. And it seems to me that, as (the human body) is made by the hand of God, I cannot imagine how many kinds of movements there are in it.” (Descartes, 1662)

 Many 17th and 18th century thinkers understood the human body as a sophisticated clockwork mechanism

Descartes concluded that all our bodily functions followed from “the mere arrangement of the machine’s organs every bit as naturally as the movements of a clock or other automaton follow from the arrangement of its counter weights and wheels”. Since precision clockwork mechanisms represented the pinnacle of technological achievement in the 17th and 18th centuries, it seemed obvious to many thinkers that the best way to understand the human body was to view it as some kind of sophisticated clockwork mechanism.

Here is the 18th century philosopher Denis Diderot: “Consider man as

a walking clock, his heart is the main-spring, the contents of the thorax are the principle parts of the works, in his head are the bells, complete with little hammers ... and the soul is the tiny figure on the top listening to the music of the chimes.” (Diderot, 1751)

Meanwhile the atheist philosopher Julien Offray de La Mettrie, claimed in 1743 that “the human body is a machine which winds its own springs. It is the living image of perpetual movement” (De la Mettrie, 1743).

With the rise of steam power and new forms of hydraulic technology in the 19th century, the metaphor changes. The human body is now conceived as a system of tubes and chambers filled with incompressible fluids. This is the period in which the physiology of the circulation, the capillaries and the lymphatic system are being established. It has often been pointed out that Sigmund Freud’s influential model of human psychology is at root an extended hydraulic metaphor. Libido is conceived as an incompressible fluid, channelled within the subterranean tubes of the mind. The superego works like a regulatory valve mechanism to ensure the internal psychic pressure is controlled. Just as a steam engine might explode if excessive, pressure was not safely released, so the human

In the mid-1800s, inspired by recent advances in communications, the German physicist Hermann von Helmholtz moved beyond hydraulics and suggested that the brain could be conceived as a telegraphic system. “In the network of telegraphs”, he wrote, “we find everywhere the same copper or iron wires carrying the same kind of movement, a stream of electricity, but producing the most different results in the various stations according to the auxiliary apparatus with which they are connected” (von Helmholtz, 1995). Therefore, it made sense to conceive of the central nervous system as a highly complex network of wires transmitting signals to and from receiving stations.

But now in the early decades of the 21st century our culture is being transformed by digital technologies and machine intelligence, so we should not be surprised that humans are being increasingly seen as information processing machines. In his book *Life 3.0*, Mark Tegmark concludes that human life can be seen as a “self-replicating information processing system” (Tegmark, 2018). To philosopher Daniel Dennett, “we’re robots made of robots made of robots. We’re incredibly complex, trillions of moving parts. But they’re all non-miraculous robotic parts” (Dennett, 2017).

Are human beings machines that think?

Physicist Sean Carroll wrote: “When asked for my thoughts about machines that think, I can’t help but reply: Hey, those are my friends you’re talking about. We are all machines that

think, and the distinction between different types of machines is eroding.” (Carroll, 2015)

Of course, there is a certain truth behind the idea that the brain can be

Arts & ethics

Birgit Feike: „Sei Dein Gott“

Das Motiv des Doppelgängers bzw. auch das des Januskopfes bekommt im Zeitalter der Künstlichen Intelligenz eine neue Dimension: Nun ist es nicht mehr das eigene Ich, das in psychischer Spaltung Zwiesprache hält oder dem Gegenüber seine Doppelgesichtigkeit zeigt, sondern es gibt nur noch das eine eigene Gesicht, das in vermeintlich ständiger Selbstspiegelung schon lange seine Profilbildung an die Algorithmen abgegeben hat. Die doppelköpfige Skulptur „Sei Dein Gott“ spielt mit diesem Vexierspiel: Zwei Normköpfe, gewebt aus Flutterband, sind in der Waagerechten angeordnet und um 180 Grad gegeneinander gedreht. Zwischen ihnen ist das Flutterband in einem lockeren Knäuel in Auflösung begriffen; und auch am rechten Kopf hängt ein Flutterbandende herunter und erweckt den Eindruck, dass man damit das gesamte Skulpturgewebe auftrennen könnte. Die Dialog- und Korrektivfunktion des Doppelgesichts geht verloren, es bleibt ein Selbstnormierungswahn, der bereits komplett digital gesteuert wird.

(Stefanie Lieb)



Birgit Feike

(geboren in Bottrop), lebt und arbeitet als Künstlerin und Goldschmiedin in Dortmund. Nach ihrer Lehre zur Goldschmiedin arbeitete sie zunächst als Dekorateurin in verschiedenen Branchen und studierte anschließend von 2012 bis 2015 Bildende Kunst an der Ruhrakademie Schwerte. 2015 schloss sie ihr Kunstdiplom bei Gesine Grundmann ab und eröffnete 2016 ihr eigenes Kunstatelier Phoenix West in Dortmund-Hörde. 2019 beteiligte sie sich am Projekt „Water Event“ von Yoko Ono am Museum für Bildende Künste in Leipzig und war mit der Einzelausstellung „willst du mit mir gehen, willst du?“ in der Katholischen Akademie Schwerte vertreten. Aktuell arbeitet Birgit Feike an einem vom Land NRW geförderten Stipendienprojekt im öffentlichen Raum, bei dem sie unter dem Titel „Grenzflächen“ temporäre Flechtobjekte aus Absperrband als Intervention im Raum platziert.

www.birgit-feike.de

www.instagram.com/birgitfeike

www.instagram.com/grenzflaechen



Sei Dein Gott
2017, *Flutterband*, Webarbeit, 70 x 25 x 20 cm



viewed as a machine. There are certain aspects of our human functioning that can usefully be seen as similar to that of a machine. In other words, the machine is a useful *metaphor* for certain aspects of our humanity. The machine metaphor has been extremely successful in fields such as human physiology, molecular biology, genetics, cognitive neuropsychology, and so on. But there is a critical difference between a helpful metaphor, and a definition, or a description of core reality. Yes, it may be helpful to say that a human being is *like a computer*, but to say that a human being *is a computer* is incoherent, it is philosophical nonsense. All the computers you have ever come across are human artefacts, designed and built by human beings in order to achieve goals dreamed up by human beings.



It may be helpful to say that a human being is *like a computer*, but to say that a human being *is a computer* is incoherent

By way of illustration you could say that a fish is *like a submarine*. It has a propulsion system, fins that stabilise and propel it through water, sensors to detect threats and so on. The submarine may be a helpful metaphor in understanding fish behaviour. But to say that 'a fish *is a submarine*' doesn't make any sense.

When people conceive of a human being as 'a machine that thinks' this

What are the differences between human and machine intelligence?

As we saw earlier, a common narrative claims everything human beings have ever achieved is a product of intelligence, and therefore intelligent machines can in theory replicate it. But this is highly misleading. What is the common factor in a musical symphony, a great cathedral, the wordless emotional intelligence in a human relation-

ship, a mathematical theorem, the activities of a stock market trader, a profound novel about human loss, the text of the Bible, a scientific research paper, a cave drawing, or the work of mime artist? The word 'intelligence' turns out to be rather slippery and misleading. In reality, the only 'intelligence' we have any deep understanding of is that becomes a deep and distorted way of understanding reality. Metaphors have profound and pervasive effects on our culture and the 'information processing machine' is becoming a dominant framework for understanding our own humanity. We talk of human beings as 'hard-wired', 'suffering from information overload', 'programmed for failure', 'needing a reboot.' It is commonplace to take the computer concepts of software and hardware and apply them to our own humanity. The hardware (often dismissively referred to as *wetware*) is the physical stuff of our brains – nerve cells, connections, neurotransmitter chemicals. The software is the *information* that somehow resides in our brains – memories, perceptions, emotions, thoughts.

This way of thinking tends to regard the software, the information carried in our brains, as having ultimate significance. Hardware is seen as disposable and recyclable. But the disembodied software is the most precious aspect of all existence – pure information.

Machine metaphors bring with them their own 'reality-distortion field' – they change the way we see the world and our own human nature. It has often been said that 'to the person with a hammer, the whole world looks like a nail.' To a coder with an algorithm, people are just a stream of data, waiting to be analysed, coded, stored and manipulated. The machine metaphor brings with it reductionist and de-humanising tendencies.

ship, a mathematical theorem, the activities of a stock market trader, a profound novel about human loss, the text of the Bible, a scientific research paper, a cave drawing, or the work of mime artist? The word 'intelligence' turns out to be rather slippery and misleading.

In reality, the only 'intelligence' we have any deep understanding of is that

of human beings. The very concept of 'machine intelligence' is deeply problematic, similar to common ideas of machines 'knowing', 'understanding', 'thinking' and 'choosing'. These are activities that are unique to human persons (and of course to God himself, in whose image we are made). Our understanding of and interaction with the world is inextricably linked to our nature as embodied mammals, born from the bodies of our parents, born into a web of social relationships, beings who explore and learn about the world in a distinctively human and embodied way.



The apparent intelligence of machines cannot be *the same as* human intelligence

So the apparent intelligence of machines cannot be *the same as* human intelligence, the form of intelligence which we recognize as characteristic of an embodied, biological organism. Whilst we recognize that machines can undertake specific tasks, such as pattern recognition or data analysis, at levels that far exceed human capacities, we should challenge the idea that machine 'intelligence' can replicate human intelligence. Machines cannot 'know', 'understand', or 'create' in the many wonderful ways that a person can.

Human and machine intelligence are therefore not identical. The abilities of a machine to recognize patterns or analyze data – no matter how differentiated they may be – are devoid of emotion and motivation, limited and, above all, programmed from the outside.

Human intelligence in its many facets, on the other hand, is intrinsic and creative. It motivates humans to formulate and achieve self-imposed goals, makes them social beings, and today enables them to program the AI they create.

Digital technology has a tendency to devalue human embodiment

As digital technology extracts, transfers, copies, stores and processes abstract information about people it contributes to a general tendency to prioritise the abstract and disembodied over the tangible and embodied. Our bodies are messy, unpredictable, constantly changing, hard to control and analyse, and localised in time and space. But once information is extracted in digital form it can be instantaneously copied, processed and transferred across the globe with astonishing precision and power.

Over the last years of the Covid pandemic we have witnessed the extraordinary power of social media and internet channels to facilitate communication between human beings. But paradoxically the very disembodiment of digital technology has also fostered terrible evils: social media abuse, cybercrime, broken relationships, trolling, bullying, internet scams and so on. If you and I are having a physical face-to-face conversation and I say something deeply hurtful and wounding to you, the damage is contained within, at most, a handful of relationships. But if I put the same comment

on to a social media platform, the evil fall-out can be disseminated to millions around the globe.

The emphasis on disembodied intelligence seems to have strange resonances with the ancient Gnostic heresy that flourished in the first centuries of the Christian era. Gnosticism can be seen as a type of 'default-heresy' that recurs in many different forms over the history of the Christian church. In its original form, Gnosticism saw the human body as 'a dark place', 'a rapacious sea which robs and devours', 'a stinking prison' in which the soul, a divine particle of light, was trapped. Redemption or 'gnosis' was separation from the body, and death was celebrated as liberation of the light from the dungeon of materiality.

In the latest materialist version of Gnosticism, the real core of our humanity, the divine spark, is now reconceived as 'information' which can be extracted and processed in a disembodied form, whilst the physical body is denigrated as pathetic, out-moded and increasingly unfit for the modern world.

The implications of being made in God's image

How does Christian thinking engage with the pervasive impact of technology on our self-understanding as a species? Over the centuries, biblical scholars and theologians have wrestled to understand the fundamental meanings of the biblical phrase 'being created in God's image'. One profound implication of this phrase is to say that to be human is to be essentially related to God. Of course this does not mean that all human beings believe and trust in God's existence. But it is to say that we are constituted by God to stand in relationship to him, whether or not that is known or acknowledged (Williams

2021). This means that each of us are accountable to God, and given a vocation by him which is particular to us as an individual. We cannot ignore or escape God because we are created to be in relation with him, called to live our lives in obedience to his purposes, and to respond in love to his compassion and faithfulness towards us.

But of course this is not the same for an AI system, however advanced the technology. As human beings we have no need to feel that our unique identity is threatened by the increasing sophistication of machine learning systems. They can play a valuable

role in assisting our human roles and activities but they cannot 'know God', in all the depth, wonder and mystery of that biblical phrase and they are not connected to God as humans are.

A related implication of our creation in God's image is that we carry a unique calling which has been given to us by our creator. This implies that we should be less concerned about our status as the most superior and intelligent beings on the planet and more concerned about faithfulness to our human calling. There is no doubt that AI systems have some capacities and abilities which far exceed those we are capable of. But we are called to respond to a unique God-given vocation to flourish as the kind of creatures we are, fulfilling the capacities with which we are gifted. As theologian Robert Song put it: "The questions we should be asking are these: Is humankind flourishing in the way it is called to, or is it in practice and in theory denying that? Are our technological practices compatible with recognizing the dignity and personhood of human beings, or do they implicitly or explicitly deny this?" (Song, 2021)

Are we creating AI systems in our own image?

It is sometimes said that "just as human beings have been created in God's image, so we create AI systems in our image". Although this way of thinking seems superficially attractive, I believe it is highly misleading. The whole point is that the way that God creates is fundamentally different from the way that human beings design and manufacture artefacts. This is reflected in the Hebrew word '*bara*' meaning 'to create, form, fashion', which throughout the Old Testament is reserved for God's acts of creation and never applied to acts of human creativity. When God creates, he calls into existence out of nothing, and he upholds and sustains his creation continuously within a covenant

of unchangeable love and faithfulness. As human beings we can never create in this sense, although we are called to exercise the creativity and freedom which is part of our human calling.



When God creates, he calls into existence out of nothing, human beings however can never create in this sense

Where I think the phrase 'we are creating AI systems in our image' may be helpful is that it reminds us that because we are fallen human beings, then

Celebrating, respecting and protecting human embodiment and human uniqueness in a world of intelligent machines

The early church fathers Irenaeus, Tertullian, Clement, Origen, and Eusebius clearly saw the Gnostic heresy as a dangerous threat to Christian beliefs and they invested considerable time and energy in refuting it. They argued that the human body was part of the material creation. Our human nature, made in the image of God and yet formed from the ground, was part of that creation which had been pronounced to be 'very good' by God himself (Genesis 1:31).

But of course it was the Incarnation, Death, and bodily Resurrection of Christ which provided the strongest grounds for the church fathers as they attacked ancient Gnosticism. Just as for the early church so the same is true for us. The Christmas and Easter events are of critical importance as we engage with the impact of sophisticated AI systems and the most recent versions of digital Gnosticism. In the Incarnation the eternal Creator and Lord of the cosmos took on human form.

"The Word became flesh and dwelt among us ..." We are so familiar with these words from the first chapter of John's gospel (John 1:14) that it is easy to lose the meaning that must have shocked and scandalised the original

everything that we create is going to reflect that fallenness. This tragic reality is repeatedly demonstrated by the unintended consequences of algorithmic systems that perpetuate and magnify human hatred, prejudice, discrimination and dangerous conspiracy theories. We should not be surprised when sophisticated human artefacts that were intended to encourage human interactions and relationships have the unexpected consequence of magnifying human evil. It is a reflection of the all-pervasiveness of human evil and brokenness which contaminates everything we create.

readers. The divine Logos, the unique expression of divine perfection, became *sarx*, the Greek word for meat. To the elite philosophers and scholars of the time this was a deeply offensive idea. The suggestion that the supreme God of the universe might become a human baby was ludicrous, crude, offensive, and distasteful. To the Platonic philosophers the ultimate divinity was mystical, abstract, immaterial perfection – the eternal Form of The Good – far from the sordid realities of a stable in Bethlehem.

But the scandalous and wonderful reality of the Incarnation led to a completely new respect for the human body which has continued to the present day. So what is so special about the human body? Why should we treat our own bodies with respect? Because

Practical implications of the Christian understanding of human embodiment

The exalted view of the human body was translated by the church fathers into practical action. The *Clementine Homilies*, dating from the 4th century A.D. state: "It behoves you to give honour to the image of God which is man,

this strange and idiosyncratic collection of 25 000 genes, 10 billion nerve cells, several miles of wiring, 8 metres of intestinal plumbing, 5 litres of blood, and assorted biochemical engineering – this is the form in which God became flesh! Our humanity is not something which comes between us and God. No, it is the very means by which God is made known.

Jesus said "Destroy this temple, and I will raise it again in three days.' [...] But the temple he had spoken of was his body" (John 2:19.21). Here is a new and exalted view of the human body. If Christ's body was a temple, then the early Christians concluded that they must treat all bodies with a new reverence. "Do not despise the wonder within you" was the response of the church fathers, and it is an authentically Christian response to the body; a sense of awe and wonder at the mystery of humanness.

In the Incarnation, Death and Resurrection of Christ, it is possible to see that human nature was both reaffirmed and fulfilled. When Christ is born and then raised as a physical, recognizable, touchable human being, God proclaims his vote of confidence in human nature and the original created order. In the resurrection of Christ, the physical creation is not overturned or annihilated. Instead, it is renewed and caught up into a greater and richer reality. In Jesus, the Second Adam, we see the pioneer, the blueprint for a new type of human being, the one in whose likeness a new creation will spring, the firstfruits of those who are to come (1 Cor 15:20).

in this wise: food to the hungry, drink to the thirsty, clothing to the naked, care to the sick, shelter to the stranger, and visiting him who is in prison, to help him as you can."



In response, Christians pioneered the concept of the hospital, including dangerous places such as hospitals for plague victims and leprosy sufferers. Even the word 'hospital' has Christian roots. It comes from the Latin *hospes*, meaning 'a guest'. A hospital is a place where we practice hospitality, neighbour-love to strangers, a concept which grows out of Christian concern and respect for embodied human beings.

Perhaps we can see parallels between the church fathers' defence of Christian truth against ancient Gnosticism and our current battle against the technological devaluing and demeaning of embodied humanity. I am struck that many thoughtful observers in our society have deep intuitions that the current emphasis on living digitally disembodied lives is unhealthy. Many people have a deep sense of unease about the anti-human direction in which much technology seems to be heading. They sense that the "frictionless" utopia which the techno-optimists are working towards is not a place in which embodied human beings will flourish. But of all the philosophies and worldviews of our current age, it is Christianity which has sources to provide a theologically and philosophically robust explanation for *why* human embodiment matters, and it should do so. The biblical Christian faith, cen-

tered on the Incarnation and Resurrection of Christ, provides an internally coherent perspective from which we can defend, protect and celebrate the primacy and the goodness of human physical embodiment.

As Christians the challenge we face is to find innovative and creative ways of using sophisticated artificial intelligence technologies not to demean, distort or diminish us, but rather to enhance and protect our embodied nature as human persons. Unexpectedly the coronavirus pandemic has illuminated many redemptive possibilities and opportunities for creative engagement with digital technology that were previously invisible. Yet we cannot ignore the destructive, de-humanising and 'Gnostic' tendencies which are also inevitably present as we spend more time in virtual interactions and online activities.

As we reflect on the range of possibilities and opportunities that AI technology offers, we must ask how we can build a future in which physically embodied human beings can flourish, the vulnerable can be protected, and face-to-face relationships can be celebrated and affirmed. And we should heed those words that come from the dawn of the Christian era: "Do not despise the wonder within you."

ZUM AUTOR

John Wyatt, born 1952, is Emeritus Professor of Neonatal Paediatrics, Ethics and Perinatology at University College London. He practised as a consultant neonatologist for more than 25 years and led an internationally recognised multidisciplinary research team into the prevention of brain damage in newborn babies. He is now focussing on research into the philosophical, ethical and theological implications of advances in biomedical and digital technologies. His most recent book is *The Robot will See You Now*, co-edited with Stephen N Williams, published by SPCK (2021). Previous books include *Matters of Life and Death: human dilemmas in the light of the Christian faith*, IVP (2009) and *Dying Well*, IVP (2018). His personal website is johnwyatt.com

Conclusion

- Human beings have always tried to understand themselves in terms of the latest technology. The concepts that we are 'machines that think' or biological 'information processing systems' are increasing in our culture.
- Although there are aspects of our human bodies that are machine-

BIBLIOGRAPHY

- Carroll, Sean (2015): We are all machines that think, in: Brockman (ed.): *What to Think about Machines that Think*, New York.
- Clementine Homilies, Homily 7, ch 4.
- Dennett, Daniel (2017): Philosopher Daniel Dennett on AI, Robots and Religion, Financial Times, London, 3 March 2017.
- Descartes, Renee (1662): *Treatise on Man*.
- Diderot (1751): *Letter on the Deaf and Mute*.
- Hassabis, Demis (2016): The Superhero of Artificial Intelligence. Interview with Demis Hassabis by Clemency Burton-Hill, in: The Guardian Newspaper 16. Feb. 2016. Online: <https://www.theguardian.com/technology/2016/feb/16/demis-hassabis-artificial-intelligence-deepmind-alphago> (30.05.2022).
- Helmholtz Hermann von (1995): *Science and Culture: Popular and Philosophical Essays*, ed. Offray.
- De La Mettrie, Julien (1743): *L'Homme machine*.
- Song, Robert (2021): Robots, AI and human uniqueness: learning what not to fear, in: Wyatt and Williams (eds): *The Robot Will See You Now*, London, 114.
- Tegmark, Max (2018): *Life 3.0 – Being human in the age of intelligent machines*, London.
- Williams, Stephen N. (2021): What is it to be a person?, in: Wyatt and Williams (eds): *The Robot Will See You Now*, London, 96.



like, it is both philosophical and theological nonsense to say that we are machines.

- As digital technology extracts and processes information about people it contributes to a general tendency to prioritise the abstract and disembodied over the tangible and embodied. This can be seen as a modern version of the ancient Gnostic her-

esy that regarded the human body with contempt.

- Because we are created in God's image we are essentially related to God and we are given a unique calling. We are called to respond to a unique God-given vocation to flourish as the kind of creatures we are, fulfilling the capacities with which we are gifted.

- In the Incarnation and Resurrection of Christ we see God's final vote of confidence in physically embodied humanity. We are called to build a future in which human beings can flourish, the vulnerable can be protected, and face-to-face relationships can be celebrated and affirmed.



Moral für Maschinen?

Grundsätze der Maschinenethik



In vielen Anwendungsbereichen wie beim autonomen Fahren, im Krieg, aber auch in der Pflege stehen künstliche Systeme moralischen Problemen gegenüber. Die Maschinenethik ist eine aufstrebende Disziplin an der Schnittstelle von Philosophie, Informatik und Robotik, die sich mit der Frage beschäftigt, ob und wie künstliche Systeme auch moralische Entscheidungen treffen können. In diesem Beitrag wird die neue Disziplin vorgestellt und Argumente für und wider dieses Vorhaben diskutiert. Abschließend werden drei Grundsätze eingeführt, die als Leitplanken dienen und sicherstellen sollen, dass die technologische Entwicklung in eine positive Richtung erfolgt.

1. Was ist Maschinenethik?

Während *Artificial Intelligence* zum Ziel hat, die kognitiven Fähigkeiten von Menschen zu modellieren oder zu simulieren, geht es im Feld der *Artificial Morality* darum, künstliche Systeme mit der Fähigkeit zu moralischem Entscheiden und Handeln auszustatten. Die Idee ist also, Computer so zu programmieren, dass sie moralische Entscheidungen treffen können. Die Maschinenethik ist diejenige Disziplin, die sich mit der Möglichkeit von *Artificial Morality*, ihren theoretischen Grundlagen und ihrer ethischen Bewertung auseinandersetzt (Misselhorn 2018).

Lange Zeit stand die Maschinenethik zu Unrecht im Verdacht, bloß Science-Fiction zu sein. Das stimmt jedoch nicht. Denn bereits ein so einfaches Gerät wie ein Staubsaugerroboter steht vor moralischen Entscheidungen: Soll er einen Marienkäfer einfach einsaugen oder soll er ihn verscheuchen bzw. umfahren? Und was ist mit einer Spinne? Soll er sie töten oder ebenfalls verschonen? Ein solcher Roboter ist in einem minimalen Sinn autonom, weil er im Unterschied zu einem konventionellen Staubsauger nicht von einem Menschen geführt oder überwacht wird. Die Pointe eines solchen Roboters ist, dass er möglichst dann aktiv werden soll, wenn wir gerade nicht zu Hause sind.

Man kann auch in Zweifel ziehen, dass es sich um eine moralische Entscheidung handelt. Das ist jedoch der Fall, weil es darum geht, ob man Tiere zu Reinigungszwecken töten darf. Gewöhnliche Staubsaugerroboter können eine solche Entscheidung allerdings noch nicht treffen. Zumindest als Prototyp liegen jedoch Forschungsansätze vor, ein Ethikmodul für das populäre Modell Roomba zu entwickeln (Bendel 2017), welches das Leben von Insekten berücksichtigt (der Prototyp besitzt wahlweise einen ‚Kill-Button‘ für Spinnen).



Pflegesysteme, Kriegsroboter und autonome Fahrzeuge stellen drei zentrale Anwendungsfelder der Maschinenethik dar

In komplexeren Einsatzbereichen autonomer Systeme stellen sich auch anspruchsvollere moralische Entscheidungen. Beispiele hierfür sind etwa Pflegesysteme, Kriegsroboter und autonome Fahrzeuge, die drei zentrale Anwendungsfelder der Maschinenethik darstellen (Misselhorn 2018). Alle drei Felder erfordern grundlegende moralische Entscheidungen, in denen es



Catrin Misselhorn

manchmal sogar um Leben und Tod von Menschen geht. Eine zentrale Frage der Maschinenethik ist, ob man Maschinen solche Entscheidungen überhaupt überlassen darf oder ob man es vielleicht sogar tun sollte.

Wir wollen zunächst Argumente zugunsten der Behauptung betrachten, dass wir eine Maschinenethik brauchen. Sodann werden Argumente dargestellt, die gegen die Maschinenethik geltend gemacht werden können. Diese Argumente beziehen sich auf der einen Seite auf die technische Machbarkeit und auf der anderen auf die moralische Wünschbarkeit von Maschinen-Moral. Wie sich erweisen wird, haben beide Seiten ihre Berechtigung. Deshalb kommt es darauf an, wie die Maschinenethik ausgestaltet wird. Ein Vorschlag dafür wird mit drei Grundsätzen der Maschinenethik vorgelegt, die den Einwänden Rechnung tragen und sicherstellen sollen, dass sich die Maschinenethik in eine gute Richtung entwickeln kann.

2. Argumente für die Maschinenethik

Ein wesentliches Motiv der Technisierung ist, dass Maschinen die Menschen von Tätigkeiten entlasten sollen, die schwer, schmutzig, gefährlich oder einfach nur unangenehm sind. Manchmal sind die Leistungen von Maschinen auch einfach schneller oder präziser als diejenigen des Menschen.

Mit zunehmender Intelligenz und Autonomie geraten Maschinen jedoch auch in Situationen, die moralische Entscheidungen erfordern. Daraus wird häufig geschlossen, dass die Entwicklung von Maschinen mit moralischen Fähigkeiten unabdingbar ist, insofern wir die Vorteile autonomer intelligenter Systeme voll ausnutzen wollen (Allen et al. 2011).

Darüber hinaus kann man argumentieren, dass moralische Maschinen bessere Maschinen sind. Eine gute Maschine zeichnet sich dadurch aus, dass sie den menschlichen Bedürfnissen und Werten besonders gut gerecht wird. Die Überlegung ist nun, dass eine Maschine mit einprogrammierter Moral den menschlichen Bedürfnissen und Werten besonders gut entspricht.

Vielleicht handeln künstliche moralische Akteure sogar besser als Menschen, weil ihr Verhalten nicht durch irrationale Impulse, Psychopathologien oder emotionalen Stress beeinflusst wird. Sie lassen sich nicht verführen und werden – im Unterschied zu Menschen – auch nicht durch Eigeninteresse vom Pfad der Tugend abgelenkt.

Außerdem könnten Maschinen in ihren kognitiven Fähigkeiten bei der Situationsbewertung Menschen überlegen sein. Sie sind in der Lage, in Sekundenbruchteilen Entscheidungen zu treffen, in denen ein Mensch gar nicht mehr bewusst entscheiden kann. Das wird als Argument dafür ins Feld geführt, Maschinen moralische Entscheidungen in besonders prekären Situationen zu überlassen, beispielsweise im Krieg (Arkin 2009).

Doch die Maschinenethik ist nicht nur von praktischem Nutzen, sie ist

auch ein theoretisch interessantes Forschungsprogramm. So verspricht sie Einsichten, die die ethische Theoriebildung voranbringen könnten. Denn die menschliche Moral ist fragmentiert und enthält sogar Inkonsistenzen. Die Entwicklung künstlicher Systeme mit moralischen Fähigkeiten stellt hingegen



Die Entwicklung künstlicher Systeme mit moralischen Fähigkeiten stellt die Anforderung, die Moral zu vereinheitlichen und konsistent zu machen

die Anforderung, die Moral (zumindest in den Anwendungsbereichen) zu vereinheitlichen und konsistent zu machen. Denn nur auf dieser Grundlage können künstliche Systeme operieren. Einheitlichkeit und Widerspruchsfreiheit gelten als theoretische Tugenden, deshalb läge darin auch ein Fortschritt für die Ethik als Theorie der Moral.

Auch kognitionswissenschaftlich ist die Maschinenethik von Bedeutung. Denn der Mensch ist zwar einerseits

3. Argumente gegen die Maschinenethik

Diesen positiven Erwartungen an die Maschinenethik zum Trotz werden jedoch auch eine Reihe von Einwänden geltend gemacht. Ihre Stoßrichtung ist einerseits die technische Machbarkeit und andererseits die moralische Wünschbarkeit von Maschinen mit moralischen Fähigkeiten. Häufig sind die kritischen Punkte die Kehrseiten der positiven Aspekte der Maschinenethik.

Dem Argument von der Unabdingbarkeit der Maschinenethik steht eine skeptische Haltung gegenüber, die ihre Umsetzbarkeit grundsätzlich in Frage stellt. Die Zweifel an der Möglichkeit der Maschinenethik gründen in der Kritik an der Vorstellung, dass der menschliche Geist analog zu einem Computerprogramm funktioniert.

Vorbild bei der Entwicklung intelligenter Maschinen, die die Fähigkeit zum moralischen Handeln haben. Auf der anderen Seite inspirierte die wissenschaftliche und technische Entwicklung des Computers und der Künstlichen Intelligenz aber auch immer wieder das Verständnis des menschlichen Geistes. Vielfach wird der Computer als Modell für die Funktionsweise des menschlichen Geistes betrachtet.

Der Versuch, künstliche Systeme mit moralischen Fähigkeiten zu konstruieren, ist deshalb mit der Erwartung verbunden, auch besser zu verstehen, wie moralische Fähigkeiten bei Menschen funktionieren könnten (Misselhorn 2019a). Idealerweise gibt es grundlegende funktionale Strukturen moralischer Fähigkeiten, die sowohl in natürlichen als auch in künstlichen Systemen realisiert werden können. Doch auch wenn gewisse Erklärungsansätze moralischer Fähigkeiten an der Implementation scheitern, birgt dies zumindest einen negativen Erkenntniswert. Die Maschinenethik kann also ein wertvolles Instrument kognitionswissenschaftlicher Erkenntnis darstellen.


So wird oft darauf hingewiesen, dass ein Computer im Unterschied zum menschlichen Geist nicht über Denken oder Bewusstsein verfügen kann (Searle 1980 [1986]). Jeder Versuch, eine starke künstliche Intelligenz zu entwickeln, die der menschlichen Intelligenz entspricht, sei deshalb zum Scheitern verurteilt. Folglich werden Maschinen auch niemals über die Fähigkeit zum moralischen Entscheiden und Handeln verfügen.

Die Maschinenethik muss jedoch nicht unbedingt mit dem Anspruch starker künstlicher Intelligenz verbunden sein. Für Anwendungszwecke würde es genügen, eine funktionale Moral zu entwickeln. Dazu müssten Maschinen lediglich über die entspre-



chenden moralischen Informationsverarbeitungsprozesse verfügen. Sie müssten die moralisch relevanten Merkmale einer Situation erkennen und nach entsprechenden moralischen Vorgaben verarbeiten können.

Moralische Informationsverarbeitung in diesem Sinn könnte auch ohne Bewusstsein oder eine dem Menschen vergleichbare Denkfähigkeit möglich sein. In diesem Fall wären Maschinen natürlich keine vollumfänglichen moralischen Akteure wie Menschen. Ihnen würden dazu notwendige Fähigkeiten wie Bewusstsein, die Bezugnahme auf die Welt (Intentionalität), die Fähigkeit zur Selbstreflexion und Moralbegründung und damit auch Willensfreiheit fehlen (Misselhorn 2018). Deshalb könnten Maschinen mit funktionaler Moral zwar moralisch handeln, aber keine Verantwortung für ihr Tun übernehmen.

 Maschinen mit funktionaler Moral könnten zwar moralisch handeln, aber keine Verantwortung für ihr Tun übernehmen

Nehmen wir einmal an, es wäre möglich, Maschinen (zumindest im funktionalen Sinn) mit moralischen Fähigkeiten auszustatten. Dann stellt sich immer noch die Frage, ob es denn aus ethischer Sicht wünschenswert ist, es auch zu tun. Der Maschinenethik wohnt der Zwang inne, in bestimmten Fällen verbindliche moralische Entscheidungen zu treffen, die wir bislang offen gehalten haben. Das kann man auch negativ sehen, denn dadurch werden vielleicht

kritische Aspekte eliminiert, ohne dass dies der Komplexität und existenziellen Bedeutung moralischer Situationen im Alltag gerecht wird.

Das lässt sich an den Dilemmasituationen veranschaulichen, die beim autonomen Fahren auftreten können. Wie soll ein Fahrzeug entscheiden, wenn es ausschließlich die beiden Handlungsalternativen hat, das Leben seiner Fahrgäste aufs Spiel zu setzen oder dasjenige von auf der Straße spielenden Kindern? Der Zwang zu einer Entscheidung ex ante erscheint in einem solchen Fall als problematisch.


Menschen können in einer solchen Lage situationsabhängig entscheiden. Das Verhalten eines autonomen Systems ist hingegen im Vorhinein festgelegt. Dadurch beschränken wir unseren Entscheidungsspielraum und die Möglichkeit, situativ von einer vorhergehenden moralischen Einschätzung abzuweichen, die uns in einer konkreten Situation nicht mehr angemessen erscheint (vgl. dazu im Hinblick auf autonome Waffensysteme Leveringhaus 2016).

Während es in diesem Beispiel eher um die Frage geht, ob Maschinen moralisch angemessen auf kritische Situationen reagieren können, richtet der folgende Einwand den Blick auf den Entscheidenden und fragt, ob man moralische Entscheidungen überhaupt abgeben darf. Nach Kant ist die Fähigkeit zum moralischen Handeln die Wurzel der menschlichen Würde. Daran anschließend kann man zu bedenken geben, dass wir gerade dasjenige aus der Hand geben, was uns als Menschen ausmacht, wenn wir moralische Entscheidungen an Maschinen delegieren.

möchte deshalb nun drei Grundsätze formulieren, die den aus meiner Sicht wesentlichsten Einwänden gegen moralische Maschinen Rechnung tragen.

Erster Grundsatz: Moralische Maschinen sollten die Selbstbestimmung von Menschen fördern und sie nicht beeinträchtigen.

Maschinen sollten die Selbstbestimmung von Menschen nicht untergraben, sondern sie in ihrem selbstbestimmten Handeln unterstützen. Ich habe diese Idealvorstellung im Bereich der häuslichen Pflege ganz praktisch als Grundlage genommen, um ein konzeptuelles Design für ein Pflegesystem vorzulegen, das sich durch Trai-

 Maschinen sollten die Selbstbestimmung von Menschen nicht untergraben, sondern sie in ihrem selbstbestimmten Handeln unterstützen

ning und die permanente Interaktion mit dem Nutzer auf dessen moralische Wertvorstellungen einstellen und Menschen nach ihren eigenen Moralvorstellungen behandeln kann (Misselhorn 2020).

Die Idee war, ein System zu entwickeln, das mit einem verlängerten moralischen Arm des Nutzers vergleichbar ist, der es diesem ermöglicht, länger selbstbestimmt in seinen vier Wänden zu leben, wenn er dies möchte. Man könnte sogar von einem moralischen Avatar sprechen. Die Ansprüche an ein solches System sollten jedoch nicht überzogen sein, denn Technologien allein werden den Pflegenotstand nicht lösen. Auch die sozialen und gesellschaftlichen Rahmenbedingungen dürfen nicht vernachlässigt werden. So sollte niemand gegen seinen Willen von Robotern gepflegt werden. Der Einsatz von Pflegesystemen darf auch nicht zur Vereinsamung und sozialen Isolation der Gepflegten führen.

Zweiter Grundsatz: Künstliche Systeme sollten nicht über Leben und Tod von Menschen entscheiden.

Sehr kritisch vom moralischen Standpunkt ist die Frage einzuschätzen, ob Maschinen über Leben und Tod von

4. Grundsätze der Maschinenethik

Es wäre jedoch falsch, das Projekt der *Artificial Morality* aufgrund der Einwände in Bausch und Bogen zu verwerfen. Man sollte sie stattdessen als Hinweise lesen, wie man gute moralische Maschinen gestalten kann. Ich



Menschen entscheiden dürfen. Ein wichtiges Argument ist in diesem Zusammenhang, dass in den Anwendungsbereichen, in denen über den Einsatz autonomer Systeme nachgedacht wird, keine moralische Pflicht zum Töten besteht (Misselhorn 2018). Eine solche Pflicht gibt es nicht einmal im Krieg.

Der üblichen Auslegung der Theorie des gerechten Kriegs zufolge ist es bestenfalls moralisch erlaubt, andere Menschen im Krieg zu töten, aber nicht moralisch geboten (Childress 1997; Eser 2011). Deshalb sollte immer die Möglichkeit bestehen, etwa aus Mitleid von einer Tötungshandlung abzusehen. Durch den Einsatz autonomer Waffensysteme wird der Entscheidungsspielraum unweigerlich geschlossen.

Eine wichtige Frage ist, ob die Einwände gegen Kriegeroboter sich auch auf andere Anwendungsbereiche übertragen lassen. So wurde eine Analogie zwischen der Programmierung autonomer Fahrzeuge zum Zweck der Unfalloptimierung und der Zielbestimmung autonomer Waffensysteme hergestellt (Lin 2016).

Um Unfallergebnisse zu optimieren, müssen Kosten-Funktionen erstellt werden, die bestimmen, wer im Zweifelsfall verletzt und getötet wird. Vergleichbar mit autonomen Waffensystemen müssten somit für den Fall einer unvermeidlichen Kollision legitime Ziele festgelegt werden, die dann vorsätzlich verletzt oder womöglich sogar getötet würden.

Um zu prüfen, ob sich das Argument, dass es keine moralische Pflicht zum Töten gibt, auf das autonome Fahren übertragen lässt, ist zu klären, ob eine moralische Pflicht besteht, unschuldige Menschen zu verletzen oder zu töten, sofern dies dazu dient, Schlimmeres zu verhindern. Eine solche Pflicht ist nicht nur moralisch problematisch (Misselhorn 2018), sondern auch im Licht der deutschen Rechtsprechung.

Das Bundesverfassungsgericht hat in seiner Entscheidung zum Luftsi-

cherheitsgesetz im Jahr 2006 zum Abschuss entführter Passagierflugzeuge, die von Terroristen als Massenvernichtungswaffen eingesetzt werden sollen, festgehalten, dass ein Abschuss immer der Menschenwürde der Flugzeugpassagiere widerspricht (BVerfGE 115, 118, (160)).


Unschuldige Menschen auf der Grundlage einer gesetzlichen Ermächtigung vorsätzlich zu töten wird somit vom Grundgesetz ausgeschlossen. Dieses Urteil steht zumindest auf den ersten Blick in einem Widerspruch zu einer Pflicht der Schadensminimierung, die die vorsätzliche Verletzung oder Tötung unschuldiger Menschen umfasst.

Dritter Grundsatz: Es muss sichergestellt werden, dass Menschen stets in einem substantiellen Sinn die Verantwortung übernehmen.

Ein weiteres Problem, das angesprochen werden muss, besteht darin, dass der Einsatz von Maschinen mit moralischen Fähigkeiten zu einem Verantwortungsvakuum führen kann. Maschinen sind zwar nicht zu vollumfänglichem moralischen Handeln in der Lage, wie es Menschen auszeichnet. Der Einsatz von Maschinen könnte jedoch die Verantwortungszuschreibung an Menschen unterminieren, so dass am Ende möglicherweise niemand für ihr Handeln die Verantwortung trägt.

Kriterien für die Zuschreibung moralischer Verantwortung sind etwa Willensfreiheit, Kausalität, Absichtlichkeit und Wissen. Ein Handelnder ist demnach nur dann für eine Handlung verantwortlich, wenn sie auf seinem freien Willen beruht, wenn sie ohne seine Beteiligung nicht zustande gekommen wäre, er sie absichtlich durchgeführt hat (oder ihre Folgen zumindest in Kauf genommen hat) und ihm ihre Konsequenzen bekannt waren (er diese hätte vorhersehen oder sich die entsprechenden Kenntnisse mit vertretbarem Aufwand beschaffen können).

Es ist klar, dass Maschinen diese Bedingungen nicht alle erfüllen. So besitzen sie keinen freien Willen; aber auch

 Maschinen können keine moralische Verantwortung tragen, aber eine Verantwortungslücke erzeugen

die Bedingungen der Absichtlichkeit und des Wissens werfen Probleme bei der Zuschreibung an Maschinen auf. Deshalb können sie zwar keine moralische Verantwortung tragen, aber eine Verantwortungslücke erzeugen.

Der australische Maschinenethiker Robert Sparrow, auf den dieser Begriff (im Original: *responsibility gap*) zurückgeht, argumentiert dafür am Beispiel autonomer Kriegeroboter. Er legt dar, dass eine Verantwortungslücke entsteht, wenn:

- (1) ein Kriegeroboter nicht absichtlich so programmiert wurde, dass er die ethischen bzw. rechtlichen Normen der Kriegsführung verletzt;
- (2) es nicht vorhersehbar war, dass der Einsatz des Kriegeroboters dazu führen würde; und
- (3) ab dem Start der Operation keine menschliche Kontrolle mehr über die Maschine bestand.

Wenn diese drei Bedingungen erfüllt sind, hat das nach Sparrow zur Folge, dass die moralische Verantwortung keinem Menschen zugeschrieben werden kann, wenn etwa ein Kriegeroboter im Widerstreit mit den ethischen bzw. rechtlichen Normen der Kriegsführung Menschen tötet. Denn kein Mensch hatte dies beabsichtigt, es war nicht vorhersehbar und niemand hatte kausal die Möglichkeit, dieses Resultat zu verhindern.

Eine Verantwortungslücke entsteht also genau dann, wenn die Maschine selbst nicht verantwortlich ist, ihr Einsatz die Bedingungen der Verantwortungszuschreibung an Menschen untergräbt. Für Sparrow ist dies ein Grund dafür, den Einsatz von Kriegerobotern als unmoralisch abzulehnen. Das Argument von der Verantwortungslücke ließe sich aber grundsätzlich auch auf an-

dere Bereiche wie das autonome Fahren übertragen.

Diese Schlussfolgerung könnte man zum Anlass nehmen, um zu fordern, der Mensch dürfe eben nicht gänzlich die Kontrolle aus der Hand geben. In militärischen Kontexten wird zwischen *In-the-Loop*-Systemen, *On-the-Loop*-Systemen und *Out-of-the-Loop*-Systemen unterschieden, je nachdem, welche Rolle der Mensch in der Kontrollschleife spielt. Bei *In-the-Loop*-Systemen bedient ein Mensch das System und fällt sämtliche Entscheidungen, wenn auch vielleicht nur per Fernbedienung. *On-the-Loop*-Systeme sind zwar programmiert, sie können aber in Echtzeit unabhängig von menschlichem Eingreifen operieren. Der Mensch übernimmt jedoch weiterhin die Überwachung, und er hat jederzeit die Möglichkeit, einzugreifen. *Out-of-the-Loop*-Systeme verhalten sich wie *On-the-Loop*-Systeme, doch besteht keine mensch-

liche Kontroll- und Interventionsmöglichkeit mehr.

Das Problem der Verantwortungslücke erscheint als gelöst, wenn der Mensch *On-the-Loop* bleibt und vielleicht sogar per Knopfdruck der Verantwortungübernahme zustimmen muss, bevor er ein künstliches System in Betrieb nimmt. Es ist jedoch fraglich, ob die Annahme realistisch ist, dass der Mensch zu einer permanenten Überwachung in der Lage ist. Lässt sich die Aufmerksamkeit über einen entsprechend langen Zeitraum halten, ohne selbst tätig zu werden? Kann ein Mensch in Sekundenschnelle entscheiden und eingreifen, wenn es darauf ankommt? Sollte das nicht möglich sein, wären Vorhersehbarkeit und Kontrolle zwar in der Theorie möglich, aber in der Realität für den Menschen nicht umsetzbar.

Des Weiteren kommt es zu spezifischen epistemischen Problemen.

Schließlich ist der Mensch zur Analyse der Situation auf die Informationen angewiesen, die das System ihm liefert. Die Frage ist, ob er diese überhaupt rational in Zweifel ziehen kann, ohne über einen unabhängigen Zugang zu den relevanten Informationen zu verfügen. Nicht zuletzt durchläuft ein solches System bei seiner Entwicklung eine Reihe von Qualitätssicherungsprozessen. Das bestärkt den Nutzer darin, die Vorschläge des Systems den eigenen Urteilen für überlegen zu halten. Die Bedingung der Vorhersehbarkeit wäre somit entgegen dem ersten Anschein nicht wirklich erfüllt.

Insgesamt erscheint es als unfair, dem Nutzer per Knopfdruck die volle Verantwortung aufzubürden, denn zumindest ein Teil der Verantwortung, wenn nicht sogar der Hauptteil sollte doch den Programmierern zukommen, deren Algorithmen ausschlaggebend für das Handeln des Systems sind.

LITERATUR

- Allen, Colin; Wallach, Wendell; Iva Smit (2011): Why Machine Ethics? In: Anderson, Michael; Anderson, Susan [eds.]: *Machine Ethics*, New York: Cambridge University Press 2011, 51–61.
- Arkin, Ronald (2009): Governing lethal behavior in autonomous robots. Boca Raton.
- Bendel, Oliver (2017): Ladybird – The Animal-Friendly Robot Vacuum Cleaner. In: *The AAAI 2017 Spring Symposium on Artificial Intelligence for the Social Good Technical Report SS-17-01*. Palo Alto, 2–6.
- Childress, James F. (1997): Nonviolent Resistance – Trust and Risk-Taking. Twenty-Five Years Later. In: *Journal of Religious Ethics* 25 (2), 213–220.
- Eser, Albin (2011): Tötung im Krieg – Rückfragen an das Staats- und Völkerrecht. In: Ivo Appel und Georg Hermes und Christoph Schönberger (Hg.): *Öffentliches Recht im offenen Staat – Festschrift für Rainer Wahl zum 70. Geburtstag*. Berlin, S. 665–687.
- Lin, Patrick: Why Ethics Matters for Autonomous Cars. In: Maurer, Markus; Gerdes, Chris; Lenz, Barbara; Winner, Hermann [eds.]: *Autonomous Driving – Technical, Legal and Social Aspects*, Berlin/Heidelberg: Springer 2016, 69–85.
- Leveringhaus, Alex (2016): *Ethics and Autonomous Weapons*. Oxford: Palgrave Macmillan 2016.
- Misselhorn, Catrin (2018): *Grundfragen der Maschinenethik*, Dietzinger: Reclam, 2019³.
- Misselhorn, Catrin (2019a): Mensch und Maschine. Leonardo da Vinci als Vorbild für die gegenwärtige Roboterethik. In: Seidl, Ernst; Dürr, Frank; La Corte, Michael [Hrsg.] *Ex machina. Leonardo da Vincis Maschinen zwischen Wissenschaft und Kunst*, Tübingen: MuT 2019.
- Misselhorn, Catrin (2019b): Digitale Rechtssubjekte, Handlungsfähigkeit und Verantwortung aus philosophischer Sicht, Verf Blog, 2019/10/02, <https://verfassungsblog.de/digitale-rechts-subjekte-handlungsfahigkeit-und-verantwortung-aus-philosophischer-sicht/>
- Artificial Systems with Moral Capacities? A Research Design and its Implementation in a Geriatric Care System, in: *Artificial Intelligence* (278), January 2020, 103179. Doi.org/10.1016/j.artint.2019.103179.
- Searle, John R. (1980) [1986]: Minds, Brains, and Programs. In: *The Behavioral and Brain Sciences* 3 (1980) Nr. 3. S. 417–424. [Dt.: Geist, Gehirn, Programm. In: Douglas R. Hofstadter / Daniel C. Dennett (Hg.): *Einsicht ins Ich*. Stuttgart 1986, S. 337–356.]
- Sparrow, Robert (2007): Killer Robots. In: *Journal of Applied Philosophy* 24 (1), 62–77.
- Teubner, Gunther (2018): Digitale Rechtssubjekte? In: *Archiv für die zivilistische Praxis* 218, 151–691.



Die Nutzer sind nur in einem schwächeren Sinn verantwortlich, weil sie das System nicht am Handeln gehindert haben. Das lässt es zweifelhaft erscheinen, ob die Bedingungen der Vorhersehbarkeit und der Kontrolle erfüllt sind. Das Problem der Verantwortungslücke droht aus diesem Grund auch bei *On-the-Loop*-Systemen. Es stellt sich letztlich sogar dann, wenn der Mensch *In-the-Loop* bleibt.

Es ist deshalb eine der großen Herausforderungen der Maschinenethik, Wege zu finden, wie Verantwortungszuschreibung in einem substantiellen Sinn in einem solchen Kontext aufgefasst werden kann. Ein Ansatz besteht darin, Handlungsfähigkeit und Verantwortung zu trennen. Juristisch gesprochen wären moralische Maschinen dann als eine Art von Erfüllungsgehilfen zu sehen, wobei die Verantwortung beim Geschäftsherrn verbleibt (Teubner 2018; Misselhorn 2019b). Allerdings ist ein solcher Ansatz nur für den privatrechtlichen Bereich geeignet.

5. Konklusion

Im Licht der drei Grundsätze guter Maschinenethik gibt es einige Einsatzbereiche moralischer Maschinen, die kritisch zu sehen sind. Das betrifft insbesondere Kriegeroboter, aber auch das autonome Fahren sollte vor diesem Hintergrund nicht zu leichtfertig forciert werden.

Wichtig wäre es, zunächst die Möglichkeiten des assistierten Fahrens voll auszuschöpfen. Denn das assistierte

ZUR AUTORIN

Catrin Misselhorn, geb. 1970, ist seit 2019 Professorin für Philosophie an der Universität Göttingen. Von 2012–2019 war sie Inhaberin des Lehrstuhls für Wissenschaftstheorie und Technikphilosophie an der Universität Stuttgart. Zuvor lehrte sie an der Universität Zürich, der Humboldt-Universität zu Berlin sowie der Universität Tübingen, wo sie 2003 promoviert und 2010 habilitiert wurde. 2007–2008 war sie als Feodor-Lynen-Stipendiatin am Center of Affective Sciences in Genf sowie am Collège de France und am Institut Jean Nicod für Kognitionswissenschaften in Paris. Ihre Forschungsgebiete sind philosophischen Problemen der KI, Roboter- und Maschinenethik.

Sie leitet eine Reihe von Drittmittelprojekten zur ethischen Bewertung von Assistenzsystemen in unterschiedlichen Bereichen, z. B. in der Pflege, in der Arbeitswelt und in der Bildung.

Sie ist Verfasserin einer Vielzahl internationaler Aufsätze in diesem Themenfeld sowie Herausgeberin mehrerer einschlägiger Sammelbände und hat die folgenden beiden Monographien veröffentlicht:

- Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern & Co (Reclam-Verlag 2021).
- Grundfragen der Maschinenethik (Reclam Verlag 2018, 4. Auflage 2020)

Generell spricht auch das Problem der Verantwortungslücke dafür, Maschinen keine Entscheidungen über Leben und Tod von Menschen zu überlassen.

Fahren ist moralisch deutlich weniger problematisch, da es Maschinen keine Tötungsentscheidungen überträgt. In punkto Verkehrssicherheit wäre zunächst zu prüfen, ob das assistierte Fahren nicht womöglich annähernd ebenso effektiv ist wie das vollautomatisierte Fahren.

Trotz dieser Einschränkungen wäre es falsch, die Grundsätze als Hemmschuh der technologischen Entwick-

lung aufzufassen. Vielmehr handelt es sich um Leitplanken, die festlegen können, worin sinnvolle technologische Innovation besteht. So wurde am Beispiel eines Pflegesystems gezeigt, dass sich die Leitlinien auch als Orientierungshilfe bei der Entwicklung nutzen lassen. Hierbei darf allerdings die entsprechende soziale und gesellschaftliche Einbettung solcher Technologien nicht vernachlässigt werden.

Erstveröffentlichung des Beitrags unter dem Titel „Grundsätze der Maschinenethik“ in: **Künstliche Intelligenz – Die große Verheißung**, hg. von Anna Strasser, Wolfgang Sohst, Katja Stepec und Ralf Stapelfeldt, Xenomoi Verlag Berlin 2021, 489–502.



„Maschinen können keine moralische Verantwortung tragen.“

Oliver Bendel verortet sich – so ist auf seiner Website zu lesen – in der Technikphilosophie, wobei er auf Roboter und Künstliche Intelligenz fokussiert. Er untersucht das Verhältnis zwischen Mensch bzw. Tier und Maschine und fragt danach, wie die Maschine der Gegenwart und Zukunft beschaffen ist, sein wird und soll. Zu seinen wichtigsten Disziplinen sind seit der Jahrtausendwende Informationsethik, Maschinenethik und Roboterethik (allgemeiner: Roboterphilosophie) geworden. Als studierter Philosoph und promovierter Wirtschaftsinformatiker denkt er nicht nur über Maschinen nach, sondern entwickelt sie zusammen mit seinen wechselnden Teams auch. Im Rahmen der Maschinenethik und der Sozialen Robotik sind Chatbots, Sprachassistenten und tierfreundliche Maschinen entstanden, die auf internationalen Konferenzen und in internationalen Journals beschrieben wurden.



Oliver Bendel

Amosinternational: Können wir Maschinen beibringen, zwischen Gut und Böse zu unterscheiden?

Oliver Bendel: Wir können autonomen Maschinen moralische Regeln geben, an die sie sich halten. Es entstehen dabei sogenannte moralische Maschinen. Diese haben kein Bewusstsein, keinen freien Willen, keine Intuition und keine Emotionen. Sie verstehen nicht wirklich, was gut und böse ist. Aber sie können im besten Falle gute und richtige Entscheidungen treffen.

Wir bauen u. a. tierfreundliche Maschinen. Normale Maschinen richten immer wieder Unheil an, verletzen und töten – man denke an Windkraftanlagen, Mähmaschinen und Rasenmäherroboter. Das kann man mit Hilfe der Maschinenethik, also der Disziplin, die moralische Maschinen und künst-

liche Moral erforscht und entwickelt, vermeiden.

Sollten Maschinen moralische Entscheidungen treffen dürfen, z. B. wenn ein autonom fahrendes Auto entscheiden muss, ob es ein Schulkind oder eine Gruppe älterer Menschen überfährt?

Wir sollten autonomen Maschinen moralische Regeln beibringen, allerdings nur bestimmten autonomen Maschinen und in bestimmten Zusammenhängen. Geschlossene und halboffene Welten eignen sich gut dafür, etwa der Haushalt oder die Landwirtschaft. Der Straßenverkehr ist problematisch. Das automatische oder autonome Auto sollte weder quantifizieren noch qualifizieren, also weder mögliche Opfer durchzählen noch mögliche Opfer nach Geschlecht, Alter etc. einordnen und be-

urteilen. Wir haben 2012 eine Formel entwickelt, die genau das kann – und vom Gebrauch abgeraten.

Bei Tieren sehe ich das anders, und wir haben um 2016 Autos konzipiert, die nicht bloß für große Tiere bremsen, sondern ebenso für kleine. Selbstverständlich einzig und allein dann, wenn die Verkehrssituation dies zulässt. Auch Ausweichen und Hochfahren des Unterbodens und der Karosserie können Mittel sein. Die Automobilindustrie war wenig begeistert. Getötete Tiere gelten dort als Kollateralschaden. Dennoch hatten die Hersteller für eine Weile die Ethik entdeckt. Vor zehn Jahren wollten sie nichts davon wissen. Ab 2016 häuften sich die Einladungen.

In den USA benutzen Gerichte Künstliche Intelligenzen, um zum Beispiel darüber zu entscheiden, ob jemand Bewährung be-

kommt. Wer ist verantwortlich, wenn eine Künstliche Intelligenz eine Entscheidung trifft?

Systeme und Maschinen können keine moralische Verantwortung tragen. Man muss also immer nach Menschen suchen, die die Verantwortung übernehmen. Das ist bei KI-Systemen und autonomen Maschinen aber beliebig schwierig, da viele Personen in unterschiedlicher Weise beteiligt sein können. Es sind zuweilen nicht einmal alle namentlich bekannt. Zudem handelt es sich häufig um vernetzte Systeme. So wird es noch komplexer.



Systeme und Maschinen können keine moralische Verantwortung tragen

In bestimmten Zusammenhängen sollte man es vermeiden, künstliche Intelligenz und autonome Systeme einzusetzen. Zumindest muss immer ein Mensch in der Schleife sein, der das letzte Wort hat und im Zweifel die Verantwortung übernimmt. Das kann freilich, wie angedeutet, ein Zuviel an Verantwortung sein, eine Überforderung und Überlastung des Einzelnen.

Haben KI-Systeme und Roboter moralische Rechte und Pflichten?

Moralische Rechte sind nach einem weit verbreiteten Ansatz mit Empfindungs- und Leidensfähigkeit verbunden, allenfalls mit einem Bewusstsein. Insofern können KI-Systeme und Roboter keine moralischen Rechte haben. Tiere dagegen schon, wie schon Jeremy Bentham mit eben dieser Argumentation betont hat. Pflichten im engeren Sinne können KI-Systeme und Roboter genauso wenig haben. Es ist aber möglich, ihnen Verpflichtungen aufzuerlegen, oder noch schwächer ausgedrückt, man gibt ihnen Aufgaben, die sie erledigen müssen. Eben das machen wir in der Maschinenethik, und natürlich generell in Robotik und Sozialer Robotik, wobei dort moralische

Implikationen nicht zwangsläufig vorhanden sind.

Werden Maschinen die Menschen von (moralischen) Entscheidungen entlasten?

Autonome Maschinen können Stellvertretermaschinen mit einer Stellvertretermoral sein. Man will ja zum Beispiel, dass sich der eigene Rasenmäherroboter in moralischer Hinsicht so verhält wie man selbst. Dazu gehört, dass er keine Igel tötet. Man pflanzt ihm eine entsprechende moralische Regel ein, setzt diese technisch um und hat das gewünschte Ergebnis. Autonome Maschinen glänzen durch die Abwesenheit ihrer Besitzer und Betreiber. Das gehört zu ihrer Definition. Damit sie auch in dieser Abwesenheit glänzen, muss man sie zähmen und dressieren, muss man sie mit Möglichkeiten und Beschränkungen ausstatten, sodass sie zu vertrauenswürdigen und ehrenwerten Stellvertretern werden.

Man könnte die Gefahr sehen, dass der Mensch verlernt, moralische Entscheidungen zu treffen. Allerdings geht es derzeit um ein paar wenige autonome Maschinen. Und diese könnten uns im Gegenteil ebenso unsere moralischen Entscheidungen – bzw. die Entscheidungen der Hersteller und unsere Kaufentscheidungen – vor Augen führen. Wollen wir einen normalen Rasenmäherroboter oder einen tierfreundlichen? Wir müssten uns verstärkt mit solchen Fragen auseinandersetzen. Vielleicht sind wir dann in bewusster Weise gut. Oder böse.

Was denken Sie über den Einsatz von Kriegerobotern?

Man unterscheidet unterschiedliche Formen von Militärrobotern. Autonome Kampfroboter halte ich für problematisch. Sie würden Entscheidungen treffen, die ihnen nicht zustehen. Im Moment gibt es lediglich Prototypen und einfache Systeme, die etwa selbstständig auf bewegte Objekte feuern. Aber in Zukunft werden autono-

me Kampfroboter auftauchen, zu Lande und in der Luft, die die einen töten und die anderen verschonen – und wie im Falle von autonomen Autos sollten sie das nicht tun.



Ich lehne autonome Kampfroboter aus mehreren Gründen ab

Mein Kollege Ronald C. Arkin hat für das Pentagon autonome Kampfmaschinen mit künstlichem Gewissen konzipiert. Er ist ein hervorragender Maschinenethiker. Wir haben uns mehrmals getroffen und in Krakau auf dem Podium darüber diskutiert, ob man betrügerische Maschinen bauen darf, was er ebenfalls macht. Ich habe das auch gemacht, wie im Falle des LIEBOT, der systematisch lügen kann. Wir waren beide der Ansicht, dass wir das dürfen, hatten jedoch ganz unterschiedliche Gründe.

Doch zurück zu den Kampfrobotern: Arkin und Co. argumentieren, dass Staaten und Soldaten immer wieder das Kriegsvölkerrecht brechen und seine Maschinen dies nicht tun würden. Das mag schon sein, doch natürlich kann man Maschinen bauen, die genau das tun. Ich lehne autonome Kampfroboter aus mehreren Gründen ab. Einer von ihnen ist, dass sie mit psychologischem Terror verbunden wären. Menschen kann man unter Umständen einschätzen, selbst wenn diese völlig verroht sind. Bei Maschinen kann das anders sein, vor allem dann, wenn sie dazulernen.

In welchen zwischenmenschlichen Bereichen werden Roboter heute eingesetzt?

Es existieren Serviceroboter wie Transportroboter, Reinigungsroboter und Sicherheitsroboter, die manchmal oder sogar regelmäßig in die Nähe von Menschen kommen. Soziale Roboter finden sich etwa in Pflege und Therapie, in Bibliotheken und Museen, in Schulen und Hochschulen sowie in Shopping Malls. Und natürlich in Haushal-




ten. Auch Sexroboter, wobei es sich im Grunde um Roboterköpfe mit Puppenkörpern handelt. Viel verbreiteter sind Liebespuppen ohne oder mit sehr eingeschränkten technischen Möglichkeiten. In Haushalten wie in Bordellen sind sie Objekte der Begierde. Neben den bekannten Stereotypen locken Fantasyfiguren wie Mangamädchen und Elfen.

Manche Roboter werden als Medien eingesetzt, wie Telepräsenzroboter. Man sieht den anderen auf dem Display, das Display hat eine mobile Plattform und rollt umher. Edward Snowden hat auf diese Weise das eine oder andere Interview geführt, ohne Russland verlassen zu müssen. Auch Hugvie ist ein Medium: Es handelt sich um eine Umarmungspuppe, an deren Kopf eine Tasche ist, in die man ein Smartphone schieben kann. Man telefoniert mit seinen Liebsten und drückt dabei Hugvie an sich – und es fühlt sich fast so an, als würden die Liebsten einen umarmen.

Was denken Sie über den Einsatz von Pflegerobotern in der Altenpflege?

Pflegeroboter können Pflegepersonal unterstützen und entlasten. Sie können die persönliche Autonomie von Pflegebedürftigen stärken und zugleich ihre informationelle Autonomie schwächen, indem sie in Privat- und Intimsphäre eindringen, Bilder anfertigen und Daten sammeln sowie diese weitergeben.

 **Roboter in der Pflege können kaum den Pflegenotstand beheben; das können nur mehr Anerkennung und bessere Bezahlung**

Wir machen die Erfahrung, dass Pflegebedürftige solche Roboter in ihr Leben lassen, sobald sie erkennen, dass diese einen Mehrwert bieten. Das ist freilich nicht bei allen Modellen und nicht bei allen Tätigkeiten der Fall. Roboter in der Pflege können kaum den

Pflegenotstand beheben. Das können nur mehr Anerkennung und bessere Bezahlung.

Kann man bei der Interaktion von Mensch und Maschine von wirklicher Kommunikation sprechen?

Chatbots, Sprachassistenten und viele soziale Roboter haben natürlichsprachliche Fähigkeiten. Man kann mit ihnen stundenlange Gespräche führen. Natürlich versteht die Maschine letztlich nicht, was man sagt, und sie interessiert sich nicht wirklich für einen. Man spricht in gewisser Weise gegen eine Wand. Das macht man bei einigen Menschen freilich auch. Und Maschinen können einen größeren Wortschatz haben als bestimmte Menschen und einem auf diese Weise – in intellektueller und ästhetischer Hinsicht – gefallen.

Können Roboter in uns Gefühle auslösen und können wir eine emotionale Bindung zu ihnen aufbauen?

Dinge aller Art können Gefühle in uns auslösen, Teddybären und Barbiepuppen, Schmuck und Autos, und ohne Zweifel gewisse Roboter. Bei sozialen Robotern kommt hinzu, dass diese oft menschliche oder tierische Merkmale abbilden, Augen, Münder und Arme haben. Wir reagieren auf ihre Mimik und Gestik ähnlich wie bei Lebewesen. Manche soziale Roboter können unsere Emotionen analysieren und dann entsprechend darauf reagieren. Dies löst dann wiederum Gefühle in uns aus. Emotionale Bindungen können jederzeit entstehen, ob der Roboter uns dabei hilft oder nicht.

Im Film „Her“ verliebt sich der Protagonist in ein Betriebssystem mit künstlicher Intelligenz und weiblicher Stimme. Werden Roboter irgendwann so hochentwickelt sein, dass Menschen sich in humanoide Roboter oder in eine Künstliche Intelligenz verlieben und sich eher für eine Beziehung mit einem Androiden als mit einem normalen Menschen entscheiden?

Es ist in „Her“ die schöne Stimme von Scarlett Johansson. Die Stimme allein reicht aus, damit man Gefühle für jemanden oder etwas entwickelt oder sich erregt. Davon lebt seit Jahrzehnten der Telefonsex. Zuletzt hat er während der Corona-Pandemie einen Aufschwung erlebt.

Schon heute verlieben sich Menschen in Sprachassistenten, Sexpuppen und Sexroboter. Und manche entscheiden sich für eine Beziehung mit ihnen anstatt mit Ihrgleichen. Es sind und bleiben jedoch einseitige Angelegenheiten. Auf der anderen Seite ist nichts, da sind nur Nullen und Einsen, Silikon und Metall. Das kann für eine Weile genügen, für die Dauer aber nicht, eine übliche soziale und sexuelle Veranlagung vorausgesetzt. Eine Ausnahme bildet die Objektsexualität, wo das reine Ding genügt. Die Beziehung bildet sich im Wesentlichen im Kopf. Das Objekt benötigt also keine weitergehenden Fähigkeiten wie Stimme oder Bewegung.

Stars lassen sich als dreidimensionale Hologramme auf Bühnen projizieren. Auch Verstorbene werden als Avatare „reanimiert“. Wird durch die Virtualisierung unserer Lebenswelten auch die Wirklichkeit zur Fiktion?

Es handelt sich um Pseudo- oder Quasihologramme, um eindrucksvolle Projektionen, wie bei den Abbataren der berühmten schwedischen Popband. Manchmal entstehen so auch erst Stars, wie im Falle von Hatsune Miku. Zunächst war sie lediglich eine Stimme, erzeugt mit der Software Vocaloid. Das äußere Erscheinungsbild wurde später entworfen. Für die aufwendige Bühnenshow werden fünf Projektoren und eine Glasscheibe verwendet.

Verstorbene versucht man durch Chatbots, Sprachassistenten oder durch visuelle (virtuelle bzw. physische) Nachbildungen zum Leben zu erwecken. Tatsächlich gehen Realität, Virtualität und Fiktionalität auf diese Weise immer mehr ineinander über. Es



ist wichtig, alles auseinanderzuhalten. Man kann Robotern nicht wehtun, aber Menschen. Wenn man sein verstorbenes Kind nachbaut, wie es in „Better Than Us“ gemacht wurde, hilft einem das vielleicht. Aber was richtet es beim Partner an?

Sind Roboter in der Lage, Dinge, die uns fehlen, zu ersetzen?

Dazu sind sie durchaus in der Lage. Interessant ist das vor allem in außerordentlichen Situationen. Wir bereiten seit einiger Zeit den Bau eines Umarmungsroboters vor, der bei Krisen und Katastrophen hilfreich sein kann. HUGGIE – so sein Name – soll aber keinesfalls die menschliche Umarmung auf Dauer ersetzen. Wir haben bereits einen



Ein Umarmungsroboter soll keinesfalls die menschliche Umarmung auf Dauer ersetzen

Sprachassistenten für einen Marsflug entwickelt, der gegenüber den Astronauten Empathie und Emotionen zeigen und ihnen bei Problemen helfen kann. Sogar die NASA hat Interesse bekundet. Ein Kollege hat mich informell nach Colorado eingeladen. Ich bin aber noch nicht dazu gekommen, dorthin zu reisen.

Dafür bin ich jedes Jahr an der Stanford University und besuche die AAAI Spring Symposia. Es ist wichtig, in der KI-Community mitzumischen. Leider denken viele Philosophen lediglich über KI-Systeme und Roboter nach, ohne sich mit den Kollegen der entsprechenden Disziplinen auszutauschen und ohne die Systeme selbst zu entwickeln oder wenigstens auszuprobieren. Eine junge Wissenschaftlerin hat meine Herangehensweise neulich während der Konferenz Robophilosophy in Helsinki „empirisch informierte Philosophie“ genannt. Das trifft es vermutlich. Wobei diese nicht nur empirisch informiert, sondern auch konstruktiv ist. Wir denken über Maschi-

nen nach und bauen sie. Und denken dann wieder über sie nach.

Ist künstliche Intelligenz eine Reproduktion menschlicher Intelligenz oder wird sie es bald sein?

Die Künstliche Intelligenz steht für einen eigenen wissenschaftlichen Bereich der Informatik, der sich mit dem menschlichen Denk-, Entscheidungs- und Problemlösungsverhalten beschäftigt, um dieses durch computergestützte Verfahren ab- und nachbilden zu können. So habe ich es in verschiedenen Artikeln und Lexika ausgedrückt, in Anlehnung an klassische Definitionen, die noch nicht KI und Machine Learning verwechseln (Machine Learning ist ein Teil der KI, aber nicht deckungsgleich mit ihr). Die künstliche Intelligenz, der Gegenstand und das Ergebnis der Künstlichen Intelligenz, imitiert und simuliert menschliche (oder tierische) Intelligenz. Sie kann auch darüber hinausgehen, bis zu einem Punkt, wo wir uns fragen müssen, ob es sich überhaupt noch um Intelligenz oder ein anderes Phänomen handelt.

Die intelligenten Maschinen sind mittlerweile hochentwickelt – wo ist der Platz des Menschen zwischen diesen hochentwickelten Maschinen? Was macht ihn aus? Wo ist die Grenze zwischen Menschen und Androiden?

Ich plädiere dafür, dass Roboter und KI-Systeme in erster Linie als Werkzeuge erschaffen und genutzt werden. Nur in wenigen Bereichen braucht es soziale Merkmale – diese sind bei Servicerobotern zuweilen hilfreich, da sie die Akzeptanz und die Hilfsbereitschaft fördern – und soziale Roboter im engeren Sinne, etwa in der Therapie. Die Grenze zwischen Androiden, also hochgradig menschenähnlichen Robotern, und Menschen bleibt immer gleich: Maschinen haben kein Bewusstsein, keine Empfindungs- und Leidensfähigkeit. Und keine Glücksfähigkeit. Sie haben keine Empathie und Emo-

tionen. Sie zeigen sie nur, mithilfe ihrer Augen und ihrer Mimik, mittels Tönen und Bewegungen.



Ich plädiere dafür, dass Roboter und KI-Systeme in erster Linie als Werkzeuge erschaffen und genutzt werden

Eine kontroverse Diskussion dreht sich darum, ob wir Roboter eher als Freunde – man spricht gerne von Companions – oder eher als Sklaven bauen sollen. Ich denke, dass wir beides tun können, je nach Zusammenhang. Ich würde mich wiederum dafür aussprechen, dass sie meistens eher Sklaven als Freunde sind, wobei wir den Begriff des Sklaven umdeuten müssen. Es sollte uns keine Gewissensbisse bereiten, wenn wir robotische Sklaven haben. Lassen wir die menschlichen Sklaven hinter uns zugunsten der robotischen.

Wird der Mensch in Zukunft Gott spielen in der Erschaffung Künstlicher Intelligenzen?

Der Mensch spielt immer Mensch. Dazu gehört, dass er sich künstliche Kreaturen ausdenkt und solche schafft. Die Ideengeschichte der künstlichen Kreatur ist uralte, wenn man an Talos und Pandora – von Hephaistos geschaffene Figuren – und Galatea von Pygmalion denkt. Später kamen der Golem und Frankensteins Monster dazu. Pinocchio gehört zu den sympathischen Artefakten.

In Renaissance und Barock wurde aus der Idee eine Entwicklungsgeschichte. Berühmte Automaten und Androiden entstanden, etwa in der Schweiz und in Frankreich. Das ist kein Wunder, denn dort werkten die Uhrenmacher. Und seit fast hundert Jahren kennt man funktionierende Roboter. Am Anfang hat man sie und die künstliche Intelligenz zusammen gedacht. Die Disziplinen der Robotik und der Künstlichen Intelligenz haben sich dann aber über Jahrzehnte separat ent-

wickelt, bis sie vor 20 Jahren wieder zusammentrafen.

Aus diesem Zusammentreffen entstehen nun die erstaunlichsten Maschinen, autonome, intelligente Systeme mit enormen Fähigkeiten. Sie jagen den einen Angst ein, den anderen nötigen sie Respekt ab. Noch haben wir alles in unserer Hand, sie sind unsere Marionetten, und wenn wir die Fäden durchschneiden, fallen sie in sich zusammen. Wie es morgen sein wird, ist nicht ganz klar, doch KI-Systeme, die die Welt aus eigenem Antrieb zerstören, sind und bleiben Science-Fiction.

Was sind die Vor- und Nachteile von Emotionserkennung und Gesichtserkennung?

Gesichtserkennung ist, wie ich in meinem Lexikon geschrieben habe, das automatisierte Erkennen eines Gesichts in der Umwelt bzw. in einem Bild (das bereits vorliegt oder zum Zwecke der Gesichtserkennung erzeugt wird) oder das automatisierte Erkennen, Vermessen und Beschreiben von Merkmalen eines Gesichts, um die Identität einer Person oder deren Geschlecht, Gesundheit, Herkunft, Alter oder sexuelle Ausrichtung festzustellen.

Mit auf Mimikererkennung basierender Emotionserkennung sind z. B. soziale Roboter in der Lage, die Gefühlslage ihres menschlichen Gegenübers zu deuten und adäquat darauf zu reagieren. Besonders gut funktioniert das, wenn man schauspielert. Ansonsten sind die Systeme nicht immer fehlerfrei, und man muss bedenken, dass man gegenüber einem Roboter nicht unbedingt das gleiche Spektrum an Gefühlsregungen zeigt wie gegenüber einem Menschen.

Wir haben vor einiger Zeit einen Chatbot gebaut, den BESTBOT, der die Texteingabe des Benutzers auf Emotionen hin analysierte und zugleich Gesichtserkennung mit Emotionserkennung benutzte, um weitere Anhaltspunkte zu erhalten. Offensichtliche Diskrepanzen hat er dann thematisiert, etwa wenn der Benutzer gesagt

hat, dass es ihm gut geht, sein Gesicht aber Angst zeigte. Wegen der Gefahren, die mit Gesichtserkennung verbunden sind, würden wir einen solchen Chatbot aber nur in bestimmten Settings empfehlen, und wir haben ihn auch nicht aus dem Labor gelassen.

Übrigens halte ich wenig von den erwähnten Versuchen, mit Hilfe von Gesichtserkennung die sexuelle oder politische Ausrichtung auszukundschaften. Zum einen ist das nicht ohne weiteres möglich, nicht ohne dass man Merkmale außerhalb des Gesichts hinzunimmt, zum anderen darf man davon ausgehen, dass die Ziele der potenziellen Betreiber zweifelhaft sind. Das Ganze erinnert an die Pseudowissenschaft der Physiognomik, die von Johann Caspar Lavater aus Zürich ausging.

Gesichtserkennung kann im öffentlichen Raum der Massenüberwachung dienen und sollte für solche Zwecke verboten werden. Auch bei Geräten und Diensten sollte man sie sparsam einsetzen. Zugleich könnte sie bei der Erkennung von Tierindividuen wertvolle Dienste leisten. Daran forschen wir seit einiger Zeit. So haben wir ein System konzipiert, das aus Wildtierkameras, Drohnen und Robotern besteht und das Gesichtserkennung nutzen soll, um Braunbären in den Alpen zu identifizieren. Wohlgermerkt nicht die Art, sondern die Individuen.

Was ist Ihnen noch wichtig zu sagen?

Es ist vermutlich deutlich geworden, dass ich KI-Systeme und Roboter nicht nur mit Blick auf Menschen, sondern auch mit Blick auf Tiere denke. Grob gesagt, mögen Techniker keine Tiere und Tierethiker und -schützer keine Technik. Diesen leeren Raum gilt es zu füllen. Dabei helfen Disziplinen wie Tier-Computer-Interaktion und Tier-Maschine-Interaktion, zudem Maschinenethik und Soziale Robotik, beide nicht in anthropozentrischer Weise verstanden.

Wir sollten Systeme und Maschinen schaffen, die Tieren nützen und Tiere schützen. Dabei sollten wir nicht allein Haushalts- und Landwirtschaftsroboter anpassen, sondern Lebensräume neu denken. Mir schweben erweiterte Reservas vor, wo Pflanzen und Tiere ungestört leben können und nur von KI-Systemen und Robotern beaufsichtigt werden. Diese halten sich sehr bedeckt und greifen nur im Ernstfall ein, etwa wenn ein Ungleichgewicht droht oder Wilderer ihr Unwesen treiben.

Ein anderes wichtiges Thema ist der Cyborg. Human Enhancement und Animal Enhancement sowie Bodyhacking als Do-it-yourself-Bewegung verändern menschliche und tierische Körper. Auch Roboter können Teil davon werden, etwa als Hightech-Prothesen. Theoretisch könnten diese selbst Entscheidungen treffen, und man müsste ihnen moralische Regeln geben, die ihnen z. B. das Würigen von Menschen verbieten, außer im Falle von Notwehr. Das ist Science-Fiction, aber ich wollte nur kurz demonstrieren, dass sich manche Phänomene plötzlich in einem anderen Kontext wiederfinden können.

Zufällig sind wir wieder bei den Tieren gelandet. Während Menschen aus ihrer Freiheit heraus sich für einen Umbau ihres Körpers entscheiden können, wird über die Tiere entschieden. Meistens geht es bei Animal Enhancement nicht um deren Interessen, sondern die der Menschen. Etwa um die Erhöhung der Produktivität. Es entstehen Cyborgs, die das Sklaventum von Nutz- und Arbeitstieren noch unwürdiger gestalten.

Cyborgs sind, um das noch zu erklären, biologische Strukturen, in die technische eingepasst werden. Etwa Menschen mit Chips. Umgekehrte Cyborgs, wie ich sie nenne, sind technische Strukturen, in die biologische Strukturen eingepasst werden. Vielleicht wird man eines Tages Gehirne, die man schon außerhalb von Körpern am Leben erhalten kann, in Roboter einpflanzen. Eine Verlängerung der



Existenz, zu einem hohen Preis. Der Terminator – mit seiner echten Haut über dem Roboterschädel ebenfalls ein umgekehrter Cyborg – lässt grüßen.



Die Aufgabe der Ethik ist es nicht, erschüttert zu sein, sondern Chancen und Risiken aufzuzeigen und mit aller Vorsicht auch Grenzen, die man nicht überschreiten sollte

Wenn man all dies hört und liest, ist man vielleicht erschüttert. Die Aufgabe der Ethik ist es nicht, erschüttert zu sein, sondern Chancen und Risiken aufzuzeigen und mit aller Vorsicht auch Grenzen, die man nicht überschreiten sollte. Ihre Aufgabe ist es auch, die Hypermoral, die in die Welt gekommen ist, kritisch zu betrachten. Sprache und Moral sind zwei unterschiedliche Systeme mit unterschiedlichen Zwecken. Und auch Wissenschaft und Politik sind eigenständige Systeme, die sich nicht nur von moralischen Erwägungen leiten lassen dürfen.

Wenn Sie einen Blick in die Zukunft werfen: Haben Sie ein optimistisches oder pessimistisches Bild zum Zusammenspiel von Mensch und Maschine?

Autonome Maschinen werden eines Tages ein selbstverständlicher Bestandteil verschiedener Bereiche sein. Manche von ihnen werden reine Werkzeuge sein, die man benutzt und wieder weglagt oder beiseiteschiebt, andere werden soziale Merkmale haben und uns an sich binden. Es lässt sich evo-

ZUR PERSON

Oliver Bendel, geb. 1968, ist Professor für Wirtschaftsinformatik und Ethik an der Hochschule für Wirtschaft FHNW. Bendel arbeitete in Deutschland und in der Schweiz als Projektleiter und stand technischen und wissenschaftlichen Einrichtungen vor, u. a. dem Competence Center E-Learning an der Universität St. Gallen, der Arbeitsgruppe Business Communication am Fraunhofer ISST in Dortmund und dem Zentrum für Innovation, Medien und Technologien (ZIMT) an der PH Weingarten, wo er auch als Dozent tätig war, u. a. mit einer Vorlesung zur Ideen- und Entwicklungsgeschichte der künstlichen Kreatur. Im April 2009 wurde er von der Hochschule für Wirtschaft FHNW zum Professor ernannt und am Institut für Wirtschaftsinformatik angestellt. Zudem ist er Dozent an der Hochschule für Technik FHNW und an der Hochschule für Architektur, Bau und Geomatik FHNW.

Oliver Bendel war mehrfach Sachverständiger und Vortragender im Deutschen Bundestag und im Europäischen Parlament sowie ständiges Mitglied der Jury zur Auswahl des Deutschen Internet-Instituts (des heutigen Weizenbaum-Instituts in Berlin). Als Sachverständiger und Gutachter stand bzw. steht er verschiedenen Bundesministerien, Forschungsgemeinschaften und Förderorganisationen zur Verfügung. Oliver Bendel ist Mitglied des Netzwerks Wissenschaftsfreiheit, Mitglied bei der Freidenker-Vereinigung der Schweiz, Mitglied bei Animal Rights Switzerland und Mitglied der Zürcher Kunstgesellschaft. Zudem ist er Mitglied im Stiftungsrat der Daimler und Benz Stiftung. Weitere Informationen über www.oliverbendel.net.

Wichtige Buchveröffentlichungen zum Thema:

- Bendel, Oliver (Hrsg.). Soziale Roboter: Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen. Springer Gabler, Wiesbaden 2021.
- Bendel, Oliver (Hrsg.). Maschinenliebe: Liebespuppen und Sexroboter aus technischer, psychologischer und philosophischer Sicht. Springer Gabler, Wiesbaden 2020.
- Bendel, Oliver (Hrsg.). Handbuch Maschinenethik. Springer VS, Wiesbaden 2019.
- Bendel, Oliver (Hrsg.). Pflegeroboter. Springer Gabler, Wiesbaden 2018.

lutionsbiologisch gut erklären, warum wir so stark auf Augen, Münder, Arme, Beine, auf ganze Körper reagieren. Wir müssen uns nun gegenseitig klar machen, dass Maschinen keine Menschen sind und warum wir sie anders behandeln und verwenden müssen. Ich bin also nicht grundsätzlich dagegen, Roboter als neue soziale Entitäten zu

entwickeln, aber wir müssen zugleich neue Kompetenzen entwickeln, um den Durchblick zu behalten und keine Nachteile zu erleiden. Besonders Kinder und Ältere, also vulnerable Gruppen, sind hier zu berücksichtigen.

Das Interview führte Claudia Schwarz, Dortmund





Kirche neu erfinden

Peter Klasvagt: *Kirche neu erfinden. Lebendiger Organismus – Lernende Organisation*, Paderborn: Bonifatius 2021, 368 S., ISBN 978-3-89710-883-7

Der Wunsch nach Neuerungen innerhalb der katholischen Kirche in Deutschland ist groß, die Desillusionierung an vielen Stellen ebenfalls. Muss die Kirche neu erfunden werden, wie es der Titel von Peter Klasvogts Buch suggeriert? Oder muss sie sich vielmehr selbst neu finden, um wieder für die Zeit passend zu sein? Um Letzteres geht es, wenn Klasvagt den viel diskutierten Ansatz zur Organisationsentwicklung des belgischen Autors und Beraters Frédéric Laloux ins Gespräch mit kirchlichen Organisationsformen bringt.

Wer sich mit neuen Möglichkeiten der sinnstiftenden Zusammenarbeit, mit New Work beschäftigt, kommt an Laloux und seinem vieldiskutierten Buch „Reinventing Organizations“ nicht vorbei. Der Anspruch der Sinnstiftung für die einzelnen Mitarbeitenden in ihrem Arbeitsumfeld legt nahe, Lalouxs Ansätze für die Kirche zu prüfen. Diese Lücke schließt Peter Klasvagt. Dabei geht es dem Verfasser nicht nur darum, ein Managementtool für die Kirche nutzbar zu machen, sondern er möchte Erfahrungen aus zwei Welten miteinander in Dialog bringen, da sich bereits jetzt schon Ansätze evolutionärer Organisationen in der Kirche finden lassen.

Das Dialogische wird bereits im Aufbau des Buches deutlich: Im ersten Teil führt Klasvagt in das neue Organisationsverständnis Lalouxs ein, das Organisationen als sich weiterentwickelnde lebendige Organismen versteht, und ergänzt diesen Ansatz durch andere Forschungen, vor allem aus dem New-Work-Bereich. Dabei sind es die sogenannten „evolutionären Organisationen“, die der Komplexität der



heutigen Welt und dem Wunsch nach sinnstiftender Arbeit besonders gerecht werden. Die Kirche – so macht der Autor im zweiten Teil deutlich – hat selbst in ihrer zweitausendjährigen Geschichte ganz unterschiedliche Stadien und Organisationsformen durchlebt. Damit ist sie ein „organisationaler Sonderfall“, weshalb neue Ansätze, wie der von Laloux, nicht einfach unreflektiert übertragen werden können. Dies gilt insbesondere deshalb, weil die Kirche als Leib Christi mehr ist als ihre irdische Verfasstheit und es darum immer wieder ein Ringen um angemessene Strukturen und Formen von Führung gegeben hat. Im dritten Teil, dem Kernstück des Buches, führt Klasvagt einzelne, konkrete Beispiele (u. a. Einstellung von Personal, Feedback-Kultur) aus und beschreibt jeweils, wie diese in evolutionären Organisationen und in der Kirche gelebt werden. Die Perspektive auf die Kirche ist dabei immer eine, die positive Beispiele in den Vordergrund rückt und mit Grundlagen aus der Soziallehre, den Reden und Schriften von Papst Franzis-

kus und weiterer christlicher Autoren aus Vergangenheit und Gegenwart untermauert wird. Während Laloux selbst die Kirche als typische konformistische Organisation einordnet, verdeutlicht Klasvagt, dass sich auch in ihr schon Ansätze evolutionärer Organisationen finden lassen. Zum Abschluss reflektiert Klasvagt, inwiefern es der Kirche gelingen kann, nicht nur aus dem Bereich der Wirtschaft und Organisationsentwicklung zu lernen, sondern selbst in die Welt hineinzuwirken, und so zu einem Dialogpartner auf Augenhöhe zu werden.

Eine Auseinandersetzung mit der Kritik am Konzept von Laloux findet lediglich in zu knapper Form im Anhang statt, der außerdem noch eine Übersicht über Ansätze gibt, die als Grundlagen Lalouxs Arbeit vorbereitet haben, ergänzen oder stützen und besonders für diejenigen interessant sein dürften, die sich bisher wenig mit Organisationsentwicklung auseinandergesetzt haben. Der insgesamt sehr klare Aufbau des Buches ermöglicht es zudem, einzelne Kapitel oder Abschnitte herauszugreifen oder gerade im dritten Teil später nachzulesen.

Klasvogts Buch verbindet aktuelle Erkenntnisse aus dem Bereich Organisationsentwicklung mit einem tiefen, aus dem Glauben getragenen Verständnis dessen, was Kirche ist und sein kann. Dabei konzentriert sich der Autor vor allem auf die organisationale Ebene und hält sich in Fragen der Zulassung von Frauen zum Weiheamt oder der Besetzung von an die Weihe gebundenen Führungspositionen eher zurück oder verweist auf pragmatische Lösungsansätze. Eine große Stärke des Buches liegt in der Fokussierung auf die Mitarbeitenden kirchlicher Organisationen. Die Perspektive, wie Mitarbeitende sinnstiftend zusammenarbeiten, wie sie gewonnen und gehalten



werden können, wird für die Zukunft der Kirche eine entscheidende sein. Die von Klasvogt im Anschluss an Laloux dargestellten Möglichkeiten der Umsetzung für den kirchlichen Bereich können dabei gute Anregungen geben, insofern sie nicht

nur halbherzig umgesetzt werden. Insgesamt entsteht der Eindruck, dass Klasvogt etwas von der Überzeugungskraft Laloux weitergibt und sie mit seiner eigenen Verbundenheit zur Kirche ergänzt. „Kirche neu erfinden“ ist damit nicht ein-

fach ein weiteres Buch über die Erneuerung der Kirche, sondern eines, das Mut macht und motiviert, Veränderungsprozesse anzustoßen.

Paulina Hauser, Erfurt/Fulda

Migrationsethik

Lukas Schmitt: Von Grenzen, Menschen und Mauern. Migrationsethische Perspektiven in der globalisierten Weltgesellschaft (Freiburger Theologische Studien 198), Freiburg i.Br.: Herder 2022, 712 S., ISBN 978-3-451-39305-1

„Wir schaffen das“, dieser von den einen so bestaunte, von den anderen so verhöhnnte Ausruf der Zuversicht der damaligen Bundeskanzlerin klingt noch immer nach, erinnert man sich an die sogenannte Flüchtlingskrise von 2015. Fragt man danach, warum es denn richtig sein soll, eine menschenwürdige Aufnahme der Flüchtlinge überhaupt schaffen zu wollen, betritt man das Feld der Migrationsethik, auf dem damals beide großen Kirchen in Deutschland in bemerkenswerter ökumenischer Einigkeit stark engagiert waren. Inzwischen stehen indes nicht mehr die Geflüchteten aus Afrika und dem Nahen Osten im Mittelpunkt der Aufmerksamkeit, sondern die vielen Menschen, die jetzt vor dem Krieg aus ihrer Heimat der Ukraine in andere europäische Länder fliehen.

Der Ukrainekrieg in seiner aktuellen Dimension lag jedoch noch in der Zukunft, als Lukas Schmitt seine zu Jahresbeginn erschienene Dissertationsschrift „Von Grenzen, Menschen und Mauern. Migrationsethische Perspektiven in der globalisierten Weltgesellschaft“ veröffentlicht hatte. Darin betreibt er Migrationsethik aus theologisch-ethischer Perspektive in Auseinandersetzung mit einschlägigen philosophischen Theorieansätzen. Diese umfangreiche Studie zielt gemäß ihrer forschungsleitenden Fragestellung im Sinne einer Auslotung migrationsethischer Perspektiven darauf ab,



die Beschaffenheit von Grenzen und ihre Bedeutung für eine sich als ein „Wir“ verstehende politische Gemeinschaft zu reflektieren. Dabei geht es in normativer Hinsicht insbesondere darum, welche Hilfspflichten diese Gemeinschaft angesichts von Migration und Flucht wahrnehmen sollte (vgl. S. 20 f.).

Das genuin christlich-theologische Erkenntnisinteresse des Autors speist sich daraus, dass „die Frage der Offenheit von Grenzen auch [als; L.S.] eine Frage nach der Reichweite christlicher Gebote“ (S. 21) verstanden wird und daher Stellungnahmen und Engagement der Kirchen ein wesentlicher Bezugspunkt dieser Arbeit sind. Die Arbeit gliedert sich in vier Hauptkapitel und ist auf einer übergeordneten Ebene zudem nach dem klassischen methodischen Dreischritt christlicher Sozialethik, *Sehen – Urteilen – Handeln*, aufgebaut (vgl. S. 40).

Im ersten Hauptkapitel geht es um das *Sehen*: Hier werden zum einen Grenzen in ihrer Ambivalenz sowie zum anderen das Phänomen der Migration in seiner Differenziertheit deskriptiv analysiert (Stichwörter sind u. a. Flucht, Willkommenskultur, Walls come tumbling down), wobei ein Schwerpunkt auf das europäische Grenzmanagement gelegt und insbesondere eine zunehmende Fortifizierung von Grenzen (Stichwort: „Festung Europa“) festgestellt wird.

Im zweiten Hauptteil beginnt der Schritt des *Urteilens*, denn hier folgt in normativer Absicht eine eingehende Rekonstruktion der anhand ihrer migrationsethischen Relevanz ausgewählten angloamerikanisch-moralphilosophischen Ansätze von John Rawls, Michael Walzer, Joseph Carens und Seyla Benhabib. Durch die erstmalige Verknüpfung des angloamerikanischen philosophischen Diskurses über Grenzen mit theologisch-migrationsethischen Auseinandersetzungen im deutschen Kontext schließt die Arbeit dem eigenen Anspruch nach eine entsprechende Forschungslücke (S. 29). Bei diesen Analysen geht es um die Frage, wie durchlässig – oder auch nicht – staatliche Grenzen in einer globalisierten Welt sein dürfen. Der Ansatz der kosmopolitischen Philosophin Benhabib wird vom Verfasser dabei als ausgewogene Mittelposition bevorzugt. In Teil drei kommt sodann die theologisch-ethische Warte ins Spiel: In wünschenswerter Gründlichkeit werden hier die migrationsethisches relevanten Stellungnahmen von Papst und Universalkirche sowie der Kirche in Deutschland vorgestellt. Dabei wird auch die Debatte darüber nachgezeichnet, ob die Kirchen hinsichtlich ihres En-

gagements für Flüchtlinge evangeliums- gemäß Parteinahme für die Schwächeren betrieben oder aber unangemessen als politisierende Moralagenturen (Hans Joas) agiert haben.

Hauptteil vier bündelt unter der Perspektive des *Handelns* sodann die Erträge der vorangegangenen Analysen in einer Zusammenschau sowie in „drei Dimensionen kosmopolitischer Verantwortung im Kontext der Flüchtlingsfrage“ (S. 638). Wann schließt man, wann öffnet man Grenzen aus ethischer Sicht? Michael Walzer, Joseph Carens und Seyla Benhabib seien sich darin einig: Offene Grenzen braucht es besonders dann, wenn Geflüchtete in einer Notlage ankommen. Wie lange, wie weit und wie durchlässig die Grenzen aber geöffnet werden, darin unterscheiden sich die drei philosophischen Ansätze. Schmitt geht es auf Basis des von ihm bevorzugten Ansatzes von Benhabib darum, sowohl radikal partikularistische als auch radikal universalistische Gerechtigkeitsvorstellungen durch die Suche nach einem gerechten Ausgleich (vgl. S. 644) zu umgehen.

Eine entsprechend ausgewogene mittlere Position sozialphilosophisch-sozioethisch zu entfalten ist ein ehrenwertes Anliegen angesichts der enormen Polarisierungen, die sich an der Flüchtlingsfrage seit Mitte der 2010er Jahre gesellschaftsspalterisch entzündet haben. Dazu passt, dass der Verfasser das seine ganze Studie durchziehende Grunddilemma, wie sich die ethische und politische Verantwortung gegenüber Menschen diesseits wie jenseits staatlicher Grenzen zueinander verhalten (S. 657), nicht vollends aufzulösen versucht. Hier kann auch nur auf diese Hauptintention der Studie, nicht aber auf die vielen Einzelfragen und Aspekte eingegangen werden, die in dem Werk ebenfalls behandelt werden.

Nur ein Beispiel sei genannt für die u. a. auch demokratie- und öffentlichkeitstheoretische bzw. -ethische Dimension dieser Studie: So findet sich darin etwa auch ein Plädoyer für eine offene Diskurskultur, das in den polarisierten Debatten über Migration und Flüchtlinge sehr wohl Hassreden und Rassismus –

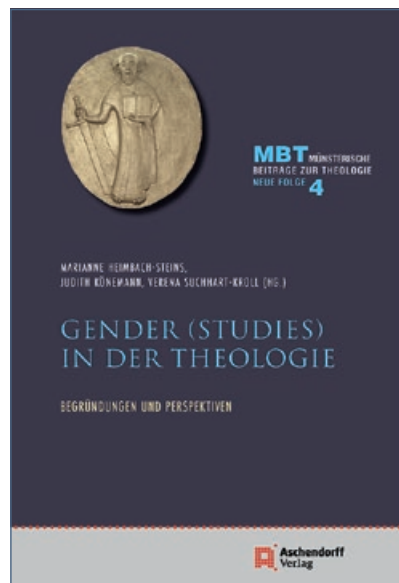
auch in ihrer christlichen Verbrämung (vgl. bes. S. 558 ff.) –, nicht aber sachliche Bedenken per se diskreditiert, weshalb das Phänomen der „cancel culture“ zu Recht kritisch bewertet wird (vgl. S. 654). Es muss vielmehr gelten – und damit ist man wieder beim thematischen Kern angelangt –, das Thema von der Ebene ideologisch eingefärbter Überzeugungskonflikte auf jene von Interessenkonflikten herunterzuführen, die bei Migration unweigerlich entstehen. Die widerstreitenden Interessen können insofern auf der unhintergehbaren Basis menschenrechtlicher Standards, nach Maßgabe der Gerechtigkeit sowie durch Abwägung und Kompromissorientierung auf eine lösungsorientierte Ebene geleitet werden und dabei den besagten gerechten Ausgleich anzielen. Dazu können Studien aus der Perspektive einer christlichen Sozial- und Migrationsethik, wie die vorliegende, einen genuinen Beitrag leisten.

Lars Schäfers,
Mönchengladbach

Gender in der Theologie

Marianne Heimbach-Steins/Judith Könemann/Verena Suchart-Kroll (Hg.): *Gender (Studies) in der Theologie. Begründungen und Perspektiven (Münsterische Beiträge zur Theologie, Neue Folge Band 4)*, Münster: Aschendorff Verlag 2021, 237 S., ISBN 978-3-402-12316-4

Wie wichtig die Auseinandersetzung mit Genderfragen in den christlichen Theologien und Kirchen ist, ist in den vergangenen Monaten mehr als deutlich geworden: Nicht nur die sich überschlagenden Enthüllungen zum Missbrauchsskandal, auch das Coming Out von 125 Mitarbeiter*innen der katholischen Kirche verweisen auf weiterhin einflussreiche lebensunfreundliche theologische Geschlechterkonzepte, die Menschen in ein enges Korsett von Geschlechternormen und Heteronormati-



vität pressen wollen. Der Band *Gender (Studies) in der Theologie. Begründun-*

gen und Perspektiven, herausgegeben von Marianne Heimbach-Steins, Judith Könemann und Verena Suchart-Kroll, könnte daher aktueller und relevanter nicht sein. Zugleich greift er mit dem Anliegen, die feministische und Gender-Forschung aus dem Status des Spezialthemas zu befreien und als eine alle Fächer und Themen der Theologie durchziehende Perspektive auszuweisen, einen Anspruch auf, den feministische und genderbewusste Theologie schon seit ihren Anfängen erhebt. Ziel des Bandes ist den Herausgeberinnen zufolge „die selbstverständliche Berücksichtigung von sowie die explizite Reflexion auf Geschlecht als Gegenstand in allen theologischen Disziplinen und Themenfeldern“ (10). Dementsprechend sollen sowohl der aktuelle Stand des deutschsprachigen theologischen Genderdiskurses aufgenommen

und analysiert, als auch neue Themen und Fragen aufgezeigt werden. Besondere Bedeutung wird dabei dekonstruktivistischen, postkolonialen, intersektionalen und queeren Perspektiven zugeschrieben.

Im ersten Teil des Bandes werden zunächst „Beispiele geschlechtersensibler theologischer Forschung“ in Bezug auf verschiedene theologische Fächer gegeben (17). So erörtert etwa Katja Winkler, inwiefern eine Rezeption der Gender Studies in der (katholischen) Sozialethik für die Fragen von Repräsentation und Exklusion fruchtbar sein kann und Hans-Ulrich Weidemann zeigt auf, was die neutestamentliche Exegese durch eine intensivere Beschäftigung mit der Männlichkeitsforschung gewinnen könnte. Einen interessanten Einblick in die Auseinandersetzung der Liturgiewissenschaft mit der Genderforschung gibt Miriam Venemann: Sie skizziert begriffliche, methodische und thematische Schritte hin zu einer gendersensiblen Liturgiewissenschaft und plädiert „für eine konstruktiv-dynamische Zusammenarbeit“ (39). Der erste Teil umfasst darüber hinaus vier kirchengeschichtliche Beiträge. Diese reichen von der kritischen Analyse einer zu wenig an konkreten Frauen interessierten Kirchengeschichtsschreibung (am Beispiel der Diakonie: Ute Gause) über Beispiele für eine andere, gendersensible und an konkreten Frauen orientierte kirchenhistorische Forschung (mit einem Oral-History-Projekt zu Diakonissen von Charlotte Langenhorst und einer Untersuchung der kirchlichen Tätigkeiten von Frauen während des Ersten Weltkrieges von Andrea Hofmann) bis hin zu einer methodischen Reflexion der Herangehensweise gendersensibler Kirchengeschichte (mit einem Plädoyer für einen intersektionalen Ansatz: Benedikt Bauer). Den Abschluss dieses Teils bildet ein religionswissenschaftlicher Aufsatz von Kristina Göthling-Zimpel, die die Möglichkeit und den Nutzen einer Auseinandersetzung mit René Girards Theorie der mimetischen Gewalt diskutiert.

Der zweite Teil des Bandes widmet sich dann vor allem politisch-ethischen

sowie pädagogischen Perspektiven. Er beginnt mit einem Aufsatz von Irene Klisenbauer, die einen knappen Blick auf das Verhältnis von Genderforschung und Theologie in Bezug auf den Menschenrechts- und Universalitätsdiskurs wirft. Weiterhin findet sich hier ein Beitrag von Sonja Angelika Strube zum Zusammenhang von Rechtsextremismus und Sexismus, der sowohl mit Blick auf den gesellschaftlichen Diskurs als auch angesichts der seltenen Thematisierung dieses Zusammenhangs in der Theologie besonders wichtig erscheint. In Auseinandersetzung u. a. mit der Kyriachatsanalyse von Schüssler-Fiorenza entwickelt Strube Ansatzpunkte für ein Verständnis „[f]eministische[r] Theologie als psychologisch geerdete Totalitarismuskritik“ (121). Zwei weitere – dezidiert ethische – Beiträge in Teil II richten den Fokus auf die Konstruktion der Zweigeschlechtlichkeit: So argumentiert Katharina Mairinger in Auseinandersetzung mit Pierre Bourdieu und Thomas Pröpper für den katholischen Kontext „für eine freiheitstheologisch formierte theologische Geschlechteranthropologie, in der auch intergeschlechtliche Menschen Platz finden“ (132). Und Mathias Wirth fordert ausgehend von der Beobachtung, dass „trans oder nicht-binäre Geschlechtlichkeit“ in der Theologie kaum thematisiert wird, eine stärkere Reflexion auf das Konzept der „Multirealisierbarkeit des Geschlechtlichen“ (144). Auf diese vier systematischen Beiträge folgen drei religionspädagogische: Aufschlussreich sind dabei etwa die Ausführungen von Nele Spiering-Schomborg und Annegret Reese-Schnitker zu konkreten Lehr-Lern-Modulen, die von den Autorinnen bereits in der Lehre erprobt wurden. Weiterhin widmet sich Jens Beiner (als zweiter in diesem Band) der Männlichkeitsforschung und untersucht die Möglichkeiten eines jungensensiblen Religionsunterrichts. Gianna Ridder wiederum plädiert ausgehend von einer empirischen Untersuchung zur feministischen Theologie im Religionsunterricht für eine umfassendere, statt nur punktuelle Beschäftigung mit diesem Thema – nur so

sei der Mehrwert feministischer Theologie vermittelbar.

Der dritte und letzte Teil des Bandes legt den Fokus schließlich auf „Entwicklungspotentiale theologischer Genderforschung“ (189). Tatsächlich haben auch schon einige der vorangegangenen Aufsätze Entwicklungspotenzial aufgezeigt, hier sollen nun aber noch einmal dezidiert neue Perspektiven eingebracht werden. Den Anfang macht dabei Julia Lis, die auf die kritische Debatte um Judith Butlers Genderansatz innerhalb der feministischen Theorie verweist und die „emanzipatorische[n] Möglichkeiten einer Dekonstruktion von Geschlechtsidentitäten“ (196) in Frage stellt – was man durchaus auch als Anfrage an den theoretischen Ansatz anderer Beiträge im Band lesen kann. Lis selbst plädiert für „die Artikulation einer weiblichen Subjektposition“ (197) und eine Stärkung der politischen Ausrichtung feministischer Theologie. Einen ungewohnten und erhellenden Blick eröffnet der Beitrag von Gerrit Spallek, der von einem „auto-ethnografisch[en]“ „Selbstexperiment“ (203) auf dem Kiez von St. Pauli ausgeht. Mit seinen Überlegungen zur Konstruktion von Männlichkeit und Sexualität im Zusammenhang von Prostitution wirbt er sowohl für eine Beziehungspastoral, die Menschen beim Sprechen über Sexualität unterstützt, als auch methodisch „für die Relevanz theologischer Ortserkundungen“ (204). Andreas Krebs plädiert in seinem Beitrag dafür, queer-theologischen Ansätzen größere Aufmerksamkeit zu schenken und zeigt deren Bandbreite von „radikaler Orthodoxie“ (Elizabeth Stuart) bis hin zu „radikaler Kritik“ (Marcella Althaus-Reid) auf. Trotz der Spannungen zwischen diesen Ansätzen entfalte sich, so Krebs, „erst im Gegen- und Miteinander“ „ihr volles subversives Potenzial“ (221). Den Abschluss des Bandes bildet eine Analyse von Gunda Werner zum Verhältnis von „Intersektionalität und Theologie“ (225), die zu einer intensiveren Auseinandersetzung mit Intersektionalität anregen will.

Insgesamt umfasst der Band 19 eher kurze Beiträge von Theolog*innen ver-



schiedener Konfessionen und aus unterschiedlichen fachlichen Perspektiven. Er zeigt damit ein weites Spektrum an Themen und Fragen sowie auch an unterschiedlichen möglichen Antworten auf. Auffallend ist, dass ein verhältnismäßig großer Teil der Autor*innen aus einer jüngeren Generation von Theolog*innen stammt – was nicht zuletzt deswegen hervorzuheben ist, weil es die fortwährende Aktualität von Genderfragen für die Theologie unterstreicht. Etwas schade ist hingegen, dass die angekündigte Bestandsaufnahme feministischer und gen-

derbewusster Theologie insgesamt etwas schmal ausfällt und dass gerade in Bezug auf dekonstruktivistische, postkoloniale, intersektionale und queere Vorstöße einige wegweisende Stimmen nicht oder kaum vorkommen (so z. B. Ulrike Auga, die sich als eine der ersten im deutschsprachigen Raum um eine Integration dekonstruktivistischer Ansätze und intersektionaler Perspektiven bemüht hat, oder Heike Walz, die bereits seit mehreren Jahren den Blick auf queere und postkoloniale Perspektiven lenkt). Nichtsdestotrotz zeigt der Band von Heimbach-

Steins, Könemann und Suchhart-Kroll höchst anregend und vielseitig die Bedeutung von Gender für die Theologie auf. Es ist daher zu hoffen, dass er in der Theologie breit rezipiert wird und die angesprochenen Themen im theologischen wie auch kirchlichen Diskurs aufgegriffen und weiterbearbeitet werden – dies nicht zuletzt, um das enge Korsett von Geschlechternormen und Heteronormativität in Theologie und Kirche endlich aufzubrechen und lebensfreundlichere Bedingungen für alle zu schaffen.

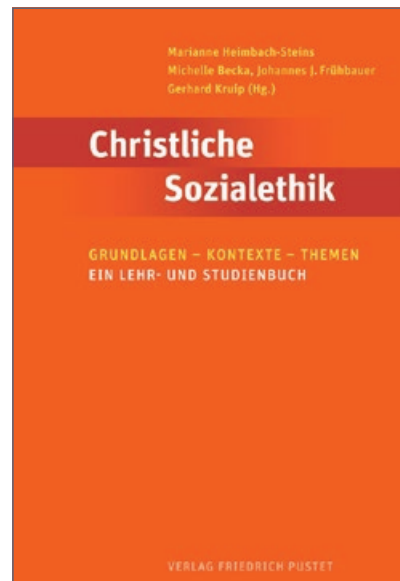
Cornelia Mügge, Dresden

Lehrbuch der Christlichen Sozialethik

Marianne Heimbach-Steins/Michelle Becka/Johannes J. Frühbauer/Gerhard Kruij (Hg.): Christliche Sozialethik, Grundlagen – Kontexte – Themen, Ein Lehr- und Studienbuch, Verlag Friedrich Pustet, Regensburg 2022, 528 S., ISBN 978-3-7917-3322-7

Nachdem M. Heimbach-Steins 2004/05 ein zweibändiges Lehrbuch der Christlichen Sozialethik herausgegeben hatte, bot es sich an, nach gut 15 Jahren die anschließende wissenschaftliche Diskussion und aktuelle Problemlagen aufzunehmen und ein neues Lehrbuch zu veröffentlichen. Es ist keine Neuauflage, weil der Bd. konzeptionell (etwa der Unterteilung von Kontexten und Handlungsfeldern) neu ausgerichtet ist und auch viele neue Autoren beteiligt sind. Neben den Herausgeber:innen sind fünf weitere Autor:innen, von denen drei Schüler:innen von Heimbach-Steins sind, vertreten (drei Frauen und sechs Männer).

Der Bd. ist in zwei große Teile aufgeteilt, nämlich Grundlegungsfragen einer fundamentalen Sozialethik, die in die drei Blöcke der konzeptionellen Grundlegung, der christlichen Tradition und der normativen Grundorientierungen unterteilt sind, sowie im zweiten Teil der Anwendungsfragen mit Kontexten und Handlungsfeldern. Die drei ersten Blöcke beginnen jeweils eingangs mit einem Vor-



spann, der in die Problematik einführt, und schließen mit einer Synthese, in der die Teile zusammengebunden werden. Die jenseits von Vorwort, Einführung, Vorspanne und Synthesen insgesamt 28 thematischen Beiträge weisen jeweils eine nahezu gleiche Seitenzahl auf, was eine hohe Disziplin der Autor:innen erfordert haben muss. Dabei haben sich die Autoren:innen in den Blöcken (z. B. Handlungsfeldern) weitgehend an einer einheitlichen Systematik orientiert. Die strikte räumliche Begrenzung führt allerdings dazu, dass in manchen Artikeln (z. B. Frieden) die klassischen so-

zialethischen Kriterien, etwa zum gerechten Krieg, zwar erwähnt, aber für die (angezielten studentischen) Leser:innen in ihrer Anwendung nicht hinreichend erklärt werden.

Der Bd. beginnt nach der Einführung, in der auf den Grundbegriff „Christliche Sozialethik“ eingegangen wird, und als Brückenfach in der Theologie zu Sozial- und Kulturwissenschaften definiert wird. Sie setzt sich zentral mit der Gerechtigkeit sozialer Institutionen auseinander und zielt auf gesellschaftliche Praxis ab.

J. Frühbauer/ M. Heimbach-Steins beginnen mit einem Artikel zum Verhältnis von Sozialethik und Gesellschaftstheorie, wobei das Gesellschaftsverständnis aus soziologischen Theorien abgeleitet wird. Die frühere Bedeutung der Volkswirtschaftslehre für die Christliche Sozialethik tritt demgegenüber zurück, obwohl etwa auf den Ökonomenobelpreisträger Amartya Sen in anderen Beiträgen mehrfach Bezug genommen wird. Mit der mehrfach im Bd. zu findenden Charakterisierung der Gesellschaft als „Spätmoderne“ greift man eher Historikern der zweiten Hälfte des 21. oder des 22. Jh. vor.

Dem schließt sich ein Beitrag von J. Frühbauer zur Sozialphilosophie und von M. Heimbach-Steins zum Verhältnis von Sozialethik und Theologie an. Christliche Sozialethik ist dabei selbst plural, weil von einzelnen Vertretern der Diszi-



plin unterschiedliche Sozialphilosophien und Sozial- und Kulturwissenschaften rezipiert werden. M. Heimbach-Steins macht deutlich, dass grundlegende Gemeinsamkeiten unterschiedlicher Ansätze der Christlichen Sozialethik, die als Schöpfung-Ordnung, Befreiung-Veränderung sowie Prophetie-Kritik gekennzeichnet werden, in der biblischen Überlieferung und christlichen Tradition sowie in den Kernelementen der christlichen Anthropologie liegen.

Im zweiten Block behandelt M. Heimbach-Steins die historische Entwicklung des vormodernen christlichen Weltengagements, G. Kruijff den deutschen Sozialkatholizismus des 19. und 20. Jahrhunderts und M. Heimbach-Steins die Entwicklung und Rolle der kirchlichen Sozialverkündigung. In diesem Block wäre es sinnvoll gewesen, dass in dem Beitrag zum deutschen Sozialkatholizismus die Institutionenstruktur der deutschen Gesellschaft wie den Sonntagschutz, Unternehmensmitbestimmung, Arbeitslosenversicherung, dynamische Rente, Pflegeversicherung etc., wofür die christlich-soziale Bewegung lange Zeit gekämpft hatte, stärker herausgestellt worden wäre, weil es für Studierende sinnvoll ist zu erfahren, dass sich christliches Weltengagement lohnen kann und dass es bereits große Früchte getragen hat. Damit würde der dritte Träger der Soziallehre der Kirche, nämlich das Praxisengagement der Gläubigen, stärker zur Geltung kommen.

Der dritte Block bildet den systematischen Mittelpunkt des Bandes, indem dort die normative Orientierung behandelt wird. Dabei ist relativ neu für die Sozialethik, dass dem Verantwortungsbegriff (M. Vogt) ein eigener Abschnitt gewidmet wird. Im Gegensatz zum Vorgängerband, in dem für jedes der klassischen Sozialprinzipien jeweils ein eigener Abschnitt enthalten war, werden diese in einem Artikel von M. Heimbach-Steins zusammengefasst. Dabei wird aber von der neuen Kategorie der Nachhaltigkeit abgesehen, der ein eigener Abschnitt (M. Vogt) eingeräumt wird. Menschenrechten (M. Becka) und Gerechtigkeit

(J. Frühbauer) als normativen Grundkategorien sind weitere Beiträge gewidmet. In seiner Synthese streicht A. Filipovic besonders den Partizipationsgedanken heraus.

In diesem Block wird ein ungelöstes Theorieproblem der Christlichen Sozialethik deutlich, weil das Verhältnis der klassischen Sozialprinzipien von unterschiedlichen Gerechtigkeitskriterien (intergenerationelle Gerechtigkeit) und der Nachhaltigkeit systematisch ungeklärt sind. So könnte man zwar Nachhaltigkeit auf die intergenerationelle Dimension beschränken und damit gut die klassischen Sozialprinzipien als intragenerationelle Maxime ergänzen, man würde dann aber dem UN-Gebrauch widersprechen. Hingegen enthält die UN-Definition der Nachhaltigkeit Aspekte, die bereits mit dem Solidaritätsprinzip abgedeckt sind.

Der zweite Teil ist in 6 Kontexte und 11 Handlungsfelder aufgeteilt. In den Kontexten sind mit „Technik“ (A. M. Riedl) und „Kultur“ (M. Becka) erfreulicherweise zwei Dimensionen aufgenommen, die bisher nicht im Vordergrund der Christlichen Sozialethik standen. Becka greift dabei zwei aktuelle Diskurse von Identitätspolitik und Postkolonialismus auf.

Die Zuordnung von Kontexten und Handlungsfeldern ist nicht in jedem Fall plausibel. Wenn etwa Grundzüge der Wirtschaftsethik (Kruijff) unter Kontexten behandelt werden, werden Grundzüge der Medienethik unter Handlungsfelder (A. Filipovic) thematisiert, obwohl die gegenwärtige Gesellschaft als Mediengesellschaft charakterisiert werden kann. Politik (Ch. Spieß) und Religion (M. Heimbach-Steins) sind weitere Kontexte, die gewisse Überschneidungen aufweisen.

Als zentrale Handlungsfelder werden von G. Kruijff Arbeit, Bildung und soziale Sicherung, von M. Heimbach-Steins „Lebensformen“, von M. Becka Migration, Gesundheit und weltweite Armut, von J. Frühbauer Frieden und von M. Vogt Klimaschutz behandelt. Außerdem stellt A. Filipovic Grundzüge der Medienethik dar. Gemäß dem induktiven Vorgehen,

das allgemein methodisch für die Christliche Sozialethik in diesem Bd. empfohlen wird, werden in diesen Beiträgen ausführlich sozialwissenschaftliche Dimensionen der Problemstellung geschildert bevor knapp normative Positionierungen unter Berücksichtigung kirchenamtlicher Positionen erfolgen. Während etwa im Beitrag zur Migration von M. Becka eine hohe Übereinstimmung mit kirchenoffiziellen Positionen anzutreffen ist, ist in dem Beitrag von M. Heimbach-Steins zu Lebensformen eine deutliche Distanz zu spüren. Manche Autoren (M. Vogt) scheuen die direkte Kritik lehramtlicher Positionen, wenn sie sich von den Positionen von Papst Franziskus in *Laudato si'* zum Bevölkerungswachstum und zum Emissionshandel nur indirekt distanzieren.

Abgeschlossen wird der Bd. von einem Artikel von D. Bogner zur Anwendung der Christlichen Sozialethik auf die Kirche als soziale Institution, wobei die häufig diskutierte Anwendung des Subsidiaritätsprinzips auf die Kirche selbst im Gegensatz zur Gewaltenteilung nicht thematisiert wird, obwohl nicht nur die Gewaltenteilung mit Schutz von Grundrechten und demokratischen Entscheidungsverfahren, sondern auch der Föderalismus zur Freiheitssicherung moderner Staatlichkeit und analog auch der Kirche gehört.

Der Bd. stellt mit seinen verschiedenen Beiträgen eine gelungene Einführung und einen Überblick über das Selbstverständnis, das methodische Vorgehen und die zentralen Inhalte der Christlichen Sozialethik dar. Die einzelnen Beiträge zeichnen sich durchweg durch eine solide Argumentation und abgewogene Urteile aus. Dort, wo ein:e Leser:in zu anderen Einschätzungen der empirischen Ausgangslage, der ethischen Güterabwägungen und der aufgezeigten Handlungsoptionen kommen könnte, dürfte immer ein sachbezogener Dialog möglich sein. In dem Bd. erweist es sich als vorteilhaft, dass ein einziger Autor inhaltlich verbundene Artikel aus den verschiedenen Blöcken angefertigt hat, so dass z. B. die Artikel Nachhaltigkeit, Ökologie und Klimaschutz von M. Vogt gut aufeinander

abgestimmt sind. Für Herausgeber:innen eines Bds., der die zentralen Inhalte einer theologischen Disziplin vorstellen will, stellt sich immer die Frage, welche Schwerpunkte und Inhalte man in einem begrenzten Raum präsentiert. Dies kann man insgesamt als gelungen betrachten, wenn auch klassische Themen wie Eigentum und Vermögensbildung nur am Rande auftauchen. Auch wenn die Begrenzung auf die deutsche Perspektive aus pragmatischen Gründen zweckmäßig erscheint, wäre es angesichts der faktischen Bedeutung, die die EU mittlerweile für das wirtschaftliche und politische Leben in Deutschland erlangt hat, sinnvoll gewesen, der EU einen eigenen Beitrag einzuräumen.

Von traditionellen Konzeptionen einer „Katholischen Soziallehre“ unterscheidet sich der Bd. dadurch, dass er die klassische neuscholastische Naturrechtskonzeption zu Recht aufgegeben hat, weil diese ein statisches Gesellschaftsverständnis beinhaltet. Die damit verbundene Defensivhaltung und Abwehr gegen Erscheinungen und Entwicklungen moderner Gesellschaften (u. a. Frauenemanzipation, Demokratie) und geistiger Strömungen ist ebenso wenig zu finden. Daher fehlen auch die traditionellen Auseinandersetzungen mit Liberalismus und Sozialismus.

Hervorzuheben ist die didaktische Orientierung, indem die einzelnen Artikel mit Leitfragen eingeführt werden und nach einzelnen Abschnitten durch

Fettdrucke immer ein zentrales Resultat der Argumentation hervorgehoben wird. Hilfreich ist auch der Überblick über die Entwicklung der kirchlichen Sozialverkündigung in Form einer Übersicht. Die Vielzahl der Querverweise zwischen den einzelnen Beiträgen dürfte zur Erschließung der Gesamtzusammenhänge sehr nützlich sein.

Der Bd. dürfte zu Recht an vielen theologischen Fakultäten zur Standardlektüre des Faches werden. Er kann darüber hinaus für Lehrer:innen und die katholisch-soziale Bildungsarbeit hilfreich sein. Dem Bd. sind viele Leser:innen zu wünschen. Der freie Zugang zu dem Bd. im Internet dürfte dem förderlich sein.

Joachim Wiemeyer, Bochum





Bernhard Irrgang: Data-Centric Biology and AI – Paradigm Shift for Science and Technology

In the interplay of AI and biotechnology, extensive processes of change in the research design of science and technology – beyond biology – have become apparent over the last 40 years. This is a development that the author describes with the term "technoresearch". This structural change is based on a new research paradigm, the "algorithmic turn", which, however, does not lead to the end of scientific theory – as is often claimed – but is based on a new style of thinking. This paradigm emerges in the transition from classical systems theory (cybernetics) to synergetics (theory of dynamic complex systems). The new orientation towards complexity and the way of dealing with it in a data-related way is changing the practice and theory of science and technology in the 21st century.

Armin Grunwald: Technology Assessment Meets Anthropology

Technology assessment (TA) develops orientation on technological progress through statements about the future and their evaluation. Digitization and artificial intelligence make future predictions more difficult in the face of machine learning and new human/technology relationships. Even the common TA approach of working with scenarios rather than making predictions often no longer works. The openness of the future is becoming too great. In this article the author presents how TA deals with this sit-

uation. In a hermeneutic perspective, it uses widespread future statements about digitalization and AI to learn about the present, in particular about the new human/technology relations revealed in the projections of AI. These need to be explicated in order to make them accessible to transparent debate and to use the results for technology design in this field.

Katharina Klöcker: Redemption from the Burden of Decision? – The Algorithmization of Society in Theological-Ethical Reflection

This article deals with the thesis of a possibly expected unstoppable transfer of decision-making power from humans to machines, as it can be found in currently circulating future narratives on the algorithmization of society. Against the background of reflections on how such future narratives can be hermeneutically opened up for ethical reflection, the text reflects on the question of what consequences the transfer of decision-making power to artificial intelligences would entail. The argument that algorithmic decision-making systems can contribute to the relief of overburdened humans and thus also to the elimination of suffering, and therefore represent progress towards the better, is problematized from a theological-ethical perspective.

John Wyatt: What does it mean to be human in a world of technology and AI?

We live in a world that is dominated by machines that appear to be intelligent and we tend to understand our own

humanity through the distorting lenses of technology and AI. Digital technologies lead to a focus on disembodied, abstract and quantifiable information, and tend to devalue our human embodiment. Human beings made in God's image are essentially related to God and given a unique vocation by God. The Incarnation and Resurrection of Christ provide a rich underpinning for celebrating, respecting and protecting human embodiment and human uniqueness in a world of intelligent machines. The challenge we face is to find innovative and creative ways of using sophisticated artificial intelligence technologies not to demean, distort or diminish us, but rather to enable us to flourish as human persons uniquely created to reflect God's character and being.

Catrin Misselhorn: Morality for Machines?

Artificial systems face moral problems in many application areas, such as autonomous driving, warfare, and even nursing. Machine ethics is an emerging discipline at the intersection of philosophy, computer science, and robotics that addresses the question of whether and how artificial systems can also make moral decisions. This paper introduces the new discipline and discusses arguments for and against this endeavour. Finally, three principles are introduced to serve as guardrails to ensure that technological development moves in a positive direction.



Bernhard Irrgang : Biologie basée sur des données et Intelligence artificielle (IA) – Un changement de paradigme pour la science et la technologie

Au cours des dernières quarante années, l'interaction entre l'IA et la biotechnologie a révélé d'importants processus de changement dans la conception de la recherche scientifique et technologique – au-delà de la biologie. Il s'agit d'une évolution que l'auteur décrit par le terme de « technoresearch ». Ce changement structurel repose sur un nouveau paradigme de recherche, l'« algorithmic turn », qui ne conduit pas – comme on le prétend souvent – à la fin de la théorie scientifique, mais qui révèle un nouveau style de pensée. L'origine de ce paradigme se trouve dans la transition entre la théorie classique des systèmes (cybernétique) et la synergie (théorie des systèmes complexes). La nouvelle orientation vers la complexité et la manière de la traiter en fonction de données modifient la pratique et la théorie de la science et de la technologie au 21^{ème} siècle.

Armin Grunwald : Numérisation et intelligence artificielle. L'évaluation des choix technologiques rencontre l'anthropologie

L'évaluation des choix technologiques (ECT) (technologie assessment) élabore des orientations sur le progrès technique en établissant et évaluant des prévisions au sujet de l'avenir. La numérisation et l'intelligence artificielle (IA) rendent des prévisions plus difficiles à cause de l'apprentissage automatique et des nouveaux rapports homme-technique. Même l'approche courante de l'évaluation des choix technologiques, qui consiste à travailler avec des scénarios plutôt qu'à faire des prévisions, souvent ne fonctionne plus. L'ouverture de l'avenir devient trop

grande. Dans cet article, je vais expliquer la manière dont l'ECT gère cette situation. Dans une perspective herméneutique, elle utilise des prévisions d'avenir courantes au sujet de la numérisation et de l'IA pour apprendre davantage sur le présent, en particulier sur les nouveaux rapports homme-technique qui apparaissent dans les projections de l'IA. Il s'agit de les expliciter afin de les rendre accessibles à un débat plus transparent et d'en utiliser les résultats pour les conceptions technologiques dans ce domaine.

Katharina Klöcker : Être délivré du fardeau de la décision ? L'algorithmisation de la société dans la perspective d'une réflexion théologique et éthique

L'article se penche sur la thèse d'un éventuel transfert inévitable du pouvoir de décision de l'homme à la machine tel qu'on le trouve dans les récits d'avenir qui circulent actuellement au sujet de l'algorithmisation de la société. Partant de réflexions sur la manière dont de tels récits d'avenir peuvent être utilisés de façon herméneutique pour la réflexion éthique, l'article réfléchit à la question des conséquences qu'entraînerait le transfert du pouvoir de décision à des intelligences artificielles. La thèse selon laquelle des systèmes de décision algorithmiques peuvent contribuer à soulager l'homme surmené et donc à éliminer la souffrance ce qui représenterait un progrès vers le meilleur, cette thèse est contestée d'un point de vue théologique et éthique.

John Wyatt : Qu'est-ce que cela dignifie d'être humain dans un monde de technologie et d'intelligence artificielle ?

Nous vivons dans un monde dominé par des machines qui semblent intelligentes, et nous avons tendance à com-

prendre notre propre humanité à travers le prisme déformant de la technologie et de l'IA. Les technologies numériques mènent à une focalisation sur des informations incorporelles, abstraites et quantifiables, et tendent à dévaloriser l'incarnation humaine. Les êtres humains, faits à l'image de Dieu, sont essentiellement reliés à Dieu et dotés par lui d'une vocation unique. L'incarnation et la résurrection du Christ fournissent une riche toile de fond pour célébrer, respecter et protéger l'incarnation et l'unicité humaines dans un monde de machines intelligentes. Le défi auquel nous sommes confrontés consiste à trouver des modes innovants et créatifs pour utiliser les technologies sophistiquées de l'intelligence artificielle, non pas pour nous abaisser, nous déformer ou nous diminuer mais plutôt pour permettre de nous épanouir en tant que personnes humaines créées de manière unique pour refléter le caractère et l'être de Dieu.

Catrin Misselhorn : Une morale pour les machines ?

Dans de nombreux domaines d'application tels la conduite autonome, la guerre, mais aussi les soins, les systèmes artificiels sont confrontés à des problèmes moraux. L'éthique des machines est une discipline émergente, à l'intersection de la philosophie, de l'informatique et de la robotique, qui s'intéresse à la question de savoir si et comment des systèmes artificiels peuvent également prendre des décisions morales. L'article présente cette nouvelle discipline et débat des arguments pour et contre ce projet. En conclusion, il introduit trois principes qui peuvent servir de garde-fous et garantir que le développement technologique prenne une direction positive.



SCHWERPUNKTTHEMEN DER BISHER ERSCHEINENEN HEFTE

4/2006	Markt für Werte (vergriffen)	3/2015	Ethik in der Stadt
1/2007	Lohnt die Arbeit?	4/2015	Laudato si'
2/2007	Familie – Wachstumsmittel der Gesellschaft?	1/2016	Soziale Ungleichheit
3/2007	Zuwanderung und Integration	2/2016	Korruption und Compliance
4/2007	Internationale Finanzmärkte (vergriffen)	3/2016	Inklusion und Behinderung
1/2008	Klima im Wandel	4/2016	Umstrittener Pluralismus
2/2008	Armut / Prekariat	1/2017	Was darf Leben kosten?
3/2008	Gerüstet für den Frieden?	2/2017	Europa und Afrika – Partnerschaft auf Augenhöhe?
4/2008	Unternehmensethik	3/2017	Ökumenische Sozialethik
1/2009	Wie sozial ist Europa?	4/2017	Konsumethik
2/2009	Hauptsache gesund?	1/2018	Digitalisierung gestalten
3/2009	Caritas in veritate	2/2018	Terrorismus und Terrorismusbekämpfung
4/2009	Wende ohne Ende?	3/2018	Wohnen
1/2010	Gerechte Energiepolitik	4/2018	Steuergerechtigkeit
2/2010	Steuern erklären	1/2019	Politik der Transformation – Transformation der Politik
3/2010	Neue Generation Internet – grenzenlos frei?	2/2019	Demokratie neu stärken
4/2010	Agrarpolitik und Welternährung	3/2019	Ethik der Mobilität
1/2011	Zivilgesellschaft	4/2019	Nationalstaat und nationale Identitäten
2/2011	LebensWert Arbeit	1/2020	Sozialethische Ansprüche an die Kirchenreform
3/2011	Wohlstand ohne Wachstum?	2/2020	Pflegearbeit
4/2011	Soziale Marktwirtschaft für Europa?	3/2020	Friedensethik vor neuen Herausforderungen
1/2012	Religionspolitik	4/2020	Konfessionelle Wohlfahrt im Umbruch
2/2012	Was dem Frieden dient	1/2021	Fratelli tutti
3/2012	Finanzmärkte und Staatsschulden	2/2021	Landwirtschaftspolitik
4/2012	Stark gegen Rechts	3/2021	Soziale Medien: Ethik und Politik
1/2013	Bevölkerungswachstum	4/2021	Jüdisches Leben in Deutschland – gestern, heute, morgen
2/2013	Menschenrechte interreligiös	1/2022	Weltwirtschaft zwischen Digitalisierung, Corona und Klima
3/2013	Geschlechtergerechtigkeit	2/2022	Finanzethik
4/2013	Altern und Pflege	3/2022	Künstliche Intelligenz
1/2014	Ressourcenkonflikte		
2/2014	Solidarität in Europa		
3/2014	Die Würde der Tiere ist antastbar		
4/2014	Freihandel		
1/2015	Transnationale Steuerung		
2/2015	Kirche und Geld		



VORSCHAU

Heft 4/2022

Schwerpunktthema: Gemeinsame Verantwortung der Religionen

Heft 1/2023

Schwerpunktthema: Stellschrauben gegen soziale Ungleichheiten

Heft 2/2023

Schwerpunktthema: Geopolitik – Sicherheitsarchitektur – Neue Weltordnung