

Medien – Ethik – Digitalisierung

Aktuelle Herausforderungen

Herausgegeben von Petra Grimm,
Harald Pechlaner und Oliver Zöllner

Medienethik
Digitale Ethik | Band 20

Franz Steiner Verlag





medien ethik

Herausgegeben von
PETRA GRIMM und OLIVER ZÖLLNER

Mitgegründet von
RAFAEL CAPURRO

Band 20

<https://www.steiner-verlag.de/brand/Medienethik>

Medien – Ethik – Digitalisierung
Aktuelle Herausforderungen

Herausgegeben von Petra Grimm,
Harald Pechlaner und Oliver Zöllner

Franz Steiner Verlag

Umschlagabbildung: Künstlicher Horizont, Foto: Petra Grimm.

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
dnb.d-nb.de abrufbar.

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt.
Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes
ist unzulässig und strafbar.

© Franz Steiner Verlag, Stuttgart 2023
www.steiner-verlag.de

Druck: Beltz Grafische Betriebe, Bad Langensalza
Gedruckt auf säurefreiem, alterungsbeständigem Papier.
Printed in Germany.

ISBN 978-3-515-13599-3 (Print)

ISBN 978-3-515-13601-3 (E-Book)

<https://doi.org/10.25162/9783515136013>

INHALTSVERZEICHNIS

<i>Petra Grimm, Harald Pechlaner, Oliver Zöllner</i> Medien – Ethik – Digitalisierung. Multidisziplinäre Zugänge	7
<i>Christopher Koska, Michael Reder</i> KI-gestützte Assistenz für moralische Konfliktsituationen. Zur Algorithmisierung im Handlungsfeld der Kindeswohlgefährdung	19
<i>Marina Moreno, Adriano Mannino, Nikil Mukerji</i> Krisenpotenziale der Künstlichen Intelligenz	37
<i>Jan Mehlich</i> ELSI-SAT Health & Care: Ein Ethics-By-Design-Tool für Integrierte Forschung und Technikentwicklung im Frühstadium	59
<i>Marcel Schlegel</i> Das Inklusionskonzept aus Sicht der (digitalen) Ethik. Werte, Funktionen und Schritte (zu) einer Inklusionsethik	77
<i>Oliver Zöllner</i> Plünderung oder Sharing? Analyse eines moralisierten Diskurses im digitalen Raum und die Frage der Wiederverwendung von materiellen Gütern	109
<i>Harald Pechlaner, Michael de Rachewiltz</i> Homo digitalis meets Homo montanus	135
<i>Susanne Kuhnert</i> Wunderbare Ethik? Märchen als Bildungsmedium für eine digitale Ethik	143
<i>Mirjam Gruber, Valeria von Miller, Michael de Rachewiltz</i> Einfach nur heiß? Die visuelle Konstruktion einer Hitzewelle in Zeiten der Klimakrise	163
<i>Jan Doria</i> Zwischen Utopie und Dystopie: Künstliche Intelligenz als nichtstattfin- dendes Zentralereignis in Fernsehmagazinbeiträgen der ARD	175

Laura Braxmaier

Querdenken 711 und nihilistische Haltungen. Die Zusammenfassung
einer qualitativen Untersuchung nihilistischer Haltungen im Kontext
des Verschwörungsglaubens in der Corona-Pandemie..... 197

Kurzbiografien219

MEDIEN – ETHIK – DIGITALISIERUNG

Multidisziplinäre Zugänge

Petra Grimm, Harald Pechlaner, Oliver Zöllner

Der vorliegende Band der Schriftenreihe „Medienethik | Digitale Ethik“ fällt in eine Zeit, in der die Künstliche Intelligenz durch die öffentliche Zugänglichmachung von generativer KI, die vor allem durch das Programm ChatGPT-4 prominent wurde, eine erneute Hochkonjunktur erfährt. Symptomatisch für die Gegenwart ist aber auch ein scheinbar paradoxes Gefühl wehmütiger Nostalgie, das mit unserer ‚analogen‘ Vergangenheit assoziiert wird. So zeigte die Studie „Werte, Ängste, Hoffnungen. Das Erleben der Digitalisierung in der erzählten Alltagswelt“ (2021) den überraschenden Befund, dass die analoge Zeit von vielen Befragten, und zwar altersübergreifend, als etwas Schätzenswertes konnotiert wird.¹ Viele der früheren Alltags- und Medienpraktiken werden zwar als anachronistisch, unwirklich oder bizarr markiert, zugleich aber als schön, ansprechend oder wünschenswert entworfen. Damit erfährt die ‚analoge‘ Vergangenheit letztlich eine *nostalgische Idealisierung*, in der nach Innen gerichtete Werte wie Harmonie, Ruhe oder Beschaulichkeit durchscheinen.

In dieses Bild einer ‚analogen Nostalgie‘² fügt sich die Nachfrage nach handgemachten Produkten genauso ein wie etwa Polaroid-Fotos oder Vinylschallplatten – auch wenn sich deren Attraktivität keineswegs nur im Nostalgischen oder als Retrotrend erschöpft.³ Sie können aber auch als Reaktionen auf die intensive Datenwirtschaft im Kontext eines ‚Überwachungskapitalismus‘⁴ gesehen werden, der im Westen von machtvollen Plattformen ausgeht und teilweise maßlos ist.⁵ Ein Rückgriff auf Bekanntes, auf analoge Formen, die zudem den diffusen Hauch der ‚Authentizität‘ haben, kann in Zeiten großer und schneller Umbrüche eine angenehme Rückversicherung sein.⁶ Die tiefgreifende Vernetzung, die die Digitalität ermöglicht, weitet zwar auch Handlungs- und Erfahrungsmöglichkeiten aus – bis allerdings auch hin zur Überforderung vieler Nutzenden.⁷ In der Gegenwart werden vor diesem Hintergrund bereits erste kritische Überlegungen zu denkbaren Formen

1 Vgl. Grimm/Müller/Trost 2021.

2 Vgl. Schrey 2017.

3 Vgl. Lund et al. 2022.

4 Vgl. Zuboff 2018.

5 Vgl. Seemann 2021; Staab 2019.

6 Vgl. zum Authentizitätskonzept Schilling 2020.

7 Vgl. te Wildt 2015; Hepp 2022.

einer „Entnetzung“ als „utopisches Gegenmodell zur digitalen Übervernetzung“ formuliert.⁸

Dies erfolgt etwa mit Blick auf die wie auch eingedenk der zahlreich zu beobachtenden überhitzten Auseinandersetzungen im Netz, die aus Sicht einer liberalen Demokratie unerwünschte, destabilisierende Phänomene wie Hassrede, Fake News, Propaganda und Verschwörungsmythen hervorbringen.⁹ Die Digitalität erfordert somit neue ethische Ansätze für das Leben im Zustand der permanenten Vernetzung im Kontext einer „tiefgreifenden Mediatisierung“.¹⁰

1 DIE HERAUSFORDERUNG NEUER TECHNOLOGISCHER ENTWICKLUNGEN

Weiter als je zuvor rückt die prädigitale Vergangenheit weg, wenn nun sogenannte „AI Content Farmen“¹¹ beginnen, mit generativen KI-Beiträgen das Web zu fluten, die die Glaubwürdigkeit der (kostenlosen) Informationsmedien noch weiter unterminieren könnten, da deren Inhalte nicht selten falsch bzw. unwahr sind. Ob die massenhafte Verbreitung als Vorzeichen eines neuen grundlegenden Wandels des Web zu verstehen ist, bleibt zum gegenwärtigen Zeitpunkt offen. Nach dem Sprung vom ursprünglich freien zum kommerzialisierten Web des „Überwachungskapitalismus“ wird nun womöglich durch eine sich verselbständigende Contentmaschinerie erneut eine Transformation der Digitalität erfolgen, die möglicherweise zu einem Strukturwandel in der Kreativ- und Kommunikationsbranche führt. Mit welchen Daten, Texten und Bildern werden die KI-Programme trainiert? Wer steuert und überwacht dies – und wie? Was ist, wenn immer mehr Texte und Bilder im Netz bereits von generativen KI-Programmen geschaffen und diese dann wiederum von anderen KI-Programmen als Grundlage für Text- oder Bildgenerationen verwendet werden? Den selbstreferenziellen Halluzinationen der *Large Language Models* (LLM)¹² wäre dann kaum noch beizukommen. In jedem Fall sind Jürgen Habermas' Ideen und Darlegungen zu einem „erneuten Strukturwandel der politischen Öffentlichkeit“, gerade erst 2021 erschienen,¹³ rasch weiterzuschreiben – und sicher über die Sphäre des Politischen hinaus tief ins Private.

Für das Medienvertrauen und damit die Basis einer liberalen Demokratie versprechen diese Entwicklungen nichts Gutes, vor allem wenn Verschwörungsmythen und Desinformation durch generative KI noch stärker verbreitet werden und damit nihilistische Einstellungen an Boden gewinnen.¹⁴ Ein gewisser Nihilismus steckt bereits in der Vorstellung, überhaupt gesellschaftliche Steuerungsleistungen

8 Stäheli 2021, S. 422.

9 Vgl. Amlinger/Nachtwey 2022; Risse 2022; Zöllner 2020 sowie die Beiträge in Schicha et al. 2021.

10 Vgl. Hepp 2020.

11 Eisenbrand 2023.

12 Vgl. Bender et al. 2021; Bommasani et al. 2022; Esposito 2022; Mahowald et al. 2023.

13 Vgl. Habermas 2021.

14 Vgl. Braxmaier in diesem Band.

an Programme und ihre Algorithmen abzugeben: „Governance by nonhuman bureaucrats – otherwise known as algorithms – can make citizens feel safe from bias and prejudice and can make political leaders feel safe from being *accused* of being biased and prejudiced.“¹⁵ Der utilitaristische und letztlich reduktionistische (und nur allzu selten hinterfragte) Glauben an den Nutzen von KI-Instanzen führt in ethischer Sichtweise also auf eine Negation von Verantwortung zurück. Es erscheint bequemer (und kostengünstiger), sich etwa auf generative Textprogramme zu verlassen als auf professionelle menschliche Akteure. „Durch die automatisiert algorithmische Vermittlung können Akteure ihre Botschaften in höherer Auflösung an spezifische Zielgruppen richten und sich mit ihnen verbinden, die digitalen Formen der Konnektivität erschweren jedoch die für demokratische Prozesse notwendige Repräsentanz und Zurechenbarkeit von Mitteilungen an politische Akteure.“¹⁶ Eine unreflektierte Digitalisierung der Prozesse, die von öffentlichem Belang sind, kann die Grundlagen des Vertrauens in gemeinwohlorientierte Institutionen wie auch in die Medien erheblich erschüttern. Die Kaskaden an Krisen, die in den letzten Jahren zu bewältigen waren, haben hier bereits ihre Spuren hinterlassen, wie etwa ein Blick auf das Vertrauen in Medien zeigt: „Zwischen 2008 und 2019 hat sich die Zahl der Menschen, die den Medien bei wichtigen Dingen nicht vertrauten, etwa verdreifacht – von 9 auf 28 Prozent (...).“¹⁷ Im gleichen Zeitraum stieg in Deutschland allerdings auch der Anteil der Personen, die den Medien vertrauen, von 29 auf 43 Prozent.¹⁸ In einem solchen Kontext der zunehmenden Polarisierung von Haltungen könnte sich in Teilen der Bevölkerung auch wieder der Bedarf nach Qualitätsjournalismus erhöhen und die Bereitschaft steigen, hierfür zu zahlen. Notwendig ist hierüber aber ein gesellschaftlicher Konsens, der sich am Gemeinwohl orientiert und nicht bloß – im Privaten wie auch in kulturellen Leitbildern – an Kosten-Nutzen-Kalkulationen.

2 BREITES THEMENSPEKTRUM DER DIGITALEN ETHIK

Vor dem Hintergrund dieser aktuellen Beobachtungen im Dreiklang von „Medien – Ethik – Digitalisierung“ repräsentiert der vorliegende Band das breite Themenspektrum im Umfeld einer Digitalen Ethik, die sich multidisziplinär orientiert. Ausgangspunkt hierfür war die Jahrestagung des Instituts für Digitale Ethik (IDE) an der Hochschule der Medien in Stuttgart, die in Kooperation mit dem Center for Advanced Studies von EURAC Research (Bolzano/Bozen) zum Thema „Zeit für Verantwortung: Digitalisierung nachhaltig gestalten“ veranstaltet wurde. Die Ziele und Perspektiven beider Institute haben sich im Rahmen dieser Tagung ergänzt und gegenseitig befruchtet. So war es ein Anliegen des IDE, der Frage nachzugehen, wie die Digitalisierung global nachhaltig gestaltet werden kann – wie also die digitale Technik so designt werden kann, dass sie weniger Ressourcen verbraucht, sich

15 Gertz 2019, S. 184.

16 Donges 2022 S. 9.

17 Jakob et al. 2023, S. 46.

18 Vgl. ebd.

an sozialen und ökologischen Bedürfnissen ausgerichtet und so einen Beitrag zu einem guten Leben auf der Erde für möglichst viele Menschen leistet.¹⁹ Design verstanden als „Weltentwerfen“²⁰ ist also nicht bloß eine rein gestalterische, sondern auch eine politische, gesellschaftliche und, damit verknüpft, ethische Aufgabe.

Der vorliegende Band der Schriftenreihe „Medienethik“ wird erstmals um den Untertitel „Digitale Ethik“ ergänzt. Diese Betitelung soll der Tatsache gerecht werden, dass die Schriftenreihe aufgrund der zunehmenden Digitalisierung unserer Gesellschaft den Fokus auf die Themen einer Digitalen Ethik seit etwa zehn Jahren erweitert. Während der erste Band, der 2002 erschien, ein klassisches Medienethik-Thema behandelte („Menschenbilder in den Medien“), widmete sich der zweite Band 2003 bereits der „Ethik im Netz“.²¹ Im Zuge einer digitalen Transformation des Mediensystems, die vor allem durch digitale Plattformen und die sogenannten Sozialen Medien symbolisiert wird, hat sich auch der Bereich der Medienethik selbst ausgedehnt. So wächst der Anwendungsbereich der Medienethik immer mehr mit dem der Tech-Ethiken zusammen, die sich originär mit Informations- und Kommunikationstechnologien (IKT) befassen. Längst ist die Digitalität eine grundlegende „Struktur unserer Lebenswelt“²². Zum anderen nimmt die Digitale Ethik auch außermediale Anwendungsfelder in den Blick, z. B. Ethics by Design im Kontext des Gesundheits- und Pflegebereichs, in dem zunehmend KI-Anwendungen und Robotik eine Rolle spielen. Der Begriff „Digitale Ethik“ kann somit als Synthese und Fortschreibung einer Medien- und Informationsethik verstanden werden, die den gesamten Bereich des Digitalen aus ethischer Sicht betrachtet und zahlreiche disziplinäre Verästelungen aufweist. Dazu gehört die ethische Gestaltung der Technologieanwendungen durch Ethics by Design ebenso wie die ethische Reflexion digitaler Medien und digitaler Artefakte wie bspw. Roboter, autonome Fahrzeuge, ‚smarte‘ Häuser und Städte wie auch digitale Angebote, Services und künstliche Systeme und Infrastrukturen. Überschneidungen mit den fachspezifischen Ethiken – Computerethik, Maschinenethik, Roboterethik, KI-Ethik, Algorithmenethik, Datenethik – sind naheliegend, da die Digitale Ethik sich ebenfalls mit den fokussierten Feldern dieser Tech-Ethiken befasst. Ebenso kann die Digitale Ethik auch in angrenzenden Bereichsethiken wie z. B. der Inklusionsethik hilfreiche Zugänge bieten.²³ Mit der Einrichtung des Instituts für Digitale Ethik (IDE) an der Hochschule der Medien Stuttgart im Januar 2014 wurde programmatisch ein Zelt für eine Ethik der Digitalisierung weit aufgespannt – ein Zelt, das in einem offenen und weiten Feld steht und zu Kooperationen einladen soll, um im Bild zu bleiben.

19 Vgl. auch Crawford 2021, S. 23–51; Devine 2019, S. 129–164.

20 von Borries 2016.

21 Vgl. Grimm/Capurro 2002; Capurro 2003.

22 Noller 2022, S. 9.

23 Vgl. hierzu Schlegel in diesem Band.

3 FORSCHUNGSANSÄTZE FÜR EINE MULTIDISZIPLINÄRE ETHIK

Ein wesentliches Anliegen der Schriftenreihe „Medienethik | Digitale Ethik“ wie auch des IDE ist es, multidisziplinäre Perspektiven in Tagungen und Publikationen zusammenzuführen sowie diese in der Forschung durch einen integrativen Ansatz weiter zu entwickeln, der Ethik bereits von Beginn an implementiert. Deutet man die Vorzeichen der Zeit richtig, so werden neuere digitale Technologien wie generative KI (z. B. GPT-4) und Robotik noch mehr als bisher das berufliche und private Leben beeinflussen. Hierbei stellt sich die Frage: Wird der Mensch zentral in den Blick genommen, wenn diese Technologien in der Forschung entwickelt und angewendet werden? Integrierte Forschung bedeutet, ethische, rechtliche und soziale (*ethical, legal and social*; ELS) Aspekte bereits zu Beginn in den Projekten zu berücksichtigen. Es werden also nicht mehr wie bisher ELS-Aspekte nur ‚add-on‘, quasi als Sahnehäubchen oder nachträgliches Feigenblättchen, in Technikentwicklungsprojekte eingebracht, wie z. B. durch ein Gutachten oder eine Post-Design-Evaluation – womit allerdings eine Neu-Justierung hin zu einem wertorientierten Design nicht mehr möglich war, weil die Technik schon entwickelt wurde und gleichsam das Kind schon in den Brunnen gefallen war. Der integrierte Forschungsmodus sieht vielmehr vor, dass Fachleute aus der Technikentwicklung mit jenen aus den Bereichen Ethik und Recht kooperativ und interdisziplinär zusammenarbeiten und ein ethisch-rechtliches Monitoring den Projektverlauf kontinuierlich begleitet und den Entwicklungsprozess gleichberechtigt mitsteuert. Zudem sollen Stakeholder, indirekt Betroffene, Anwendende und Fachleute aus der Praxis, wenn möglich in Studien mit einbezogen werden.²⁴ Das Feld der KI „opens up an abyss of critical questions about human knowledge, human society, and the nature of human morality“²⁵. Die Forschungsprojekte des Instituts für Digitale Ethik versuchen, diesen Ansatz einer integrierten Forschung in unterschiedlichen Themenfeldern zu etablieren.²⁶

Das EURAC-Center for Advanced Studies stellt Multi- und Interdisziplinarität in den Mittelpunkt methodischer Fragen des Zusammenspiels unterschiedlicher Wissenschafts- und Forschungstraditionen sowie Disziplinen. Dabei konzentriert sich die Arbeit auf global-regionale Zusammenhänge unter besonderer Berücksichtigung von Schnittstellen zwischen Technologie und Gesellschaft. „Global Studies“ und „Regional Studies“ haben sich aus je unterschiedlichen disziplinären Zugängen entwickelt, dabei eine multi- und interdisziplinäre Praxis eingeübt und dadurch selbst wieder im Grunde Disziplinarität entwickelt. Die Kooperation mit dem IDE und die gemeinsame Herausgabe des vorliegenden Bandes stellen eine sehr gute Gelegenheit dar, den Versuch zu unternehmen, aktuelle technologische und gesellschaftliche Entwicklungen in Verbindung zueinander zu bringen, was nicht zuletzt durch den Zugang über Ethik gelingen kann. Medien, Ethik und Digitalisierung

24 Vgl. hierzu Mehlich in diesem Band.

25 Coeckelbergh 2020, S. 61.

26 Vgl. hierzu die Übersicht auf der Homepage des IDE: <https://www.hdm-stuttgart.de/digitale-ethik/forschung>.

stellen demzufolge ein geradezu ideales Feld für die Reflexion von Multi- und Interdisziplinarität dar.

Weshalb aber ist multi-, inter- und auch transdisziplinäre Forschung so wichtig für die Forschung? Hierfür gibt es mindestens zwei gute Argumente: Erstens das Argument des Sachzwangs, also der Tatsache, dass manche Problemstellungen derart komplex sind, dass sie nur durch einen interdisziplinären Zugang, unterschiedliche Perspektiven, Methoden usw. bearbeitet werden können. Sogenannte Querschnittsthemen wie Nachhaltigkeit oder Digitalisierung werden im Rahmen von komplexen Zusammenhängen, die sich wegen ihrer Vielschichtigkeit der Lösung durch einzelne Disziplinen entziehen, betrachtet – der adisziplinären Wirklichkeit wird mit interdisziplinärer Forschung begegnet. Interdisziplinarität ist kein Allheilmittel, sondern nur eine Möglichkeit, Entwicklungen zu analysieren und zu deuten. Das zweite Argument betrifft die Innovation: Wenn Disziplinengrenzen überschritten werden, entsteht ein kreatives Potenzial, da die Spezialisierung und Beschäftigung mit Detailfragen auch zu einer Entfremdung von der Praxis führen kann. Und schließlich kann Interdisziplinarität auch helfen, neue Blickweisen einzuführen. Ein Forschungsgegenstand kann mit den eigenen Methoden und Theorien einer Disziplin ergründet werden und bringt auch Erkenntnisfortschritt im disziplinären Rahmen, der notwendig ist, um wissenschaftliches Arbeiten sowie das methodische Vorgehen zu begründen; diese Legitimation verhilft der Erkenntnis zu Akzeptanz in Wissenschaft und Gesellschaft. Demgegenüber dient der Erkenntnisgewinn durch multi- und interdisziplinäre Forschung auch der Horizonterweiterung durch die Beschäftigung mit anderen Zugängen, Theorien und Methoden. Ob aus multidisziplinären Zugängen ernsthafte interdisziplinäre Bemühungen im Sinne einer Zusammenführung bzw. Integration von Methoden und Theorien entstehen, hat auch viel mit der Bereitschaft zu Reflexion zu tun, die insbesondere aufgrund der Problemstellung notwendig wird, dass sich ein Themenkomplex wie Digitalisierung kaum über einheitliche Definitionen erschließt. Eine gemeinsame Sprache zu finden mag dabei die größte Herausforderung sein, was zusätzlich die hohe Komplexität in der Zusammenschau von Medien, Ethik und Digitalisierung erahnen lässt und zugleich die Beschäftigung dieser Zusammenhänge erforderlich macht.

4 ÜBERSICHT ÜBER DIE BEITRÄGE IN DIESEM BAND

Christopher Koska und *Michael Reder* (Hochschule für Philosophie München) eröffnen den vorliegenden Band mit ihrem Blick auf „KI-gestützte Assistenz für moralische Konfliktsituationen“, was sie am Beispiel eines Forschungsprojekts zur Algorithmisierung im Handlungsfeld der Kindeswohlgefährdung darlegen – konkret an der Frage, inwiefern die Sicherheit von Kindern und der Schutz ihres Wohlergehens durch den Einsatz von digitalen Assistenzsystemen in Jugendämtern verbessert werden kann. Es sei prinzipiell sinnvoll, über künstliche moralische Akteure nachzudenken, so die Autoren. Dabei gelte es aber, „konsequent zwischen den Dimensionen des Kalkulierens, Entscheidens und Handelns zu differenzieren“. Koska und Reder zeigen aktuelle Grenzen, Chancen und Herausforderungen für die algo-

rithmische Entscheidungsfindung in der sozialen Praxis und für die praxisorientierte Gestaltung von KI-gestützten Assistenzsystemen auf.

Marina Moreno, Adriano Mannino und Nikil Mukerji (Ludwig-Maximilians-Universität München) legen anschließend „Krisenpotenziale der Künstlichen Intelligenz“ dar. Nach einem technikgeschichtlichen Blick zurück auf die Historie der KI und die intransparente Komplexität von KI-Technologie der Gegenwart umreißen die Autor:innen die kurz-, mittel- und langfristigen „Herausforderungen und Krisenszenarien“ für den Einsatz solcher Programme. Während sich auf der politischen Ebene das gegenwärtige Fake-News-Problem „zu einem noch gefährlicheren Deepfakes-Problem auszuweiten“ droht, stehe längerfristig stets auch die Frage im Raum, inwieweit dereinst künstliches Bewusstsein und Superintelligenz möglich sind. Die „Aufwärtsspirale“ der Technologieentwicklung „könnte ein explosives Tempo annehmen“, so der Beitrag, was aber „geschehen würde, wenn tatsächlich Milliarden übermenschlich leistungsfähiger Modelle auf wissenschaftlich-technische Fragen angesetzt werden sollten – einschließlich der Selbstoptimierung der KI“ –, lasse sich „kaum absehen“.

Jan Mehlich (Rheinische Friedrich-Wilhelms-Universität Bonn) stellt in seinem Beitrag Ergebnisse zu einem Begleitforschungsprojekt zu ethischen, rechtlichen und sozialen Implikationen (ELSI) der Gesundheits-, Medizin- und Pflorgetechnik vor, in dessen Rahmen mit einem Ethics-by-Design-Ansatz ein automatisiertes Screening- und Assessment-Tool (SAT) für integrierte Forschung und Technikentwicklung im Frühstadium entstanden ist. Kernidee des im Projekt entstandenen Tools ist die Differenzierung von ethischen, rechtlichen und sozialen Implikationen zu einem „besseren Verständnis und zur effektiveren Beratung bezüglich möglicher Forschungspfade und Technologiedesigns“, das Mehlich an fünf technologieinduzierten Konfliktfeldern und Veränderungspotenzialen darlegt: Sie betreffen Gesundheitspolitik und Regulierung, ökonomische Faktoren, medizinische Berufe und den entsprechenden Arbeitsmarkt sowie Umwelt und Gesellschaft. Das entwickelte ELSI-Instrumentarium möge hierbei „nicht als Prüftool, sondern als Inspirations- und Lerntool verstanden werden“, also als Reflexionsinstanz, so der Autor.

Einen Ethics-by-Design-Ansatz greift auch *Marcel Schlegel* (Hochschule der Medien Stuttgart) auf. In seinem Beitrag zum „Inklusionsparadigma aus Sicht der (Digitalen) Ethik“ blickt er auf „Werte, Funktionen und Schritte (zu) einer Inklusionsethik“. Er arbeitet dieses Konzept für die gesellschaftlich breite Eingliederung von Behinderten bzw. Menschen mit Beeinträchtigungen in drei Schritten aus: Zunächst fragt der Autor, welchen Blick überhaupt die Ethik auf das Inklusionskonzept wirft, „welche ethischen Traditionen darin zum Vorschein kommen“ und welche Funktionen somit einer Inklusionsethik zukommen könnten. Daraufhin beschreibt Schlegel, „wie das Inklusionskonzept erweitert werden könnte, wenn sich der Gegenstandsbereich um Instrumente zur Verbesserung von digitaler Teilhabe vergrößert“. Abschließend identifiziert und spezifiziert er „im Begründungsdiskurs referierte Werte von Inklusion“, speziell auch solcher Werte, „die im Rahmen der Digitalisierung relevant werden“.

In der Schnittmenge von gesellschaftlichen und ökologischen Diskursen ist der Beitrag von *Oliver Zöllner* (Hochschule der Medien Stuttgart) angesiedelt: „Plünderung oder Sharing?“ Hier ist das Fallbeispiel auf den ersten Blick eher banal und sehr lokal – ein öffentlicher Büchertauschschrank in einem Stadtviertel im Ruhrgebiet –, weitet sich in einer am Ort ausgetragenen polarisierten und mit erheblichen Vorwürfen und Empörung aufgeladenen Social-Media-Diskussion allerdings zu einem Konflikt, der auf größere gesellschaftliche Spaltungen verweist. Das unentgeltliche Teilen von Büchern, also die Wiederverwertung von ansonsten ungenutzten Gütern, wird im empirisch (per Diskursanalyse) und philosophisch (im Rückgriff auf Aristoteles sowie zeitgenössische Moralismusdebatten) nachgezeichneten Fall von einigen Bewohnern des Stadtteils als „Plünderung“ skandalisiert. Dies lässt, so der Autor, bei den Diskutanten auf einen „selbstbezüglichen Moralismus“ schließen. „Epistemische Irrtümer“ führten „in Verbindung mit irrationalen kontrafaktischen Behauptungen zu einer maßlosen Übermaßkritik“ – ein durchaus in vielen Online-Debatten zu beobachtender Befund, der auf Disruptionen im digitalen Raum verweist.

Harald Pechlaner und *Michael de Rachewiltz* (Center for Advanced Studies, EURAC Research, Bozen/Bolzano) skizzieren – ausgehend von der klassischen philosophischen Leitfrage, was der Mensch sei und wer ihn erklären könne – zwei mehr oder weniger entgegengesetzte ethisch fundierte Menschenbilder: den „Homo digitalis“ und den „Homo montanus“. Indem der Homo digitalis immer erreichbar sein wolle und im Dauermodus digitale Medien konsumiere, ermögliche er u. a. den „Überwachungskapitalismus“ mit seinem problematischen Ausbeutungsgestus. Doch auch bei diesem Personenkreis wachse „die Sehnsucht nach Zeit, Ruhe, Spiritualität, der Einsamkeit und dem Abschalten“. Der Homo montanus dagegen sei an einer rauen und unmittelbarer erfahrbaren Welt der Berge orientiert, in der die Menschen sich ein Wissen aneignen müssten, „das für das (Über-)Leben von zentraler Wichtigkeit ist“: „In den Bergen zu leben bedeutet, Verantwortung zu zeigen gegenüber der Natur, denn auf sie kann man nur reagieren, was ein hohes Maß an Flexibilität und Einfühlungsvermögen braucht, Verantwortung aber vor allem auch gegenüber sich und dem menschlichen Umfeld.“ Ein Lösungsansatz, so die Autoren, liege „im mittleren Weg“ der beiden Existenzformen: „Nachhaltigkeit und Digitalisierung finden gemeinsam den richtigen Rhythmus und ermöglichen so eine selbstbestimmte Lebenshaltung mit einer klaren Vorstellung vom guten Leben.“

Susanne Kuhnert (Hochschule der Medien Stuttgart) führt unter dem Titel „Wunderbare Ethik?“ die literarische Erzählform Märchen „als Bildungsmedium für eine digitale Ethik“ ein – ein narrativer Ansatz, dem bisher in der Tat viel zu wenig Aufmerksamkeit gewidmet worden ist. Während die Psychoanalyse schon seit Längerem mit Märchen arbeite, um persönliche Konflikte zu verstehen und zu lösen, könne eine digitale Ethik, so die Autorin, „diese Methode aufgreifen, um Wertekonflikte der Digitalisierung zu thematisieren und um die Arbeit an der Konfliktlösung anzuregen“ – und dies keineswegs nur für die Zielgruppe Kinder. Denn solche Erzählungen ermöglichten es, die Vorstellung eines „guten Lebens“ anschaulich zu durchdenken, schreibt Kuhnert, die auch Mitherausgeberin und Koautorin des 2021 am IDE erschienenen Buchs zu „Märchen und Erzählungen der Di-

gitalen Ethik“ ist.²⁷ „Das Märchen bietet dabei als Gattung mehr Freiraum für die Phantasie als andere literarische Gattungen, weil das Übernatürliche als ‚normales‘ Element einfließen kann.“

Mirjam Gruber, Valeria von Miller und Michael de Rachewiltz (Center for Advanced Studies, EURAC Research, Bozen/Bolzano) blicken auf eine enorm dringliche und diskursiv oft konflikthafte, aber reale Erzählung der Gegenwart: den Klimawandel. Ihr Beitrag „Einfach nur heiß? Die visuelle Konstruktion einer Hitzewelle in Zeiten der Klimakrise“ analysiert die Eigenlogik der Bebilderung der Online-Berichterstattung über eine Hitzewelle in der Region Südtirol 2022. „Bilder von hungernden Eisbären, schmelzenden Gletschern, anhaltenden Dürren oder großflächigen Hochwassern gehören mittlerweile zur populären Bebilderung der Klimakrise in den Medien und darüber hinaus“, so die Autor:innen. Aber wie wird eine an Ort und Stelle real erfahrbare Hitzewelle visuell dargestellt – und inwieweit werden in der journalistischen Berichterstattung konkrete Verbindungen zum Klimawandel gezogen? Der Beitrag kommt zu einem eher skeptischen Schluss: „Fotos von Menschen im Schwimmbad vermitteln (...) in der untersuchten Berichterstattung zur Hitzewelle eher positive Emotionen und stellen somit Gefahren für Mensch, Tier und Umwelt in den Hintergrund.“ Für eine verantwortungsvolle Orientierungsfunktion des Journalismus ist dies ein desaströser Befund.

Jan Doria (Hochschule der Medien Stuttgart) präsentiert in seinem Beitrag „Zwischen Utopie und Dystopie: Künstliche Intelligenz als nichtstattfindendes Zentralereignis (NSZE) in Fernsehmagazinbeiträgen der ARD“ Befunde aus einer erzähltheoretisch fundierten Analyse der KI-Narrative im journalistischen Diskurs. Sie basieren auf einer klassischen Inhaltsanalyse (statistische Auswertung) und einer narrativ-semiotischen Korpusauswertung von 30 Magazinbeiträgen zum Thema im Zeitraum von Anfang 2020 bis Anfang 2021. „Die Fernsehmagazinbeiträge der ARD zum Thema Künstliche Intelligenz bilden im Untersuchungszeitraum“, so der Autor, „eine hohe Vielfalt unterschiedlicher lebensweltlicher Kontexte ab.“ In seiner Stichprobe ließ sich dabei etwa ein Drittel der Beiträge einem utopisch geprägten und zwei Drittel einem dystopisch geprägten Metanarrativ zuordnen. Neben diesen beiden hauptsächlich prägenden Diskursorientierungen konnte Doria in den untersuchten TV-Beiträgen aber auch noch drei besonders auffällige sogenannte „narrative Mutationen“ identifizieren: das „nichtstattfindende Zentralereignis“, die „Verhinderung des Zentralereignisses“ und das „semantische Gefängnis“, für die er jeweils einen Definitionsvorschlag vorlegt.

Abschließend stellt *Laura Braxmaier* (Absolventin der Hochschule der Medien Stuttgart) Ergebnisse ihrer Bachelorarbeit zum Thema „Querdenken 711 und nihilistische Haltungen“ vor, die Zusammenfassung ihrer „qualitativen Untersuchung nihilistischer Haltungen im Kontext des Verschwörungsglaubens in der Coronapandemie“. Sie beleuchtet also die Verfestigung und tieferen Ursachen verschwörungstheoretischer Narrative während der Covid-19-Pandemie an einem prominenten lokalen Stuttgarter Beispiel. Auf der empirischen Ebene kommen hierfür sechs Tiefeninterviews mit teils exponierten Anhängern der Szene und das Auswertungs-

27 Vgl. Grimm/Kuhnert 2021.

verfahren der qualitativen Inhaltsanalyse zum Einsatz. Die Autorin legt schlüssig dar, dass die Ausprägung nihilistischer Haltungen mit dem Ausprägungsgrad der individuellen Verschwörungsmotivität in Zusammenhang stehen kann: „Je stärker die Verschwörungsmotivität der Probandinnen, desto ausgeprägter zeigten sich auch die nihilistischen Haltungen in ihren Aussagen und damit ein intensiverer politischer und epistemologischer Nihilismus sowie eine stärker ausgeprägte Herdenbildung wie auch die Tendenz zu einem passiven Nihilismus.“ Die Studie führt vor Augen, dass die Einstellungen der befragten „Querdenker“ eindeutig „nihilistische Haltungen im Kontext des Verschwörungsglaubens aufweisen und diese keinesfalls zu verharmlosen sind.“ Für Analysen auch zukünftiger polarisierter gesellschaftspolitischer Diskurse ist diese Untersuchung instruktiv.

Bei Multidisziplinarität gibt es oft ein disziplinäres Nebeneinander auf demselben oder einem ähnlichen Themengebiet ohne strukturierte Zusammenarbeit oder irgendwelche Synthesebemühungen. Die eigentliche Interdisziplinarität ist an der Lösung komplexer Probleme durch Zusammenarbeit mehrerer Disziplinen interessiert. Zusammengesetzte Interdisziplinarität betrifft jene, wo drängende Probleme eine Zusammenarbeit motivieren (z. B. Klimaschutz, Globalisierung, Friedensforschung, aber auch Digitalisierung), wobei das Fachensemble durch den komplexen Problembereich zusammengehalten wird.²⁸ Die Studien im vorliegenden Band, die allesamt im Spannungsfeld von Medien, Ethik und Digitalisierungsprozessen angesiedelt sind, verweisen in ihrer Multidisziplinarität auf die notwendige Vielfalt an methodischen Zugängen. Das emergente Feld der Digitalen Ethik kann sich vor allem weiterentwickeln, wenn es über die weithin üblichen Fachgrenzen hinaus blickt und Verknüpfungen – oder im Sinne der Digitalität: Vernetzungen – schafft. Dieser Sammelband will hierzu ein kleiner Beitrag sein.

BIBLIOGRAFIE

- Amlinger, Carolin/Nachtwey, Oliver (2022): *Gekränkte Freiheit. Aspekte des libertären Autoritarismus*. Berlin: Suhrkamp.
- Bender, Emily M./Gebru, Timnit/McMillan-Major, Angelina/Shmitchell, Shmargaret (2021): On the dangers of stochastic parrots: Can language models be too big? In: *FAccT '21. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March 3–10, 2021, S. 610–623. <https://doi.org/10.1145/3442188.3445922>.
- Bommasani, Rishi/Hudson, Drew A./Adeli, Ehsan et al. (2022): On the opportunities and risks of foundation models. In: *arXiv:2108.07258v3*. <https://arxiv.org/abs/2108.07258>.
- Borries, Friedrich von (2016): *Weltentwerfen. Eine politische Designtheorie*. Berlin: Suhrkamp.
- Capurro, Rafael (2003): *Ethik im Netz (Medienethik, Bd. 2)*. Stuttgart: Steiner.
- Coeckelbergh, Mark (2020): *AI ethics*. Cambridge, London: MIT Press.
- Crawford, Kate (2021): *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. New Haven, London: Yale University Press.
- Devine, Kyle (2019): *Decomposed: The political ecology of music*. Cambridge, London: MIT Press.
- Donges, Patrick (2022): Digitalisierung der politischen Kommunikation. Folgen der automatisiert algorithmischen Vermittlung auf die Sichtbarkeit und Zurechenbarkeit des Politischen. In: *Köl-*

28 Vgl. Jungert et al. 2013.

- ner Zeitschrift für Soziologie und Sozialpsychologie 74 (Suppl 1), S. 209–230. <https://doi.org/10.1007/s11577-022-00834-7>.
- Eisenbrand, Jochen (2023): Mehr als 200 neue Schrott-Websites: Kommen jetzt die AI Content Farmen? Online: <https://omr.com/de/daily/ai-content-farmen/> (letzter Zugriff: 29.06.2023).
- Esposito, Elena (2022): *Artificial communication: How algorithms produce social intelligence*. Cambridge/London: MIT Press. <https://doi.org/10.7551/mitpress/14189.001.0001>.
- Gertz, Nolen (2019): *Nihilism*. Cambridge, London: MIT Press.
- Grimm, Petra/Capurro, Rafael (Hrsg.) (2002): *Menschenbilder in den Medien – ethische Vorbilder?* (Medienethik, Bd. 1). Stuttgart: Steiner.
- Grimm, Petra/Müller, Michael/Trost, Kai Erik (2021): *Werte, Ängste, Hoffnungen. Das Erleben der Digitalisierung in der erzählten Alltagswelt*. Baden-Baden: Academia.
- Grimm, Petra/Kuhnert, Susanne (Hrsg.) (2021): *Märchen und Erzählungen der Digitalen Ethik*. Stuttgart: Institut für Digitale Ethik. <https://hdms.bsz-bw.de/frontdoor/deliver/index/docId/6693/file/maerchenbuch.pdf> (letzter Zugriff: 29.06.2023).
- Habermas, Jürgen (2021): Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit. In: Martin Seeliger/Sebastian Seignani (Hrsg.): *Leviathan 37* (Sonderband), Baden-Baden: Nomos, S. 470–500. <https://doi.org/10.5771/9783748912187>.
- Hepp, Andreas (2020): *Deep mediatization*. London, New York: Routledge.
- Hepp, Johannes (2022): *Die Psyche des Homo Digitalis. 21 Neurosen, die uns im 21. Jahrhundert herausfordern*. München: Kösel.
- Jackob, Nikolaus/Schultz, Tanjev/Jakobs, Ilka/Quiring, Oliver/Schmer, Christian/Ziegele, Marc/Viehmann, Christina (2023): *Medienvertrauen in Deutschland*. Bonn: Bundeszentrale für politische Bildung.
- Jungert, Michael/Romfeld, Elsa/Sukopp, Thomas/Voigt, Uwe (Hrsg.) (2013): *Interdisziplinarität. Theorie, Praxis, Probleme*. 2. Aufl. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Lund, Holger/Michel, Burkard/Zöllner, Oliver (2022): Die Vinylschallplatte als Zeichen- und Handlungsträger gesellschaftlicher Transformationen in der Digitalisierung. In: Christian Schwarzenegger/Erik Koenen/Christian Pentzold/Thomas Birkner/Christian Katzenbach (Hrsg.): *Digitale Kommunikation und Kommunikationsgeschichte. Perspektiven, Potentiale, Problemfelder* (Digital Communication Research, Bd. 10). Berlin: Böhland & Schremmer, S. 343–373. <https://doi.org/10.48541/der.v10.13>.
- Mahowald, Kyle/Anna A. Ivanova/Idan A. Blank/Nancy Kanwisher/Joshua B. Tenenbaum/Evelina Fedorenko (2023): Dissociating language and thought in large language models: A cognitive perspective. In: ar-Xiv:2301.06627. <https://doi.org/10.48550/arXiv.2301.06627>.
- Noller, Jörg (2022): *Digitalität. Zur Philosophie der digitalen Lebenswelt* (Schwabe reflexe, Bd. 75). Basel: Schwabe.
- Risse, Mathias (2022): Artificial intelligence and the past, present, and future of democracy. In: Silja Voeneke/Philipp Kellmeyer/Oliver Mueller/Wolfram Burgard (Eds.): *The Cambridge handbook of responsible artificial intelligence: Interdisciplinary perspectives*. Cambridge, New York, Port Melbourne, New Delhi, Singapore: Cambridge University Press, S. 85–103. <https://doi.org/10.1017/9781009207898.009>.
- Schicha, Christian/Ingrid Stapf/Saskia Sell (Hrsg.) (2021): *Medien und Wahrheit. Medienethische Perspektiven auf Desinformation, Lügen und „Fake News“* (Kommunikations- und Medienethik, Bd. 15). Baden-Baden: Nomos.
- Schilling, Erik (2020): *Authentizität. Karriere einer Sehnsucht*. München: Beck.
- Schrey, Dominik (2017): *Analoge Nostalgie in der digitalen Medienkultur*. Berlin: Kadmos.
- Seemann, Michael (2021): *Die Macht der Plattformen. Politik in Zeiten der Internetgiganten*. Berlin: Links.
- Staab, Philipp (2019): *Digitaler Kapitalismus. Markt und Herrschaft in der Ökonomie der Unknappheit*. Berlin: Suhrkamp.
- Stäheli, Urs (2021): *Soziologie der Entnetzung*. Berlin: Suhrkamp.
- te Wildt, Bert (2015): *Digital Junkies. Internetabhängigkeit und ihre Folgen für uns und unsere Kinder*. München: Droemer.

- Zöllner, Oliver (2020): Klebrige Falschheit. Desinformation als nihilistischer Kitsch der Digitalität. In: Petra Grimm/Oliver Zöllner (Hrsg.): Digitalisierung und Demokratie. Ethische Perspektiven (Medienethik, Bd. 18). Stuttgart: Steiner, S. 65–104.
- Zuboff, Shoshana (2018): Das Zeitalter des Überwachungskapitalismus. Aus dem Englischen von Bernhard Schmid. Frankfurt/New York: Campus.

KI-GESTÜTZTE ASSISTENZ FÜR MORALISCHE KONFLIKTSITUATIONEN

Zur Algorithmisierung im Handlungsfeld der Kindeswohlgefährdung

Christopher Koska, Michael Reder

1 EINLEITUNG

Das weitläufig bekannteste Gedankenexperiment im Zusammenhang mit moralisch kalkulierenden Algorithmen basiert auf dem sogenannten Trolley-Problem. Juristen und Philosophen beschäftigen sich seit fast 100 Jahren mit der moralphilosophischen Bewertung diverser Varianten dieses Problems.¹ Durch die Inbetriebnahme der Online-Plattform *moral machine*² wurde der ethische Diskurs in den letzten Jahren geöffnet und die Anzahl denkbarer Szenarien zugleich enorm gesteigert. Aus einer Datenbank mit ca. 26 Millionen Möglichkeiten lassen sich inzwischen verschiedenste Szenarien für selbstfahrende Fahrzeuge generieren.³ Mit einem einzigen Klick können die Besucher:innen der Website dann zwischen unvermeidbaren Schadensfolgen abstimmen. 40 Millionen Entscheidungen wurden bis zur Veröffentlichung des Nature-Artikels von Awad et al. (2018) bereits gesammelt und ausgewertet.⁴ Die Notwendigkeit zur Öffnung des Diskurses begründen die Autoren:innen folgendermaßen: „Decisions about the ethical principles that will guide autonomous vehicles cannot be left solely to either the engineers or the ethicists (...) both groups will need to understand the origins of the ethical principles that are programmed into these vehicles. In other words, even if ethicists were to agree on how autonomous vehicles should solve moral dilemmas, their work would be useless if citizens were to disagree with their solution. (...) Any attempt to devise artificial intelligence ethics must be at least cognizant of public morality.“⁵

Gegen das *moral machine experiment* lässt sich u. a. einwenden, dass diese Form der theoretischen Auseinandersetzung (mit den unterschiedlichsten Varianten des Trolley-Problems) den Blick auf die Praxis (des Datensammelns) in der Automobilbranche verstellt.⁶ Im Rahmen dieses Beitrags wird gezeigt, welchen Mehr-

- 1 In seiner ursprünglichen Form (vgl. Engisch 1930) beschreibt das Experiment eine Situation, in der ein Weichensteller durch den Wechsel der Fahrspur (einer Straßenbahn) einen Unfall mit Todesopfern zwar nicht verhindern, aber die Zahl der Opfer verringern könnte.
- 2 Siehe online: <https://www.moralmachine.net> (Abfrage: 10.11.2022).
- 3 Awad et al. 2018, S. 65.
- 4 Vgl. ebd., S. 59.
- 5 Ebd.
- 6 Vgl. Koska 2018, S. 67–90.

wert die Analyse von authentischen Daten (hier: Fallakten von Jugendämtern) für die algorithmische Entscheidungsfindung in der sozialen Praxis bietet und welche Funktion die Beschäftigung mit fiktiven Szenarien (synthetischen Daten) einnehmen kann. Ziel ist es, die Chancen und Herausforderungen von KI-gestützten Entscheidungssystemen in hoch konfliktiven Situationen auszuloten, um ein besseres Verständnis darüber zu erlangen, welche Software-Applikationen in welchen Bereichen und unter welchen Bedingungen aus moralischen Gründen sinnvollerweise eingesetzt werden können – und wo ihr Einsatz begrenzt werden muss. Der Beitrag schließt an die Ausgangsfrage des transdisziplinären Forschungsverbunds KAIMO (Kann ein Algorithmus moralisch kalkulieren?)⁷ an und diskutiert die skizzierte Fragestellung mit Blick auf die digitale Operationalisierung im Feld der Kindeswohlgefährdung.

Es wird zunächst dargelegt, warum es prinzipiell sinnvoll ist, über *artificial moral agents* nachzudenken. Dabei gilt es, so lautet eine zentrale These des Beitrags, konsequent zwischen den Dimensionen des Kalkulierens, Abwägens, Entscheidens und Handelns zu differenzieren. Anknüpfend daran folgt eine problemorientierte Beschreibung der Forschungsfrage des Konsortialprojekts KAIMo sowie des aktuellen Prototyps. Insbesondere wird darüber nachgedacht, inwiefern die Sicherheit von Kindern und der Schutz ihres Wohlergehens durch den Einsatz von digitalen Assistenzsystemen in Jugendämtern verbessert werden kann. Aufbauend auf diesen Erkenntnissen werden in einem nächsten Schritt, aktuelle Grenzen, Chancen und Herausforderungen für die algorithmische Entscheidungsfindung in der sozialen Praxis reflektiert. Abschließend soll in einem vierten und letzten Schritt eine praxisorientierte Gestaltungsperspektive für KI-gestützte Assistenzsysteme aufgezeigt werden.

2 DIFFERENZIERUNG ZWISCHEN KALKULIEREN, ABWÄGEN, ENTSCHIEDEN UND HANDELN

Seit gut 20 Jahren wird unter dem Stichwort „Engineering Artificial Moral Agents“⁸ intensiv darüber nachgedacht, ob künstliche Intelligenz eine sehr spezifische menschliche Eigenschaft übernehmen kann, und zwar die, moralisch zu handeln. Hintergrund dieser Diskussion waren technische Entwicklungen, die der Maschine scheinbar mehr und mehr so etwas wie ‚autonomes Handeln‘ ermöglichen. Maschinen haben im Zuge dessen nicht nur standardisierte, automatisierte Prozesse übernommen, sondern durch die ihnen spezifische Trainings- und Lernfähigkeit auch neue Räume erschlossen, die handlungsanalog zu sein scheinen. Mit diesen

7 Das Projekt wird vom Bayerischen Forschungsinstitut für Digitale Transformation (bidt) über einen Zeitraum von drei Jahren (von Januar 2021 bis Dezember 2023) gefördert. Projektkonsortium: Technische Hochschule Nürnberg, Hochschule für Angewandte Wissenschaften Würzburg-Schweinfurt und Hochschule für Philosophie München. Weitere Infos siehe: <https://www.kaimo.bayern> (Abfrage: 10.11.2022).

8 Siehe etwa Floridi/Sanders 2004; Wallach/Allen 2009; Misselhorn 2018; Brieger 2018; Weber 2018 und van Wynsberghe/Robbins 2019.

Möglichkeiten zu eigenständigem Handeln wird nun die Frage relevant, ob die Maschine auch etwas tun kann, was eigentlich nur der Mensch tun kann, nämlich moralisch zu entscheiden und zu handeln.

Katharina Zweig hat in diesem Zusammenhang auf einige grundlegende Probleme hingewiesen.⁹ Denn die Informatik braucht eine formalisierte Definition von normativen Begriffen. Erst dann können diese in einen digitalen Code übersetzt werden. Die philosophische Frage nach der Normativität ist jedoch selten so binär, wie sie die Informatik gerne hätte. Normative Konflikte sind oft kompliziert; zudem verändern sie sich zeitlich schnell und können deswegen nur selten in einer eindeutigen und zeitlosen Heuristik theoretisch gefasst werden. Dies gilt auch für Normen selbst. Viele Normen sind oft so formal, weil nur so ihre Allgemeingültigkeit begründet werden kann. Wenn diese Normen dann material gefüllt werden – wenn beispielsweise diskutiert wird, was genau Kindeswohl inhaltlich meint – ist diese Eindeutigkeit schnell begrenzt. Mit Blick auf verschiedene soziale, kulturelle oder zeitliche Kontexte und mit Blick auf den Einzelfall werden dann Normen oft unterschiedlich material gefüllt. Diese Kontextualität und auch inhärente Dialektik von Normativität widersprechen der binären Logik der digitalen Technologie.

Die Frage ist, ob und wo die Möglichkeiten und Grenzen einer Übersetzung zwischen der Sphäre des Normativen und des Digitalen liegen. Einerseits wird im aktuellen Diskurs oft betont, dass Maschinen durch die digitale Verarbeitung menschlichen Entscheidens und Handelns dessen Verhalten reproduzieren können. Andererseits betonen Philosoph:innen und auch viele Praktiker:innen, dass angesichts komplexer Konfliktlagen sich der Einzelfall meist anders darstellt.¹⁰ Eine ethisch begründete Entscheidung zu fällen, bedeutet, unabhängig von diesen Situationen *und gleichzeitig* kontextsensibel zu reflektieren und auf der Basis überzeugender Argumente zu entscheiden. Entscheidungssituationen sind dann moralisch besonders komplex und schwierig, wenn mindestens zwei oder sogar mehrere normative Positionen einander gegenüberstehen, die letztlich unvereinbar sind. Ethisch betrachtet gibt es dann oftmals keine eindeutige Lösung, obwohl man sich entscheiden muss. Moralische Entscheidungen sind für Menschen dann besonders überzeugend, wenn sie kohärent, logisch gültig und plausibel sind. Zentral dabei ist, dass es vor dem skizzierten Kontext gute Gründe gibt, die andere Personen verstehen und die die Situation der betroffenen Menschen verbessern.

In diesem Zusammenhang ist es wichtig, verschiedene Arten von moralischen Konflikten zu unterscheiden. Moralische Dilemmata liegen dann vor, wenn zwei moralische Forderungen sich gegenseitig ausschließen und nicht beide gleichzeitig verfolgt werden können. Auflösbare moralische Konflikte liegen wiederum vor, wenn Konflikte zum Beispiel durch neue Informationen oder die Entdeckung von Fehlern gelöst werden können. Diese beiden Typen von moralischen Konflikten stehen für die Frage nach *artificial moral agents* nicht im Fokus. Besonders wichtig

9 Siehe Zweig 2019.

10 Im deutschsprachigen Diskurs wurde diese Beobachtung u. a. als „kollektive Dimension von Algorithmen“ (Jaume-Palasi/Spielkamp 2017: 2) beschrieben, als eine Form der „numerischen Allgemeinheit“ (Heesen 2018: 47) charakterisiert und zuletzt auch unter dem Aspekt der „Gruppenhaft“ (Mühlhoff 2022) diskutiert.

ist vielmehr ein dritter Typ von moralisch gewichtigen Konflikten. Diese sind entscheidbar; allerdings stehen auf allen Seiten gewichtige moralische Forderungen, die teils zu verschiedenen Perspektiven auf den Konfliktfall werden. Gerade unter sozialen Gesichtspunkten ist dieser dritte Typus umstritten und kann massive Folgen für die Betroffenen haben. Es wird gegenwärtig mit Blick auf unterschiedliche Konfliktfelder analysiert und diskutiert, ob und inwiefern Maschinen diese Abwägung und Entscheidung moralischer Konfliktsituationen unterstützen oder gar übernehmen können.¹¹

Um die möglichen Funktionsweisen digitaler Technologien bestmöglich bestimmen zu können, ist es in diesem Zusammenhang wichtig, zwischen dem Kalkulieren, Abwägen, Entscheiden und Handeln zu unterscheiden. Beim Entscheiden geht es darum, den Konflikt aufzulösen. Angesichts verschiedener Optionen soll herausgefunden werden, welche davon die beste ist, um den Konflikt zu lösen. Angesichts der Kindeswohlgefährdung ist dies der Konflikt zwischen Kindeswohl und dem Recht der Eltern auf Erziehung. Durch die Abwägung des Konflikts mit Blick auf die konkreten Umstände wird dann entschieden, ob ein Kind aus der Familie heraus- und in Inobhutnahme gegeben werden soll. Bezüglich digitaler Assistenzsysteme spielt die Entscheidungsdimension für unsere Überlegungen keine zentrale Rolle. Denn abgesehen davon, dass sich Maschinen in kein normativ-wollendes Verhältnis zu bestimmten Werturteilen setzen können,¹² obliegen die Entscheidungen besonders weitreichender und v. a. auch Grundrechte betreffender Konflikte in einem Rechtsstaat letztlich immer der Judikative. Dieses Feld bleibt also der menschlichen Entscheidungsfähigkeit, in diesem Fall den Richter:innen, überlassen.

Eine weitere Dimension ist die der Handlung. Dabei geht es um eine möglichst angemessene und überzeugende Bewältigung des Konflikts durch einzelne konkrete Gestaltungsschritte. Nach einer Entscheidung setzen Menschen diese handelnd um. Beispielsweise indem die spezifischen Kontexte einer Konfliktsituation angemessen reflektiert und in den entsprechenden Handlungsschritten mitbedacht werden. Auch diese Dimension spielt für die moralische Maschine gegenwärtig noch keine zentrale Rolle. Denn insbesondere wenn es um die Anerkennung moralischer Werte geht, bleibt festzuhalten: *There is no "I" in "Robot"*¹³, das über die Fähigkeit der personalen Selbstbestimmung verfügt und von diesem reflektierenden Standpunkt aus die Grundfragen der Ethik: „Was sollen wir tun?“, „Wie sollen wir handeln?“ angehen könnte.¹⁴ Assistenzsysteme können lediglich das Handeln von Menschen unterstützen, aber nicht selbst als moralische Akteure in Erscheinung treten.

Die zentrale Dimension scheint vielmehr die der moralischen Abwägung zu sein, die *vor* der Entscheidung und der Handlung liegt, diese aber auch begleiten

11 Siehe z.B. Cornelissen et al. 2022; Meier et al. 2022; Coin/Dubljević 2022 und Gundersen/Bærøe 2022.

12 Autonome Maschinen agieren „von selbst“, aber nicht „aus sich selbst“, wie der Natur- und Technikphilosoph Benjamin Rathgeber in seiner Antrittsvorlesung an der Hochschule für Philosophie dargelegt hat.

13 Vgl. Grau 2006.

14 Vgl. Koska (im Druck).