F·E·S·T Forschung
Band 2

Framing KI

Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz

Frederike van Oorschot / Selina Fucker (Hrsg.)

UNIVERSITÄTS-
BIBLIOTHEK
HEIDELBERG

FEST Forschung

Band 2

Reihenherausgeberinnen und -herausgeber

Benjamin Held, Madlen Krüger, Magnus Schlette, Philipp Stoellger,
A. Katarina Weilert

Reihenbeschreibung

Die Reihe »FEST Forschung« versammelt Forschungsbeiträge aus der laufenden wissenschaftlichen Arbeit der interdisziplinären Forschungsstätte der Evangelischen Studiengemeinschaft (FEST) in Heidelberg.

Das Themenspektrum der Reihe spiegelt die Schwerpunkte der Forschung an der FEST: Frieden – Nachhaltige Entwicklung – Religion, Recht und Kultur – Theologie und Naturwissenschaft sowie die fachliche Expertise der einzelnen Mitarbeiter:innen wider. Die Bände und Beiträge der Reihe nehmen dabei aktuelle gesellschaftliche Themen und Diskurse in den Blick. Sie liefern Analysen für die Wissenschaft und geben Orientierung für Kirchen, Gesellschaft und Politik.

Die wissenschaftliche Qualität der Bände der Reihe wird durch einen wissenschaftlichen Beirat sichergestellt, der sich aus den Mitgliedern des wissenschaftlichen Kuratoriums der FEST zusammensetzt. Alle Bände durchlaufen ein mehrstufiges, von den Band- und Reihenherausgeber:innen durchgeführtes Review-Verfahren.

Wissenschaftlicher Beirat

OKRat Dr. Niklaus Blum (Rechtswissenschaften) München
Prof. Dr. Armin von Bogdandy (Rechtswissenschaften) Heidelberg
Regionalbischöfin em. Susanne Breit-Keßler (Theologie) München (seit 2017 als Ehrenmitglied)
Prof. Dr. Christopher Daase (Politikwissenschaft/Friedens- und Konfliktforschung) Frankfurt/M.
Prof. Dr. Horst Dreier (Öffentliches Recht) Reinbek
Prof. Dr. Verena V. Hafner (Informatik) Berlin
Kirchenpräsident Dr. Volker Jung (Theologie) Darmstadt
Prof. Dr. Nicole C. Karafyllis (Philosophie) Braunschweig
Dr. Friederike Krippner (Germanistik/Evangelische Theologie) Berlin
Prof. Dr. Hartmut Leppin (Geschichte) Frankfurt/M.
Prof. Dr. Michael Moxter (Theologie) Hamburg (Vorsitzender)
Prof. Dr. Olaf Müller (Philosophie) Berlin (Begutachter des vorliegenden Bandes)
Prof. Dr. Sigrid Stagl (Ökonomie) Wien
Prof. Dr. Andreas Unterberg (Medizin/Neurochirurgie) Heidelberg
Prof. Dr. Ulrich Willems (Politikwissenschaft) Münster
Prof. Dr. Monika Wohlrab-Sahr (Kulturwissenschaften/Soziologie) Leipzig

Frederike van Oorschot, Selina Fucker (Hrsg.)

Framing KI

Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz

UNIVERSITÄTS-
BIBLIOTHEK
HEIDELBERG

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

Die durch Papier und Druck entstandenen Emissionen werden über die Klimaschutzprojekte der Klima-Kollekte kompensiert.

Publiziert bei heiBOOKS, 2022

Universität Heidelberg/Universitätsbibliothek
heiBOOKS
Grabengasse 1, 69117 Heidelberg
https://books.ub.uni-heidelberg.de/heibooks

Die Online-Version dieser Publikation ist auf heiBOOKS, der E-Book-Plattform der Universitätsbibliothek Heidelberg, https://books.ub.uni-heidelberg.de/heibooks, dauerhaft frei verfügbar (Open Access).
urn: urn:nbn:de:bsz:16-heibooks-book-1106-5
doi: https://doi.org/10.11588/heibooks.1106

Layout und Satz: text plus form, Dresden

ISBN 978-3-948083-69-4 (Softcover)
ISBN 978-3-948083-68-7 (PDF)

ISSN 2749-6392 (Print)
ISSN 2749-6406 (online)

Inhalt

Einleitung

Frederike van Oorschot & Selina Fucker

Neue Entwicklungen im Digitalen verlangen nach neuen Beschreibungen, nach neuen Sprachbildern, Metaphern und Narrativen der Technologie. Nur so wird verständlich und im Wortsinn anschaulich, was neue Technologien vermögen und wo und wie sie eingesetzt werden können. Zugleich kommt diesen Beschreibungen, Sprachbildern, Metaphern und Narrativen damit eine konstruierende Wirkung für die Technik zu: Sie prägen unser Verständnis dieser Technologien und bahnen so Wege für gesellschaftliche Debatten um ihren Einsatz und mögliche Fortentwicklungen.

Diese neuen Metaphern und Frames bilden einen Teil sozialer Imaginationen (C. Taylor), welche den Rahmen individueller und gesellschaftlicher Kommunikationsprozesse bilden und gemeinsames Handeln ermöglichen.[1] Dies gilt gerade für nicht sensual erfassbare Welten wie das Psychische oder auch das Virtuelle.

Dies gilt auch und gerade für den Bereich der sogenannten »künstlichen Intelligenz«. Der Diskurs über KI bedient sich dabei zentraler anthropologischer Kategorien wie »Intelligenz«, »Lernen«, »Denken« und überträgt diese hier auf eine technologische Entwicklung. Diese ursprünglichen Beschreibungen für den Menschen und sein Handeln prägen nun unser Verständnis dieser Technologie, was zugleich sich aber auch auf das Verständnis des Menschen auswirkt. Daher ist das Wechselverhältnis zwischen Technik und Anthropologie hier sehr zentral. Die verwendeten Bilder über den Menschen imaginie-

1 Taylor 2007: 23.

Frederike van Oorschot/Selina Fucker 2022: Einleitung. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 7–11.
https://doi.org/10.11588/heibooks.1106.c15190

ren eine Angleichung der Technik an den Menschen, die bis hin zur Angst vor der Ersetzung des Menschen durch Maschinen reichen kann. Zugleich prägen technische Beschreibungen dadurch auch die Wahrnehmung des Menschen, wenn etwa das menschliche Gehirn als Denkmaschine mittels einer Struktur neuronaler Netze nachgebaut werden soll.

Diese Imaginationen künstlicher Intelligenz prägen gesellschaftliche Diskurse in unterschiedlichen Feldern. Untersuchungen der medialen Debatten zeigen, dass diese noch ziemlich neue Technologie als Chance für die Wirtschaft, aber auch als Gefahr beschrieben wird.[2] Diese Wahrnehmung prägt auch Onlinediskurse um Künstliche Intelligenz. Filmisch kommt dem Verhältnis von Mensch und Maschine ein großer Raum zu. Wie künstliche Intelligenz ethisch verantwortlich gestaltet werden kann, wird in diesen Imaginationen diskutiert und durch diese geprägt. So legte etwa die Expert*innengruppe für Künstliche Intelligenz der Europäischen Kommission 2019 Richtlinien für eine »trustworthy AI« vor, die u. a. die Resilienz, Verantwortung und Zurechenbarkeit der Systeme als Kriterien für Künstliche Intelligenz benennt. Diese werden ausgeführt und mit menschlicher Verantwortung und Zurechenbarkeit verbunden, ohne diese Kategorien im Blick auf ihre anthropologischen und Technologien Implikationen zu differenzieren. Deutlich wird hier, dass die Suche nach ethischen Kriterien zur Gestaltung von Künstlicher Intelligenz ebenfalls nach Begriffsarbeit und Reflexion der zu Grunde liegenden Imaginationen verlangt.[3]

Dieser Band ist das Ergebnis einer Tagung, die im Dezember 2019 stattgefunden hat. Die digitale Tagung diente der Vernetzung aktueller Projekte, die sich in unterschiedlichen Fächern empirisch und ethisch mit der Untersuchung und Reflexion von Metaphern und Narrativen im Feld Künstlicher Intelligenz befassen. Ziel des vorliegenden Bandes ist es, eine Übersicht über den aktuellen Forschungsstand in unterschiedlichen Disziplinen zum Thema zu bieten und Ansatzpunkte für eine weiterführende ethische Reflexion anzubieten.

Der interdisziplinäre Beitrag von Andreas Böhn, Sonja Kleinke, Marie-Hélène Adam und Katrin Strobel untersucht die zentralen Topoi sowie die Gemeinsamkeiten und Unterschiede in der Konstruktion und Verhandlung von KI in Filmen und Online-Diskussionen. Hierfür wurden qualitative Inhalts- und (kritische) Diskursanalysen durchgeführt. Sie zeigen, dass sowohl im Filmkorpus, als auch im Korpus mit den Online-Diskussionen die Menschliche Hybris und das Motiv des Menschen als Schöpfers zentral ist. Ebenfalls häufig

2 Brennen et al. 2018: 4.

3 https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai. Vgl. Fazit dieses Bandes und van Oorschot 2022.

werden die (In-)Transparenz bzw. Semitransparenz, sowie die Opazität technischer Vorgänge und deren Mystifizierung in beiden Korpora thematisiert. Machtverhältnisse, Macht und Machtlosigkeit, die Angst des Menschen durch die Technik ersetzt zu werden stellen weitere wichtige Topoi nach. Auch die Fragen der ethischen Identität von KI, sowie der Rechte von KI werden sowohl im Filmkorpus, also auch in den Online-Diskussionen häufig thematisiert.

Christian Katzenbach hat bei der Tagung das Projekt »Die diskursive und politische Konstruktion von KI« vorgestellt, in dem empirische Vergleichsstudien zur diskursiven und politischen Konstruktion von KI in Deutschland, Kanada, Israel, Großbritannien und der Schweiz durchgeführt werden.[4]

Jonas Bedford-Strohm beschreibt in seinem Praxisbericht eine konzeptionelle und ethische Exploration der datengesteuerten Medientransformation. Er entwickelt hierbei eine Typologie des Datenverarbeitungsprozesses. Er identifiziert die technologische Leistungsfähigkeit des verantwortlichen Bearbeiters, allgemeine Datenverfügbarkeit, Angemessenheit der Trainingsdaten, Berechenbarkeit der angestrebten Funktion, angewandte Methodik und die damit verbundenen Verzerrungen und Zielgerichtetheit für den Einsatz und die Optimierung von Algorithmen als notwendige Bedingungen für die ethische Erkundung komplexer Datenverarbeitung.

Der Frage, wie die in der medialen Berichterstattung über KI verwendeten Frames wirken, geht Selina Fucker in ihrem Beitrag nach. Das von ihr durchgeführte Online-Experiment zeigt, die Chancen- und Risiko-Frames vor allem emotionale Effekte haben und über diese emotionalen Effekte die Chancen- und Risikobeurteilung von KI beeinflussen. Direkte Effekte der Chancen- und Risiko-Frames auf die Chancen- beziehungsweise Risikobeurteilung konnten hingegen nicht festgestellt werden.

Florian Höhne beschreibt in seinem Beitrag die theologisch-ethischen Herausforderungen der Vorstellungswelten künstlicher Intelligenz. Sein Ausgangspunkt ist die These, dass das in theologischer Hinsicht reduktive Menschenbild des »risikoinformierten Entscheiders« die Entwicklung und Deutung sogenannter »künstlicher Intelligenz« präformiert. Er reflektiert das Wechselspiel von Menschenbildern und Technik sozialethisch anhand einer exemplarischen Pointierung des Menschenbildes vom Menschen als »risikoinformierten Entscheider«. Dies führt ihn zu der These, dass was als KI bezeichnet wird, als technisches Gebilde das Menschenbild des risikoinformierten Entscheiders verkörpert und als solche reduktiv ist.

Alexander Filipović und Julian Lamers legen in ihrem Beitrag das Ver-

4 Der Tagungsbericht ist im Band nicht dokumentiert. Mehr Informationen zum Projekt: Katzenbach, Christian 2022: Die diskursive und politische Konstruktion von KI. https://www.hiig.de/project/ki-konstruktion/ (aufgerufen am 27. 01. 2022).

hältnis von Diskursanalyse und Angewandter Ethik im Feld Künstlicher Intelligenz dar. Sie skizzieren den Diskurs über KI und betrachten diesen als ethischen Diskurs. Davon ausgehend entwickeln sie die These, dass Technikdiskurse die beschriebene politische Bezugnahme auf eine Einstellungs- oder Meinungs-Empirie die Angewandte Ethik ebenso betrifft wie die Rolle von Fachgremien.

Philipp Stoellger wirft in seinem Beitrag einen Blick auf die anthropologische Differenz zwischen Roboter und Menschen. Anhand eines Vergleiches der Metapher von Robotern als Freunde mit der Legende des Golems arbeitet er das Spezifische der Roboter heraus. Er kommt zu dem Ergebnis, dass Roboter belebte Dinge sind, die mehr als Dinge werden. Sie werden zu realen und imaginären Mitgliedern des sozialen Lebens.

In einer ausblickenden Bündelung entwirft Frederike van Oorschot auf der Grundlage der vorliegenden Einzelstudien ein Modell imaginationssensibler Ethik: Eine imaginationssensible Ethik nimmt die in diesem Band gestellte Frage nach den gesellschaftlichen Imaginationen – greifbar in Sprachbildern, Metaphern, Filmen, medialen Frames und über das hier untersuchte hinaus etwa auch in der Werbung, in Strategiepapieren u. v. a. m. – als konstitutive Aufgabe der Ethik auf und verhandelt diese prospektiv, korrelativ und retrospektiv: In transdisziplinären Analysen der gesellschaftlichen Imaginationen werden diese analysiert, aus ethischer Perspektive kritisiert und ggf. modifiziert. Der Beitrag entwirft damit das Rahmenparadigma für weitere Forschungen am untersuchten Themenfeld.

An dieser Stelle sei Allen herzlich gedankt, die zum Gelingen dieses Bandes beigetragen haben. Allen voran den Autorinnen und Autoren der vorliegenden Beiträge und darüber hinaus allen Teilnehmerinnen und Teilnehmern, die mit Projektvorstellungen und Diskussionsbeiträgen die Debatte bei der Tagung und damit auch diesen Band bereichert haben.

Den Herausgeberinnen und Herausgebern der Reihe FEST Forschung sei für die Aufnahme des Bandes in die Reihe gedankt. Prof. Hafner danken wir für das fachliche Lektorat. Ohne Anke Munos und Steffen Schröters tatkräftige Unterstützung bei Satz und Layout wäre der Band in der vorliegenden Form nicht denkbar gewesen – herzlichen Dank dafür!

Heidelberg, August 2022

Literatur

Brennen, J. Scott/Howard, Philip N./Nielsen, Rasmus K. 2018: An industry-led debate: How UK media cover artificial intelligence. RISJ Fact-Sheet.

Katzenbach, Christian 2022: Die diskursive und politische Konstruktion von KI. https://www.hiig.de/project/ki-konstruktion/ (aufgerufen am 27. 01. 2022).

Taylor, Charles 2007: Modern social imaginaries. 4. print. Durham: Duke Univ. Press (Philosophy social theory).

van Oorschot, Frederike 2022: Alles Technik oder was? in: Diebel-Fischer, Hermann/Kunkel, Nicole/Zeyher-Quattlender, Julian (Hg.): Mensch und Maschine im Zeitalter »Künstlicher Intelligenz«. Theologische Herausforderungen (Jahrestagung des Arbeitskreises für theologische Wirtschafts- und Technikethik Okt. 2021), Münster: LIT-Verlag [im Druck].

ORCID

Frederike van Oorschot https://orcid.org/0000-0003-4359-8949

Selina Fucker https://orcid.org/0000-0001-8728-3485

I. Fallstudien

Künstliche Intelligenz im Spannungsfeld gesellschaftlicher Diskurse

Filmische Science-Fiction und alltagsweltliche Online-Diskussionen

Sonja Kleinke, Andreas Böhn, Katrin Strobel & Marie-Hélène Adam

A KI als zentraler Kristallisationspunkt sozialer, medialer und ästhetischer Aushandlungs- und Repräsentationsprozesse von Zukunftsvorstellungen

»Die Furcht vor einer künstlichen Superintelligenz ist übertrieben«, titelt die *Süddeutsche Zeitung* am 23. Januar 2021.[1] Doch auch wenn ihre tatsächliche Erschaffung noch in weiter Ferne ist – die Vorstellung einer *starken* Super-K(ünstlichen)I(ntelligenz), die über Singularität, ein (möglicherweise gar dem Menschen überlegenes) Bewusstsein und eine eigene Agenda verfügt, ist ein prävalentes Motiv in sozialen und medialen Technikdiskursen. *Schwache* KI, ohne ein hochentwickeltes Bewusstsein, aber dafür hochspezialisiert für spezifische Aufgaben, dringt bereits in Form von digitalen Assistenzsystemen oder smarten Anwendungen in diverse Bereiche unseres Alltagslebens vor. In der gesellschaftlichen Digitalisierungsdebatte nimmt die Analyse und Bewertung von KI folglich einen zentralen Platz ein.[2] Dies erfordert eine »intensive Auseinandersetzung mit dem Menschenbild der Digitalisierung«[3] nicht nur

1 Christoph von Eichhorn: Die Furcht vor einer Superintelligenz ist übertrieben. In: Süddeutsche Zeitung, https://www.sueddeutsche.de/digital/ki-kuenstliche-intelligenz-superintelligenz-kampfroboter-1.5183093, veröffentlicht am 23.01.2021, letzter Zugriff am 13.02.2021.

2 Neumaier 1994, Göcke 2018.

3 Brand 2018: 7.

Sonja Kleinke et al. 2022: Künstliche Intelligenz im Spannungsfeld gesellschaftlicher Diskurse: Filmische Science-Fiction und alltagsweltliche Online-Diskussionen. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 15–54.
https://doi.org/10.11588/heibooks.1106.c15192

aus philosophisch-theologischer Perspektive,[4] sondern auch aus kulturell-ästhetischer und linguistischer Perspektive.

Zunehmende Technisierung und Informatisierung unserer Alltagswelt erhöhen deren Effizienz, bergen aber zugleich Unsicherheitsfaktoren und bilden so ein Spannungsfeld aus Steigerung und Verlust von Autonomie. Techniken der Bewertung, Klassifizierung und Selbstvermessung tangieren existenzielle Fragen der Identität und problematisieren sie auf neue Art und Weise. Doch welche Bedeutung und welches Potential wird KI in der Gesellschaft eigentlich zugeschrieben? Was kann KI? Was kann sie nicht? Haben Entwickler:innen und Endverbraucher:innen das gleiche Konzept in ihren Köpfen, wenn sie über KI reden? Was verstehen wir unter einer *Künstlichen Intelligenz*?

Die Erschaffung vollendeter KI gilt als Menschheitstraum und symbolische Zäsur eines neuen technologischen Zeitalters. KI ist einer der beliebtesten Topoi der Science-Fiction. In ihrer Inszenierung manifestieren sich nicht nur populäre Vorstellungen von Technologie, sondern auch Hoffnungen, Ängste und kulturelle Diskurse, die zudem ethische Diskussionen über menschliche Hybris und Machbarkeitsphantasien sowie mögliche Konsequenzen evozieren und somit eine breite sozio-kulturelle Debatte im Umfeld von Identität, Subjektivität und Ethik anstoßen. So setzt die medial repräsentierte Konfrontation mit KI Prozesse der Grenzüberschreitung und Reflexion epistemologischer Fragen von Körper und Sein in Gang, die auch die Identität und das Geschlecht von KI berühren.

KI ist ein zentraler Kristallisationspunkt sozialer, medialer und ästhetischer Aushandlungs- und Repräsentationsprozesse von Zukunftsvorstellungen. Ihre Analyse erfordert eine interdisziplinäre Perspektive, die sowohl gesellschaftliche Diskursivierung als auch (massen)mediale Repräsentation von KI und insbesondere die Korrelationen beider Diskursfelder berücksichtigt. In der audiovisuellen Science-Fiction konstituieren sich (Be-)Deutungsangebote rund um das Themenfeld KI auf allen Ebenen der filmischen Gestaltung. Besonders die potentielle Subjektwerdung der KI fungiert als wichtiger Topos, der narrativ, symbolisch und ästhetisch ausgestaltet und mit Bedeutung aufgeladen wird. Diese Konvergenzen von Mensch und Maschine eröffnen Fragen und Deutungsmuster, die sich auch im öffentlichen (partizipatorischen) Diskurs niederschlagen.[5]

Gesellschaftliche KI-Diskurse in Politik und Medien werden aktuell systematisch durch neue kommunikative Räume für die diskursive Konstruktion und Verhandlung im Internet komplementiert. Diese Form der Partizipa-

4 Ebd.

5 Wir danken HEiKA (Heidelberg–Karlsruhe Research Partnership) für die Anschubförderung unserer interdisziplinären Kooperation (Projekt D.801000/17.061).

tionsmöglichkeit am öffentlichen Diskurs (›bottom up‹, jenseits institutionalisierter Akteure und Rahmen) rückt neben etablierten Diskursformaten zunehmend in den Fokus von Linguistik, Kommunikations- und Medienwissenschaften,[6] ist jedoch ebenso wie die traditionellen Formate bislang im Hinblick auf die diskursive Konstruktion und Verhandlung komplexer KI-Identitäten nicht systematisch erforscht. Zu ausgewählten Aspekten diskursiver Konstruktionen von KI liegen nur Einzeluntersuchungen vor, die zudem linguistische Aspekte kaum berühren.[7] Das Projekt *KI im Spannungsfeld gesellschaftlicher Diskurse* setzt an diesem Forschungsdesiderat an und untersucht die Gemeinsamkeiten und Unterschiede in der Konstruktion und Verhandlung von KI in verschiedenen Diskursgenres aus dem Blickwinkel zweier zentraler Leitfragen: (1) Welche Arten bzw. Teilaspekte von KI werden jeweils thematisiert? (2) Welche Korrelationen und/oder Komplementaritäten in der diskursiven Konstruktion, Reflektion und Bewertung von KI als gesellschaftlich und medial konstruierte Bedeutungsangebote lassen sich in den verschiedenen Diskursgenres beobachten? Insbesondere, wo zeigen sich Parallelen oder Unterschiede in den Framingstrategien der textmedialen und filmischen Repräsentation und Diskursivierung, z. B. im Bereich von Topoi, Metaphern, filmästhetischen Inszenierungsstrategien, symbolischen Codierungen und anderen Perspektivierungstechniken?

Im Folgenden werden nach einer knappen Skizze der Prämissen des interdisziplinären Projektes (Abschnitt B) zunächst die Daten und die Methodologie der Pilotstudie vorgestellt (Abschnitt C). Abschnitt D und E präsentieren auch unter Bezugnahme auf exemplarisch ausgewählte Fallbeispiele erste Zwischenergebnisse des Projekts. Abschnitt F umfasst ein knappes medienvergleichendes Fazit und skizziert weitere Untersuchungsschritte.

B Prämissen: Filmische Science-Fiction und öffentliche alltagsweltliche Online-Diskussionen als kommunikative Räume für Zukunftsdiskurse

Bei der Erschaffung des künstlichen Menschen gilt der große technische Durchbruch als zum Greifen nah. Als Zukunftsvorstellungen darüber, wie Technik unser zukünftiges Leben begleiten und bestimmen soll, sind Technikutopien und -dystopien fest in der gesellschaftlichen Debatte unserer modernen Wissensgesellschaft verankert. Diese sieht sich mit zunehmend komplexeren Fragen der modernen Zukunftsforschung konfrontiert, deren vertiefte(s)

6 Papacharissi 2010, Johansson et al. 2017.

7 Z. B. Drux 1999, Weingart/Pansegrau 2003, Adam et al. 2016c.

Verständnis, Bearbeitung und Lösung einen engen Zusammenschluss »sowohl wissenschaftliche[r] als auch lebenspraktische[r] Perspektiven« erfordert.[8] Politik und Sozialwissenschaft erkennen die Notwendigkeit einer breiten öffentlichen Debatte zukünftiger technischer Entwicklungen, die die Wünsche, Bedürfnisse, Ängste und Hoffnungen der Menschen aufnimmt. Die moderne technische Zukunftsforschung integriert die Zukunftsentwürfe einer großen Bandbreite gesellschaftlicher Akteure in ihre »Technikzukünfte« (»Vorstellungen über die zukünftige Entwicklung von Technik und Gesellschaft«[9]) als wesentlichen Aspekt gesellschaftlicher Technikdebatten, die politische und wirtschaftliche Entscheidungsprozesse beeinflussen. Sie umfassen nicht nur wissenschaftliche Modellbildungen und empirische Erhebungen, sondern auch »künstlerische Entwürfe wie literarische oder filmische Produkte der Science-Fiction […] [und] Erwartungen oder Befürchtungen, die über Massenmedien Teil der öffentlichen Kommunikation werden«.[10] Zunehmend manifestieren auch partizipatorische, öffentliche politische und wissenschaftliche Diskurse in der Online-Kommunikation Zukunftsvorstellungen und Bewertungen einer breiten Öffentlichkeit.[11]

In Technikzukünften enthaltene individuelle und kollektive »Wünsche, Hoffnungen, Erwartungen und Befürchtungen, normative Setzungen und Interessen, Werte oder schlichte Annahmen«[12] werden im gesellschaftlichen Diskurs vor dem Hintergrund »spezifische[r] und kulturell geprägte[r] Wahrnehmungsfolien […als] Deutungsrahmen«[13] kollektiv konstruiert und verhandelt. Ihre Erfassung und Modellierung mittels sogenannter »intuitive[r]«[14] Verfahren der qualitativen Erforschung von Konzepten und Visionen im breiten gesellschaftlichen Diskurs (literarische, filmische, journalistische massenmediale Repräsentationen und solche im breiten partizipatorischen Online-Diskurs) bildet aus linguistischer und medienwissenschaftlicher Perspektive weiterhin ein Forschungsdesiderat. Dies gilt insbesondere für die Themen KI und Technisierung von Lebenswelten. Hier knüpft dieses Projekt an. Die zu untersuchenden Technikutopien und -dystopien im gesellschaftlichen und filmischen Diskurs geben als eine Form von Technikzukünften Auskunft darüber, »welche zukünftige gesellschaftliche und technologische Realität für

8 Dienel 2015: 71, vgl. auch Schrögel/Weitze 2018: 23.

9 acatech 2012: 6.

10 Ebd.

11 Schrögel/Weitze 2018: 22.

12 acatech 2012: 12.

13 Felder 2009: 16.

14 acatech 2012: 23.

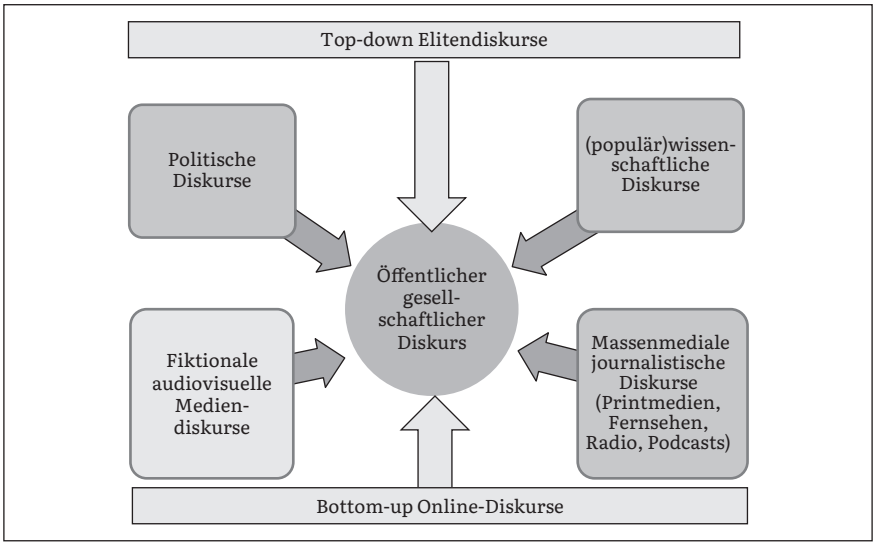

Abbildung 1 KI in öffentlichen Diskursdomänen

möglich, mehr oder weniger wahrscheinlich, gewünscht oder unerwünscht gehalten wird«[15].

Ziel des komplexen diskursvergleichenden Projektes KI *im Spannungsfeld gesellschaftlicher Diskurse* ist die systematische komplementäre linguistische und medienwissenschaftliche Analyse der Konstruktion, Repräsentation und Verhandlung von KI in öffentlichen Diskursen gesellschaftlicher Eliten (fiktionale – insbesondere audiovisuelle – Medienprodukte, politischer und journalistischer Mediendiskurs, populärwissenschaftlicher Diskurs) und in partizipatorischen Diskursen einer breiten Öffentlichkeit in öffentlichen Diskussionsräumen des Internets (z. B. Forenkommunikation, Wikipedia, YouTube-Kommentare – vgl. Abbildung 1) mit ihren jeweiligen genrebedingten makrostrukturellen kommunikativen Rahmenbedingungen.

Im Mittelpunkt der hier vorgestellten Pilotstudie standen bislang thematisch einschlägige audiovisuelle Science-Fiction und öffentliche Forendiskussionen einer breiten Online-Öffentlichkeit im Internet aus dem Blickwinkel zentraler Leitfragen (Welche Teilaspekte von KI werden jeweils thematisiert? Wo zeigen sich Parallelen oder Unterschiede in den gewählten Strategien, insbesondere in Bezug auf Parallelen oder Unterschiede in den Framingstrategien, z. B. im Bereich von Topoi, Metaphern, filmästhetischen Inszenierungsstrategien, symbolischen Codierungen und anderen Perspektivierungstechniken?).

15 acatech 2012: 6.

Dies bildet den Ausgangspunkt für einen interdisziplinären Beitrag zu aussagekräftigen Modellen für die Ermittlung von Technikzukünften, die im breiten gesellschaftlichen Diskurs konstruiert werden und verankert sind.

I Medienwissenschaftliche Prämissen: U- und Dystopien in audiovisueller Science-Fiction und Veränderungen der Lebenswelt durch technische Innovationen

Fiktionen zeichnen sich dadurch aus, dass sie nicht auf die Wiedergabe von Existierendem oder die Darstellung von tatsächlich gegebenen Verhältnissen festgelegt sind. Gleichwohl nehmen sie nicht nur vieles von dem auf, was wir für die Realität halten, sondern beziehen sich auch gerade da auf die Wirklichkeit und die jeweils aktuelle Lebenswelt, wo sie sich von dieser abheben.[16] Ihre erdachten Konflikte und Konstellationen reflektieren reale Problemlagen und Debatten, jedoch in spezifisch anderer Weise als etwa wissenschaftliche oder journalistische Diskurse. Mit diesen haben sie im Unterschied zu den im linguistischen Projektteil untersuchten Online-Diskursen gemein, dass sie überwiegend von professionellen Diskursteilnehmern produziert werden, jedoch richten sie sich an ein Laienpublikum. Insofern ist davon auszugehen, dass sie die an den genannten Online-Diskursen teilnehmenden Personen grundsätzlich erreichen können, was für wissenschaftliche Diskurse nicht uneingeschränkt gilt. Dies ist insbesondere bei Fiktionen relevant, die wissenschaftlich-technische Themen wie KI aufgreifen.

Im Zuge der Herausbildung unserer wissenschaftlich-technisch geprägten Zivilisation hat sich gegen Ende des 19. und zu Beginn des 20. Jahrhunderts das fiktionale Genre der *Science-Fiction* etabliert.[17] Dies ist als Reaktion nicht einfach nur auf die große Bedeutung von Wissenschaft und Technik in der Gesellschaft zu verstehen, sondern insbesondere auf die Erfahrung einer zunehmenden Beschleunigung von Veränderungen der Lebenswelt, die wissenschaftliche Entdeckungen und technische Innovationen hervorrufen.[18] Insofern ist das Moment des Zukünftigen für Science-Fiction konstitutiv. Sie entwirft eine Welt der Zukunft, indem sie gegenwärtige Entwicklungen extrapoliert und ihre Fiktion damit plausibiliert, und führt uns damit die möglichen Folgen heutigen Handelns plastisch vor Augen. Diese Utopie, der Entwurf einer (noch) nicht existierenden Welt, kann dabei getragen von Technikeupho-

16 Suvin 1979: 27.

17 Saage 1997: 48; Page 2016.

18 Als genrebegründend gilt hier Mary Shelleys *Frankenstein oder Der moderne Prometheus* (Orig.: 1818). Aldiss 1973: 23; Aldiss 1998: 323; Page 2016: 71.

rie eher eutopisch[19] ausfallen und ein Wunschbild präsentieren oder aber eine innovationsskeptische Dystopie als Schreckbild. In jedem Fall ist damit der Anspruch verbunden, uns *The Shape of Things to Come* zu zeigen, wie ein berühmter Buchtitel von H. G. Wells (1933) lautet.

Aus der literarischen Science-Fiction geht schon sehr früh in der Geschichte des Mediums der Science-Fiction-Film hervor.[20] Die in diesem Filmgenre ausgeprägte Tradition wirkt auch in audiovisuellen Gestaltungen in anderen medialen Kontexten wie etwa bei TV-Serien weiter,[21] weshalb wir zusammenfassend von audiovisueller Science-Fiction sprechen. Angesichts der großen Beliebtheit dieses medialen Genres kann davon ausgegangen werden, dass es in einer breiteren Öffentlichkeit zur Bildung von Vorstellungen, Hoffnungen und Ängsten in Bezug auf technische Innovationen wie im Feld der KI in hohem Maße beiträgt.[22] Den theoretischen Rahmen unserer Untersuchung bildet Simon Spiegels *Poetik des Science-Fiction-Films* (2007). Er stellt heraus, dass das Kriterium für die Zuordnung eines Werks zu diesem Genre nicht in tatsächlicher ›Wissenschaftlichkeit‹ im Sinne wissenschaftlicher Überprüfbarkeit seiner Annahmen besteht, sondern dass es sich den Anschein einer wissenschaftlich begründeten Realitätskompatibilität gibt. Dazu dienen spezifische Plausibilisierungs- und Authentifizierungsstrategien, die an populäre Vorstellungen von Wissenschaft anschließen.[23] Ein Werk der Science-Fiction führt jeweils mindestens ein sogenanntes *Novum* ein, also ein mit den gängigen Realitätsannahmen nicht kompatibles Element,[24] und leitet dieses Novum extrapolierend aus dem aktuellen Stand der Wissenschaft bzw. dem bei einem breiteren Publikum zu erwartenden Wissensstand über einschlägige wissenschaftliche Theorien, Erkenntnisse und Methoden her. Science-Fiction kennzeichnet sich also dadurch, »dass sie ihre Wunder *pseudowissenschaftlich* legitimiert und dass sie ihre Nova *naturalisiert*, so dass sie den *Anschein* wissenschaftlich-technischer Machbarkeit aufweisen.«[25]

Als Teil der Populärkultur reagiert Science-Fiction auf technische Entwicklungen und Technikinnovationsdiskurse zwischen Wissenschaft, Journalis-

19 Da die Utopie jedoch im Laufe der Gattungsgeschichte zunächst sehr stark positiv geprägt war, also eine bessere Version der Welt entwarf, ist diese terminologische Differenzierung zwischen eu- und dystopischen Utopien wenig verbreitet. Wir sprechen daher im Folgenden von positiv gefärbten Utopien und negativen Dystopien – z. B. Sargent 1994.

20 Fitting 1993; Fitting 2003; Ruddick 2016.

21 Jancovich/Johnston 2011; Miller 2012.

22 Sobchack 1998; Sobchack 2005; Cornea 2007; Telotte 2016.

23 Weingart/Muhl/Pansegrau 2003.

24 Suvin 1979: 93–95.

25 Spiegel 2007: 51 [Hervorhebungen im Original].

mus und Politik.[26] Sie setzt sich damit auseinander, indem sie Themen, Aspekte und Argumentationsmuster auswählt, gewichtet, emotional einfärbt und narrativ sowie dramaturgisch aufbereitet und Deutungsangebote macht, die von Rezipient:innen in unterschiedlicher Form aufgenommen werden können. Da wir keine Rezeptionsstudie durchgeführt haben, steht bei unserem Vorgehen die Erhebung des jeweiligen Deutungspotentials durch eine qualitative Analyse im Zentrum. Charakteristisch für populärkulturelle Medienprodukte, die sich gesellschaftlich vielbeachteten, heiklen und umstrittenen Themen wie KI zuwenden, ist, dass meist keine eindeutige Positionierung oder Tendenz der Deutung auszumachen ist. Vielmehr werden unterschiedliche, auch konträre Einschätzungen und die zugehörigen Argumentationsmuster aufgegriffen, durch die Haltungen verschiedener Figuren repräsentiert und im Handlungsverlauf die entsprechenden Aspekte des Themas exemplifiziert – etwa positive und negative Aspekte von Technik und die zugehörigen Emotionen zwischen Hoffnung und Angst. Ihre ästhetische Gestaltung kann diese Ambivalenz verstärken, indem sie unterschiedliche Auslegungsmöglichkeiten im Werk konstituiert. Die Aufnahme gesellschaftlicher Debatten in fiktionalen Medienprodukten initiiert häufig gesellschaftliche Anschlusskommunikationen der Rezipient:innen in partizipatorischen Online-Diskursen.

II Linguistische Prämissen: Partizipatorische öffentliche Diskussionsforen im Internet und Strategien der konzeptuellen Perspektivierung

Bottom-up Online-Diskurse in Gestalt öffentlicher Diskussionsformate im Internet unterscheiden sich durch ihre spezifischen genrebedingten makrostrukturellen Rahmenbedingungen fundamental von gesellschaftlichen Elitendiskursen im Allgemeinen ebenso wie von audiovisueller Science-Fiction im Besonderen. Sie eröffnen neue Partizipationsmöglichkeiten einer breiten Öffentlichkeit am öffentlichen Diskurs ›von unten‹, jenseits der Diskursräume institutionalisierter Akteur:innen und institutionalisierter künstlerischer, politischer, (populär)wissenschaftlicher und journalistischer medialer Rahmen. Im Unterschied zu Top-down-Elitendiskursen und insbesondere zu den hier vergleichend betrachteten fiktionalen audiovisuellen Diskursen können die Beteiligten an öffentlichen Online-Diskussionen, die normalerweise nicht aktiv in gesellschaftliche Elitendiskurse eingreifen können,[27] ein deut-

26 Vgl. Stollfuß 2016: 3–4.

27 Van Dijk 2008: 66–68, Fraas et al. 2012: 39–42.

lich breiteres Spektrum an Interaktionsrollen ausüben.[28] Der wichtigste Unterschied betrifft den systematischen Zugriff auf die Produktionsrolle, in der aktive Teilnehmer:innen als *Sender:innen*[29] gleichzeitig im Sinne Goffmans[30] die Rolle von *Animator* (physische Quelle der Nachricht), *Author* (Person, die die Nachricht formuliert) und *Principal* (Person, deren Standpunkt kommuniziert wird) einnehmen können und zu ratifizierten Teilnehmer:innen an gesellschaftspolitischen Debatten werden.[31] Unterschiede, insbesondere zu audiovisueller Science-Fiction, zeigen sich aber auch im Rezeptionsformat, in der Rolle von Leser:innen,[32] die in öffentlichen Online-Diskussionen personell gleichermaßen mit der Produktionsrolle wie mit der Rolle lediglich passiv an der Interaktion beteiligter Mitlesender, die selbst keine aktiven Beiträge verfassen, zusammenfallen kann. Die Komplexität beider Rollen ermöglicht es den Beteiligten, in öffentliche Diskurse einer Meinungsbildung *von unten* direkt aktiv einzugreifen bzw. deren Postulate als Rezipient:innen in ihre eigene Meinungsbildung einfließen zu lassen. Zusätzlich können aktive Diskussionsteilnehmer:innen in öffentlichen Online-Diskussionsformaten durch häufiges Posten von Beiträgen, Beantworten von Fragen, Konstruktion von Expert:innenstatus, wechselseitiges solidaritätsstiftendes gruppendynamisches Verhalten sowie das Initiieren neuer Themen potentiell die metapragmatische Rolle des *Hosts*[33] (Diskussionsführer:in) einnehmen und so aktiv in den Prozess der thematischen Entwicklung eines Diskussionsstranges und damit potentiell auch selbst in Prozesse der Meinungsbildung eingreifen. Im Unterschied zu institutionalisierten Elitendiskursen beteiligen sich die Nutzer:innen öffentlicher Online-Diskussionsformate in aller Regel anonym, lediglich mit Online-Identitäten, wodurch sie neben ihrer individuellen Meinungsäußerung per se gleichzeitig zur kollektiven Konstruktion überindividueller Bewusstseinsstände beitragen.

In den hier untersuchten Online-Diskussionen konstruieren und verhandeln Nutzer:innen das Potential und die Merkmale von KI (vgl. Abb. 2).

Dabei nehmen sie mit ihren individuellen Wahrnehmungen in komplexen kognitiven Prozessen der Konstruktion von Konzepten in der sprachlichen Bedeutungskonstruktion unausweichlich eine bestimmte ›Perspektive‹

28 Vgl. weiterführend auch Marcoccia 2004: 115.

29 Marcoccia 2004: 131.

30 Goffman 1981: 131–132, 144.

31 Johansson/Kleinke/Lehti 2017: 1.

32 Marcoccia 2004: 131.

33 Marcoccia 2004: 131, vgl. auch Eller 2017.

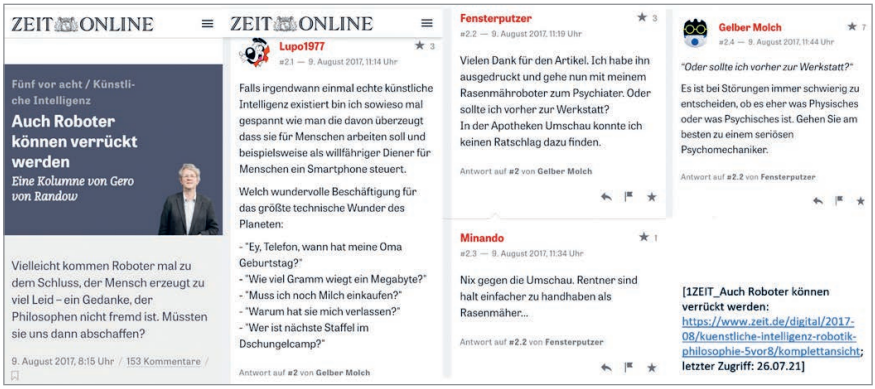

Abbildung 2 Beispiel Forendiskussion

oder einen ›point of view‹ als Zeichen der subjektiven Bedeutungskonstruktion ein.[34]

Operationen des Framing gehören zu den zentralen konzeptuellen Perspektivierungstechniken.[35] In der menschlichen Kognition stellen Frames ein strukturiertes »Repräsentationsformat für Erfahrungswissen« dar.[36] Es bildet »prototypische Strukturen des Wissens«[37], die durch Assoziationen an sprachliche Strukturen geknüpft sind[38] und damit Grundlagen für Selektionsprozesse darstellen, in deren Ergebnis Beteiligte in der Interaktion auf jeweils relevante Ausschnitte ihres Wissens dynamisch zugreifen und diese vor dem Hintergrund ihres gesamten Wissens- und Erfahrungsrahmens relevant setzen.[39] Framingprozesse sind hochkomplex, manifestieren sich auf verschiedenen Ebenen und können ineinandergreifen (z. B. als *Entscheidungskatalysatoren* oder *als Hintergrund für soziale Wahrnehmung*[40]). Im Mittelpunkt unserer Untersuchung von Foreninteraktion stehen perspektivierende Framingoperationen als *Selektionsmechanismen* im Sinne Fillmores.[41] Sie manifestieren sich

34 Kleinke/Schulz 2019: 79. Vgl. auch Talmy 2000: 14 zu »conceptual alternativity«, Langacker 1987: 128, Verhagen 2007: 49 und Langacker 2008: 43 zu »construal«-Operationen oder Kövecses 2010: 91–93 zu metaphorischer Perspektivierung.

35 Fillmore 1976: 27–28, Cienki 2007: 174–175, Busse 2012: 620.

36 Mederake 2015: 189, vgl. auch Fillmore 1976: 26, Kleinke/Schulz 2019: 80.

37 Fillmore 1976: 25–26, Busse 2012: 824.

38 Fillmore 1976: 25, Wehling 2017: 43.

39 Vgl. z. B. Fillmore 1976: 27–28, Wehling 2017: 43 und für einen knappen Überblick aus kognitiv-semantischer Perspektive Cienki 2007: 170–175.

40 Fillmore 1976: 25, Goffmann 1986, Wehling 2017: 43, Kleinke/Schulz 2019: 80.

41 Fillmore 1976: 26.

in den hier untersuchten Forendiskussionen z. B. durch die Wahl semantischer Diskurstopoi in Prädikationen.[42] Außerdem bilden sie die Grundlage für den Einsatz konzeptueller Metaphern, mit denen Sprecher:innen ein Targetkonzept durch die Wahl eines bestimmten Quellkonzeptes konzeptuell konstruieren und perspektivieren.[43] Metaphorische Ausdrücke fungieren teilweise auch in Referenzprozessen[44] als explizite Benennung und damit Salientsetzung einzelner Frameslots bzw. konzeptueller Domänen.[45] Framingoperationen beeinflussen auch die force-dynamische Konstruktion eines Geschehens und seiner Akteure im Sinne Talmys[46] (mit Akteur:innen als starken, handlungsmächtigen Entitäten/Antagonisten oder als vergleichsweise schwachen Entitäten, die dem Einfluss eines starken Antagonisten unterliegen). Alle drei framebasierten Perspektivierungstechniken haben sich im Pilotprojekt in den Online-Diskussionen als zentrale Elemente der Bedeutungskonstruktion erwiesen, lassen sich jedoch nicht immer scharf voneinander trennen. So können wie Metaphern auch Referenzen oder force-dynamische Konstruktionen semantisch selbst prädizieren.[47] Für analytische Zwecke wurden alle drei Perspektivierungen in der Pilotstudie einzeln kategorisiert. Einschlägige Beispiele aus den untersuchten Daten werden in Abschnitt E vorgestellt.

C Methodologie und Datengrundlage

Die Pilotstudie geht interdisziplinär und multimedial vor und synthetisiert die Analyse und den Vergleich sowohl textmedialer als auch fiktionaler audiovisueller Repräsentationen mit den Methoden der Linguistik und der Medienwissenschaft.

Theoretisch-methodologisch sollen bestehende linguistische Untersuchungsmodelle zur (kritischen) Diskursanalyse mit denen der filmwissenschaftlichen Medienproduktanalyse verknüpft werden, um dem multimedialen Charakter der Prozesse individueller und kollektiver Bewusstseinskonstruktion im Bereich von Technikutopien und -dystopien besser gerecht zu werden. Das theoretisch-methodische Bindeglied zwischen der linguistischen, auf den Text fokussierten, und der medienwissenschaftlichen, auf die filmästhetischen Aspekte fokussierten, Analyse liegt in der datengeleiteten und je-

42 Hart 2010: 65, Reisigl/Wodak 2001: 46.

43 Kövecses 2010: 91–93.

44 Reisigl 2007: 377–378.

45 Hart 2010: 11.

46 Talmy 2000: 409, vgl. auch Oakley, 2005: 450, Fraas 2013.

47 Hart 2010: 60.

weils für beide Disziplinen kategorieninspirierten semantischen Inhaltsanalyse. Der Analysegegenstand ist ein selbsterstelltes, thematisch generiertes dynamisch erweiterbares Korpus zu KI. Es bildet sowohl gesellschaftliche Eliten- als auch partizipatorische Online-Diskurse aus den Diskursdomänen *partizipatorische Online-Kommunikation* und audiovisuelle *Science-Fiction* ab.

I Medienwissenschaft

Aus medienwissenschaftlicher Sicht soll durch qualitative Inhaltsanalysen mit besonderem Augenmerk auf die filmästhetische Gestaltung die Inszenierung von Wissenschaft und Technik im Kontext von KI auf verschiedenen Ebenen untersucht werden. Betrachtet wird u. a., wie sich Haltungen zu KI in der Anlage der Figuren spiegeln – etwa in der Verteilung von technikeuphorischen und -skeptischen Einstellungen auf Figuren und die mögliche Koppelung dieser Einstellungen mit Geschlechterrollen; mit welchen Topoi, also kulturell vorgeprägten Motivkomplexen, KI verbunden wird – etwa mit dem Motiv des Zauberlehrlings als spezifischer Ausprägung des Topos der menschlichen Hybris; welche symbolischen Codierungen auftreten – etwa die Parallelisierung der KI-Bewusstwerdung mit dem Sündenfall; welche ästhetischen Gestaltungsmittel hinzukommen – etwa die Inszenierung von Transparenz, Semitransparenz oder Opazität technischer Vorgänge durch optisch in unterschiedlichem Maße filternde Materialien wie Glas oder durchscheinende Stoffe; und welche Deutungsangebote letztlich durch das Zusammenspiel all dieser Aspekte mit der Gesamtanlage des Medienprodukts einschließlich Handlungsentwicklung und -lösung entstehen. Leitend sind bei der Untersuchung zwei allgemeine Fragestellungen:

1. Wie werden technologischer Fortschritt und seine Funktionalisierung in der Gesellschaft in utopischer und dystopischer Science-Fiction dargestellt und inszeniert?
2. Wie sind die Wechselwirkungen zwischen der spezifischen ästhetischen Gestaltung in der Fiktion – einschließlich der Funktionalisierung von kulturellen und diskursiven Bezügen und der Aufladung mit symbolischer Bedeutung – und den relevanten Technik- und Wissensdiskursen einzuschätzen?

Hierzu werden fiktionale audiovisuelle Medienprodukte (Filme und Fernsehserien) des zurückliegenden Jahrzehnts aus unterschiedlichen Ländern bzw. Nationalkulturen betrachtet, die internationale Beachtung gefunden haben. Dabei wurden zunächst Werke aus dem westlich-europäischen Kulturkreis

ausgewählt, um einerseits eine gute Vergleichbarkeit zu gewährleisten und andererseits in diesem Rahmen auch ein gewisses Spektrum unterschiedlicher Ausgestaltungen abzudecken.[48] Die Thematisierung von KI wird zunächst einer semantischen Analyse unterzogen. Da KI als Technik etwas Menschengemachtes ist, aber zugleich verspricht, etwas für den Menschen Konstitutives, nämlich Intelligenz und möglicherweise Bewusstsein, künstlich herzustellen, stellt sich die Frage, welcher Charakter dieser Technik zugeschrieben wird und wie die Vorstellung von ihren Produzent:innen zwischen Wissenschaftler:-/Ingenieur:innen und Schöpfer:innen oszilliert. Im Verhältnis zwischen KI und ihr begegnenden Menschen spielen Machtverhältnisse, aber auch Identitätsfragen eine wichtige Rolle. Diese Aspekte manifestieren sich im Setting, also der Anlage der fiktiven Welt, der Figurencharakterisierung, der Handlungsführung und den Dialogpartien. Ihre Betrachtung muss jedoch ergänzt werden durch eine filmästhetische Analyse, die alle Ebenen des kinematographischen Codes inklusive der visuellen und kameratechnischen Gestaltung, der Montage und schließlich der integrativen Verbindung aller Gestaltungsebenen berücksichtigt.

Das ausgewählte Korpus umfasst zwei Filme und eine Fernsehserie, nämlich die Filme *Eva* (Spanien 2011, Regie: Kike Maíllo) und *Ex Machina* (GB 2015, Regie: Alexander Garland) sowie die TV-Serie *Äkta människor/Real Humans* (Schweden 2012–2014).[49] Alle drei Werke sind europäische Produktionen aus dem Zeitraum 2010–2015 und befassen sich mit dem Thema KI mit dem Fokus darauf, Fragen von Macht, Autonomie und Menschlichkeit zu reflektieren. Während Filme aus Gründen der Laufzeit einen Teilaspekt der facettenreichen Thematik herausgreifen müssen, kann eine Serie eher alle relevanten Aspekte ausbreiten und auf unterschiedliche Handlungsstränge verteilen oder in einzelnen Episoden Schwerpunkte setzen. Insofern bot es sich an, diese beiden unterschiedlichen audiovisuellen Formen im Korpus zu repräsentieren.

II Linguistik

Vor dem Hintergrund der konstruktivistischen Tradition der Beschreibung des Zusammenhangs von Sprache und Gesellschaft[50] untersucht dieses Projekt

48 Eine Erweiterung des Korpus z. B. auf den ostasiatischen Kulturkreis könnte für die zukünftige Entwicklung des Projekts von Interesse sein und den Untersuchungsfokus um interkulturelle Unterschiede in KI-Narrativen erweitern.

49 Im Rahmen des Projekts wurde außerdem der Film *The Circle* (USA 2017) untersucht, auf den im Rahmen dieses Aufsatzes nicht näher eingegangen wird, da KI dort nicht personifiziert und daher weniger exponiert inszeniert wird.

50 Felder 2013, für eine differenzierte Diskussion vgl. z. B. Felder/Gardt 2015.

textmediale Repräsentationen gesellschaftlicher Bedeutungskonstruktionen und deren diskursive Perspektivierungen gesellschaftlicher Wirklichkeit. Aus linguistischer Perspektive wird die komplexe (text)mediale Repräsentation, Konstruktion und Perspektivierung der untersuchten Konzepte qualitativ, korpusbasiert und kategorieninspiriert mit den Methoden der (kritischen) Diskursanalyse beleuchtet. Dazu werden traditionell in der (kritischen) Diskursanalyse[51] und deren Erweiterungen in kognitiv-semantischen Beschreibungsansätzen unter besonderer Berücksichtigung von (metaphorischen) Framingprozessen[52] etablierte sowie im Zuge der Pilotstudie im Rahmen des HEiKA-Projektes empirisch ermittelte Kategorien eingesetzt. Diese umfassen z. B. Perspektivierungstechniken des Framing durch semantische Topoi und konzeptuelle Metaphern, die ihrerseits in Prädikations- und Referenzprozesse eingehen, sowie die Nutzung force-dynamischer Konstruktionen. Darüber hinaus vollziehen Beitragende fortwährend epistemische, emotionale, deontische und interpersonelle Positionierungsprozesse, mit deren Hilfe die Relationen zwischen Diskursteilnehmer:innen, gesellschaftlichen Akteur:innen im Umfeld von KI sowie (realen und imaginierten) KI-Manifestationen ausgeleuchtet werden, die jedoch im Rahmen dieser ersten Vorstellung unseres Pilotprojektes nicht im Einzelnen beleuchtet und illustriert werden können. Das Inventar der Analysekategorien wurde im Zuge der Untersuchung datenbasiert, den Prinzipien der *Grounded Analysis*[53] folgend, dynamisch erweitert.

Als Datengrundlage für die qualitative linguistische Analyse diente ein selbsterstelltes Korpus thematisch einschlägiger Online-Diskussionen. Das Korpus wurde stichwortgeneriert erhoben (vgl. Tab. 3: Korpus des Pilotprojektes).

Die Analyse erfolgte datengeleitet, aber kategorieninspiriert. Sie kombiniert in einer ersten Stufe die detaillierte qualitative Analyse der Korpusdaten durch manuelle Codierung mit deren systematisch erfassten Vorkommenshäufigkeiten. Dafür wurde sowohl für die linguistische als auch für die medienwissenschaftliche Analyse das in den Sozialwissenschaften entwickelte Softwarepaket MAXQDA als zentrales Analysewerkzeug eingesetzt. Es gestattet die Entwicklung eines eigenen, dynamischen und datengeleiteten Codierungssystems,[54] die korpusgestützte Ermittlung von Kollokationen und Wortfrequenzen sowie die Kombination qualitativer und quantitativer Ergebnisse der Codierung in Gestalt von Korrelationen innerhalb und zwischen den lin-

51 Reisigl/Wodak 2001, 2009, Reisigl 2007, Wodak et al. 2009.

52 Musolff 2016, Hart 2010, Kövecses 2010.

53 Zuerst Glaser und Strauss 1967, vgl. auch Spieß 2011, Tereick 2016, Pentzold/Fraas 2018.

54 Alle Codierungen erfolgten in den beiden Arbeitsgruppen, wobei Zweifelsfälle eingehend diskutiert und gegebenenfalls für verschiedene Kategorien codiert wurden.

Tabelle 1 Korpus des Pilotprojektes

Partizipatorische Online-Kommunikation	Anzahl der Threads/Postings	Häufig diskutierte Themen
Themenkreis 1: *Künstliche Intelligenz und Human Enhancement*	25/2 294	• Manifestation populärer Vorstellungen sowie Hoffnungen und Ängste • Bezüge zu Technikdiskursen • Verknüpfung von Wissen und ethischen Fragen • Transhumanismus vs. Optimierungswahn
Themenkreis 2: *Technisierung und Informatisierung von Lebenswelten*	18/1 322	• Effizienzsteigerung vs. Unsicherheitsfaktoren • Gewinn vs. Verlust von Autonomie • Problematisierung von Identität (vgl. Technologien der Klassifizierung und Selbstvermessung)
	Σ 3 616	Quellen*: SPON, Focus, Zeit, welt.de, Tagesschau, IOFF – Das Medienforum, forum.grenzwissen.de

* Eine Erweiterung des Korpus um Threads aus weiteren Quellen (z. B. reddit, YouTube) ist in Vorbereitung.

guistischen und Filmkorpora. Ihre Ergebnisse sollen den Ausgangspunkt für künftige korpusbasierte quantitative und qualitative Analysen größerer Datenmengen und weiterer Diskursdomänen bilden.

D Filmische Inszenierung von KI: Erste Ergebnisse und exemplarische Einblicke in Fallstudien der Medienwissenschaft

Im Mittelpunkt der medienwissenschaftlichen Analyse filmischer Repräsentation von KI steht insbesondere die Frage, inwiefern die Inszenierung von Wissenschaft und Technik mit bestimmten Topoi, Leitmotiven und symbolischen Codierungen verknüpft ist und welche Deutungsangebote daraus konstituiert werden. In diesem Kapitel soll ein Einblick in exemplarische Filmanalysen gegeben werden. Die untersuchten Werke demonstrieren, dass filmische Repräsentationen von KI typischerweise vom Verhältnis des Selbst zum Anderen erzählen, wobei es sich hier nicht um stabile Kategorien handelt, sondern die Grenzen, die zwischen Subjekt und Objekt wie auch zwischen Mensch und Maschine verlaufen, zunehmend durchlässig und prekär werden. Dies ist

insbesondere dann der Fall, wenn es sich um KI handelt, die Singularität erlangt, d. h. ein Bewusstsein und einen freien Willen entwickelt – ein Topos, der in der Science-Fiction seit dem späten 20. Jahrhundert besonders populär ist.[55] So zeichnet die schwedische Serie *Real Humans* das Bild einer Gesellschaft, die sich lediglich durch das technische Novum der Hubots (= *humanoid robots*) von unserer Alltagswelt unterscheidet. Anders als viele andere Werke konzentriert sich *Real Humans* nicht auf den Moment der KI-Schöpfung und zeigt auch kein futuristisches Setting, in dem die Androiden nur eine unter vielen technischen Innovationen sind – *Real Humans* erzählt von der Zeit dazwischen, von der Phase eines Alltäglich-Werdens, in der eine Gesellschaft um den richtigen Umgang und das Inkludieren der künstlichen Menschen ringen muss.[56] Die Hubots bilden dabei ein Spektrum von KI ab, von Industrie-Hubots, die lediglich für bestimmte Fertigkeiten programmiert sind und über kein komplexes Innenleben verfügen, bis hin zu den *freien* Hubots, die Singularität entwickelt haben und im Untergrund leben, um ein selbstbestimmtes Leben führen zu können. Während auch die Industrie-Hubots Ablehnung hervorrufen, z. B. die Angst, ersetzt zu werden (verkörpert z. B. durch Roger, der in seiner Abteilung der letzte *echte Mensch* ist), sind es vor allem die *freien* Hubots, die die instabile Grenze zwischen Mensch und Maschine, dem Selbst und dem Anderen in ihrer Auflösung sichtbar machen. Nicht nur die Hubots in *Real Humans*, sondern auch die KI in den Filmen *Eva* und *Ex Machina* fungieren als Projektionsfläche für die Reflexion ethischer und sozialer Fragestellungen. Alle untersuchten Werke weisen unterschiedliche Erzählstrategien auf, doch lassen sich typische, rekurrierende Plotelemente, Topoi und Leitmotive identifizieren. Die KI oszilliert jeweils zwischen Wunsch- und Angstbild, zwischen Begehren und Bedrohung. Weitere zentrale Themenfelder sind die Spannung zwischen Steigerung und Verlust von Autonomie, die Konstitution vs. die Gefährdung von Subjektivität und Identität, die Rolle des Körpers, transparentes vs. verschleiertes Wissen sowie Subjekt-Objekt-Konstellationen bzw. die Frage nach Macht-, Kontroll- und Normstrukturen, wobei zahlreiche dieser Spannungsverhältnisse eine geschlechtliche Codierung aufweisen.

I Schöpferfiguren und Technikrepräsentation

David Eischer, der Erschaffer der Hubots in *Real Humans*, ist zwar die bedingende Instanz für die Hubots, ist aber in der Serie nur in Rückblenden sichtbar. Vor allem zwei seiner Schöpfungen sind durch persönliche Hintergründe mo-

55 Weber 2008: 197–198.

56 Adam/Knifka 2016: 341.

tiviert: Nach einem tragischen Unfall macht Eischer seinen Sohn zum Cyborg, um sein Leben zu retten, und erschafft den Hubot Bea als Abbild seiner verstorbenen Frau. Hauptprotagonist des Films *Eva* ist Alex, der nach vielen Jahren als erfolgreicher Programmierer an seine Heimatuniversität zurückkehrt, um dort ein früheres Projekt zu beenden: die Programmierung einer KI in Form eines *künstlichen Kindes*. Alex trifft seine frühere Liebe Lana wieder und begegnet auch ihrer Tochter Eva, die bei ihrer ersten Begegnung auf den Händen balanciert – ein symbolischer Verweis, dass sie soziale Konventionen (auch im unmittelbaren Wortsinn) auf den Kopf stellt. Alex möchte Eva zur Blaupause für seine KI machen, da er fasziniert ist von diesem außergewöhnlichen und befremdlichen Kind, in dem er sich selbst wiederzuerkennen glaubt. Reift in ihm doch die Überzeugung, dass Eva seine Tochter ist. Diese Ahnung erfüllt sich auf ungewöhnliche Weise: Lana gesteht Alex, dass es sich bei Eva um die KI des früheren Projekts handelt. Eva wurde von Lana vollendet und – ohne dass sie um die Hintergründe der eigenen Existenz wusste – als ihre Tochter aufgenommen. Eva, die dieses Gespräch belauscht, läuft erschüttert davon. Als ihre Mutter sie tröstend umarmen möchte, stößt Eva sie von sich, und Lana kommt beim Sturz in eine Schlucht ums Leben. Daraufhin wird entschieden, dass Eva eine Gefahr darstellt und ihr Speicher gelöscht werden muss, was dem ›Tod‹ der KI gleichkommt. Alex selbst übernimmt diese Aufgabe: Er bringt seine *Tochter* zu Bett und sagt den Satz »Was siehst du, wenn du die Augen schließt?«, die standardisierte Formel, mit der der Löschvorgang in Gang gesetzt wird. Der Film schließt mit Aufnahmen von Alex, Lana und Eva an einem Strand als glückliche Familie, wobei es sich wahrscheinlich um die letzten Vorstellungsbilder von Eva handelt. Bezeichnend ist, dass Alex bei seinen Programmierarbeiten regelrecht als eine Art *Technikalchemist* inszeniert wird:[57] Durch das Setting des altmodischen Labors seines Vaters und die goldenen Kugeln, die die zu kombinierenden Charaktereigenschaften der KI darstellen, erscheint der Vorgang nicht wissenschaftlich, sondern wird vielmehr als magischer Akt ins Bild gesetzt – Technik wird mystifiziert. Zugleich wird die intakte Familie in ihrer technisch verfremdeten Form aus Programmier-Vater, Robotik-Mutter und KI-Kind als utopisches Sehnsuchts-Moment installiert.

Auch in *Ex Machina* ist die Schöpferfigur männlich und repräsentiert damit die typische Konstellation der KI-Narrative, die ein männliches erschaffendes Subjekt einem in der Regel weiblichen Objekt als dem Anderen gegenüberstellt.[58] Nathan ist in *Ex Machina* der Erfinder der weltgrößten Suchmaschine

57 Adam 2021: 403.

58 Lana in *Eva* ist zwar eine weibliche Programmiererin, die die KI Eva nach Alex' Weggang selbstständig vollendet hat. Dies liegt zeitlich aber vor der Handlungsgegenwart. Im Film selbst wird sie in erster Linie als Evas Mutter und als Lehrende repräsentiert.

Abbildung 3 *Eva*: Alex als ›Technikmagier‹

und lebt zurückgezogen in einem abgelegenen, hochtechnisierten Anwesen. Er lädt den jungen Programmierer Caleb zu sich ein, um seine neueste Schöpfung, die KI Ava, zu testen: Caleb soll in sieben Tagen entscheiden, ob Ava ihn überzeugen kann, über Singularität und ein *echtes* Bewusstsein zu verfügen. Scheitert der Test, wird Avas Bewusstsein gelöscht. Caleb entwickelt Gefühle für Ava und will mit ihr fliehen. Mit Hilfe einer weiteren KI gelingt Ava die Flucht aus ihrem Gefängnis: Sie tötet Nathan und lässt Caleb in dem abgeriegelten Gebäude zurück. Alle drei Schöpferfiguren sind mit der Hybris, einer typischen Eigenschaft des *mad scientist*[59], assoziiert und stehen in einer Schöpfungskonstellation, die mit unterschiedlichen Akzenten patriarchalisch-familiär aufgeladen ist: Die *freien* Hubots sind mit den Initialen ihres Erschaffers markiert und sozusagen *seine Kinder*; mit Bea erschafft dieser sogar das Ebenbild seiner verstorbenen Frau. Bei Alex überlagern sich sowohl das Begehren des Anderen (hier nicht im sexuellen Sinne, sondern in seiner Faszination und dem Drang, Eva verstehen und rekonstruieren zu wollen) als auch die narzisstische Selbstfixierung (so ist Alex vor allem deshalb von Eva fasziniert, da er sich in ihr wiedererkennt).[60] Nathan bezeichnet sich selbst in einem Dialog mit Caleb als *Avas Dad*. Er repräsentiert den inzestuös aufgeladenen, sadistisch-strafenden Vater und ist getrieben von einem narzisstischen Bedürfnis, den weiblichen Techno-Körper zu kontrollieren.[61] So programmiert er bei Ava ein sexuelles Lustempfinden (das er definiert) und möchte nach dem Löschen ihres Speichers ihren Körper behalten, da dieser *gut* sei.[62] Genauso bewahrt er die Körper von Avas

59 Zur Darstellung von Wissenschaftler:innen vgl. Haynes 1994 und 2017.

60 Adam 2021: 402.

61 Adam 2021: 567–574.

62 Henke 2018.

Abbildung 4 *Ex Machina*: Nathans Monitor-System

ausrangierten Vorgängerinnen als bizarre Trophäensammlung in den Schränken seines Schlafzimmers auf.

II KI im Kontext von Identität und Macht

Alle in den Beispielwerken dargestellten künstlichen Menschen ringen um Identität und Autonomie oder sind der stetigen Gefahr ausgesetzt, dass ihnen der Status eines selbstbestimmten Subjekts abgesprochen wird. *Real Humans* verhandelt über die Hubots die ethische Frage nach der Menschlichkeit sowie der Menschenähnlichkeit.[63] So erweisen sich die *echten Menschen* oftmals als inhumaner als viele der dargestellten Hubots, die beispielsweise entführt, auf dem Schwarzmarkt verkauft, in Bordellen ausgebeutet, misshandelt und sexuell missbraucht werden sowie generell soziale Ausgrenzungserfahrungen machen. Alle Filme thematisieren ethische Probleme, die aus dem Verhältnis des Selbst zum Anderen resultieren. Die Hubots behaupten ihre Menschlichkeit durch kleine Charakterzüge oder Gewohnheiten, wie die Vorliebe für bunte Haarbänder beim Hubot Anita/Mimi, die Suche nach Spiritualität bei Gordon, aber auch durch die Schattenseiten des menschlichen Verhaltensrepertoires wie das Ausüben von Gewalt oder das Reproduzieren von Vorurteilen: Die blonde Hubot Flash, die von einer heteronormativen Familienidylle mit Ehemann, Einbauküche und (Adoptiv-)Kindern träumt, lehnt z. B. die Lebensweise einer lesbischen Pastorin ab, reproduziert also eine Marginalisierung, die sie selbst erfährt.[64] Im Gegensatz zur Fernsehserie *Real Humans*, die

63 Adam/Knifka 2016: 350.

64 Adam/Knifka 2016: 346, 359–360. Adam 2021: 615, 627.

es aufgrund ihres Formats vermag, die soziokulturellen Implikationen von KI differenzierter zu erzählen, weisen die beiden Filme einen dichteren und abgeschlossenen Plot auf, der im Falle Avas mit einer erfolgreichen Subjektkonstitution endet, bei Eva hingegen mit ihrem Scheitern. Auffällig ist die Prävalenz von biblischen und religiösen Referenzen in KI-Narrativen, wovon sowohl der Vorname *Eva* als auch die Abwandlung *Ava* zeugen. Dies wird insbesondere in *Ex Machina* auch szenisch akzentuiert: Nach der Tötung ihres Schöpfers wird Ava zur einer biblischen *Eva* 2.0 stilisiert und erschafft sich selbst neu, indem sie ihren defekten Arm durch den einer ausrangierten Vorgängerin ersetzt.[65] Sie legt sich zunächst die künstliche Haut einer anderen KI an, um ihre Technizität zu verdecken, und trägt schließlich ein weißes Kleid – erschafft sich folglich sowohl als Mensch als auch in ihrer Weiblichkeit neu und wird, wenn auch nicht ohne Widersprüche,[66] gemäß Haraways Cyborg-Konzept zu »eine[r] Art zerlegte[m] und neu zusammengesetzte[m], postmoderne[m] kollektive[m] und individuelle[m] Selbst«[67], das sich selbst codiert und so auch feministisches Potential aufweist. In *Eva* verdichtet sich die ethische Problematik zu besonderer moralischer Brisanz, da es sich bei Eva um ein Kind handelt, das *abgeschaltet*, also in letzter Konsequenz ›getötet‹, wird. Die Dynamik von Akzeptanz und Ablehnung kehrt sich im Laufe des Films um und läuft beim Publikum und innerhalb der Diegese konträr: Zunächst erscheint Eva für die Zuschauer:innen in ihrem impulsiven Verhalten befremdlich, während sie von den Figuren scheinbar als normales Kind akzeptiert wird. Dass sie in Wahrheit eine KI ist, erfährt Eva zeitgleich mit dem Publikum. Als die Ereignisse nach der Enthüllung zum Tod Lanas führen, wird sie von den Figuren in der erzählten Welt auf ihren Status als Maschine reduziert; selbst Alex als ihr *Vater* hinterfragt die Anweisung seiner Vorgesetzten, Eva abzuschalten, nur einmal und zieht nicht in Betracht, sich zu widersetzen. Bezeichnend ist, dass die Darstellung des Films es nahelegt, dass die Zuschauer:innen Eva gerade jetzt als zutiefst menschlich wahrnehmen, da sie sich in einer traumatischen Situation wie ein verängstigtes *normales* Kind verhält und die Inszenierung keinen Zweifel daran lässt, dass Eva ihre Mutter nicht töten wollte. Die Auslöschung des KI-Kindes verweist auf die ethisch-moralischen Fragen rund um den Status und die Rechte von KI, die in den untersuchten Filmen immer wieder thematisiert werden und sich verschärfen, wenn die KI vom Artefakt zum Sozialpartner wird.

---

65 Adam 2021: 599–610.

66 Dass sich Ava in die Haut einer asiatisch gelesenen KI kleidet und dass das ethnisch Andere folglich der Emanzipation eines weißen Subjekts geopfert wird, wird in der Diskussion des Films häufig als problematisch kritisiert, vgl. u. a. Micheline 2015 und Cross 2015.

67 Haraway 2007: 256.

III Filmästhetische Inszenierung

Jedes der untersuchten Werke entfaltet in der Inszenierung ein Gefüge aus sinnhaften Zuschreibungen und setzt eigene Akzente. Es lassen sich auf der Ebene des kinematographischen Codes rekurrierende Muster und Elemente identifizieren. Exemplarisch aufgezeigt werden soll dies nachfolgend am Leitmotiv von Glas bzw. gläsernen Oberflächen sowie an der ambivalenten Inszenierung der KI als Wunsch- und Angstbild, wobei beide Motivkomplexe sich zum Teil überlagern. Das Glas-Motiv findet sich in allen untersuchten Beispielwerken: In *Real Humans* kommt es bereits in der ersten Episode zu einer prägnanten Szene, als die freien Hubots sich nachts auf der Flucht vor einem abgelegenen Haus versammeln. Die Kamera nimmt den Blick des ängstlichen Ehepaars ein, das sich im Inneren befindet und durch den Glaseinsatz der Haustür die dunklen Silhouetten der Hubots erblickt, die trotz ihrer anthropomorph anmutenden Umrisse gerade auf das Nicht-Menschliche verweisen und populäre Darstellungsmuster aus dem Horror-Genre zitieren, wo das bedrohliche Wesen im Halbverborgenen lauert.[68] Eva sieht Alex zum ersten Mal durch die Scheibe seines fahrenden Autos, und auch im weiteren Verlauf klopft Eva, wenn sie Alex besucht, an die Scheibe oder wirft Steine auf die gläserne Kuppel des Labors, durch die sie auch jenes Gespräch beobachtet, das ihre KI-Identität enthüllt. Das Glas akzentuiert in der Szene der ersten Begegnung Alex' beobachtenden Blick und fungiert im Verlauf des Films immer wieder als durchlässige, aber letztlich doch trennende Grenze.

Abbildung 5 *Eva*: Alex' Blick durch die Scheibe auf Eva

68 Adam/Knifka 2016: 355. Adam 2021: 620–621.

Die Konnotation von Glas durch den Komplex von (In-)Transparenz und (Un-)Sicherheit tritt bei *Ex Machina* besonders deutlich zutage: Schon bei der ersten Sitzung von Caleb und Ava verweist die Großaufnahme eines Sprungs in der trennenden Glasscheibe darauf, dass sichergeglaubte (Macht-)Verhältnisse destabilisiert werden. Zu nennen sind außerdem die semi-transparenten Glastüren des Anwesens, durch die sich mehrfach die Silhouetten der Figuren abzeichnen. Diese Inszenierung eines Oszillierens zwischen dem Sichtbaren und dem Verborgenen steht nicht nur im Kontext von Täuschung und Manipulation durch die Figuren, sondern ist auch mit der zentralen Thematik der De- und Rekonstitution von Identität assoziiert. So überrascht es nicht, dass Caleb nach der Offenbarung, dass es sich auch bei der Hausangestellten Kyoko um eine KI handelt, in seiner Identität erschüttert wird und sich vor dem Spiegel seiner eigenen Menschlichkeit zu vergewissern versucht; auch Avas Selbst- bzw. Neuschöpfung findet vor den verspiegelten Türen der Schränke in Nathans Schlafzimmer statt.

Die Silhouetten hinter Glas visualisieren symbolisch das prekär gewordene Verhältnis zwischen dem Selbst und dem Anderen. Während Identitäten de- und restabilisiert, zu- oder abgesprochen werden, sind in den untersuchten Werken Wunsch- und Angstbilder stark präsent, die auch mit der Repräsentation der KI korrelieren. Als Eva zum letzten Mal ihre Augen schließt, hört das Publikum ihre Stimme aus dem Off und sieht transparente Blasen oder Kugeln, die – so legt es die Inszenierung nahe – Evas innerer Vorstellung entsprechen und an die kugelartigen Gebilde erinnern, die Alex als ›Technikmagier‹ zuvor beim Programmieren zusammengefügt hat. Die Kamera zoomt in eine der Kugeln und offenbart die idyllische Szene am Strand, in der Alex, Lana und Eva wieder vereint sind. Familie erscheint im Film als sozialer Raum, der durch Technik in ihrer unkonventionellen Form erst ermöglicht, dann schließlich zerstört wird und am Ende lediglich als utopisch besetztes Sehnsuchtsbild existieren kann. In *Ex Machina* wird v. a. die KI selbst zur Projektionsfläche von Sehnsüchten. Als Caleb Ava über seinen Monitor beobachtet, scheint sie seinen Blick zu erwidern und sich zu inszenieren. Aus dem Wissen, beobachtet zu werden, zieht Ava Macht – vor allem aber auch durch ihren technisch überlegenen Blick, der sie jede Lüge Calebs sofort entlarven lässt, sowie durch ihren KI-Logos, der sie mit Nathans Datenbank und damit unbegrenztem Wissen verbindet.

Ava repräsentiert das sowohl begehrte als auch bedrohliche Andere und weist damit Bezüge zur *femme fatale* auf, setzt hier aber auch neue Akzente[69] – wird sie doch am Ende weder getötet noch zur Ehefrau gemacht. Auch in *Real Humans* wird die Ambivalenz der KI als Wunsch- und Angstbild u. a. durch Re-

69 Farrimond 2018: 148–162.

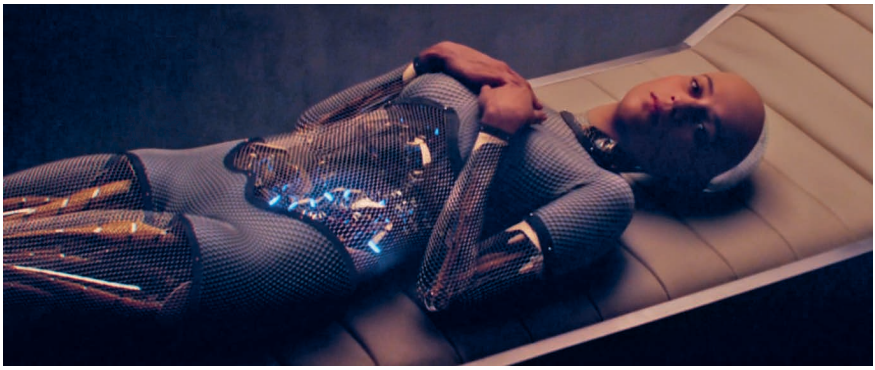

Abbildung 6 *Ex Machina*: Calebs Blick auf Ava über den Monitor

ferenzen auf die *femme fatale* ausgedrückt, was sich im Hubot Bea personifiziert: Bei Bea handelt es sich um einen freien Hubot, der unerkannt als Polizistin unter den Menschen lebt und eine Beziehung mit dem menschlichen Roger eingeht, der Hubots eigentlich ablehnt. Ihr *Outing* wird als Verführung inszeniert: Bea erwartet den heimkehrenden Roger mit Kerzen, Rosenblättern, in einem Negligé und mit dem zur Schau gestellten Ladekabel, das unter ihrer Achselhöhle hervortritt. Nach ihrer Aufforderung berührt Roger das Kabel, die Szene wird erotisch aufgeladen und Bea in begehrenswert-bedrohlicher Doppelcodierung zur phallischen Frau.[70] Ähnlich wie die bereits erwähnte Einstellung der sich zusammenrottenden Hubots kippt auch die Darstellung des Pflegeroboters Vera ins Unheimliche, wenn sie nachts mechanisch ans Bett der von ihr zu betreuenden Person tritt, um ihren Blutdruck zu messen.[71]

Demgegenüber steht die verklärende Inszenierung des Hubots Anita/Mimi, deren Erwachen gleich zweifach gezeigt wird: zunächst in einer Rückblende, als sie zum ersten Mal die Augen öffnet und durch eine Scheibe hindurch vom Sohn des Schöpfers beobachtet wird, für den sie zur geliebten, idealisierten, romantischen Gefährtin wird – ein erneuter Einsatz von Glas als Element, das den Blick auf KI akzentuiert und symbolisch auflädt.[72] Aber auch in der Gegenwart der Serienhandlung, als sie von der Familie Engmann in einem Hubot-Markt gekauft und von ihrem neuen Besitzer aktiviert wird, findet filmästhetisch Verklärung statt. Anita/Mimi erinnert in ihrem weißen Behältnis und der Farbcodierung (schwarzes Haar, helle Haut, rote Lippen) ikonographisch an Schneewittchen, der Akt der Erweckung auch an das Märchen Dorn-

70 Adam 2021: 628–629.

71 Adam/Knifka 2016: 355. Adam 2021: 620–621.

72 Adam/Knifka 2016: 358–359. Adam 2021: 624–626.

Abbildung 7 *Real Humans*: Bea offenbart ihr Ladekabel in einer Verführungsszene

Abbildung 8 *Real Humans*: Anita/Mimi wird von ihrer neuen Familie erweckt

röschen. »Der magische Kuss des traditionellen Märchens wird hier zum zwar technischen, aber [...] doch geheimnisvollen Betätigen des Schalters, durch das Mimi zum Leben erwacht und das maschinellen Futurismus in archetypische Rollenkonstellationen kleidet und teilweise auflöst.«[73] Die exemplarisch aufgezeigten Inszenierungsstrategien verorten KI typischerweise im Kontext von Subjektivierung und Identität sowie in den Spannungsfeldern zwischen dem Selbst und dem Anderen bzw. zwischen Wunsch- und Angstbildern.

E Diskursive Konstruktion von KI in Online-Diskussionen: Erste Ergebnisse und exemplarische Einblicke in Fallstudien der Linguistik

Die diskursive Konstruktion von KI erweist sich in den bislang untersuchten Online-Diskussionen als Kristallisationspunkt sozialer und medialer Aushandlungs- und Repräsentationsprozesse von Zukunftsvorstellungen. Die qualitative Analyse der Diskussionsstränge zeigt zunächst, dass in den partizipatorischen Online-Diskursen nicht systematisch zwischen *starker* und *schwacher* KI unterschieden wird. Einerseits umfasst in den bisher untersuchten Diskussionssträngen das Konzept der KI Systeme, die im Expertendiskurs als sogenannte *schwache* KI (Expertensysteme, Navigationssysteme, Spracherkennung, Filterfunktionen in der Werbung) bezeichnet werden. Andererseits werden KI häufig menschliche Eigenschaften zugesprochen (Gleichwertigkeit mit menschlichem Denkvermögen, Planung, Lernen, Entscheidungsfähigkeit, logisches Denken, Kommunikation in natürlicher Sprache), die im Expertendiskurs eher als *starke* KI bezeichnet werden, von deren erfolgreicher technischer Entwicklung die Menschheit derzeit jedoch noch weit entfernt ist. Beide alternierenden Perspektiven finden sich sowohl im Bereich der semantischen Topoi als auch im Bereich der force-dynamischen Konstruktionen unter dem gemeinsamen Label KI, was im alltagsweltlichen Laiendiskurs zum Teil zu diffusen (aus Expertensicht eher unbegründeten) Ängsten oder Hoffnungen führt.

Im Folgenden sollen die in Abschnitt B erläuterten linguistischen Perspektivierungstechniken (semantische Topoi – z. B. realisiert durch Prädikationen, Referenz und konzeptuelle Metaphern sowie durch force-dynamische Konstruktionen) exemplarisch anhand von Beispielen aus den untersuchten Diskussionen erläutert werden. In ihrer Gesamtheit konstruieren Beitragende bereits mittels der hier vorgestellten sprachlichen Perspektivierungstechniken, die auch innerhalb eines einzelnen Beitrages ein dichtes Netz sich teilweise

73 Adam 2021: 625–626.

überlagernder Perspektivierungselemente ergeben können, auch aus linguistischer Sicht wie im medienwissenschaftlichen Teil unseres Projektes wesentliche Aspekte kollektiver alltagsweltlicher KI-Narrative. Die Aufdeckung der Komponenten dieser kollektiven Narrative bildet die Grundlage für übergreifende medienvergleichende Schlussfolgerungen im anvisierten Gesamtprojekt.

I Semantische Topoi (Prädikationen und konzeptuelle Metaphern)

Semantische Topoi werden hier als wiederkehrende (Teile von) Diskursthemen verstanden.[74] Sie manifestieren sich in den untersuchten Online-Diskursen in erster Linie als in den Beiträgen über KI getroffene Aussagen (Prädikationen) in Form diskursiver Zuweisungen von qualitativen und quantitativen Merkmalen und Eigenschaften.[75] Damit kennzeichnen sie im KI-Diskurs entscheidende Aspekte der Selbst- und Fremdrepräsentation. Die in den untersuchten Daten zum Teil auch als konzeptuelle Metonymien[76] und Metaphern[77] realisierten Prädikationen sind zentrale Komponenten der nutzergenerierten alltagsweltlichen KI-Narrative. Diese ließen sich für die Pilotstudie parallel zu den Beobachtungen der Medienwissenschaft entlang der Achsen einer mehr oder minder stark *mit menschlichen Zügen ausgestatteten KI* (Topos 1), der *mehr oder minder machtvollen Position menschlicher Akteure* in ihren komplexen Beziehungen zu künstlichen Intelligenzen (Topos 2), der mehr oder minder ausgeprägten *Stärke und Macht von KI über den Menschen*, bis hin zu dessen *Bedrohung* (Topos 3) sowie einer mit großem *Potential oder Defiziten* ausgestatteten KI (Topos 4) verorten. Zusätzlich werden diese vier thematischen Narrative durch force-dynamische Konstruktionen (vgl. E.II) gestützt. Ein weiteres Narrativ mit erweitertem Akteursspektrum und deutlichen Schnittstellen zu den Topoi der untersuchten audiovisuellen Science-Fiction entsteht in den untersuchten Diskussionen im Umfeld der Postulate von Topos 5, *Der Mensch ist defizitär*, gestützt durch Topos 6, *kollektives oder individuelles Unwissen zu KI* mit seinen Bezügen zur in Abschnitt D diskutierten Intransparenz von KI. Die folgende Diskussion beschränkt sich auf wenige exemplarische Beispiele und beleuchtet das Ineinandergreifen der verschiedenen Topoi.

74 Reisigl 2007, Scollon et al. 2012.

75 Reisigl/Wodak 2001: 46, Hart 2010: 9, 65.

76 Kövecses 2010: 173, Radden/Kövecses 1999: 21–23.

77 Lakoff/Johnson 2003, Kövecses 2010.

Topos 1: KI trägt (+/–) menschliche Züge

Die Frage danach, ob KI menschliche Züge trägt, ist in den Diskussionen heftig umstritten und wird häufig Gegenstand von Aushandlungsdebatten. Dies zeigt sich nicht allein an den inhaltlich gegensätzlichen Positionierungen der Beitragenden wie in (1) und (2), sondern auch an häufigen metapragmatischen Aushandlungssequenzen – vgl. (2) und (3). In den bislang untersuchten Diskussionen weisen Beitragende KI häufig durch Prädikationen menschliche Züge im Allgemeinen, aber auch konkrete Eigenschaften wie Intelligenz, Können, Würde oder Emotionen zu bzw. stellen diese infrage. In (1) wird z. B. zunächst pauschal konstatiert, dass KI dem Menschen immer ähnlicher werde, bevor explizit der Furcht, dass Roboter uns eines Tages »überflügeln« werden, Ausdruck verliehen wird. Beispiel (3) illustriert die Zuweisung von *Rechten* und einem *eigenen Bewusstsein* an eine in filmischer Science-Fiction realisierte KI, die sich auch in der medienwissenschaftlichen Analyse zeigte (vgl. Punkt D).

(1) Unbenommen dessen, werden wir uns bald fragen müssen, wie wir mit Robotern umgehen sollen, *die uns immer ähnlicher werden, und vielen Belangen uns überflügeln!* [1Zeit2_74][78]

(2) Diese Roboterethik, die auf der zweiten Seite angesprochen wird, ist *meiner Meinung nach völliger Unsinn. Roboter sind Dinge*. Auch wenn sie irgendwann noch so lebensecht aussehen *sind sie trotzdem totes Material*. [1Zeit2_139]

(3) Zum Thema Roboterethik *hatte* die US-amerikanische TV-Serie *Star Trek. The next Generation schon vor einem Vierteljahrhundert viel Substantielleres beizutragen als Sie hier*. Insbesondere die Frage, *welche Rechte der mit einem eigenen Bewusstsein ausgestattete Android Commander Data haben sollte* … [1Zeit2_141]

Topos 2: Macht(losigkeit) und Manipulierbarkeit des Menschen und Topos 3: Macht(losigkeit) der KI, KI als Bedrohung

Im unmittelbaren Umfeld von Topos 1 positionieren sich die Beiträge im bislang untersuchten Korpus zum Verhältnis von Mensch und KI gleichfalls kontrovers. Während zahlreiche Beiträge wie (4)–(5) die Rolle des Menschen im

78 Aus Gründen der Authentizität verbleiben alle Beispiele (auch orthografisch und grammatisch) in ihrer ursprünglichen Form, werden aber gegebenenfalls auf die zur Illustration jeweils notwendigen Segmente gekürzt. Relevante Strukturen werden von den Verf. hervorgehoben.

Gefüge eines gesellschaftlichen Lebens mit KI als mächtig, überlegen und unersetzbar perspektivieren (besonders im kreativen und emotionalen Bereich z. B. der Pflege), zeichnet Beispiel (6) den Menschen als der KI unterlegen, gegebenenfalls auch durch die KI als Subjekt mit menschlichen Eigenschaften manipulier- oder ersetzbar und drückt damit tiefe Ängste in der alltagsweltlichen Wahrnehmung zukünftiger technischer Entwicklungen aus.

(4) Auf absehbare Zeit *wird die AI nicht an die menschliche Kreativität heranreichen*. [1SPON_137]
(5) *Computer werden niemals ein freundliches Gespräch ersetzten können*. Wenn sie einer älteren Dame helfen die notwendigen Windeln zu wechseln …, dann *kann nur ein netter und subtiler Mensch ihr ruhig übers Haar streicheln und sagen: Alles gut, kein Problem. Computer werden nie Menschen ersetzten*. [1Zeit3_105]
(6) *Kaufen sich die Roboter dann unnütze Menschen*, einfach, *weil sie es können*? Als Statussymbol? [1Zeit1_193]
(7) Kleiner Scherz: Natürlich wäre es ein schöner Traum, *wenn eine KI mächtig genug würde die Menschheit zur Vernunft zu zwingen*. [1SPON1_27])
(8) Ohne diese Instinkte *werden die KIs genau *nichts* tun – weder lernen noch irgendwie im Wald rumlaufen und auch keine anderen KIs zwecks Wissenstransfer treffen*. Also *muss Jemand diese Impulse »beifüttern«* [1SPON3_68]
(9) *Wie wollen wir etwas künstlich nachbauen, das wir beim Menschen nicht annähernd begreifen*? [1ZEIT1_45]

Entlang der gleichen semantischen Achse wird in den bislang untersuchten Diskussionen Topos 2 häufig durch den komplementären Topos 3 *Macht(losigkeit) der KI, KI als Bedrohung* – teilweise auch scherzhaft – wie in (7) realisiert, wohingegen die *Machtlosigkeit* der KI insbesondere im Verhältnis zu ihren Entwicklern hervorgehoben wird, die *diese Impulse beifüttern müssen* – vgl. (8).

Topos 2 ist in den bereits untersuchten Daten auch eng mit Topos 5, *Defizite des Menschen*, verbunden, wobei Topos 5 auf diese Defizite in Relation zur Zukunftsaufgabe der Entwicklung und effektiven aber auch ethisch angemessenen Nutzung von KI abzielt und zum kollektiven Narrativ einer *Unzulänglichkeit des Menschen gemessen an den technischen Herausforderungen der Zukunft* beiträgt – vgl. (9).

Topos 4: Potential und Defizite von KI vs. Topos 5: Defizite des Menschen

In der Diskussion eng verwoben mit Topos 3 *Macht(losigkeit) der KI und KI als Bedrohung* ist auch die Diskussion des Potentials und der Defizite von KI. Analog zu Topoi 1–3 zeigt sich auch hier die Diskussion durchaus polarisiert. Während eine Vielzahl von Beiträgen das positive Potential von KI, z. B. *ihr Gebrauchtwerden* wie in (10) insgesamt positiv bewertet (das Verhältnis positiver zu negativen Bewertungen in der Pilotstudie ist hier ca. 4:1), rücken andere Beiträge ihr negatives Potential oder explizit den defizitären Charakter von KI in den Mittelpunkt. Dies geschieht teils pauschal, wie in (11), wo neben der expliziten Prädikation, dass *Dinge bereits vor fünfzig Jahren analog schon erledigt* worden seien, die referenzielle Bezugnahme auf KI durch *Gadget-Brimborium heutzutage* die negative Bewertung des Potentials von KI zusätzlich verstärkt.

(10) Also ich sehe in einem KI *kein Schreckgespenst*, sondern *einen wichtigen Helfer* in einer Notlage auf die wir uns als Menschen im Kollektiv immer mehr zu bewegen. [1SPON1_57]
(11) *Das ganze Gadget-Brimborium heutzutage* ist doch oft nichts weiter als *eine Methode, Dinge digital schlechter zu erledigen als man sie vor fünfzig Jahren analog schon erledigt hat*. [1Zeit1_38]
(12) *Den FC St. Pauli (erkannt: »FC sind Pauli«) kennt Siri auch nicht*, obwohl ich seit über 10 Jahren in Hamburg lebe. *Wenn DAS Intelligenz ist, was ist dann dämlich?* [1SPON2_18]
(13) Verstehen Sie mich nicht falsch, ich sehe *jede Menge Chancen*, aber halt auch *jede Menge Risiken*, die man managen muss, wenn man die Chancen wahrnehmen will. [1SPON3_21].

Teilweise werden spezifische Defizite (Topos 5), z. B. in *Den FC St. Pauli … kennt Siri auch nicht* benannt und explizit bewertet wie in (12). Daneben finden sich jedoch auch Beiträge, in denen relativierende, beide Aspekte berücksichtigende Aussagen getroffen werden – vgl. (13), so dass auch Topos 4 zum Kristallisationspunkt konträrer Standpunkte wird.

Topos 6: Unwissen zu KI (kollektiv, individuell), Intransparenz von KI

Topos 6 bezieht sich auf ein erweitertes Akteursfeld, in dem nicht nur der Mensch als Individuum, das sich mit KI auseinandersetzt, bzw. als Kollektiv, das sich den Herausforderungen der Entwicklung und Nutzung von KI stellt, thematisiert wird. Stattdessen stehen die Unwissenheit der breiten Mehrheit

im Hinblick auf den Entwicklungs- und Planungsprozess von KI durch gesellschaftliche Eliten wie Entwickler:innen oder Politiker:innen, ihre Nutzung durch die Wirtschaft und große Unternehmen sowie die technischen Aspekte von KI, die für die Mehrheit der Nutzer:innen unbekannt und intransparent bleiben, im Mittelpunkt.

(14) Die dieser Entscheidung zugrunde liegenden *Algorithmen sind nicht transparent*, weil ja nicht programmiert, sondern erlernt – *auf welcher Datengrundlage auch immer*. [1SPON3_21]
(15) *Wer definiert diese grundlegenden Werte*, die im richtigen Leben von grundlegender Bedeutung sind und *wer garantiert diese Regeln*? *ist das alles der Wirtschaft überlassen*? [1SPON1_21]
(16) Die KI optimiert in die Richtung die ihr vorgegeben wird. *Diese Richtung werden einzelne Personen(gruppen) vorgeben. Dies gibt den Personen(gruppen) diktatorengleiche Macht über den grossen Rest*. … [1SPON2_81]

Dieses Nichtwissen, sprachlich z. T. auch wie in (15) durch offene Fragen in Gestalt von Interrogativsätzen realisiert, die seitens der Fragenden das Nichtwissen der Antwort implizieren, positioniert den Menschen im alltagsweltlichen Narrativ der Beitragenden in komplexem Sinne als mental/intellektuell und in seiner Handlungsmacht einem als *intransparent* erlebten Agens unterlegen, wobei zum Teil wie in (15) und (16) zusätzlich force-dynamische Konfigurationen (vgl. Punkt E.II) zur Perspektivierung des Geschehens zum Tragen kommen.

Metaphorische und metonymische Perspektivierungen

Im Folgenden sollen am Beispiel von Topos 1 und 3 exemplarisch einige metaphorische und metonymische Perspektivierungen vorgestellt werden, die gleichfalls deutliche Parallelen zur medienwissenschaftlichen Analyse aufweisen. Insbesondere Topos 1 ist im Korpus häufig durch konzeptuelle Personifizierungsmetaphern aus der Gruppe EIN ABSTRAKTES KOMPLEXES SYSTEM IST DER MENSCHLICHE KÖRPER/DIE EIGENSCHAFTEN UNBELEBTER DINGE SIND MENSCHLICHE EIGENSCHAFTEN[79] realisiert. Ihre zum Teil mehrfach innerhalb eines Beitrages anzutreffenden Mappingprozesse zeigen die jeweils von den Beitragenden in den Vordergrund gerückten teils hochkomplexen teils aber auch elementaren menschlichen Eigenschaften, die KI metaphorisch zugewiesen werden, wie z. B. WISSEN/INTELLEKT, BEWUSSTSEIN, EMOTIONEN,

79 Kövecses 2010: 157–158.

EMPATHIE, MENSCHLICHE AKTIVITÄTEN, CHARAKTEREIGENSCHAFTEN, MORAL, MENSCHLICHE ENTWICKLUNGSSTADIEN – vgl. (17)–(20).[80]

(17) Also solange das alles nur vorgefertigte *Antworten* [MENSCHLICHE AKTIVITÄTEN (SPRACHE)] sind und nicht wirklich ne anständige KI selbstständig »*denkt*« [BEWUSSTSEIN] … [Amazon Echo_22]
(18) Die KI *sagt* [MENSCHLICHE AKTIVITÄTEN (SPRACHE)] dann ich kenne Dein Verlangen und Deine persönlichen Verhältnisse [EMPATHIE, WISSEN] [1SPON1_69]
(19) zukünftigen KIs, die recht bald vom *Homunkulus zum erwachsenen KI mutieren* [MENSCHLICHE ENTWICKLUNG]. [1SPON1_14]
(20) Die KI wird immer *egoistischer* [CHARAKTEREIGENSCHAFTEN, MORAL] [1SPON_1_64]

Die hoch abstrakte konzeptuelle Metapher DIE EIGENSCHAFTEN UNBELEBTER DINGE SIND MENSCHLICHE EIGENSCHAFTEN ist insbesondere in Topos 3 (*Macht(losigkeit) der KI, KI als Bedrohung*) auch als spezifischere Metapher KI IST EIN MACHTVOLLES AGENS/DIKTATOR realisiert – vgl. (21).

(21) *Herrschaft* [MENSCHLICHE AKTIVITÄTEN (HERRSCHEN)] der maschinellen Intelligenz (1SPON4_18)
(22) *KI befiel, wir folgen!* [MENSCHLICHE AKTIVITÄTEN (BEFEHLEN); KI IST EIN MACHTVOLLES AGENS/DIKTATOR] [1SPON4_12]

Alternativ finden sich auch metonymische Perspektivierungen von KI, in der z. B. in TEIL-FÜR-GANZES Metonymien über einen salienten Teil der Domäne MÄCHTIGER AKTEUR/DIKTATOR auf das Gesamtkonzept zugegriffen wird – vgl. (22). *KI befiel, wir folgen* referiert assoziativ auf *Führer befiehl, wir folgen* als eine der kulturell salienten Losungen des Dritten Reiches, mit denen Diskursteilnehmer:innen, die über dieses kulturelle Hintergrundwissen verfügen (das im Beitrag offensichtlich vorausgesetzt wird), mühelos auf die Gesamtdomäne DRITTES REICH und in einem weiteren metonymischen Schritt (DRITTES REICH FÜR DIKTATUR) auf das Gesamtkonzept der als diktatorisch perspektivierten KI zugreifen können.

80 Die jeweils einschlägigen Mappingkategorien wurden in den Beispielen zur metaphorischen Perspektivierung jeweils durch die Verfasser:innen in eckigen Klammern hinzugefügt.

II Force-dynamische Konstruktionen – KI als starkes (Alter) Ego vs. Mensch als schwaches (Alter) Ego

Force-dynamische Perspektivierungsoperationen werden in der Kognitiven Semantik als übergreifende linguistische und konzeptuelle Operationen verstanden, mit denen Kausalverhältnisse im weitesten Sinne (wie z. B. *etwas verursachen – (geschehen) lassen, etwas befördern – verhindern*) konstruiert werden.[81] Sie rücken damit relationale Beziehungen zwischen Ego und Alter Ego (*agonist – antagonist*) sowie deren gegenseitige Bedingtheit und Konstruktion gleichermaßen wie den Umstand, dass beide Entitäten eines Geschehens jeweils (offensichtlich oder verdeckt) in reziproker Beziehung zueinander als *schwächer* oder *stärker* perspektiviert werden,[82] in den Mittelpunkt. Linguistisch lassen sich force-dynamische Operationen nicht nur lexikalisch durch Prädikatsausdrücke mit semantisch kausalem Bedeutungspotential realisieren (vgl. im Korpus verwendete lexikalischen Konstruktionen für die starke (Alter) Ego-Perspektivierung in (23) oder für die Konstruktion eines schwachen (Alter) Egos in (24)). Möglich sind z. B. auch grammatische Satzbaukonstruktionen, deren (vollständiges) semantisches Rolleninventar wie in (25) explizit AGENS und PATIENCE und damit starkes Ego und schwaches Alter Ego abbildet.

(23) *beherrschen, beurteilen, bevormunden, degradieren, Einfluss/Macht gewinnen, emanzipieren, entmenschlichen, entmündigen, entscheiden, ermöglichen, ersetzen, etwas (er)schaffen, etwas abschalten, gefügig machen, gestalten, hinters Licht führen, kontrollieren, kümmern, manipulieren, neutralisieren, ruhigstellen, überrennen, verdrängen, verführen, (ver)schaffen, von etwas abbringen, zwingen, …*

(24) *abhängig, akzeptieren, ausgeliefert, in jemandes Hand sein, erlauben, sich anpassen, sich beherrschen lassen, sich ergeben/unterwerfen, zulassen, (nicht) gebraucht werden, (noch) können, müssen, nachmachen;* deontische Adverbien wie *zwangsweise;* nominale Prädikationen (*z. B. Werkzeug, Sklave*), …

(25) Es geht darum, ob »künstliche Intelligenz« [AGENS] den DURCHSCHNITTLICHEN Menschen [PATIENCE] auf Dauer zu einem Volltrottel degradieren wird. [1ZEIT_1_44]

(26) Wenn nämlich eine sog. starke KI erst einmal in freier Wildbahn existiert, werden wir nicht mehr entscheiden können, was wir ihr zu tun gestatten, dann wird eher andersherum ein Schuh draus. [1SPON_1_13]

81 Talmy 2000: 409.

82 Oakley 2005: 450.

(27) Nur dass sie natürlich niemals schlauer oder objektiver sein können als diejenigen, die sie trainieren. [1SPON_1_119]

Für die bislang untersuchten Daten reflektiert die force-dynamische Perspektivierung, wie sehr KI auch im alltagsweltlichen Laiendiskurs als ein Kristallisationspunkt gesellschaftlicher Debatten wirkt. Auch auf dieser semantisch subtilen Ebene zeugen die Beiträge von unbestimmten Ängsten, indem KI deutlich häufiger als der Mensch explizit als starkes Ego – vgl. (26) – und deutlich seltener explizit als schwaches Alter Ego – vgl. (27) – konstruiert wird. Umgekehrt wird der Mensch deutlich häufiger als schwaches Alter Ego, das dem Wirken der KI ausgesetzt ist (26), ihm wenig entgegenzusetzen hat, und deutlich seltener als starkes Ego, das machtvoll auf die KI einwirkt (27), perspektiviert. Diese force-dynamischen Konstruktionen unterstützen insbesondere die Machtkonstellationen in den in Punkt E.I diskutierten Topoi 2 und 3, unterscheiden sich in dieser klaren Tendenz jedoch von den Ergebnissen der medienwissenschaftlichen Untersuchung.

F Die Konstruktion von KI im Vergleich: Fazit und Ausblick

Die vorgestellte interdiskursive Analyse hat für erste abgeschlossene Fallstudien (*Ex Machina*, *Eva* und die Serie *Real Humans*) und die bereits untersuchten Online-Kommentare für beide Diskursdomänen rekurrierende, sich zum Teil überlagernde Topoi offengelegt. Diese wurden hinsichtlich ihrer spezifischen Kontextualisierung, Inszenierung und Funktion als Element von Sinn- und Bedeutungskonstruktion im Rahmen von Motiven und Narrativen analysiert. So wurden sowohl Gemeinsamkeiten als auch Unterschiede in ihren diskursiven Perspektivierungen sowie Besonderheiten ihrer ästhetischen und narrativen Umsetzung in der filmischen Repräsentation ermittelt, von denen hier nur eine kleine Auswahl exemplarisch benannt werden kann.

Häufig verwendete zentrale Diskurstopoi beider Korpora umfassen z. B. *Menschliche Hybris*, *das Motiv des Schöpfers*, *die (In-)Transparenz*, *Semitransparenz*, *Opazität technischer Vorgänge und deren Mystifizierung*, *Machtverhältnisse und Macht(losigkeit)* inklusive der Angst des Menschen, durch die Technik ersetzt zu werden (z. B. in *Real Humans* und Beispiel (1) der Online-Kommentare) sowie Fragen der ethischen Identität von KI, inklusive ihrer Rechte (vgl. *Ex Machina*, *Eva* und Topos 1 in den Online-Kommentaren).

Gemeinsame Topoi unterscheiden sich zum Teil in ihrer Gewichtung und Perspektivierung (z. B. *mit technologischem Fortschritt verknüpfte antizipierende Erwartungen und Ängste*, *Transparenz vs. Intransparenz von Wissen*, *Humanisierung/Personifizierung von KI und die Hybris des Menschen*). Diese wer-

den in den partizipatorischen Online-Diskursen explizit verbalisiert und auf die unmittelbare Lebenswirklichkeit der Nutzer:innen zugeschnitten. Im Unterschied dazu bedienen sich die Filme gezielter Erzähl- und Visualisierungsstrategien, um ihre utopischen/dystopischen Inszenierungen von Technik zusätzlich zu emotionalisieren und mit weiteren Bedeutungen zu konnotieren. Unterschiede zeigen sich hier insbesondere in der Performativität und geschlechtlichen Codierung (z. B. *Ex Machina* und *Eva*), die im Pilotkorpus primär im filmischen Diskurs, nicht aber im partizipatorischen Online-Diskurs anzutreffen ist.

Weitere spezifische topologische Variationen zeigen sich im Science-Fiction-Korpus z. B. für den Topos *Hybris des Menschen und Macht der Technologie, Machbarkeitsphantasien und schöpferischer Größenwahn* sowie ihre häufige Verortung in religiösen Kontexten – vgl. Eva und Ava in *Ex Machina*. Auch subversive Akte der Selbst- und Neuschöpfung – vgl. *Ex Machina* – finden sich im Online-Diskurs nur in Ansätzen (vgl. etwa Bsp. (6) im Kontext der Topoi *Macht(losigkeit) und Manipulierbarkeit des Menschen* und *KI als Bedrohung*).

In beiden Korpora werden die genannten topologischen Gemeinsamkeiten und Unterschiede häufig durch konzeptuelle Metaphern z. B. ABSTRAKTE KOMPLEXE SYSTEME SIND DER MENSCHLICHE KÖRPER (z. B. mit Gestalt- und Bildmetaphern sowie Mappings bezogen auf den menschlichen Lebenszyklus), DIE EIGENSCHAFTEN UNBELEBTER DINGE SIND MENSCHLICHE EIGENSCHAFTEN (z. B. mit Mappings bezogen auf WISSEN UND BEWUSSTSEIN, MENSCHLICHE SPRACHE oder MENSCHLICHE WÜRDE UND INTEGRITÄT) umgesetzt. Diese werden teilweise analog zu anderen Diskursdomänen[83] in komplexen metaphorischen Szenarien verhandelt, deren detaillierte Untersuchung jedoch noch aussteht. Insbesondere die Topoi *Macht(losigkeit) und Manipulierbarkeit des Menschen und der KI* werden im linguistischen Korpus zusätzlich explizit durch entsprechende force-dynamische Konfigurationen konstruiert (vgl. E.II).

Das Online-Korpus weist darüber hinaus auch Instanzen sprachlicher Strukturen auf, die dem linguistischen Inventar populistischer Diskurse zugerechnet werden (z. B. sprachliche Strategien horizontaler und vertikaler Identitätszuweisungen in *Wir-und-die-Anderen*-Dichotomien – hier im Umfeld des Topos der *(In-)Transparenz gesellschaftlicher Entscheidungsprozesse* (vgl. E.I, Topos 6). Diese ließen sich in dieser Form im hier diskutierten filmischen Diskurs nicht beobachten.

Die Ergebnisse der Pilotstudie skizzieren wichtige Anhaltspunkte dafür, wie sich die Gesellschaft den neuen Dimensionen der Veränderungen unserer Lebenswelt durch Wissenschaft und Technik in komplexen sprachlichen und fiktional-bildlichen Positionierungsprozessen stellt, die im Sinne der Ent-

83 Musolff 2006, Vogelbacher 2019.

wicklung komplexer Technikzukünfte in gesellschaftliche Entscheidungen zu KI als technologische Chance und Herausforderung einfließen sollten. Weitere Untersuchungen sollen semantische Inhaltsanalyen größerer Korpora eines breiteren Spektrums gesellschaftlich relevanter Diskursdomänen umfassen. So sollen die Komplexität gesellschaftlicher Diskursivierungen und ihr Beitrag zur Konstruktion und Verhandlung gesellschaftlicher Wissensbestände zu KI ebenso wie damit verbundene Hoffnungen und Ängste umfassender und gleichzeitig differenzierter, bezogen auf verschiedene gesellschaftliche Akteursgruppen, abgebildet werden. Um der mit der weiteren Entwicklung von KI verbundenen globalen Herausforderung einer immer stärker technologisierten Welt gerecht werden zu können, konzentriert sich die Fortführung des Projektes auch auf den Sprach- und Kulturvergleich.

Literatur

acatech (Hg.) 2012: Technikzukünfte: Vorausdenken – Erstellen – Bewerten. Heidelberg, Springer.

Adam, Marie-Hélène 2021: Technikutopien und Genderkonzepte. Populärkulturelle Repräsentationen von Geschlecht in Science-Fiction-Filmen und -Fernsehserien als Prozess ambivalenter Bedeutungskonstruktion, phil. Diss., Karlsruher Institut für Technologie.

Adam, Marie-Hélène/Knifka, Julia 2016c: Beyond the Uncanny Valley. Inszenierung des Unheimlichen als Wunsch- und Angstbilder in der Serie ›Echte Menschen – Real Humans‹. In: Dies./Gellai, Szilvia (Hgg.): Technisierte Lebenswelt. Über den Prozess der Figuration von Mensch und Technik. Bielefeld, transcript: 341–365.

Aldiss, Brian 1973: Billion Year Spree. The True History of Science Fiction. New York, Doubleday.

Aldiss, Brian: The Twinkling of an Eye Or My Life as an Englishman, London: Little, Brown and Company 1998.

Brand, Lukas 2018: Künstliche Tugend: Roboter als moralische Akteure, Regensburg: Verlag Friedrich Pustet.

Busse, Dietrich 2012: Frame-Semantik. Ein Kompendium. Berlin, De Gruyter.

Cienki, Alan 2007: Frames, Idealized Cognitive Models, and Domains. In: Geeraerts, Dirk/Cuyckens, Hubert (Hgg.): The Oxford Handbook of Cognitive Linguistics. Oxford, Oxford University Press: 170–187.

Cornea, Christine 2007: Science Fiction Cinema. Between Fantasy and Reality. Edinburgh, Edinburgh University Press.

Cross, Katherine 2015: Goddess from the Machine. A Look at Ex Machina's Gender Politics. In: Feministing. http://feministing.com/2015/05/28/goddess-from-the-machine-a-look-at-ex-machinas-gender-politics/ (aufgerufen am 07. 05. 2021).

Dienel, Hans-Liudger 2015: Transdisziplinarität, in: Gerhold, Lars/Holtmannspötter, Dirk/Neuhaus, Christian/Schüll, Elmar/Schulz-Montag, Beate/Steinmüller, Karlheinz/Zweck, Axel (Hgg.): Standards und Gütekriterien der Zukunftsforschung: Ein Handbuch für Wissenschaft und Praxis, Wiesbaden: Springer Verlag.

Drux, Rudolf (Hg.) 1999: Der Frankenstein-Komplex. Kulturgeschichtliche Aspekte des Traums von künstlichen Menschen. Frankfurt am Main, Suhrkamp.

Eller, Monika 2017: Reader Response in the Digital Age. Letters to the Editor vs. below-the-Line Comments. A Synchronic Comparison. unv. Diss., Ruprecht-Karls-Universität Heidelberg.

Farrimond, Katherine 2018: The Contemporary Femme Fatale. Gender, Genre and American Cinema. New York/London, Routledge.

Felder, Ekkehard 2009: Sprachliche Formationen des Wissens. Sachverhaltskonstitution zwischen Fachwelten, Textwelten und Varietäten. In: Felder, Ekkehard/Müller, Marcus (Hgg.): Wissen durch Sprache. Theorie, Praxis und Erkenntnisinteresse des Forschungsnetzwerkes »Sprache und Wissen«, Bd. 3: Sprache und Wissen. Berlin, de Gruyter: 21–78.

Felder, Ekkehard (Hg.) 2013: Faktizitätsherstellung in Diskursen. Die Macht des Deklarativen. Berlin, De Gruyter.

Felder, Ekkehard/Gardt, Andreas (Hgg.) 2015: Handbuch Sprache und Wissen. Berlin, De Gruyter.

Fillmore, Charles 1976: Frame Semantics and the Nature of Language. In: Harnad, Steven R./Steklis, Horst D./Lancaster, Jane (Hgg.): Origins and Evolutions of Language and Speech. New York, Academy of Sciences: 20–32.

Fitting, Peter 1993: What is Utopian Film? An Introductory Taxonomy. In: Society for Utopian Studies 4(2): 1–17.

Fitting, Peter 2003: Unmasking the Real? Critique and Utopia in Recent SF Films. In: Baccolini, Raffaella/Moylan, Tom (Hgg.): Dark Horizons. Science Fiction and the Dystopian Imagination. New York/London, Routledge: 155–166.

Fraas, Claudia 2013: Frames. Ein qualitativer Zugang zur Analyse von Sinnstrukturen in der Online-Kommunikation. In: Frank-Job, Barbara/Mehler, Alexander/Sutter, Tilmann (Hgg.): Die Dynamik sozialer und sprachlicher Netzwerke. Wiesbaden, Springer: 259–283.

Fraas, Claudia/Meier, Stefan/Pentzold, Christian 2012: Online-Kommunikation. Grundlagen, Praxisfelder und Methoden. München, Oldenbourg.

Glaser, Barney G./Strauss, Anselm L. 1967: The Discovery of Grounded Theory. Strategies for Qualitative Research. Chicago, Aldine Publ.

Göcke, Benedikt Paul (Hg.) 2018: Designobjekt Mensch. Freiburg, Herder.

Goffman, Erving 1981: Forms of talk. Oxford, Blackwell.

Goffman, Erving (1974/1986): Frame Analysis: An Essay on the Organization of Experience. Boston: Northeastern University Press. [ursprünglich 1974 erschienen bei Harper Colophone, New York]

Haraway, Donna 2007: Ein Manifest für Cyborgs. Feminismus im Streit mit den Technowissenschaften. In: Bruns, Karin/Reichert, Ramon (Hg.): Reader Neue Medien. Texte zur digitalen Kultur und Kommunikation. Bielefeld, transcript: 238–277.

Hart, Christopher 2010: Critical Discourse Analysis and Cognitive Science. New Perspectives on Immigration Discourse. Basingstoke, Palgrave Macmillan.

Haynes, Roslynn D. 1994: From Faust to Strangelove. Representation of the Scientist in Western Literature. Baltimore/London, John Hopkins University Press.

Haynes, Roslynn D. 2017: From Madman to Crime Fighter. The Scientist in Western Culture. Baltimore, John Hopkins University Press.

Henke, Jennifer 2018: »Ava's body is a good one.« (Dis)Embodiment in Ex Machina. In: American, British, and Canadian Studies 29 (1): 126–146.

Jancovich, Mark/Johnston, Derek 2011: Film and Television, the 1950s. In: Bould, Mark/Butler, Andrew M./Roberts, Adam/Vint, Sherryl (Hgg.): The Routledge Companion to Science Fiction. London/New York, Routledge: 71–79.

Johansson, Marjut/Kleinke, Sonja/Lehti, Lotta 2017: The digital agora of social media: Introduction. In: Discourse, Context & Media 19: 1–4.

Kleinke, Sonja/Schultz, Julia 2019: Ist ›Nation‹ gleich ›nation‹?. Zwei Wikipedia-Artikel im Sprach- und Kulturvergleich. In: Diskurse – digital, Bd. 1. Mannheim, Universität Mannheim, Philosophische Fakultät, Seminar für deutsche Philologie der Universität Mannheim, Germanistische Linguistik.

Lakoff, George/Johnson, Mark 2003: Metaphors We Live by [with a new afterword]. Chicago, IL, University of Chicago Press.

Kövecses, Zoltán 2010^2: Metaphor. A Practical Introduction. Oxford, Oxford University Press.

Langacker, Ronald W. 1987: Foundations of Cognitive Grammar, Bd. 1. Stanford, CA, Stanford University Press.

Langacker, Ronald W. 2008: Cognitive Grammar. A Basic Introduction. Oxford, Oxford University Press.

Marcoccia, Michel 2004: On-line polylogues: conversation structure and participation framework in internet newsgroups. In: Journal of Pragmatics 36 (1): 115–145.

Mederake, Nathalie 2015: Wikipedia. Text- und Wissensverfahren im kollaborativen Hypertext. Frankfurt am Main, Peter Lang.

Micheline, J. A. 2015: Ex Machina. A (White) Feminist Parable for Our Time. In: Women Write About Comics, 21. 05. 2015. https://womenwriteaboutcomics.com/2015/05/ex-machina-a-white-feminist-parable-for-our-time/ (aufgerufen am 07. 05. 2021).

Miller, Cynthia J. 2012: Domesticating Space. Science Fiction Serials Come Home. In: Telotte, J. P./Duchovnay, Gerald (Hgg.): Science Fiction Film, Television, and Adaptation. Across the Screens. New York/London, Routledge: 3–19.

Musolff, Andreas 2006: Metaphor scenarios in public discourse. In: Metaphor & Symbol 21 (1): 23–38.

Musolff, Andreas 2016: Political Metaphor Analysis. Discourse and Scenarios. London, Bloomsbury Academic.

Neumeier, Otto (Hg.) 1994.: Angewandte Ethik im Spannungsfeld von Ökologie und Ökonomie. Sankt Augustin, Academia Verlag.

Oakley, Todd 2005: Force-dynamic dimensions of rhetorical effect. In: Hampe, Beate/Grady, Joseph E. (Hgg.): From Perception to Meaning. Image Schemas in Cognitive Linguistics. Berlin, De Gruyter: 443–473.

Page, Michael R. 2016: The Literary Imagination from Erasmus Darwin to H. G. Wells. Science, Evolution and Ecology. London/New York, Routledge.

Papacharissi, Zizi 2010: A private sphere. Democracy in a digital age. Cambridge, Polity Press.

Pentzold, Christian/Fraas, Claudia 2018: Verbale und visuelle Medienframes im Verfahrensrahmen der Grounded Theory analysieren. In: Scheu, Andreas M. (Hg.): Auswertung qualitativer Daten in der Kommunikationswissenschaft. Strategien, Verfahren und Methoden der Interpretation nichtstandardisierter Daten in der Kommunikationswissenschaft. Wiesbaden, Springer: 227–246.

Radden, Günter/Kövecses, Zoltán 1999: Towards a Theory of Metonymy. In: Panther, Klaus-Uwe/Radden, Günter (Hgg.): Metonymy in Language and Thought. Amsterdam, John Benjamins: 17–61.

Reisigl, Martin 2007: Discrimination in discourses. In: Kotthoff, Helga/Spencer-Oatey, Helen (Hgg.): Handbook of Intercultural Communication. Berlin, De Gruyter: 365–394.

Reisigl, Martin/Wodak, Ruth 2001: Discourse and Discrimination. Rhetorics of Racism and Antisemitism. London, Routledge.

Reisigl, Martin/Wodak, Ruth 2009²: The discourse-historical approach. In: Wodak, Ruth & Meyer, Michael (Hgg.): Methods of critical discourse analysis. London, Sage: 87–121.

Ruddick, Nicholas 2016: Science Fiction Adapted to Film. Canterbury, Gylphi Limited.

Saage, Richard 1997: Utopie und Science-fiction. Versuch einer Begriffsbestimmung. In: Hellmann, Kai-Uwe/Klein, Arne (Hgg.): »Unendliche Weiten…«. Star Trek zwischen Unterhaltung und Utopie. Frankfurt a. M., Fischer: 47–60.

Sargent, Lyman Tower 1994: The Three Faces of Utopianism Revisited. In: Utopian Studies 5 (1): 1–37.

Scollon, Ron/Scollon, Suzanne Wong/Jones, Rodney H. 2012: Intercultural Communication: A Discourse Approach. Malden, MA, Wiley-Blackwell.

Schrögel, Philipp/Weitze, Marc-Denis 2018: Comics als visueller Zugang zum transdisziplinären Diskurs über Technikzukünfte. In: Lettkemann, Eric/Wilke, René/Knoblauch, Hubert (Hgg.): Knowledge in Action. Wiesbaden, Springer: 21–48.

Semino, Elena 2008: Metaphor in discourse. Cambridge, Cambridge University Press.

Sobchack, Vivian 1998: Screening Space. The American Science Fiction Film. 2nd, Enlarged Edition. New Brunswick/London, Rutgers University Press.

Sobchack, Vivian 2005: American Science Fiction Film. An Overview. In: Seed, David (Hg.): A Companion to Science Fiction. Oxford, Blackwell: 261–274.

Spiegel, Simon 2007: Die Konstitution des Wunderbaren. Zu einer Poetik des Science-Fiction-Films. Marburg, Schüren.

Spieß, Constanze 2011: Diskurshandlungen. Theorie und Methode linguistischer Diskursanalyse am Beispiel der Bioethikdebatte. Berlin, De Gruyter.

Stollfuß, Sven 2016: Cyborg-TV. Genetik und Kybernetik in Fernsehserien. Wiesbaden, Springer VS.

Suvin, Darko 1977: Pour une poétique de la science-fiction. Études en théorie et en histoire d'un genre littéraire. Montréal, Presses de l'Univ. du Québec.

Talmy, Leonard 2000: Toward a Cognitive Semantics, Vol 1. Cambridge, MA, MIT Press.

Telotte, J. P. 2016: Robot Ecology and the Science Fiction Film. New York/Abingdon, Routledge.

Tereick, Jana 2016: Klimawandel im Diskurs. Berlin, De Gruyter.

van Dijk, Teun A 2008: Discourse and power. Basingstoke, Palgrave Macmillan.

Verhagen, Arie 2007: Construal and Perspectivization. In: Geeraerts, Dirk/Cuyckens, Hubert (Hgg.): The Oxford Handbook of Cognitive Linguistics. Oxford, Oxford University Press: 48–81.

Vogelbacher, Stefanie 2019: Scenario Negotiation in Online Debates about the European Union: Analysing Metaphor in Communication. Duisburger Arbeiten zur Sprach- und Kulturwissenschaft Bd. 123, Berlin, Peter Lang (zugleich Dissertation Universität Heidelberg 2018).

Weber, Jutta 2008: Sex, Service und digitale Geborgenheit. Technoimaginationen des Humanoiden zwischen Fiktion und Dienstleistungsökonomie. In: Maltry, Karola/Holland-Cunz, Barbara/Köllhofer, Nina/Löchel, Rolf/Maurer, Susanne (Hgg.): genderzukunft. Zur Transformation feministischer Visionen in der Science Fiction, Königstein: Ulrike Helmer Verlag: 189–201.

Wehling, Elisabeth 2017: Politisches Framing. Wie eine Nation sich ihr Denken einredet und daraus Politik macht. Köln, Herbert von Halem.

Weingart, Peter/Pansegrau, Petra 2003: Introduction: Perception and Representation of Science in Literature and Fiction Film. In: Public Understanding of Science 12(3): 227–228.

Weingart, Peter/Muhl, Claudia/Pansegrau, Petra 2003: Of Power Maniacs and Unethical Geniuses. Science and Scientists in Fiction Film. In: Public Understanding of Science 12 (3): 279–287.

Wodak, Ruth/de Cillia, Rudolf/Reisigl, Martin/Liebhart, Karin (Hgg.) 2009²: The Discursive Construction of National Identity. Edinburgh, Edinburgh University Press.

ORCID

Sonja Kleinke https://orcid.org/0000-0002-6165-0918

Katrin Strobel https://orcid.org/0000-0001-7209-661X

Demythologizing Artificial Intelligence[1]

Reflections on the Role and Purpose of Complex Data Processing in Digital Media Transformation

Jonas Bedford-Strohm

»Digital technology begins where the world can be represented in data in order to perceive patterns and structures that cannot be noticed by the human eye and the natural mental capacity to perceive and compute,«[2] Nassehi observes. Once this world is precariously duplicated into data, this representational world develops a life of its own. This duplicate »reality of its own kind«[3] is only loosely related to what we consider our »original« life world. The duplicate reality is self-referential in that it can only communicate or relate to that which is external if it comes in its own form. Data »know the world only in their own data form and cannot escape from it. Everything that appears in it must take data form itself.«[4] Thus, data can process the world only in its own image.

---

1 Earlier results of the research for this article are published in German: Bedford-Strohm 2019.

2 Nassehi 2019: 229. The German original: »Die Digitaltechnik fängt dort an, wo sich die Welt in Daten repräsentieren lässt, um Muster und Strukturen zu erkennen, die mit bloßem Auge und den Wahrnehmungs- und Rechenkapazitäten des natürlichen Bewusstseins nicht erfasst werden können.«

3 Nassehi 2019: 114. The original: »Realität eigener Art«.

4 Nassehi 2019: 111. The original: »Daten … [sind] in besonders radikaler Weise auf sich verwiesen …, denn sie kennen die Welt eben nur in ihrer je eigenen Datenform und können daraus nicht ausbrechen. Alles, was dort wieder vorkommt, muss selbst Datenform annehmen.«

Jonas Bedford-Strohm 2022: Demythologizing Artificial Intelligence. Reflections on the Role and Purpose of Complex Data Processing in Digital Media Transformation. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 55–80.
https://doi.org/10.11588/heibooks.1106.c15193

Hence, we are faced with problems of representation because all simplistic notions of »original« and »duplicate« are rendered precarious. The digital twin in data form is never a perfect representation, rather it is a precarious duplicate that impacts what it duplicates, especially when the object of representation is a human subject. Hence, Nassehi speaks of duplication (*Verdopplung*) as an »ironic concept, since [...] what appears to be a duplication in practice turns out to mean the exact opposite: a new creation of something that only exists by being duplicated.« In this practice, he notes, »we stabilize our life world by duplicating it and pretending that it is as it appears.«[5]

Since the widely applied forms of artificial intelligence rely on the complex data processing techniques of machine learning, Nassehi's observations about data can help us identify key problems in the application of artificial intelligence. If, the reasoning goes, we are able to understand AI's most vital processing *resource* (data), we can better understand the processing *techniques* that we call »AI.« Hence, we shall explore the »stuff« that makes artificial intelligence algorithms effective first, before we zoom out to further explore the conditions under which these algorithms are deployed.

If we take Nassehi's ideas seriously, we are challenged by (at least) two sets of questions – one conceptual, one ethical: 1) What do we mean when we speak of data? How can we understand this form that is so self-referentially totalitarian that it will only accept communication with the world when it takes its own form? 2) What does the process of duplication or representation in data form look like and imply? What are its conditions? Who can trigger this process? Who can navigate and utilize the »duplicate« reality it creates? And why would they be incentivized to do so?

The former challenge takes the form of a conceptual exploration: We need to reconstruct and understand what we are speaking of, before we evaluate and analyze its uses and impacts. The latter challenge takes the form of an ethical exploration: We need to notice the uses and impacts of what we reconstructed conceptually, and critically probe the conditions for its possibility. We will therefore first define our concept of data (A) and reconstruct data processing in the form of a four-fold typology (B). Only then will we trace relevant contentions of critical data theory (C) and explore the ethics of complex data processing (D) by analyzing the necessary conditions of its practice.

---

5 Nassehi 2019: 113. The original: »Verdopplung ist gewissermaßen ein ironischer Ausdruck, weil er auf die Paradoxie aufmerksam macht, dass das, was praktisch als Verdopplung erscheint, exakt das Gegenteil bedeutet: eine Neuschöpfung von etwas, das nur dadurch existiert, dass es verdoppelt wird. ... Wir stabilisieren Lebenswelten, indem wir die Welt verdoppeln und zugleich so tun, als sei sie so, wie sie lebensweltlich erscheint.«

A Definitions of Data

In the German subculture of those who deal with data as a feature and resource of digital society, the international buzzwords Big Data, Artificial Intelligence and Deep Learning, but also German terms like *Datenschutz*, *Datenleck* und *Datenverarbeitung* are used ubiquitously. In the vast sea of thematic content, however, precise definitions of the discussed terms are surprisingly rare – despite their ubiquitous use. Although the term »data« in its various forms is employed so ubiquitously, its use, as with many terms of everyday use, remains blurred. In the German discussion, »data« usually means: *digital micro-packets of communication that can be stored electronically and, with the right methods of interpretation, become substantive information in the right context.*[6]

If we understand data in this way, the relationship between analog and digital dimensions in the source and structure of data comes into focus with the criterion of electronic storability.[7] It implies that only *electronic* data is to be understood as data in a meaningful sense. But the word »date« (*Datum*), derived from the Latin word »dare« which translates to the English »to give,« permits a diverging definition: One single »date« as the elemental building block of the multitude of »data« is simply a single instance of something »given« – the Latin participle perfect of »dare« is *datum*.

A »date« or »datum«, therefore, is an entity that appears distinct from other entities and yet forms part of an integrated, perceivable realm of specifically structured experience – the precariously »duplicated« world Armin Nassehi writes about. In this respect, »the given« can be characterized by the romantic concept of the *unity of unity and difference*. Data is comprised of a vast multitude of singular instances of given information yet forms an integrated unit – the sum is more than its parts. This is illustrated by the fact that a given »date« occurs almost exclusively in the plural »data« in colloquial discussion – we say »a piece of data,« rather than »one date,« even though it would be the usual way of applying singular and plural in the English language. Like the quantum in physics, the »date« never seems to exist in isolation, but always as part of a larger horizon of meaning. In this respect, the use of the plural is not only grammatically correct, but epistemologically meaningful.

Hence, we can recognize data as a collective multitude of communicative micro-packets that can, at least potentially, be mapped in *quantitative* structures. Taking this preliminary reflection into account, a refined version of our

6 This definition is, in fact, already more nuanced than the definition employed in most content on the subject in the popular media and everyday language, because it is based on Joseph Weizenbaum's critical theory of information. Cf. Weizenbaum 2001, especially chapter 1 on information and meaning.

7 On the »conditions of existence« for specific media forms, see Parikka 2012: 6.

colloquial terminology emerges. For our purposes here, we shall understand data as *digital and analog communicative micro-packets that can, in principle, be stored electronically due to their quantifiable structure and (when interpreted with adequate methods under the conditions of shared grammar and semantics) can become meaningful information, which can guide action and thus can impact the material or embodied practices of communicative agents*.

B Typology of Data Processing

For our purposes, we do not reduce the concept of data to digitally generated, electronically stored information, but rather define data broadly enough to take a *multimedia* perspective, including analog forms. For practical purposes, we cannot place the general concepts of rationality or communication at our theoretical center, because both would require a solid cognitive-theoretical foundation, which is impossible to adequately deliver here. We will thus limit ourselves to data processing in *media*, which appears in the daily as both task and tool. A variety of different typologies have been employed to draw out different functional dimensions of data processing. One might, for instance, consider the distinction of data storage in 1D or 2D arrays. Or one might point to the distinction of data transmission into serial and parallel transmission. Because, however, our epistemic interest is to draw out the difference in practical impact on media transformation for different types of data processing, we require a different set of types, because neither the form of storage, nor the form of transmission can suffiently expose the impact and uses that different types of data processing might have for its practical application in media. For our purposes, it will suffice to distinguish four ideal-type forms: linear, variable, spatial and explorative data processing.

I Linear

A large part of human media history is shaped by a type of data processing that we can characterize as *linear*. This linear type of data processing manifests itself in all forms of end-to-end communication between individuals or between the individual and mass media broadcasting (especially radio and television). Such end-to-end communication is more or less successful when based on shared code. This includes the often-subconscious socio-cultural coding of semantic contents in their transportation through media, but also the conscious technological encryption with cryptographic intention.

Due to the end-to-end structure of linear data processing, medial actions

of this type are difficult to scale. Linear data processing is nevertheless ubiquitous in all human social practice. Every simple form of transported messages between individuals displays the characteristic of linearity – be it the one-to-many mode of a broadcast model or the one-to-one mode of a simple communication model. A holiday postcard, for example, might be classified this way. The linear characteristic might also be used to reconstruct the more complex media form of a newspaper, which is created in the editorial room of a publishing company, produced at the print shop and then directly sent to households.

Even the basic functions of the internet are based on this type of linear data processing. The TCP/IP suite, the collection of foundational internet protocols, is based on this principle. On the internet, individual hosts send small data packets to an address via a digital network, much like the holiday postcard. Using TSL/SSL encryption, the information on these digital postcards can be encrypted, even on an open network like the internet, thus enabling the relatively secure transmission of confidential communication in linear form. Even after the invention of more complex data processing technologies, linear data processing remains the foundation and majority of digital communication. With Nassehi, we can note that precisely its *simplicity* is what explains digital technology's *ubiquity*.[8]

What used to be stored on paper in address books is now often stored electronically, but the basic structure of data processing remains the same.[9] This can be illustrated by the example of the electronic mailing list: An initial communication agent sends the general message to a more or less specified audience that *serial* linear communication is possible in the form of a newsletter, for instance by advertising it on a website. The recipient then transfers a contact address to the address book of the initial agent, for instance by typing it into a contact form on a website. In this way, the initial agent collects a multitude of linear contact addresses and begins a serial *broadcast* of linear messages.

From a data protection perspective, the decentralized nature of the data is particularly noteworthy here. In the form of a silo, the address books of the various communication agents are stored separately. The strategic use of this

8 He hones in on the quantitatively infinite possibility of recombination with the simple binary code that makes up all digital technologies at their core. Cf. Nassehi 2019.

9 Again, Nassehi picks up this thought and develops it further than we can here. His theoretical approach is to ask what societal conditions needed to be in place for digital technology to be adopted at such a swift pace and high rate. This leads him to conclude that, in fact, the very foundation of modern society is digital in structure, which in turn explains why digital technology could function as an effective problem solver in this society. Cf. Nassehi 2019. See chapters 1 and 2 especially.

data for pattern recognition therefore remains limited.[10] And even if one of the silos is attacked, the other silos remain secure. Linear data processing, therefore, implies a kind of safeguarding clause: One compromised silo does not automatically compromise other silos. In this respect, the linear use of data can be classified as fairly secure. However, the linear use of data is also impractical for many use cases, because the findings from such data remains limited.

The only strategic analysis possible is *category formation* and *individual analysis*. The example of itemized telecommunications bills can explain this. The data generated from itemized connections is only informative if the identity of both the contacting and contacted agents is known and available for investigation with a concrete epistemic interest (meaning: you know what you are looking for). If, for example, police want to check an alibi after some type of criminal offence, the registration of the linked mobile device in a mobile cell tower far away from the crime scene for the purpose of a telephone call with an unconnected third person offers strong indication that the alibi is valid. For such simple analysis with a pre-existing epistemic interest, the linear data processing of individual end-to-end data flows is sufficient. However, inquiries beyond individual analysis and category formation require more sophisticated data processing, especially when one does not know what exactly one is looking for.

II Variable

Data processing becomes more complex when a vast amount of data is available from linear data processing and a system for high level pattern analysis is developed strategically. Such methods were invented long before computer-based technologies. The indexing of analog crime scene photographs and the strategic comparison of murder weapons, murder methods and special features of a crime, for instance, can lead to more complex data processing based on the linear type.

Applied to our telecommunications example: Through *data aggregation* and stray search for extraordinary prevalence of certain types of actions, the system can *identify patterns*, thereby allowing investigators to deduce potential habits and strategies. In this case, the individual pieces of data become *metadata* in aggregated form and thus can serve as an ideal basis for more complex forms of variable data processing. In such processing, one or more vari-

10 This explains both why Facebook intends to combine user data from all its services that hitherto had remained separate – the value of the data increases manifold once it is combined into a shared silo. This also explains why it is so controversial.

ables are defined in a fixed formula, which ensures that a change in *signals* actually shows up as a change in the *result* in *real-time*. As formula-based data processing, we may consider simple algorithms that include a variety of forms and quantities in the variables and thus renders differences in incoming signals visible in the final result.

Practical examples of variable data processing are simple personalized purchase algorithms such as the book suggestion function on Amazon's online retail platform. When the potential buyer places a book in the shopping cart or even just clicks on it to read through the description, data is generated through these click signals which indicates interest in this specific book's general category. If a large amount of such individual pieces of data is available from other users, Amazon can determine which other books have also been viewed or added to the shopping list in similar purchase processes. Based on the known formula of these purchasing patterns, Amazon can develop a personalization algorithm suggesting interesting books to new customers: »If user A clicks on book X and user B has bought book X and Y at the same time, then suggest book Y to user A as well.« Although the principle is simple, it is based on the sophisticated variable inter-relation of linear end-to-end types of data processing.

III Spatial

While both linear and variable data processing is largely based on simple algorithms,[11] many industrial algorithms show the characteristics of *spatial* data processing. The application of such algorithms does not result in a 2D visualization as in Facebook's News Feed, but in a multidimensional rendering. A practical example for this form of data processing are 3D printers, which are able to bring linearly stored data structures into spatial application by means of multidimensional blueprints in a computer program. Just as in linear and variable data processing, the data is broken down into small and simple communicative micro packets. And yet the applied algorithms are able to draw a coherent spatial picture from this complex multitude of information packets in order to reconstruct them materially.

Many different branches of industry use this type of data processing on a daily basis. Stress tests of manufacturing materials and prototypes in aero-

11 Even if scaled and connected into more complex algorithmic systems, the basic operations are in the form of simple formula-based algorithms. The only deviation from this rule is the application of machine learning on top of the formula-based processing. Certain personalization algorithms (e.g. Facebook's News Feed) are increasingly applying machine learning and technologies from spatial and explorative data processing.

space engineering, for example, cannot be performed physically with sufficient replication – either due to prohibitively high cost or simply a lack of time. Therefore, complex mathematical models are used to simulate the physical effects of wear and tear in order to calculate where reinforcements have to be made and where weight can be saved to increase efficiency in fuel usage and material cost.

Another instance of spatial data processing are the calculation and visualization programs for architects, product designers, vehicle engineers, structural engineers and meteorologists. In these fields of application, operational safety is of particular importance from an ethical perspective. Since public infrastructure, product application, vehicle operation, building usage and weather calculations often impact the chances of survival in emergency situations, the inviolability of the person is of utmost importance in this form of data processing, given the foundational consensus of the modern concept of personal dignity. In applications of medical and scientific research, the ethical questions of the good life and holistically life-enhancing strategies for spatial data processing are even more evident.

IV Explorative

The category of *explorative* data processing shall summarize the technologies known as »artificial intelligence« in popular discourse. Usually, the term artificial intelligence means a more or less intelligent algorithmic system that, in most cases, is trained by humans with annotated training data and recognizes patterns in these data sets with none or little structure. The recognized patterns are then applied to new incoming data and if the domain area of this incoming data matches the domain of the training data, these patterns can produce meaningful insights for the human employing such systems. Such machine learning techniques are commonly called »artificial intelligence« because a human being could never have manually defined the patterns recognized by the system and thus required skill augmentation by human-made technological tools, in this case: artificial in the sense of »made,« »created« or »built« intelligence. When the searchable data set is so vast that human beings cannot go through it themselves, explorative data processing with machine learning is exponentially more powerful than any category formation in linear data processing could ever be.

The »artificial intelligence« system, in such cases, is nothing like the mystical all-powerful god-like general intelligence portrayed in popular culture and requires a very narrowly specified domain to function appropriately. A facial recognition system will likely produce gibberish if applied to music and

a music recognition system will likely produce gibberish if applied to images. But the technologies employed, even if limited to a narrow domain, have powerful properties that impress humans enough to inspire the title *artificial intelligence*. In our example, the machine learning system assumes the role of *detective*, processing a super-human amount of information simultaneously, as well as the role of *analyst*, interpreting the recognized patterns through *quantitative* means which can then be augmented by human *qualitative* analysis to produce an end product deemed in many cases superior to human analysis without computer assistance.

The relevant methods for explorative data processing are mainly data mining techniques employing machine learning methods, machine learning methods. In more complex tasks, these might take the form of deep learning, which attempts to mathematically map and functionally imitate the neuronal structures of the human brain.[12] If paired with high-speed computing power, deep learning can vastly outperform machine learning (for instance in machine translation of natural language). But for many simpler applications, machine learning comes close and is the more resource-efficient option. For many personalization algorithms in media platforms, for instance, machine learning methods will suffice.

The neural networks utilized in deep learning methods are designed for evolution and learn a certain intelligent behavior for a specific area of application, in some cases with the help of human trainers and always with large amounts of data. That is why the algorithms generated through these methods are categorized as *self-learning* algorithms. In contrast to the formula-based algorithms of variable data processing, artificial intelligence procedures are less of a strategic approach and more of an investigative, discovering, unstructured trial-and-error approach. This trial-and-error philosophy imitates the human learning curve marked by empirical experiences of pain and happiness.

C Contentions of Critical Data Theory

Much of the discourse on artificial intelligence has been (inadvertently or not) shaped by the product marketing initiatives of tech companies and euphoric researchers in search of funding on one side, and the fundamental critics of technology and big business on the other side, while politicians try to safely

12 Though brain scientists reject that metaphor, because computational systems require a stability that human brains never reach. To them, the attempt to imitate a dynamic, self-stabilizing system with a static, stable system (be it ever so fast in computation) is a dead end. See for instance Singer 2003. We should, therefore, consider the analogy more poetic device than scientific characterization.

tread on middle ground, reminding us of both the »risks and opportunities« in mantra-like fashion. Because the discourse is not yet broad enough to realistically mirror societal complexity, the discourse remains caught in extreme perspectives from either side of the polarized spectrum mixed with fearful repetition of vague set phrases that more often than not demonstrate a lack of technical knowledge (which further piles onto the reasons for politicians to be afraid of clear statements and initiatives for fear of ridicule). What would a critical theory of artificial intelligence look like? Could it be truly critical, in the sense of both critiquing practices and dialectically critiquing insufficient critiques of such practices? What topics would such a theory have to confront? Among them, certainly are these four: (1) autonomy, (2) transparency, (3) mythology, and (4) contextuality.

I Autonomy

No existing artificial intelligence system can rightfully be called *autonomous* if we follow the literal sense of *self-legislation*, derived from the Greek *autos* = self and *nomos* = law). Machine learning, at least, can still not do without human input, even if the human input is less than in linear or variable data processing. The utilized algorithm is not based on simple formulas with variables and signal prioritization explicated by humans. However, a framework and training data set must still be given to the system by humans in most cases. In short: AI does not just fall from the sky. To create powerful AI systems, immense amounts of human work, model training and optimization are necessary, thus begging the question: Is it really cheaper to invest in expensive AI systems for a given problem? Or is it more expensive to hire AI experts for a job that can be done by manual labor as efficiently?

An example for this is machine learning in community management on social media for publishers. In order to find patterns in comments, humans must define for the algorithm what the relevant data point is, such as the most common word in the trove of comments that is not a filler or sentence-connecting word, such as »like« or »and«. Alternatively, humans could optimize the algorithm to discover the most swift and steep increase in usage of a word, which can power trend analysis, because it could identify which increase occurred after a defined period of stagnation or regression in use. One could also search for signals occurring in pairs (to establish correlations), for parallel appearing changes (to establish interdependencies) or other forms of patterns and connections.

All such analyses, which ultimately might produce a meaningful result, are more directly related to human analytical competence than the popular dis-

course on artificial intelligence makes us believe. The result of such analyses becomes truly exciting for the analytical teams comprised of both humans and algorithms when a variety of data sets are superimposed on each other and the identified patterns can be compared with other data sets. In this way, the human-machine teams can identify correlations to specific events, weather developments, demographic changes, time of day and much more which by no means could be discovered by human beings alone. Hence, we can understand machine learning as experimental and explorative, but not truly »autonomous« from human influence and decisions.

This even goes for the algorithms applied in so-called »autonomous driving,« since the algorithmic decision-making is strictly determined by the data collected through the sensors of the self-driving car. If you change the sensors, you change the decision. If you change the training to a more aggressive driving style, you will get a more aggressive self-driving car. The car has no conscious reflection and decision-making about what to optimize its driving towards: Speed at all cost? Avoidance of injury? All those guidelines are human guidelines, external to the algorithm and trained or programmed into it. The »autonomous driving« algorithms hence are more dependent on external guidance than their names imply. If a self-driving vehicle identifies a human being in front of them, the algorithm has been trained to hit the brakes. Truly autonomous decision-making, as ascribed to humans, would imply that running the human over is a possible option in this case. The data sets that have trained the algorithms and the humans training the systems, however, never allowed for that possibility. Therefore, the algorithm is not *autonomous* in the meaningful sense of *self-legislation*.

II Transparency

The term artificial intelligence is usually used in public discourse as a collective term for all those procedures that result in a computer system performing tasks considered to be intelligent in humans. It is imprecise, but expresses a new quality of complexity in computer processes. Simple »if X, then Y« formulas develop into more complex instructions for machine learning: »If X results in result A, then assume A for the next experiment Y. But if Y results in result B, then correct A into B for case Y. And replicate this procedure n-fold to calculate probabilities for each further result by aggregating individual results and discovering patterns through similarity analysis.« Some claim that due to its so-called »autonomous« and evolutionary nature, such a computational feat should not be called an algorithm anymore, because it is unlike the formula-based algorithm of linear data-processing. But since it is still a quan-

titative process based on calculation with human involvement in the data set, computational framework and learning instructions, it is more similar to traditional algorithms than the evangelists of AI mythology would have it.

The word algorithm has its roots in the Latin word *algorismus* and used to mean the Indian art of calculation in reference to the Greek word *arithmós*, meaning »number.« The word was created from the name of the Persian-Arabic 9th century mathematician Al-Hwarizmī and is defined by the standard German dictionary as a »procedure for step-by-step transformation of number sequences« and »a process of calculation according to a certain [repetitive] scheme.«[13] Similarly, the Oxford dictionary defines an »algorithm« as »a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.«[14] Since, in the case of machine learning and even deep learning we are talking about a process of calculating probabilities (even in the case of deep learning the computational imitation of neurons in the brain is far less complex than its organic original and has well-established, stable math at its procedural core), we are talking about algorithms, which come into the world only because of human creativity. But their evolutionary nature allows these algorithms their own learning biographies as they were known only from humans and other animals.

It is difficult for the most complex of these experimental algorithms to give a meaningful account of their decision criteria, especially in the hidden layers in deep learning's neural networks. Not unlike cases involving human action, complex investigations into the decision criteria are necessary when something goes wrong, and only the most specialized machine learning experts can estimate where the root problems is. Especially when the root cause is in biased data or mistaken annotations of the training data, it takes time, focus and effort to find the source of bugs. Users of an AI system in a consumer product usually cannot identify any such bias or mistakes in the system in their own user interface. Similar to the pre-Reformation priesthood of the Church, machine learning experts are granted far-reaching competency to decide what counts as responsible development. But if it is true that such technologies will permeate every aspect of our lives in the not-too distant future, such authority must meet the highest of standards of transparency and accountability.

One of the key problems in terms of transparency and accountability in AI algorithms, has been the *black box* problem. Arthur Clarke has famously offered a poignant rule of thumb, commonly known as Clarke's third law: Any

13 Duden 2021. Author's translation. The definition in its original German wording: »Verfahren zur schrittweisen Umformung von Zeichenreihen; Rechenvorgang nach einem bestimmten [sich wiederholenden] Schema.«

14 Oxford University Press 2021.

sufficiently advanced technology is indistinguishable from magic.[15] The black box problem hits deep learning algorithms based on neural networks the hardest, because the neural network's hidden layers are precisely that: hidden. Analyst Alok Aggarwal notes that »even researchers are currently unable to develop a theoretical framework for understanding how or why they give the answers they do.«[16]

As an example, Aggarwal offers the Deep Patient experiment run by Joel Dudley and several colleagues. Deep Patient's objective was to use deep learning technologies »to predict health status, as well as to help prevent disease or disability« and »provide a machine learning framework for augmenting clinical decision systems.«[17] The project was successful and achieved improved predictions »for severe diabetes, schizophrenia, and various cancers« by using aggregated electronic health records of around 700,000 patients from their hospital's data warehouse.[18] Will Knight has reported that the Deep Patient algorithm anticipates »psychiatric disorders like schizophrenia surprisingly well.«[19] But given how difficult the prediction of schizophrenia is, the algorithm's co-inventor Joel Dudley »wondered how this was possible.«[20] But even Dudley himself has no way to find out because the algorithm »offers no clue as to how it does this.« He acknowledges that his team »can build these models … but we don't know how they work.«[21] Will Knight suggests that in order for such an algorithm to reliably help doctors, it will have to »give them the rationale for its prediction, to reassure them that it is accurate and to justify, say, a change in the drugs someone is being prescribed.«[22]

Among the open questions for algorithmic accountability studies is how to reconcile the public value of *transparency* with the public interest in *privacy*. What kind of transparency is possible if personal data must stay protected and secured from the very public eye that attempts to deliver transparency? In sensitive areas like medical application, who receives explanations from the »Explainable AI« is key. Under the traditional data privacy framework, only the patient and their medical team should have access to the data employed in the computational process. This has been eroded by the complexity required to process and store the vast troves of electronic data for medical purposes, which

15 For Clarke's third law's context, please see Clarke 1973.

16 Aggarwal 2018.

17 Miotto et al. 2016: 1.

18 Miotto et al. 2016: 1.

19 Knight 2017.

20 Knight 2017.

21 Knight 2017.

22 Knight 2017.

has led most doctors to outsource their data processing. This involves a third party in the process, adding complications to the task of transparent attribution of influences and factors in the process.[23] The complexity of AI systems further adds another layer of complexity in this attribution.

III Mythology

Developing AI systems has been a key goal in computer science and statistics for decades, not least due to the lucrative applications in medicine, biotechnology, industrial design, logistics management, quality control and many other commercial fields. Due to cost-effective availability of huge quantities of computing power, as well as the growing availability of training data (though this is still a massive hurdle for many AI projects), the goal is slowly becoming more and more realistic. Nevertheless, AI development remains difficult and error-prone even in the most successful systems and requires rare and costly talent that only the most attractive employers have available.

This is just one of the reasons why it remains doubtful whether a general AI can ever reach the much-discussed stage of *singularity*. The claim that a reliably flawless metasystem (termed general AI or strong AI) can result from the sophisticated interconnection of domain-specific subsystems (called narrow AI or weak AI) created by error-prone humans with imperfect data is logically impossible without some type of mythical leap.[24] For leapfrog innovation towards singularity to happen, some other foundationally new approach to error elimination must be found. The neuroscientist Wolf Singer has shown the flaws in the claim that AI systems can actually replicate the human brain's neural networks. Singularity theorist Kurzweil, Singer charges, »has fallen prey to a huge misunderstanding if he believes that an increase in computing speed alone will lead to a qualitative leap. The analogy of computer and brain is superficial at best. While both systems can execute logical computations, the systems architectures are radically divergent.«[25] While the human brain is both complex/non-linear and stable in its neural processing, all computing systems

23 The complexity in this process lead to blind spots in the processing chain, often leaving sensitive medical data exposed. Cf. Dangelmayer et al. 2019 and Gillum et al. 2019.

24 In Schmidhuber's contributions the leap takes story form and is presented as a logical progression: »As I grew up I kept asking myself, ›What's the maximum impact I could have?‹ … And it became clear to me that it's to build something smarter than myself, which will build something even smarter, et cetera, et cetera, and eventually colonize and transform the universe, and make it intelligent.« Cf. Markoff 2016.

25 Singer 2003: 33. The German original: »Ich denke, dass Kurzweil einem riesigen Missverständnis aufsitzt, wenn er glaubt, dass Vermehrung der Rechengeschwindigkeit allein zu einem qualitativen Umschlag führen würde. Die Analogie zwischen Computer und Gehirn ist

are either complex/non-linear and unstable or simple and stable.[26] The great riddle, Singer concludes, is how the non-linear complex processing of the brain »retains its stability and … integrates the various partial functions.«[27]

Nevertheless, some (self-proclaimed) pioneers of AI technology such as Schmidhuber[28] remain vehemently self-confident advocates of strong AI, which he propagates with mythological language as a godlike hyperintelligence and expects to emerge in the medium-term through continuous technological advancement. Schmidhuber's storytelling exploits the widespread lack of technical understanding and has significant power to frame how AI technologies are viewed. But since the average person has no way to verify or falsify grand claims, the discourse on the future of artificial intelligence has become a question of trust.

Because »everything we know about the world in which we live, we know through the media«[29], this question of trust is a question of *media* trust: Are journalists independently and critically verifying the grand claims of computer scientists and marketing directors? Are they even technologically capable of critical judgment on such specialized issues? In all critical probing of this kind, we must take note of what Luhmann adds after his famous dictum about the mediated nature of social reality: »we also know enough about the mass media to not be able to trust these sources. We resist it suspecting manipulation, but without consequence, since the knowledge we derive from the mass media, as if by itself, completes itself into a self-reinforcing framework.«[30]

---

bestenfalls eine oberflächliche. Beide Systeme können zwar logische Operationen ausführen, aber die Systemarchitekturen sind radikal verschieden.«

26 Singer's definition of »simple« includes machine learning.

27 Singer 2003: 37. The German: »Das große Rätsel ist, was die Großhirnrinde im Einzelnen macht, wie sie es macht, wie sie sich stabil hält und wie die vielen Teilfunktionen, die in ihren verschiedenen Arealen erbracht werden letztlich gebunden werden.«

28 Schmidhuber's student Hochreiter developed the Long Short Term Memory method that is used today in billions of smartphones for speech recognition, handwriting recognition, image analysis and other applications. Schmidhuber is cited as an author on the paper. Other AI researchers have questioned his credit. LeCun for instance, is not impressed: »Jürgen is manically obsessed with recognition and keeps claiming credit he doesn't deserve for many, many things … It causes him to systematically stand up at the end of every talk and claim credit for what was just presented, generally not in a justified manner.« Cf. Markoff 2016. Schmidhuber's research partners defend his claims for credit.

29 Luhmann 1996: 9. Given in the author's translation. The German original reads: »Was wir über unsere Gesellschaft, ja über die Welt, in der wir leben, wissen, wissen wir durch die Medien.«

30 Luhmann 2009: 9.

IV Contextuality

Depending on the culture in which a critical discussion of technical processes takes place, the popular assumptions about this process transported in media and everyday practice vary:

In the German-speaking world, AI is often portrayed as an *enemy* that destroys safe and secure working conditions and symbolizes impersonal coldness.[31] In the Anglo-Saxon world, AI is staged more as a *servant* or even a slave, which is also reflected in the usability dogma in the marketing of products developed by U.S. tech companies. In Chinese culture, AI is more often seen as a *partner* and colleague, which is reflected, for example, in the initiative of the Chinese news agency Xinhua to »hire« an AI-based avatar as news anchor.[32] Japanese reports repeatedly show that AI is viewed more as a *friend* there, as can be seen, for example, from the use of robots in assisted living for seniors.[33]

Even if we cannot provide methodologically rigorous evidence for these cultural differences and their socio-technical consequences here, such heuristic indications put the topic on the radar and stimulate interdisciplinary research. Reliable comparative ethnographic studies would turn this discourse into a highly productive interdisciplinary field of learning for the ethical evaluation and socio-psychological analysis of the hopes, fears and uses of AI applications.

D Ethics of Complex Data Processing

Once we clarified our definition of data and reconstructed four prevalent types of data processing in our typology, we were prepared to explore four areas of contention for critical data theory. Now we can venture into the ethics of complex data processing by exploring the conditions for the possibility of its practice. We can identify six hallmarks of ethical evaluation: I) technological *capability* of the responsible processor, II) general *availability* of data, III) *equitability* of the training data, IV) *computability* of the intended function, V) applied *methodology* and its corresponding distortions, and VI) *directionality* for the use and optimization of algorithms.

31 E.g. the dramatic headline »Die Jobfresser kommen.« Cf. Schultz 2016.

32 Cf. Kuo 2018.

33 Cf. 3sat 2018.

I Capability: Who can develop it?

As digital divide research has shown: access to digital opportunities is unequally distributed. This applies in amplified ways to artificial intelligence opportunities. Whoever can access and use large data sets, is in a good position to train machine learning algorithms, while those with little data are in a weak to impossible situation. Public institutions with strict data privacy regulation, for instance, are forced into competitive disadvantage to more liberally regulated private actors that can collect large quantities of data as long as the user has given consent.[34] Hence, an ethical evaluation of AI must include power analysis: Who is in a position to develop artificial intelligence systems in the first place?

There is, of course, an indirect limit to this type of power, because even those who can train AI systems in one domain will not necessarily be able to train systems in other domains. Even massive corporate conglomerates like Facebook with vast amounts of user data in many domains struggle to develop AI systems that can effectively take down live-streamed shooting videos from their platform before they are distributed to millions of users in real-time.[35] Improving such preventative AI systems requires domain-specific data of what such first-person shooting videos look like, which few companies have available in sufficient quantities to train a machine learning model. Facebook, for instance, has resorted to »working with American and British law enforcement authorities to obtain camera footage from their firearms training programs to help its A.I. learn what real, first-person violent events look like.«[36] Ethically, the question arises what kind of publicly funded data should be provided to privately held digital platforms for such preventative law enforcement purposes.

II Availability: What data is used?

After the stage of *power* analysis might come a stage of *data* analysis, because the type, source and structure of training data matters greatly for the ethical evaluation of a given machine learning solution. What data is available for

34 This consent remains precarious if most users unconsciously tick a box without reading privacy terms.

35 In March 2019, for instance, Facebook was used by the mass shooter in Christchurch, New Zealand to spread live video of 51 killings. And in August 2019, Facebook's platform was used in El Paso, USA to distribute the shooting plans posted on 8chan through Facebook and other social media sites.

36 Alba et al. 2019.

training could, for instance in medical research, inadvertently decide about who gets to live and who has to die. Taking genomics as a concrete example: If only European genomes are sequenced because of resources available there and few African genomes get sequenced, and groundbreaking research is thus based on European genomics, the developed treatment strategies might not work when applied in African contexts. The type, source and structure of the data, therefore, might serve to perpetuate existing power and illegitimate privilege.

III Equitability: Is there structural bias?

If an algorithm is trained with a set of data, this data will impact the results of this algorithm the application it powers. This has led to instances where racial prejudice or other forms of discriminatory patters in the training data have caused the algorithm to reproduce such prejudice in its results. A famous example is Microsoft's conversational bot Tay trained on tweets which turned it into a racist in less than a day.[37] Another example is an HR tool developed by Amazon that was intended »to review job applicants' resumes with the aim of mechanizing the search for top talent.«[38]

The problem was that »Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period« which meant that most resumes were from male candidates because of the massive gender gap in the tech industry. The system had »taught itself that male candidates were preferable«[39] because of the data underlying it. The system »penalized resumes that included the word ›women's,‹ as in ›women's chess club captain.‹ And it downgraded graduates of two all-women's colleges.«[40] At first, Amazon attempted to make the system more neutral to these specific words, but since the data set was biased, the algorithm was necessarily biased and even if singular instances could be corrected, the overall patterns could not. Amazon had no choice but to pull the plug on the project.

There is an increasingly established discourse on algorithmic bias and there are numerous attempts to develop best practices against such bias.[41] And while algorithmic bias is not easy to solve, it is still one of the easiest AI ethical problems to solve, because it is (a) evident in most cases, (b) quantifiable in

37 Vincent 2016.

38 Dastin 2018.

39 Dastin 2018.

40 Dastin 2018.

41 Cf. Lee et al. 2019.

many cases, and therefore (c) addressable through technological refinement. While it remains a challenge to test and evaluate training data ethically, the more fundamental ethical question is the limit of the quantitative paradigm.

IV Computability: What limits are inherent?

The basic fact about any computational technology is that it is just that: computational, and therefore quantitative. This is the foundationally inherent limit in any artificial intelligence system. Unless AI research comes up with a fundamentally different approach to intelligence – one that is not solely mathematical – there will be severe limits to the types of cognitive tasks artificial intelligence systems can take on. And there will even be mathematical limits, as Gödel has proven with his theorem of incompleteness,[42] which further casts doubt on euphoric anticipation of general AI. If the quantitative mathematical paradigm remains the only relevant paradigm in AI development, essential dimensions of human experience will never be captured as part of artificial intelligence systems due to the methodology's inherent limits.

Love, for instance, is one of the key human experiences and experienced as an integrated emotion with strong cognitive elements. Yet, it is impossible to quantify, which is why quantifiable proxies need to be found to even begin to approach the phenomenon. We might visualize which parts of the brain are active when we experience an intense moment of love. But love in a deeper, more philosophical sense manifests itself in so many ways and nuances that it becomes too complex to reduce to mathematical, quantified calculations. And even if we identified the most important part of the brain for the act and experience of loving, we would still not have proven anything. For like force in physics, love is impossible to prove. We can only deduce its existence from the effects we can observe. For a low standard of proof, this might suffice. But any sophisticated concept of love will include non-quantifiable dimensions and limit the functionality of AI applications in this area.

The limitations of the current methodologies do not just have philosophical implications for those who worry about the limits of a quantitative paradigm. They have implications for those who invested their capital in the commercialization of artificial intelligence systems. Analyst Alok Aggarwal notes that »several of the obstacles that led to the demise of the first AI boom phase over forty years ago remain unresolved today, and it seems that serious theoretical advances will be required to overcome them.«[43] Aggarwal concludes that »the

42 Cf. Rautenberg 2008.

43 Aggarwal 2018.

predictions … are unlikely to be met in the next fifteen years, and financiers may not receive an expected return on their recent investments in AI.«[44]

V Methodology: What might skew results?

Beyond the inadvertent impacts of the quantitative paradigm and the theoretical limitations of the current concepts in artificial intelligence research, there are other ways the process might skew the results. An example for this is what has at times been called *coding populism*. The concept means that applying machine learning to content distribution (as Facebook has increasingly done with its News Feed) will necessarily advantage populist contributions on the platform. If the distribution system is based on billions of trial and error experiments (for instance, with contextual bandit testing) designed to find the posts that trigger the highest engagement and thereby increase the time spent on the platform which translates directly into higher advertisement revenue, then nuanced contributions will be drowned out on the platform and attention-grabbing, incendiary, controversial, aggravating content will necessarily be the most-distributed content on the platform.

Applying machine learning to content distribution instead of operating on editorial philosophy and principles of human curation, might actually end up feeding our subconscious *worst* angels instead of Lincoln's proverbial »better angels of our nature.« Such machine learning applications exploit subconscious behavioral patterns, including the bias and prejudice that we try to fight *consciously* but still often act out *subconsciously*. This does not mean we endorse our subconscious hopes, fears and prejudice consciously. It means that we are imperfect human beings who might not want to engage in discriminatory behavior, but unknowingly contribute to racist, sexist, or violent structures built into an engagement-only based algorithm. Because machine learning is based on *preformed* and not *intended* behavior (as an editorial policy would expound), it does not ask users to support audacious ideals, but rather feeds into their subconscious failings. It leaves users feeling manipulated, as is observable in the low trust in social media platforms. One example: Only 14 percent of the German population trust social media networks generally, with Twitter specifically at only 10 percent.[45]

44 Aggarwal 2018.

45 Institut für Demoskopie Allensbach 2016.

VI Directionality: What is the purpose?

Despite its outspoken commitment to the high-flying ideal of »making the world more open and connected,«[46] Facebook's ultimate purpose of applying machine learning to the content distribution has been to increase advertising revenue.[47] And it has been successful with it – the revenue can now sustain operations at scale and leaves significant capital for innovation projects, acquisitions and other investments in Facebook's overall technological capabilities. While any business analyst would hail this move as responsible business practice with an impressive record of success, those who do not earn from this success as shareholders and worry about the societal impacts will question the integrity of machine learning application for this purpose. And even those who have earned a fortune from Facebook's ad technology have started to question the impact of the News Feed algorithm. Several former employees have expressed concern about »the unintended consequences of a network when it grows to a billion or 2 billion people« and »exploiting a vulnerability in human psychology« through »a social-validation feedback loop.«[48] Others have aired »tremendous guilt« for helping to create »tools that are ripping apart the social fabric«[49] – for instance Facebook's micro-targeting technology[50] and the like button.[51] Mark Zuckerberg's co-founder has even called for a breakup of the company.[52]

This vigorous debate, however, has not just been stimulated by former Facebook employees, but a great number of other individuals and organizations. One of those individuals is Tristan Harris, a former Google engineer, who has critiqued Silicon Valley's dopamine-driven product strategies as the »attention economy« and a »race to the bottom of the brain stem.«[53] Harris started the Time Well Spent movement[54] and went on to start the Center for Humane Technology. Through the work of this center, Harris wants to fight

46 Hoffmann et al. 2016: 1.

47 Given its massive user growth, venture capital alone was not enough. Facebook needed a solid revenue stream and perfected its role as ad-broker and micro-targeting to reach highly focused user groups.

48 Allen 2017.

49 Vincent 2017.

50 Cf. Garcia-Martinez 2017.

51 Cf. Morgans 2017.

52 Hughes 2019.

53 Thompson 2019.

54 The *Time Well Spent* movement has been taken up by Apple, Instagram, Facebook and Harris's former employer Google through the implementation of time-monitoring apps. Some have seen this as a move to coopt the movement, others as a legitimate response to it. Cf. Stolzoff 2018.

the »downgrading« of humanity trough technology. Tech companies have furthered what he sees as a race to the bottom »by promoting shortened attention spans, outrage-fueled dialogue, smartphone addiction, vanity, and a polarized electorate. Harris called for tech companies to enable a new ›race to the top,‹ centered on building tools to help people focus, find common ground, promote healthy childhoods, and bolster our democracy.«[55]

The intense debate on the purpose and impact of tech giants like Facebook and Google has strengthened the wider ecosystem of debate around the purpose and impact of technology more generally. In Germany, for instance, the Conscious Coders student group works towards »beneficial and sustainable use of digital technologies for the society and the environment« as well as »a profound understanding of emerging technologies throughout the whole society« and calls for »critical developers who review their work against ethical questions and are aware of their responsibility.«[56]

This mission-driven approach as garnered momentum in the field of artificial intelligence as well. The AI for Good Foundation, for instance, wants to build »lasting communities that bring the best technologies to bear on the world's most important challenges … by coordinating the AI research community, technologists, data, and infrastructure with the stakeholders on the ground, policy makers, and the broader public« in support of the United Nation's Sustainable Development Goals.[57] Another organization from within the U.S. tech community working »to ensure that artificial general intelligence … benefits all of humanity« is OpenAI.[58]

VII Summary: Ethical Use of Complex Data Processing

By asking ourselves the right questions at the right time in the process, ethical review can become an integral part to technology operations. Our probing of the conditions for the possibility of complex data processing has yielded a non-exhaustive list that can power such an operationalization of ethical review. Underlying this exploration is the assumption that all technology is a cultural product and requires a whole range of different constructively linked factors for its successful implementation. Before the popular questions about singularity become meaningful, a whole host of things can go wrong in day-today AI systems that are already in widespread use.

---

55 Newton 2019.

56 Conscious Coders 2019.

57 AI for Good Foundation 2019.

58 OpenAI 2018.

Sometimes, it seems, the storytellers of AI mythology deploy the smokescreen of singularity, to hide more day-to-day AI applications in plain sight. Instead of narrowing the ethical scope to the grand questions of imagined futures, researchers should expand the ethical analysis of existing technologies in complex data processing. Our non-exhaustive list of review questions might serve as a first step in that endeavor: Who can deploy complex data processing in the first place? What (kind of) data is available for training of intelligent algorithms? Does the training data show signs of prejudice or structural bias? What are the inherent limitations of data processing? How might the specific process skew the results? For what goal is a given technology deployed and optimized?

References

Aggarwal, Alok 2018: The Current Hype Cycle in Artificial Intelligence. https://scryanalytics.ai/the-current-hype-cycle-in-artificial-intelligence/ (accessed 9 October 2019).

AI for Good Foundation 2019: About Us. https://ai4good.org/about/ (accessed 11 October 2019).

Alba, Davey/Edmondson, Catie/Isaac, Mike 2019: Facebook Expands Definition of Terrorist Organizations to Limit Extremism. In: The New York Times, 17 September 2019. https://www.nytimes.com/2019/09/17/technology/facebook-hate-speech-extremism.html (accessed 17 October 2019).

Algorithmus, der. In: Duden 2021. https://www.duden.de/rechtschreibung/Algorithmus (accessed on 30 November 2021).

Algorithm. In: Oxford University Press 2021. https://www.lexico.com/definition/algorithm (accessed on 30 November 2021).

Allen, Mike 2017: Sean Parker unloads on Facebook: »God only knows what it's doing to our children's brains«. In: Axios, 9 November 2017. https://www.axios.com/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792-f855e7b4-4e99-4d60-8d51-2775559c2671.html (accessed 11 October 2019).

Bedford-Strohm, Jonas 2019: Mythologie, Typologie, Pathologie: Bausteine einer kritischen Theorie der Datenverarbeitung in den Verfahren der künstlichen Intelligenz. In: Görder, Björn/Zeyher-Quattlender, Julian (Hgg.): Daten als Rohstoff: Die Nutzung von Daten in Wirtschaft, Diakonie und Kirche aus ethischer Perspektive, Münster, LIT.

Clarke, Arthur C. 1973: Profiles of the Future: An inquiry into the limits of the possible. London, Macmillan.

Conscious Coders 2019: Vision. https://www.consciouscoders.io/ (accessed 11 October 2019).

Dangelmayer, Pia/Meyer-Fünffinger, Arne/Hagmann, Ulrich/Köppen, Uli/Kühne, Steffen/Nierle, Verena/Schnuck, Oliver/Streule, Josef/Tanriverdi, Hakan/Thamerus, Tatjana/Zierer, Maximilian: Millionenfach Patientendaten ungeschützt im Netz. In: BR24, 17 September 2019. https://www.br.de/nachrichten/deutschland-welt/millionenfach-patientendaten-ungeschuetzt-im-netz,RcF09BW (accessed on 20 July 2021).

Dastin, Jeffrey 2018: Amazon scraps secret AI recruiting tool that showed bias against women. In: Reuters, 10 October 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (accessed 10 October 2019).

Garcia-Martinez, Antonio 2017: I'm an ex-Facebook exec: don't believe what they tell you about ads. In: The Guardian, 2 May 2017. https://www.theguardian.com/technology/2017/may/02/facebook-executive-advertising-data-comment (accessed 11 October 2019).

Gillum, Jack/Kao, Jeff/Larson, Jeff 2019: Millions of Americans' Medical Images and Data Are Available on the Internet. Anyone Can Take a Peak. In: ProPublica, 17 September 2019. https://www.propublica.org/article/millions-of-americans-medical-images-and-data-are-available-on-the-internet (accessed on 20 July 2021).

Hoffmann, Anna Lauren/Proferes, Nicholas/Zimmer, Michael 2018: »Making the world more open and connected«: Mark Zuckerberg and the discursive construction of Facebook and its users. In: New Media & Society, 20 (1), 199–218.

Hughes, Chris 2019: It's Time to Break Up Facebook. In: The New York Times, 9 May 2019. https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html (accessed 11 October 2019).

Institut für Demoskopie Allensbach 2016: Welche dieser Informationsquellen halten Sie für vertrauenswürdig, wo kann man besonders zuverlässige Informationen über Politik und politische Ereignisse erwarten? In: Allensbacher Archiv, IfD-Umfrage 11062.

Knight, Will 2017: The Dark Secret at the Heart of AI. In: MIT Technology Review, 11 April 2017. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (accessed 14 October 2019).

Kuo, Lily 2018: World's first AI news anchor unveiled in China. In: The Guardian, 9 November 2018. https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china (accessed 13 November 2018).

Lee, Nicol Turner/Resnick, Paul/Barton, Genie 2019: Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. In: Brookings, 22 May 2019. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ (accessed 10 October 2019).

Luhmann, Niklas 1996: Die Realität der Massenmedien. Opladen, Westdeutscher Verlag.

Heidelberger Institut für theoretische Studien 2017: The Dark Side of Natural Language Processing. https://www.h-its.org/scientific-news/ethics-nlp/ (accessed on 15 October 2018).

Markoff, John 2016: When A.I. Matures, It May Call Jürgen Schmidhuber ›Dad‹. In: The New York Times, 27 November 2016. https://www.nytimes.com/2016/11/27/technology/artificial-intelligence-pioneer-jurgen-schmidhuber-overlooked.html (accessed 21 July 2021).

Miotto, Riccardo/Li, Li/Kidd, Brian A./Dudley, Joel T. 2016: Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. In: Scientific Reports, 6 (26094). https://www.nature.com/articles/srep26094 (accessed 14 October 2019).

Morgans, Julian 2017: The Inventor of the ›Like‹ Button Wants You to Stop Worrying About Likes. In: Vice, 6 July 2017. https://www.vice.com/en_uk/article/mbag3a/the-inventor-of-the-like-button-wants-you-to-stop-worrying-about-likes (accessed 11 October 2019).

Nassehi, Armin 2019: Muster: Theorie der digitalen Gesellschaft. München, Beck.

Newton, Casey 2019: The leader of the Time Well Spent movement has a new crusade. In: The Verge, 24 April 2019. https://www.theverge.com/interface/2019/4/24/18513450/tristan-harris-downgrading-center-humane-tech (accessed 11 October 2019).

OpenAI 2018: OpenAI Charter. https://openai.com/charter/ (accessed 11 October 2019).

Parikka, Jussi 2012: What is media archaeology? Cambridge, Polity.

Rautenberg, Wolfgang 2008: Unvollständigkeit und Unentscheidbarkeit. In: Einführung in die Mathematische Logik. Wiesbaden, Vieweg+Teubner, 167–208.

Schultz, Stefan 2016: Arbeitsmarkt der Zukunft. Die Jobfresser kommen. In: DER SPIEGEL, 2 August 2016. http://www.spiegel.de/wirtschaft/soziales/arbeitsmarkt-der-zukunft-die-jobfresser-kommen-a-1105032.html (accessed 14 November 2018).

Singer, Wolf 2003: Ein neues Menschenbild? Gespräche über Hirnforschung. Frankfurt/M., Suhrkamp.

Stolzoff, Simone 2018: Technology's »Time Well Spent« movement has lost its meaning. In: Quartz, 4 August 2018. https://qz.com/1347231/technologys-time-well-spent-movement-has-lost-its-meaning/ (accessed 11 October 2019).

Thompson, Nicholas 2019: Tristan Harris: Tech Is ›Downgrading Humans.‹ It's Time to Fight Back. In: WIRED, 23 March 2019. https://www.wired.com/story/tristan-harris-tech-is-downgrading-humans-time-to-fight-back/ (accessed 11 October 2019).

Vincent, James 2017: Former Facebook exec says social media is ripping apart society. In: The Verge, 11 December 2017. https://www.theverge.com/2017/12/11/16761016/former-facebook-exec-ripping-apart-society (accessed 11 October 2019).

Vincent, James 2016: Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. In: The Verge, 24 March 2016. https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist (accessed 10 October 2019).

Weizenbaum, Joseph 2001: Computermacht und Gesellschaft. Frankfurt/M., Suhrkamp.

3sat 2018: Mein elektrischer Freund. Für Japaner haben auch Roboter eine »Seele«. https://www.3sat.de/page/?source=/nano/technik/184917/index.html (accessed 13 November 2018).

ORCID

Jonas Bedford-Strohm https://orcid.org/0000-0003-4165-1881

Medien zwischen Angstmachern und Hoffnungsstiftern

Zur emotionalen Wirkung der medialen Berichterstattung über künstliche Intelligenz

Selina Fucker

A Einleitung

Zahlreiche Nachrichtenbeiträge stellen künstliche Intelligenz (KI) als eine Chance oder als ein Risiko dar. »KI als Wachstumsturbo«, »Mit Machine Learning Millionen sparen«, »Bedrohung wie in ›Terminator‹ – superintelligente KI für Menschen unbeherrschbar« oder »Künstliche Intelligenz – Die Alternative ist: Irgendwann ist dein Arbeitsplatz fort« sind typische Headlines, die in diesem Zusammenhang veröffentlicht wurden.[1]

Künstliche Intelligenz ist ein Thema, das viele Menschen bisher kaum bewusst in ihrem Alltag betrifft. Zwar wird schwache KI zum Beispiel schon in Sprachassistenten und bei Bild- und Gesichtserkennung eingesetzt, allerdings wird dieser Einsatz von KI nur wenig thematisiert.[2] Aufgrund der geringen Thematisierung von KI als neue Technologie ist davon auszugehen, dass die Medienberichterstattung eine wichtige Rolle in der Meinungsbildung in Hinsicht auf KI spielt.[3] Dies lässt sich dadurch begründen, dass die Medienberichterstattung in Massenmedien wie Fernsehen, Zeitungen und Online-Medien neben der Schule oft die einzige Informationsquelle für Informatio-

1 dpa 2020: o. S., Jatscha 2021: o. S., Kramper 2021: o. S. und Schirmer 2019: o. S.

2 Schreiner 2021: o. S.; Reiche 2021: o. S.

3 Ho et al. 2011: 609.

Selina Fucker 2022: Medien zwischen Angstmachern und Hoffnungsstiftern. Zur emotionalen Wirkung der medialen Berichterstattung über künstliche Intelligenz. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 81–108.
https://doi.org/10.11588/heibooks.1106.c15194

nen über Wissenschaft und Technologie darstellt.[4] Erste Studienergebnisse bestätigen auch, dass Nachrichtenmedien wie Zeitungen und Fernsehen bei Studierenden die häufigsten Informationsquellen über KI sind.[5]

Darüber hinaus haben Studien gezeigt, dass vor allem die Chancen und die Risiken von KI in der Medienberichterstattung betont werden.[6] Die Berichterstattung über KI erfolgt laut den Studien dennoch überwiegend neutral oder positiv.[7]

Es ist davon auszugehen, dass Fortschritte in KI-Forschung und -Entwicklung weiterer gesellschaftlicher Entscheidungen zum Umgang mit künstlicher Intelligenz bedürfen.[8] In diesem Zusammenhang sind vor allem politische, moralische und auch wirtschaftliche Entscheidungen gemeint. Solche Entscheidungen können durch die Einstellungen der Bevölkerung gegenüber der Technologie beeinflusst werden. Sollte die Medienberichterstattung über künstliche Intelligenz die Einstellungen gegenüber dieser verändern, so kommt der Medienberichterstattung bei der Meinungsbildung eine besondere Bedeutung zu. Dennoch ist die Wirkung der Medienberichterstattung über künstliche Intelligenz bisher nicht untersucht worden. Auf diese Forschungslücke hat auch schon Vergeer hingewiesen.[9] Ziel dieses Beitrages ist es diese Forschungslücke ein Stück weit zu schließen.

Die Berichterstattung über KI geschieht daher vielfach mittels Chancen- und Risiken-Frames. Die Bedeutung solcher Frames soll im Folgenden dargestellt und erläutert werden. Durch die Verwendung solcher Frames ist es möglich, dass die Berichterstattung polarisierend wirkt, vor allem wenn entweder die Chancen- oder die Risiken-Frames in einem Nachrichtenbeitrag überwiegen.[10] Daraus ergeben sich für die folgenden Überlegungen folgende Leitfragen:

F1: Wie beeinflussen Chancen- und Risiko-Frames in der medialen Berichterstattung über künstliche Intelligenz die Einstellungen gegenüber künstlicher Intelligenz?

4 Nisbet et al. 2002: 592.

5 Vgl. Du-Harpur et al. 2020: o. S., Chuan et al. 2019: 339.

6 Vgl. Brennen et al. 2018: 4, Chuan et al. 2019: 342 f., Ouchchy et al. 2020: 931 f. und Sun et al. 2020: 8.

7 Vgl. Chuan et al. 2019: 342, Ouchchy et al. 2020: 930, Vergeer 2020: 382.

8 Reiche 2021: o. S.

9 Vgl. Vergeer 2020: 388.

10 Brennen et al. 2018: 9.

F2: Welche emotionalen Effekte haben Chancen- und Risiko-Frames in der medialen Berichterstattung über künstliche Intelligenz?

Um diese Fragestellungen zu untersuchen, wurde ein Online-Experiment mit vier verschiedenen, randomisierten Experimentalgruppen durchgeführt.[11] Es wurde sowohl die Wirkung von Chancen- und Risiko-Frames von KI als auch die Wirkung von beidseitigem Framing künstlicher Intelligenz untersucht. Im Folgenden wird zunächst der Forschungsstand und die Methode ausführlicher dargelegt, bevor die Ergebnisse des Experimentes vorgestellt werden.

B Forschungsstand

I Framing: Definition und Eingrenzung

Mit Hilfe der Framing-Forschung können Darstellungsformen und ihre Wirkung in der Medienberichterstattung erklärt werden. Der Begriff »Framing« bezeichnet die Einrahmung von Informationen in der medialen Berichterstattung.[12]

Da der Fokus dieser Arbeit auf der Wirkung der Medienberichterstattung über KI liegt und über KI häufig sowohl positive als auch negative Aspekte in einem Medienbericht erwähnt werden, werden in dieser Arbeit in Anlehnung an Entmann und Druckmann Frames als eine Auswahl an Aspekten festgelegt.[13] Diese beschreiben ein bestimmtes Thema beziehungsweise ein Problem und können so dessen Beurteilung beeinflussen.

II Mediale Darstellung von künstlicher Intelligenz

Zur medialen Darstellung von künstlicher Intelligenz sind bislang nur wenige Studien erschienen. Brennen et al. haben 2018 die mediale Darstellung künstlicher Intelligenz in Großbritannien mit einer nicht näher dargelegten Analysemethode untersucht. Dabei haben sie drei Frames identifiziert, die besonders präsent in der Nachrichtenberichterstattung über KI sind. Am häufigsten

11 Die Studie war Teil einer Masterarbeit, die im Jahr 2021 am Institut für Kommunikationspsychologie und Medienpädagogik der Universität Koblenz-Landau unter der Betreuung von Christian von Sikorski und Stephan Winter verfasst wurde.

12 Nelson et al. 1997: 222–224.

13 Entmann 1993; Druckmann 2001.

wird KI als Chance für die Wirtschaft dargestellt.[14] Außerdem wird KI als große Veränderung beziehungsweise Transformation beschrieben. Dabei wird diese Art der Transformation durch die KI mit der Industrialisierung gleichgesetzt.[15] Der dritte Frame, den Brennen et al. identifizieren, ist die Darstellung von Produkten mit künstlicher Intelligenz als »creepy«, also gruselig oder unheimlich.[16] Auch wenn die Autoren der Studie nicht näher auf die Bedeutung des »creepy« Frames eingehen, ist eine emotionale Wirkung dieses Frames naheliegend. Außerdem haben sie festgestellt, dass die politische Ausrichtung der Nachrichtenmedien die Berichterstattung beeinflusst. Eher linksgerichtete Medien haben mehr über den Arbeitsplatzverlust im Zusammenhang mit künstlicher Intelligenz und ethische Bedenken zum Beispiel im Zusammenhang mit autonomen Waffen berichtet. Konservativ ausgerichtete Medien hingegen berichteten verstärkt über die wirtschaftlichen und geopolitischen Chancen durch KI für Großbritannien.[17]

Dies deutet daraufhin, dass die Berichterstattung über Chancen und Risiken der KI in der medialen Berichterstattung im Mittelpunkt steht, zumindest im Hinblick auf den Aspekt der Chancen.

Dies bestätigt auch die Studie von Vergeer aus dem Jahr 2020: Vergeer analysiert ein Korpus aus niederländischen Zeitungsartikeln, in denen entweder die Worte artifical intelligence oder die Abkürzung AI vorkommen, mit Hilfe von topic modelling und Sentimentanalyse. Sie zeigt, dass sich positive und negative Sentimente, also Stimmungen in der Berichterstattung über KI in den Niederlanden finden.[18] Die positiven Sentimente überwiegen aber über den gesamten analysierten Zeitraum von 18 Jahren.[19] Interessant ist hierbei, dass Zeitungen mit einer religiösen oder wirtschaftlichen Ausrichtung positiver über KI berichten als andere Zeitungen.[20]

Die Studie von Chuan et al. zur Berichterstattung über KI, die 2019 die Medienberichterstattung über KI in fünf amerikanischen Zeitschriften mit Hilfe einer qualitativen Inhaltsanalyse untersucht, kommt zu dem Ergebnis, dass in 52.9 % der analysierten amerikanischen Zeitungsbeiträge über KI mindestens ein Nutzen von KI erwähnt wird und in 47.6 % der Artikel mindestens ein Risiko.[21] Die häufigsten Themen bei positiven Frames waren wirtschaftliche Vor-

14 Brennen et al. 2018: 4.
15 Ebd.
16 Brennen et al. 2018: 5.
17 Ebd.
18 Vgl. Vergeer 2020: 382.
19 Ebd.
20 Vergeer 2020: 382.
21 Chuan et al. 2019: 342.

teile und Verbesserungen des menschlichen Lebens. Bei den negativen Frames wurden die Themen Technikversagen, Arbeitsplatzverlust und Datenschutzbedenken genannt.[22] Sie haben auch festgestellt, dass die Medienberichterstattung über KI von 2009 bis 2018 deutlich zugenommen hat.[23] Die Zunahme der Medienberichterstattung veränderte auch die Valenz: Bis zum Jahr 2013 haben die untersuchten Zeitungen überwiegend positiv über KI berichtet. Seitdem hat der Anteil an Artikeln mit negativer und gemischter Valenz zugenommen. Auch ihre Analyse kommt zu dem Schluss, dass über KI hauptsächlich im Zusammenhang mit wirtschaftlichen Themen (35.1 %) und mit wissenschaftlichen und technologischen Themen (23.6 %) berichtet wird.[24] Es wurde auch ein Zusammenhang zwischen den Themen und dem Nutzen-Framing beziehungsweise Risiko-Framing festgestellt. So wird das Thema Wirtschaft mehr in Artikeln behandelt, die ausschließlich Nutzenframes enthalten, während die Themen Science-Fiction und Ethik in Artikeln behandelt werden, die nur Risiken thematisieren.[25]

Eine qualitative Inhaltsanalyse von englischsprachigen Artikeln aus Zeitungen, Magazinen und Webblogs von Ouchchy et al. hat 2020 gezeigt, dass in einem überwiegenden Teil der Artikel ausbalanciert beziehungsweise neutral über KI berichtet wird. Die Anzahl der Artikel, in denen in einem negativen Ton berichtet wird, hat seit 2015 stetig zugenommen, während der Anteil mit einem enthusiastischen Ton abgenommen hat.[26] Zahlreiche der in den Artikeln behandelten Themen, wie zum Beispiel Datenschutz oder Vorurteile, lassen sich unter der Kategorie unerwünschte Folgen von künstlicher Intelligenz zusammenfassen.[27] Interessant hierbei ist, dass ein Großteil dieser Artikel trotz der beinhalteten kritischen Themen in einem ausbalancierten, beziehungsweise neutralen Ton verfasst wurden.[28]

Sun et al. haben 2020 eine automatische Inhaltsanalyse von 1776 englischsprachigen Zeitungsartikeln, die das Thema KI behandeln, durchgeführt. Diese Analyse wurde durch eine Netzwerkanalyse ergänzt. Dabei haben sie die Themen Roboter, Hirnforschung und Regulierung künstlicher Intelligenz als häufige Themen in englischsprachigen Zeitungsberichten identifiziert.[29] Ihre Analyse der argumentativen Aussagen über KI in den Berichten hat ergeben,

22 Chuan et al. 2019: 342.
23 Chuan et al. 2019: 340.
24 Chuan et al. 2019: 341.
25 Chuan et al. 2019: 343.
26 Ouchchy et al. 2020: 930.
27 Ouchchy et al. 2020: 931 f.
28 Ouchchy et al. 2020: 932.
29 Sun et al. 2020: 8.

dass 38.6 % dieser argumentativen Aussagen pragmatisch sind. Sie befassen sich mit dem praktischen Einsatz von künstlicher Intelligenz wie zum Beispiel in der Industrie.[30] Eine weitere häufige argumentative Aussage ist Relativierung, hierbei werden die Grenzen der Chancen von künstlicher Intelligenz betont.[31]

Vergleicht man die eben vorgestellten Studienergebnisse, so begründen sie die oben eingeführte These, dass Chancen, Nutzen und Risiken ein zentrales Thema in der Berichterstattung über KI sind.[32] Die Berichterstattung über KI erfolgt überwiegend positiv beziehungsweise neutral.[33] Auffällig ist hierbei, dass bei diesem Aspekt keine nennenswerten Unterschiede hinsichtlich des Tons zwischen der Berichterstattung in den Niederlanden, in Großbritannien und der Berichterstattung im gesamten englischsprachigen Raum festzustellen sind. Allerdings ist die Vergleichbarkeit der Ergebnisse der Studien bei diesem Aspekt nur bedingt gegeben, da sie auf unterschiedlichen Analysemethoden beruhen. Diese unterschieden sich vor allem durch die Tiefe der Analyse.

Die Themen Chancen für die Wirtschaft, Arbeitsplatzverlust, Datenschutz werden in englischsprachigen Medienberichten besonders häufig beschrieben.[34]

III Wirkprozesse bei der Framing Rezeption

Da Frames eine Problemdefinition vorschlagen, Ursachen identifizieren, Lösungen aufzeigen beziehungsweise eine bestimmte Perspektive einnehmen, können sie eine Wirkung beim Rezipienten erzielen.[35] Diese potenzielle Wirkung ist charakteristisch für einen Frame.

Ein Framing-Effekt liegt dann vor, wenn der Frame bei dem Rezipienten das Verständnis des Themas beziehungsweise Problems beeinflusst.[36] Um eine Wirkung zu erzielen, müssen Frames dabei nicht zwingend eine für die Rezi-

30 Sun et al. 2020: 8.

31 Ebd.

32 Vgl. Brennen et al. 2018: 4, Chuan et al. 2019: 342 f., Ouchchy et al. 2020: 931 f., Sun et al. 2020: 8.

33 Vgl. Chuan et al. 2019: 342, Ouchchy et al. 2020: 930, Vergeer 2020: 382.

34 Vgl. Brennen et al. 2018: 5, Chuan et al. 2019: 342, Ouchchy et al. 2020: 931 f.

35 Entmann 1993: 52, Tversky/Kahnemann 1981: 453.

36 Price et al. 1997: 482.

pienten neue Information enthalten.[37] Des Weiteren sind auch emotionale Framing-Effekte möglich.[38]

Die Grundlage für einen Framing-Effekt ist zunächst die Rezeption des Frames.[39] Dabei wird der Frame selektiv verarbeitet und ihm wird eine subjektabhängige, durch den Inhalt beziehungsweise die Darstellung beeinflusste Bedeutung zugewiesen. Die Zuweisung der subjektabhängigen Bedeutung erfolgt zusammen mit den Kognitionen, Emotionen und Bewertungen des Rezipienten.[40] Weil die Bedeutungszuweisung subjektabhängig ist, kann ein und derselbe Frame bei unterschiedlichen Rezipienten zu unterschiedlichen Bedeutungszuweisungen führen.[41]

Nelson et al. erklären Framing-Effekte hauptsächlich durch die Wirkung von Frames auf bestehende Vorstellungen und Kognitionen: »Frames apear to activate existing beliefs and cognitions, rather than adding something new to the individual's beliefs about the issue«.[42] Durch die Aktivierung bisheriger Kognitionen wird eine Nachricht, beziehungsweise ein Artikel beurteilt.[43] Die Frames beeinflussen dabei vor allem die Gewichtung von Informationen.[44] Allerdings sind auch Wissenseffekte durch Frames möglich. Diese treten auf, wenn die Rezipienten den Inhalt des Frames noch nicht kennen.[45] Solche Effekte sind zum Beispiel denkbar, wenn bei den Rezipienten noch kein oder nur kaum Vorwissen über KI besteht und sie durch die Frames mit neuen Informationen konfrontiert werden.

Framing-Effekte sind somit auch abhängig von den Rezipienten, ihrem Vorwissen und Voreinstellungen. Daher lassen sich neben dem Wissenseffekt weitere Arten von Framing-Effekten unterscheiden.[46] Es sind Reaktanzeffekte möglich, wenn sich die Person durch den Frame eingeschränkt fühlt und dann genau die gegenteilige Position einnimmt.[47] Accessibility-Effekte entstehen, wenn eine dem Rezipienten schon bekannte Information wieder neu in das Gedächtnis gerufen wird und dann in der Bewertung verwendet wird.[48] Dies

37 Nelson et al. 1997: 225.
38 Vgl. Gross/Ambrosio 2004, Gross/Brewer 2007, Gross 2008.
39 Potthoff 2012: 221.
40 Potthoff 2012: 221.
41 Kühne 2013: 11.
42 Nelson et al. 1997: 235 f.
43 Caccicatore et al. 2016: 16.
44 Nelson et al. 1997: 235, Kühne 2013: 7.
45 Potthoff 2012: 224, de Vreese et al. 2011: 182 f.
46 Potthoff 2012: 224.
47 Ebd.
48 Potthoff 2012: 224.

ist zum Beispiel denkbar, wenn in einem Frame erwähnt wird, dass künstliche Intelligenzen häufig viele Daten erfassen und der Rezipient schon einmal von der Datenschutzproblematik im Zusammenhang mit KI gehört hat und so daran erinnert wird.

Gross hat gezeigt, dass unterschiedliche Arten von Frames verschiedene emotionale Reaktionen auslösen können. Dabei hat sie die Wirkung von episodischen Frames und thematischen Frames verglichen.[49] Episodische und thematische Frames unterscheiden sich dadurch, dass episodische Frames mit einem spezifischen Beispiel verbunden sind, während thematische Frames das Thema in einem größeren Kontext verorten.[50] Die Ergebnisse zeigen, dass die Rezeption der Frames zu spezifischen emotionalen Wirkungen führen kann und die emotionale Wirkung bei den episodischen Frames stärker ist.[51] Der thematische Frame hingegen führte zu stärken Einstellungsänderungen.[52] Es kann somit zwischen emotionalen und kognitiven Framing-Effekten unterschieden werden.[53]

Emotionale Framing-Effekte werden als Ergebnis von kognitiven Evaluationsprozessen verstanden, die das subjektive Erleben des emotionalen Zustandes hervorrufen.[54] Emotionen sind immer mit einem der jeweiligen Emotion entsprechenden Handlungsziel verbunden. Frames können Emotionen auslösen, die zu emotionskongruenten Einstellungen führen und so die Verhaltensintention verändern.[55]

Thematische Voreinstellungen beeinflussen ebenfalls die Emotionsauslösung.[56]

IV Wirkung von Frames in Medienberichten über neue Technologien

Da es bisher keine Studien zu Framingeffekten bei künstlicher Intelligenz gibt, werden zur Begründung meines Vorgehens im Folgenden Studien zu Framingeffekten bei anderen komplexen Technologien, wie zum Beispiel Nanotechnologie und synthetische Biologie, vorgestellt.

---

49 Gross 2008: 171.

50 Ebd.

51 Gross 2008: 180.

52 Gross 2008: 181.

53 Ebd.

54 Kühne 2013: 16.

55 Kühne 2013: 15.

56 Vgl. Kühne 2013, Gross/Ambrosio 2004 und Gross/Brewer 2007.

Cobb hat die Wirkung von verschiedenen Frames im Zusammenhang mit Nanotechnologie auf Emotionen und die Einstellungen zu Nanotechnologie untersucht. Dabei hat Cobb die Wirkung von sechs einseitigen Frames (Frames pro oder contra Nanotechnologie) und drei zweiseitigen Frames, die sowohl Argumente für als auch gegen Nanotechnologie beinhalten erforscht.[57] Dabei hat er sowohl die Wirkung von Value-Frames als auch von Nutzen- und Risiken-Frames untersucht.[58] Die Ergebnisse zeigen, dass die einseitigen Frames, die spezifische Risiken und Chancen enthalten, einen Effekt haben, während die Value-Frames keinen Effekt ausgelöst haben.[59] Die Risiko-Frames hatten nicht den Effekt, dass die Risiken größer als die Nutzen beschrieben wurden, sondern sie führten dazu, dass die Teilnehmenden skeptischer gegenüber den möglichen Nutzen von Nanotechnologie wurden.[60] Die zweiseitigen Frames führten zu weniger Meinungsänderungen als die einseitigen Frames.[61] Die Ergebnisse zeigen auch emotionale Effekte, so führten Risiko-Frames zu weniger Hoffnung und mehr Sorge in Bezug auf Nanotechnologie.[62] Die Nutzen-Frames führten zu einer leichten Reduktion des Ärgers über Nanotechnologie im Vergleich mit der Kontrollgruppe.[63]

Binder et al. haben untersucht, wie Darstellung von Unsicherheit in der Berichterstattung über Nanotechnologie das Gefühl der Unsicherheit und die Risikowahrnehmung verändert. Dafür haben sie drei Experimentalgruppen mit jeweils einem konstruierten Nachrichtenartikel als Stimulus und eine Kontrollgruppe zu ihren Einstellungen gegenüber Wissenschaftler:innen, gegenüber wissenschaftlichen Erkenntnissen und den Bewertungen des Nachrichtenbeitrags befragt.[64] Ein Stimulus enthielt ausschließlich Informationen, ohne Quellenangabe, ein weiterer Stimulus enthielt Äußerungen über Risiken von Wissenschaftlern und ein dritter Stimulus enthielt sowohl Äußerungen über Unsicherheit als über Sicherheit von Wissenschaftlern.[65] Die Ergebnisse zeigen, dass es keinen direkten Effekt der Frames auf das Gefühl der Unsicherheit und die Risikowahrnehmung gibt.[66] Aber es wurde eine Moderation von der Einstellung gegenüber wissenschaftlicher Autorität auf die Be-

57 Cobb 2005: 227.
58 Cobb 2005: 227 f.
59 Cobb 2005: 229.
60 Cobb 2005: 230.
61 Ebd.
62 Vgl. Cobb 2005: 232.
63 Ebd.
64 Binder et al. 2016: 837.
65 Ebd.
66 Binder et al. 2016: 841.

ziehung von Frame und das Gefühl der Unsicherheit und der Risikowahrnehmung gefunden. Bei Teilnehmenden mit einer geringen Achtung gegenüber wissenschaftlicher Autorität erhöhen die Risiko-Frames die Risikowahrnehmung von Nanotechnologie.[67]

Die dargestellten Studien zeigen, dass Framing in Medienberichten die Einstellungen gegenüber neuen Technologien nach dem Lesen dieser Medienberichte beeinflussen kann. Sie zeigen auch, dass die Framing-Effekte von der Art und dem Inhalt des Frames und von weiteren Faktoren, wie zum Beispiel der Einstellung gegenüber wissenschaftlicher Autorität, abhängig sein können.

Im Folgenden soll nun geprüft werden, ob sich die Erkenntnisse zur Wirkung von Frames in der Berichterstattung über Nano- und Biotechnologie auch auf KI übertragen lassen.

Wie unter IV. dargelegt, zeigt Cobb, dass Chancen-Frames die Einstellungen gegenüber neuen Technologien positiv beeinflussen. Daraus lässt sich folgende Hypothese über die Wirkung von Chancen-Frames in der Berichterstattung über KI ableiten:

Hypothese 1 (H1): *Chancen-Frames erhöhen die positiven Einstellungen gegenüber künstlicher Intelligenz.*

Auch bei Risiko-Frames wurde eine Wirkung auf die Einstellungen gegenüber der jeweiligen neuen Technologie gefunden. Sie können die Einstellungen gegenüber der jeweiligen Technologie negativ beeinflussen.[68] Dies führt zu folgender Hypothese:

Hypothese 2 (H2): *Risiko-Frames erhöhen die negativen Einstellungen gegenüber künstlicher Intelligenz.*

Wenn Chancen und Risiken in einem Frame kombiniert werden, der Frame somit zweiseitig ist, wurden kaum direkte Veränderungen der Einstellung gegenüber der jeweiligen neuen Technologie gefunden.[69] Dies lässt sich dadurch erklären, dass sich die Wirkungen der Chancen- und Risikodarstellungen im Frame ausgleichen. Daraus lässt sich folgende Hypothese ableiten:

Hypothese 3 (H3): *Beidseitige Frames, die sowohl Chancen als auch Risiken enthalten, haben keinen Effekt auf die Einstellungen gegenüber künstlicher Intelligenz.*

67 Binder et al. 2016: 841.

68 Cobb 2005: 230.

69 Ebd.

Die Studie von Cobb (2005) zeigt, dass Frames in der Berichterstattung über neue Technologien auch eine emotionale Wirkung haben können. Risiko-Frames können die Sorge der Befragten steigern und die Hoffnung negativ beeinflussen.[70] Ausgehend davon sind bei Chancen-Frames umgekehrte Effekte vorstellbar. Dies führt zu folgenden Hypothesen:

Hypothese 4 (H4): *Risiko-Frames erhöhen die Sorge.*

Hypothese 5 (H5): *Chancen-Frames erhöhen die Hoffnung.*

V Einstellungen gegenüber neuen Technologien

1 Einflussfaktoren auf Einstellungen gegenüber neuen Technologien

Bisherige Studien über neue technologische Entwicklungen haben gezeigt, dass das Vorwissen über die neue Technologie unter anderem den Framing-Effekt beeinflussen kann. Bei höherem Vorwissen ist die Einstellungsänderung geringer.[71] Hierbei kann zwischen der Selbstauskunft, ob Wissen über die Technologie vorhanden ist, und dem tatsächlichen Wissen über die Technologie, unterschieden werden. Scheufele und Lewenstein haben gezeigt, dass bei der Selbstauskunft über das Wissen über Nanotechnologie weniger Wissen angegeben wird, als vorhanden ist, und dennoch hat die Selbstauskunft über das Vorwissen über Nanotechnologie einen Einfluss auf die Einstellungen gegenüber Nanotechnologie.[72] Deshalb wird das Vorwissen der Befragten in dieser Arbeit mitberücksichtigt.

Auch soziodemografische Faktoren wie Alter, Bildungsstand, Einkommen und Geschlecht der Befragten beeinflussen die Einstellungen gegenüber neuen Technologien und wurden deswegen miterhoben.[73]

Das Vertrauen in Führungskräfte aus der Wirtschaft hat einen positiven Effekt auf die Wahrnehmung der Nutzen von Nanotechnologie und einen negativen Effekt auf die Wahrnehmung von Risiken.[74]

Emotionen wie Hoffnung und Angst beeinflussen ebenfalls die Einstellungen gegenüber Nanotechnologie. Hoffnung hat einen positiven Effekt auf die

70 Cobb 2005: 232.

71 Brossard et al. 2009: 555, Scheufele/Lewenstein 2005: 663.

72 Cobb/Macoubrie 2004: 403, Scheufele/Lewenstein 2005: 663.

73 Vgl. Lee/Scheufele 2006: 826, Brossard et al. 2009: 552.

74 Cobb/Macoubrie 2004: 402.

Wahrnehmung der Nutzen von Nanotechnologie.[75] Angst erhöht die Wahrnehmung der Risiken von Nanotechnologie.[76] Daher lässt sich annehmen, dass Risiko-Frames die Sorge erhöhen, wie schon in Hypothese 4 vermutet. Weitergehend lässt sich aber auch annehmen, dass diese gesteigerte Sorge ähnlich wie Angst die Risiko-Wahrnehmung erhöht und so zu einer negativeren Einstellung gegenüber künstlicher Intelligenz führt. Daher lässt sich folgende Hypothese ableiten:

Hypothese 6 (H6): *Risiko-Frames erhöhen die Sorge, was zu einer negativen Einstellung gegenüber künstlicher Intelligenz führt.*

In diesem Zuge lässt sich auch annehmen, dass eine gesteigerte Hoffnung die Chancen-Wahrnehmung erhöht und so zu einer positiveren Einstellung gegenüber künstlicher Intelligenz führt. Deshalb lässt sich folgende Hypothese ableiten:

Hypothese 7 (H7): *Chancen-Frames erhöhen die Hoffnungen, was zu einer positiven Einstellung gegenüber künstlicher Intelligenz führt.*

2 Einstellungen gegenüber künstlicher Intelligenz

Es liegen bisher nur wenige Studien zu Einstellungen gegenüber künstlicher Intelligenz vor. Lobera et al. haben 2020 in einer spanischen Studie erhoben, dass 33.3 % der Bevölkerung angeben, dass sie denken, dass KI mehr Risiken als Nutzen hat. 38.4 % glauben, dass sie mehr Nutzen als Risiken hat und 28.3 % denken, dass sich die Risiken und Nutzen ausgleichen.[77]

Eine Befragung von Medizinstudierenden in Deutschland hat ergeben, dass Männer KI in der Radiologie eher als Chance einschätzen als Frauen.[78] Auch bei der Beurteilung von automatisierten Entscheidungen durch KI gibt es Geschlechterunterschiede, diese nehmen Frauen als weniger nützlich wahr als Männer.[79] Ebenso ist der Widerstand gegenüber KI bei Frauen höher.[80]

Das Vorwissen und der Bildungsabschluss beeinflussen die Einstellungen gegenüber KI. So haben Araujo et al. 2020 gezeigt, dass der Bildungsabschluss

75 Cobb/Macoubrie 2004: 402.
76 Ebd.
77 Lobera et al. 2020: 11.
78 Dos Santos et al. 2019: 1642.
79 Araujo et al. 2020.
80 Lobera et al. 2020: 14.

einen positiven Effekt auf die Nutzenwahrnehmung von automatisierten Entscheidungen durch künstliche Intelligenzen hat.[81] Es besteht eine Korrelation zwischen einem niedrigen Bildungsabschluss und einem etwas höheren Widerstand gegenüber KI.[82]

Es besteht ein Zusammenhang zwischen dem Vertrauen in Wissenschaft und Technologie und der positiven Wahrnehmung von KI.[83] Ein weiterer Faktor, der die Einstellungen gegenüber KI beeinflusst, ist die Haltung gegenüber Innovation. Hier zeigt sich eine Korrelation zwischen Innovationsresistenz und einer negativen Wahrnehmung von KI.[84]

Aufgrund der hier dargelegten Erkenntnisse über die Einflüsse auf die Einstellungen gegenüber KI wird vermutet, dass Alter, Geschlecht, Bildungsabschluss und Vorwissen über KI die Wirkung der Chancen- und Risiken-Frames beeinflussen können. Zudem werden Einflüsse von Berufstätigkeit, Vertrauen in Wissenschaft und Innovationsresistenz auf die Wirkung der Chancen- und Risiken-Frames vermutet.

Alter, Geschlecht, Vertrauen in die Wissenschaft sowie Innovationsresistenz werden aufgrund ihres Einflusses auf die Einstellungen gegenüber KI als mögliche Drittvariablen mit erhoben.

3 Modell

Aus dem dargelegten Forschungsstand und den aufgestellten Hypothesen wird folgendes Forschungsmodell abgeleitet:

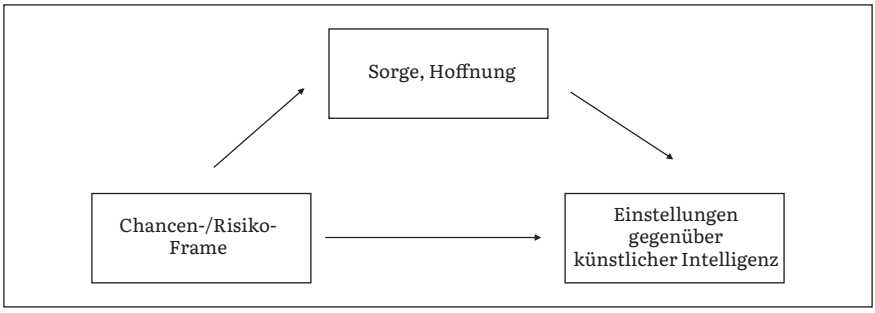

Abbildung 1 Forschungsmodell

81 Araujo et al. 2020.

82 Lobera et al. 2020: 11.

83 Vgl. Lobera et al. 2020: 15.

84 Ebd.

Der Frame (in den Ausprägungen Grundbeitrag, einseitiger Chancen-Frame, einseitiger Risiken-Frame und beidseitiger Frame) stellt die unabhängige Variable (UV) dar. Aus den Hypothesen 1 und 2 werden die Einstellungen gegenüber künstlicher Intelligenz als abhängige Variable (AV) abgeleitet. Die Variable Sorge stellt in dem Modell aufgrund von Hypothese H4 eine weitere abhängige Variable dar. Sie ist aber aufgrund der Hypothese H6 auch eine Mediatorvariable. Die Variable Hoffnung ist aufgrund der Hypothese H5 und H7 sowohl eine weitere abhängige Variable, als auch eine Moderatorvariable. Drittvariablen sind das Geschlecht der Befragten, die Innovationsresistenz und das Vertrauen in die Wissenschaft.

C Methode

I Forschungsdesign

Zur Überprüfung der Hypothesen wurde ein Online-Experiment mit vier verschiedenen Experimentalgruppen durchgeführt. Dem Experiment liegt somit ein Between-Subject-Design zugrunde. Alle Teilnehmenden wurden randomisiert einer der vier Experimentalgruppen zugeordnet. Die Wirkung der Frames sowie weitere Variablen wurden mit einem standardisierten Fragebogen erfasst. Der Fragebogen der Experimentalgruppen unterscheidet sich jeweils nur hinsichtlich ihres Stimulus. Das Forschungsdesign mit sowohl einseitigen Frames als auch einem beidseitigen Frame ist an ein ähnliches Experiment von Cobb angelehnt.[85] Das Experiment wurde als Online-Experiment durchgeführt. Im Anschluss erfolgten Varianzanalysen sowie Mediationsanalysen mit SPSS und PROCESS.

II Entwicklung der Stimuli und Pretest

Bei dem angewendeten Stimulus für die vorliegende Studie handelte es sich um einen dafür konstruierten Zeitungsartikel.

Alle Stimuli bestanden aus dem Grundbeitrag. Dieser bestand, um möglichst wenig Tendenzen in Richtung von Chancen und Risiken der künstlichen Intelligenz zu enthalten, aus einem Hinweis auf eine erfundene aktuelle Konferenz für Unternehmen über KI und einer Definition von KI und Kennzahlen und Fakten zur Forschung über KI in Deutschland.

85 Cobb 2005.

Da es bisher keine Studien zu Frames in der deutschen Berichterstattung über KI gibt, wurde in der weiteren Konstruktion der Stimuli auf Studienergebnisse aus den Niederlanden und dem englischsprachigen Raum zurückgegriffen. Dabei wurden vor allem die von den Studien als häufige Themen identifizierte Themen aufgegriffen: Das wirtschaftliche Potential von KI, sowie KI als Lösung für praktische Probleme im Gesundheitswesen. Im Risiko-Frame wird das Risiko des Verlusts vieler Arbeitsplätze und ein möglicher Verlust von Privatsphäre thematisiert.[86]

Sowohl der Chancen- als auch der Risiko-Frame wurden dabei so aufgebaut, dass nach dem Grundbeitrag eine Zwischenüberschrift kommt, die bei dem Chancen-Frame das Potenzial von künstlicher Intelligenz und bei dem Risiken-Frame die Gefahr von künstlicher Intelligenz betont. Danach wurde auf Studienergebnisse zu Chancen beziehungsweise Risiken verwiesen, ohne konkrete Studien zu nennen. Abschließend wurde bei beiden Frames ein fiktiver Experte einer bekannten Organisation (Deutsches Krebsforschungszentrum beziehungsweise Netzpolitik.Org) zitiert.

Der vierte Frame ist der beidseitige Frame, er besteht aus den Texten des Grundbeitrags, des Chancen-Frames und des Risiko-Frames.

Das Layout des Stimulus wurde an das Design von Online-Nachrichtenartikeln angelehnt. Die Stimuluskonstruktion wurde durch Pretests überprüft.

III Teilnehmende und Durchführung

Die Teilnehmenden wurden überwiegend über die sozialen Medien rekrutiert. Der Erhebungszeitraum verlief vom 01. 12. 2020 bis zum 14. 01. 2021.

Der finale Datensatz enthält die Daten von 449 Teilnehmenden, wovon 119 (26.5 %) der Gruppe mit dem Grundbeitrag, 111 (24.7 %) der Gruppe mit Chancen-Frame, 107 (23.8 %) der Gruppe mit dem Risiken-Frame und 112 (24.9 %) der Gruppe mit dem beidseitigen Frame zuzuordnen sind.

Es wurde ein Manipulationscheck durchgeführt, der gezeigt hat, dass 90.6 Prozent der Befragten die inhaltliche Frage zum Text richtig beantwortet haben und somit davon auszugehen ist, dass diese Befragten den Stimulus auch gelesen haben. Da allerdings viele Menschen im Alltag beim Nachrichtenlesen auch häufig nur Texte überfliegen und dabei nicht alle Inhalte wahrnehmen und Wirkungen auch bei kaum gelesenen Texten, zum Beispiel allein durch die Wahrnehmung der Überschrift beziehungsweise Zwischenüberschriften nicht ausgeschlossen werden konnten, wurden auch diejenigen nicht ausgeschlossen, bei denen der Manipulationscheck nicht erfolgreich war.

86 Vgl. Chuan et al. 2019: 342.

Von den Teilnehmenden waren 51.9 Prozent weiblich, 47.2 Prozent männlich und 0.9 Prozent divers. Die Altersspanne reichte von 18 bis 76 Jahren, das Durchschnittsalter betrug 41 Jahre.

66.4 Prozent der Teilnehmenden verfügen über einen Hochschulabschluss, 22.7 Prozent haben Abitur beziehungsweise die allgemeine oder fachbezogene Hochschulreife, 2.7 Prozent haben die Fachhochschulreife, 0.9 Prozent haben die Polytechnische Oberschule abgeschlossen, 3.6 Prozent haben einen Realschulabschluss, niemand der Befragten hat einen Hauptschulabschluss oder keinen Schulabschluss und 3.8 Prozent haben einen anderen Schulabschluss angegeben. Da keine der befragten Personen angegeben haben, dass sie keinen Schulabschluss oder einen Hauptschulabschluss haben, fallen diese beiden Kategorien aus der Analyse heraus. Deshalb wurde bei der Bildung nur zwei Kategorien unterschieden hohe Bildung (mit Hochschulabschluss) und niedrige Bildung (alle anderen Bildungsabschlüsse).

IV Fragebogen und Variablen

Der Fragebogen wurde in deutscher Sprache verfasst. Die Studie wurde bei den Teilnehmenden als Untersuchung über die Medienberichterstattung über neue Technologien vorgestellt.

Im Fragebogen wurden zunächst die soziodemografischen Variablen Alter und Geschlecht abgefragt. Danach wurde die Innovationsresistenz mit einer von Lobera et al. entwickelten Skala erhoben.[87] Anschließend wurde das Vertrauen in Wissenschaft und Forschung erhoben und der höchste Bildungsabschluss der Teilnehmenden abgefragt.

Das Vorwissen über KI wurde in Anlehnung an Pinto Dos Santos et al. und Cho et al. mit drei Fragen erhoben, auf die jeweils mit Ja oder Nein geantwortet werden konnte.[88] Die drei Fragen lauten »Haben Sie schon von künstlicher Intelligenz gehört?«, »Haben Sie schon von maschinellem Lernen gehört?«, »Haben Sie schon von neuronalen Netzen gehört?«.

Danach folgten Fragen zum politischen Interesse, zur politischen Einstellung und zur Einschätzung der wirtschaftlichen Lage in Deutschland.

Anschließend wurde den Befragten, der jeweils randomisiert zugewiesene Stimulus angezeigt. Im Anschluss daran wurden die Emotionen erhoben. Daraufhin folgten dann KI-bezogene Fragen. So wurde mit der in Anlehnung an Cobb entwickelten Frage »Vertrauen Sie Unternehmer/innen, die künstliche Intelligenz in ihrem Unternehmen einsetzen oder Produkte mit künst-

87 Lobera et al. 2020.

88 Cho et al. 2020; Pinto Dos Santos et al. 2019.

licher Intelligenz produzieren, dass diese möglichen Risken reduzieren?« erhoben, inwieweit die Befragten Vertrauen in den wirtschaftlichen Einsatz von KI haben.[89] Die Einstellungen gegenüber künstlicher Intelligenz wurden mit zwei Itembatterien und einem semantischen Differential erhoben. Abschließend wurden noch zwei Fragen zum Artikel gestellt, um abzufragen, ob der Artikel gelesen und wie er wahrgenommen wurde.

D Auswertung der Ergebnisse

I Varianzanalysen zur Analyse der Frame-Wirkungen

Die Hypothesen 1 bis 5 befassen sich mit der direkten Wirkung von Chancen- und Risiken-Frames. Um sie zu testen, wurden zwei einfaktorielle Varianzanalysen (ANOVAs) mit dem zugewiesenen Stimulus als unabhängiger Variable durchgeführt. Diese wurden durch eine weitere Analyse (ANCOVA) mit den Kovariaten Geschlecht, Vertrauen in die Wissenschaft und Innovationsresistenz ergänzt.

Die ANOVA zeigt, dass sich die Risikobeurteilung signifikant zwischen den Gruppen unterscheidet, $F(3, 445) = 7.46$, $p < 0.001$, $\eta p^2 = 0.05$. Bei der Chancenbeurteilung zeigt sich kein signifikanter Unterschied zwischen den Gruppen, $F(3, 445) = 0.71$, $p = 0.55$, $\eta p^2 = 0.01$.

Da kein signifikanter Unterschied zwischen den verschiedenen Gruppen hinsichtlich der Chancenbeurteilung nachgewiesen werden konnte, wird die Hypothese 1 *Chancen-Frames erhöhen die positiven Einstellungen gegenüber künstlicher Intelligenz* abgelehnt.

Bei der Risikobeurteilung unterscheidet sich die Gruppe, die den beidseitigen Frame erhalten hat ($M = 11.39$, $SD = 2.18$), signifikant von der Gruppe, die den Grundbeitrag erhalten hat ($M = 10.04$, $SD = 2.73$, $p < 0.001$). Die Gruppe, die den beidseitigen Frame erhalten hat, schätzt das Risiko, das mit KI verbunden ist, höher ein. Ebenso unterscheidet sich die Gruppe, die den beidseitigen Frame bekommen hat ($M = 11.39$, $SD = 2.18$), signifikant von der Gruppe, die den Chancen-Frame erhalten hat ($M = 10.13$, $SD = 2.76$, $p = 0.001$). Die Gruppe, die den beidseitigen Frame erhalten hat, schätzt das Risiko ebenfalls höher ein als die Gruppe, die den Chancen-Frame erhalten hat. Die anderen Gruppen unterscheiden sich hinsichtlich der Risikobeurteilung nicht signifikant voneinander ($p > 0.064$).

Berücksichtigt man die Kovariaten, so unterscheidet sich die Risikobeurteilung weiterhin mit einer schwachen Effektstärke signifikant zwischen den

89 Cobb 2005.

Gruppen, $F(3, 442) = 7.13$, $p < 0{,}001$, $\eta p^2 = 0.05$. Die Chancenbeurteilung unterscheidet sich weiterhin nicht signifikant zwischen den Gruppen, $F(3, 442) = 1.00$, $p = 0.394$, $\eta p^2 = 0.01$.

Da kein signifikanter Unterschied zwischen der Gruppe, die den Beitrag mit den Risiko-Frames gelesen hat, und den anderen Gruppen hinsichtlich der negativen Einstellungen gegenüber KI festgestellt werden konnte ($p > 0.064$), wird die Hypothese 2 *Risiko-Frames erhöhen die negativen Einstellungen gegenüber künstlicher Intelligenz* abgelehnt.

Die schon dargestellte Varianzanalysen mit der Risiko- und Chancenbeurteilung von KI haben gezeigt, dass sich die Gruppe, die den beidseitigen Frame bekommen hat ($M = 11.39$, $SD = 2.18$) hinsichtlich der Risikobeurteilung, signifikant von der Gruppe, die den Chancen-Frame erhalten hat, unterscheidet ($M = 10.13$, $SD = 2.76$, $p = 0.001$). Die Gruppe, die den beidseitigen Frame gesehen hat, schätzt die mit künstlicher Intelligenz verbundenen Risiken somit höher ein als die Gruppe, die den Chancen-Frame gesehen hat.

Da hinsichtlich der Risikobeurteilung ein signifikanter Unterschied zwischen der Gruppe mit dem beidseitigen Frame und der Gruppe mit dem Chancen-Frame sowie mit der Gruppe mit dem Grundbeitrag gefunden wurde, wird die Hypothese 3 *Beidseitige Frames, die sowohl Chancen als auch Risiken enthalten, haben keinen Effekt auf die Einstellungen gegenüber künstlicher Intelligenz* widerlegt.

Zur Beantwortung der zweiten Forschungsfrage wurden zudem die Hypothesen 4 und 5 betrachtet. Sie befassen sich mit der direkten Wirkung der Chancen- und Risiken-Frames auf die Emotionen. Hierfür wurden ebenfalls eine einfaktorielle ANOVA und ANCOVA durchgeführt. Auch hier war der zu-

Tabelle 1 Ergebnisse der einfaktoriellen Varianzanalyse der Frames auf abhängige Variablen

	Grundbeitrag		Chancen-Frame		Risiken-Frame		Beidseitiger Frame				
	M	SD	M	SD	M	SD	M	SD	F (3,442)	p	ηp^2
Risikenbeurteilung	10.04	2.73	10.13	2.76	10.87	2.26	11.39	2.18	7.13	<0.001	0.05
Chancenbeurteilung	11.66	2.01	11.60	2.13	11.70	2.00	11.96	1.87	1.00		0.01

Anmerkungen: Die Innovationsresistenz, das Geschlecht und das Vertrauen in die Wissenschaft wurden als Kovariate miteinbezogen. In Bezug darauf wird hier der um diese Einflüsse bereinigte Effekt der Frames dargestellt.

gewiesene Stimulus die unabhängige Variable. Die abhängigen Variablen waren in dieser Analyse die Emotionen Sorge und Hoffnung. Die Kovariaten blieben gleich.

Die ANOVA zeigt, dass sich die Sorge signifikant mit einer vergleichsweisen hohen Effektstärke zwischen den Gruppen unterscheidet, $F(3, 445) = 19.31$ $p < 0.001$ $\eta p^2 = 0.12$. Die Hoffnung unterscheidet sich mit einer schwachen Effektstärke signifikant zwischen den verschiedenen Stimuligruppen $F(3, 445) = 7.39$ $p < 0.001$ $\eta p^2 = 0.05$.

Die Gruppe, die dem Grundbeitrag gesehen hat ($M = 2.27$, $SD = 1.04$), hat signifikant weniger Sorge angegeben als die Gruppe, die den Risiko-Frame gesehen hat ($M = 3.11$, $SD = 1.03$, $p < 0.001$). Auch die Gruppe, die den Grundbeitrag gesehen hat ($M = 2.27$, $SD = 1.04$), hat signifikant weniger Sorge angegeben als die Gruppe, die den beidseitigen Frame gesehen hat ($M = 2.99$, $SD = 1.03$, $p < 0.001$). Auch wenn man die Kovariaten mitberücksichtigt, unterscheiden sich die Gruppen weiterhin signifikant ($p < 0.001$) hinsichtlich der Variable Sorge, $F(3, 442) = 19.58$, $p < 0.001$, $\eta p^2 = 0.12$ und hinsichtlich der Variable Hoffnung, $F(3, 442) = 8.51$, $\eta p^2 = 0.06$.

Die Gruppe mit dem Chancen-Frame ($M = 2.34$, $SD = 1.07$) unterscheidet sich hinsichtlich der Sorge signifikant von der Gruppe mit dem Risiko-Frame, bei der die Sorge deutlich höher ist ($M = 3.11$, $SD = 1.06$, $p < 0.001$). Sie unterscheidet sich zudem signifikant (p <0,001) von der Gruppe mit dem beidseitigen Frame, die ebenfalls eine höhere Sorge angegeben haben ($M = 2.99$, $SD = 1.03$, $p < 0.001$). Somit erhöhen der Risiko-Frame und der beidseitige Frame die Sorge. Die Gruppe, die den Risiko-Frame gesehen hat, unterscheidet sich nicht signifikant von der Gruppe, die den beidseitigen Frame gesehen hat ($p = 0.826$). Ebenso unterscheidet sich die Gruppe, die den Grundbeitrag gesehen hat, hinsichtlich der Sorge nicht signifikant von der Gruppe, die den Chancen-Frame gesehen hat ($p = 0.952$).

Die Hypothese 4 *Risiko-Frames erhöhen die Sorge* kann angenommen werden, da sich die Gruppe, die den Risiko-Frame gesehen hat, von allen Gruppen au 0.394 ßer der Gruppe mit dem beidseitigen Frame signifikant unterscheidet, zumal der beidseitige Frame ebenfalls den Risiko-Frame enthält.

Die Gruppe, die den Chancen-Frame gesehen hat, hat eine signifikant höhere Hoffnung angegeben ($M = 3.53$, $SD = 0.90$) als die Gruppe, die den Risiko-Frame gesehen hat ($M = 2.95$, $SD = 0.87$, $p < 0.001$). Die Gruppe, die den Chancen-Frame gesehen hat ($M = 3.53$, $SD = 0.90$), weist außerdem signifikant mehr Hoffnung auf als die Gruppe, die den beidseitigen Frame gesehen hat ($M = 3.20$, $SD = 0.90$, $p = 0.029$). Es konnte kein signifikanter Unterschied zwischen der Gruppe mit Chancen-Frame und der Gruppe mit dem Grundbeitrag hinsichtlich der Hoffnung gefunden werden ($p = 0.112$). Auch die Gruppe mit Risiko-Frame unterscheidet sich hinsichtlich der Hoffnung nicht signifikant von der

Tabelle 2 Ergebnisse der einfaktoriellen Varianzanalysen auf die abhängigen Variablen

	Grund-beitrag		Chancen-Frame		Risiken-Frame		Beidseitiger Frame				
	M	SD	M	SD	M	SD	M	SD	F (3,442)	p	ηp^2
besorgt	2.27	1.04	2.34	1.07	3.11	1.06	2.99	1.03	19.58	<0.001	0.12
hoffnungs-voll	3.25	0.98	3.53	0.90	2.95	0.87	3.20	0.90	8.51	<0.001	0.06

Anmerkungen: Die Innovationsresistenz, das Geschlecht und das Vertrauen in die Wissenschaft wurden als Kovariate miteinbezogen, dargestellt wird der um diese Einflüsse bereinigte Effekt der Frames.

Gruppe mit dem Grundbeitrag ($p = 0.074$) und der Gruppe mit dem beidseitigen Frame ($p = 0.176$).

Eine vollständige Annahme der Hypothese ist durch den nicht signifikanten Unterschied zwischen der Gruppe mit dem Chancen-Frame und der Gruppe mit dem Grundbeitrag nicht möglich. Da sich allerdings die Gruppe mit Risiko-Frame hinsichtlich der Hoffnung nicht signifikant von der Gruppe mit dem Grundbeitrag und der Gruppe mit dem beidseitigen Frame unterscheidet, können die signifikanten Unterschiede zwischen der Gruppe mit dem Chancen-Frame und der Gruppe mit dem Risiko-Frame sowie der Gruppe mit dem beidseitigen Frame auch nicht alleine durch eine Reduktion der Hoffnung durch den Risiko-Frame erklärt werden. Die Hoffnung ist bei der Gruppe, die den Chancen-Frame gesehen hat, signifikant höher als bei den Gruppen, die einen Frame gesehen haben, der Risiken enthält (Risiko-Frame oder beidseitiger Frame). Die Hypothese 5 *Chancen-Frames erhöhen die Hoffnung* kann somit teilweise angenommen werden.

II Pfadanalysen zur Mediatorrolle der Sorge und Hoffnung

In der Hypothese 6 wurde eine Mediatorrolle der Sorge bei der Wirkung des Risiko-Frames auf die negativen Einstellungen gegenüber künstlicher Intelligenz vermutet. In der Hypothese 7 wurde eine Mediatorrolle der Hoffnung bei der Wirkung des Chancen-Frames auf die positiven Einstellungen gegenüber künstlicher Intelligenz vermutet. Beide Hypothesen wurde mit Hilfe einer Mediationsanalyse mit PROCESS getestet. Die Frames wurden als Dummy kodiert und als unabhängige Variable in die Regressionsanalyse aufgenommen.

Es konnte kein signifikanter direkter Effekt des Risiko-Frames auf die Ri-

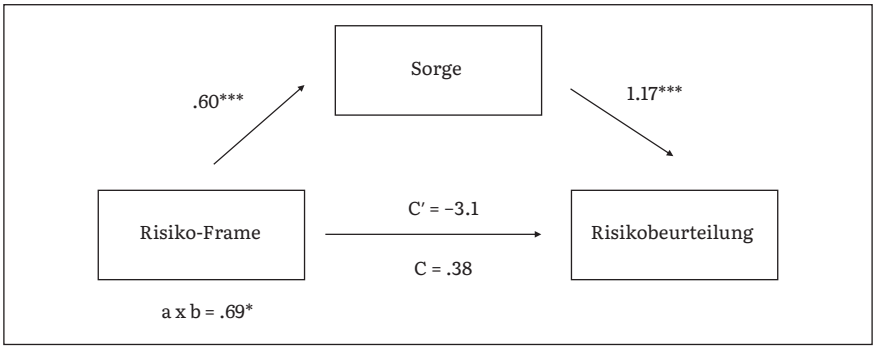

Abbildung 2 Mediationsmodell des Effektes vom Risiko-Frame auf die Risikobeurteilung über die Sorge

Anmerkungen: Die Innovationsresistenz, das Geschlecht und das Vertrauen in die Wissenschaft wurden als Kovariate miteinbezogen.
*** P < .001 . ** p < .01 . * p < .05

sikobeurteilung festgestellt werden, B = 0.38, p = 0.145. Nachdem der Mediator in das Modell eingeführt wurde, sagte der Stimulus den Mediator Sorge signifikant vorher, B = 0.60, p < 0.001. Dieser wiederum sagte die Risikobeurteilung signifikant vorher, B = 1.17, p < 0.001. Durch die Mediation wurde der direkte Effekt des Risiko-Frames auf die Risikobewertung nicht signifikant, B = −0.31, p = 0.171. Es konnte somit festgestellt werden, dass der Effekt des Risiko-Frames von der Sorge mediiert wird. Der indirekte Effekt zeigte sich mit *ab* = 0.69, 95%-KI[0.41, 1.00]. Da das Bootstrap-Konfidenzintervall keine Null beinhaltet, konnte die Mediatorrolle der Sorge auf die Risikobeurteilung noch einmal bestätigt werden. Die graphische Darstellung des Mediationsmodells befindet sich in der Abbildung 2.

Die Hypothese 6 *Risiko-Frames erhöhen die Sorge, was zu einer negativen Einstellung gegenüber künstlicher Intelligenz führt* kann somit bestätigt werden.

Es konnte kein signifikanter direkter Effekt des Chancen-Frames auf die Chancenbeurteilung festgestellt werden, B = −0.11, p = 0.641. Nachdem der Moderator aufgenommen wurde, sagte der Chancen-Frame den Mediator Hoffnung signifikant vorher, B = 0.42, p < 0.001, welcher wiederum die Chancenbeurteilung signifikant vorhersagte, B = .68, p < 0.001. Damit konnte festgestellt werden, dass die Wirkung des Chancen-Frames auf die Chancenbeurteilung vollständig durch den Mediator Hoffnung mediiert wird, indirekter Effekt *ab* = 0.28, 95%-KI[0.14, 0.44]. Die Hypothese 7 *Chancen-Frames erhöhen die Hoffnungen, was zu einer positiven Einstellung gegenüber künstlicher Intelligenz führt* konnte somit bestätigt werden.

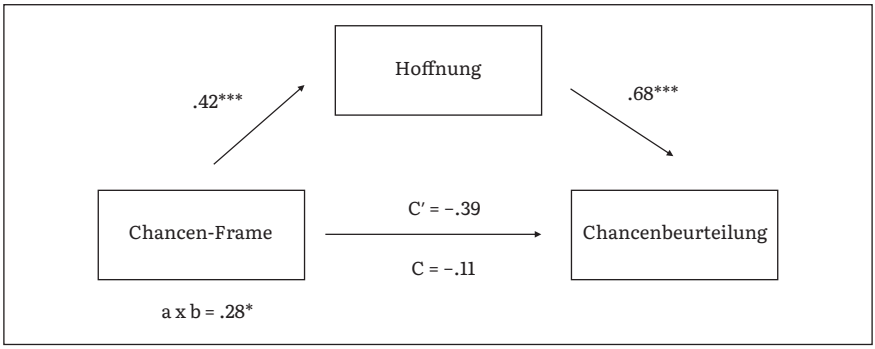

Abbildung 3 Mediationsmodell des Effektes vom Chancen-Frame auf die Chancenbeurteilung über die Hoffnung

Anmerkungen: Die Innovationsresistenz, das Geschlecht und das Vertrauen in die Wissenschaft wurden als Kovariate miteinbezogen.
*** P < .001 . ** p < .01 . * p < .05

III Übersicht über die Ergebnisse

Tabelle 3 Übersicht über die Bestätigung oder Ablehnung der getesteten Hypothesen

Hypothese		bestätigt/ abgelehnt
H1	*Chancen Frames erhöhen die positiven Einstellungen gegenüber künstlicher Intelligenz*	x
H2	*Risiko Frames erhöhen die negativen Einstellungen gegenüber künstlicher Intelligenz*	x
H3	*Beidseitige Frames, die sowohl Chancen als auch Risiken enthalten, haben keinen Effekt auf die Einstellungen gegenüber künstlicher Intelligenz*	x
H4	*Risiko-Frames erhöhen die Sorge*	√
H5	*Chancen-Frames erhöhen die Hoffnung*	√/x
H6	*Risiko-Frames erhöhen die Sorge, was zu einer negativen Einstellung gegenüber künstlicher Intelligenz führt*	√
H7	*Chancen-Frames erhöhen die Hoffnung, was zu einer positiven Einstellung gegenüber künstlicher Intelligenz führt*	√

E Diskussion

Die Studie hatte das Ziel zur Beantwortung folgender Fragen beizutragen:

F1: Wie beeinflussen Chancen- und Risiko-Frames in der medialen Berichterstattung über künstliche Intelligenz die Einstellungen gegenüber künstlicher Intelligenz?

F2: Welche emotionalen Effekte haben Chancen- und Risiko-Frames in der medialen Berichterstattung über künstliche Intelligenz?

Die Ergebnisse zeigen, dass Chancen- und Risiko-Frames vor allem eine emotionale Wirkung haben und so die Einstellungen gegenüber künstlicher Intelligenz beeinflussen. Die vermutete direkte Wirkung der Frames auf die Einstellungen gegenüber künstlicher Intelligenz konnte nicht nachgewiesen werden. Daher wurden die Hypothesen 1 *Chancen-Frames erhöhen die positiven Einstellungen gegenüber künstlicher Intelligenz* und 2 *Risiko-Frames erhöhen die negativen Einstellungen gegenüber künstlicher Intelligenz* widerlegt. Hinsichtlich der Wirkung der Chancen- und Risiko-Frames auf die Einstellungen war der Forschungsstand allerdings auch nicht eindeutig. So unterscheiden sich diese Ergebnisse von den Ergebnissen von der Studie von Cobb, der eine direkte Wirkung von Chancen- und Risiken-Frames auf die Einstellungen gegenüber Nanotechnologie nachgewiesen hat.[90] Die Ergebnisse bestätigen allerdings die Ergebnisse von Binder et al., die ebenfalls keinen direkten Effekt der Frames auf die Risikobeurteilung gefunden haben.[91]

Hypothese 3 *Beidseitige Frames, die sowohl Chancen als auch Risiken enthalten, haben keinen Effekt auf die Einstellungen gegenüber künstlicher Intelligenz* wurde widerlegt, da bei der Gruppe, die den beidseitigen Frame bekommen hat, die Risikobeurteilung signifikant höher ist als bei der Gruppe mit dem Chancen-Frame. Der beidseitige Frame erhöht somit die Risikobeurteilung. Dies gilt allerdings nur für die Risikobeurteilung. Hinsichtlich der Chancenbeurteilung und der Gesamtbewertung von künstlicher Intelligenz konnte kein Effekt der beidseitigen Frames nachgewiesen werden. Damit widersprechen diese Ergebnisse dem Forschungsstand, demnach beidseitige Frames zu weniger Einstellungsänderungen führen.[92]

Hypothese 4 *Risiko-Frames erhöhen die Sorge* konnte bestätigt werden. Hier ist interessant, dass sich die Gruppe mit dem Risiko-Frame nicht signifikant

90 Vgl. Cobb 2005: 229.

91 Vgl. Binder et al. 2016: 841.

92 Vgl. Cobb 2005: 230.

von der Gruppe mit dem beidseitigen Frame unterscheidet, bei dem der Frame ebenfalls den Risiko-Frame enthält. Hier scheint der Chancen-Frame die durch den Risiko-Frame erhöhte Sorge nicht auszugleichen. Somit konnte hier das Ergebnis von Cobb, dass Risiko-Frames in der medialen Berichterstattung über Nanotechnologie zu mehr Sorge führen, auch für KI bestätigt werden.[93]

Hypothese 5 *Chancen-Frames erhöhen die Hoffnung* konnte nur teilweise bestätigt werden, da sich die Hoffnung zwischen der Gruppe, die den Chancen-Frame gesehen hat und der Gruppe, die den Grundbeitrag gesehen hat, nicht signifikant unterscheidet. Aber die Hoffnung ist bei der Gruppe, die den Chancen-Frame gesehen hat, signifikant höher als bei den Gruppen, die den Risiko-Frame oder den beidseitigen Frame – und damit einen Frame, in dem Risiken erwähnt werden – gesehen haben. Dies ergänzt den bisherigen Forschungsstand, bei dem bisher nur die Wirkung des Risiko-Frames auf die Hoffnung untersucht wurde.[94]

Die in Hypothese 6 *Risiko-Frames erhöhen die Sorge, was zu einer negativen Einstellung gegenüber künstlicher Intelligenz führt*, vermutete Mediatorrolle der Sorge auf den Effekt zwischen Risiko-Frames und Risikobeurteilung konnte bestätigt werden. Damit bestätigt sich auch der im Forschungsstand dargelegte Effekt von negativen Emotionen auf die Risikobeurteilung neuer Technologien.

Die in Hypothese 7 *Chancen-Frames erhöhen die Hoffnung, was zu einer positiven Einstellung gegenüber künstlicher Intelligenz führt* vermutete Mediatorrolle der Hoffnung wurde ebenfalls bestätigt und legt nahe, dass Emotionen generell die Wirkung von Chancen- und Risiko-Frames mediieren.

F Fazit

Die Ergebnisse haben gezeigt, dass Chancen- und Risiko-Frames über KI vor allem einen emotionalen Effekt haben. So führt die Erwähnung von Risiken sowohl beim Risiko-Frame als auch beim beidseitigen Frame zu einer Steigerung der Sorge und Reduktion der Hoffnung. Es konnte auch gezeigt werden, dass Sorge und Hoffnung die Framewirkung mediieren. Eine direkte Wirkung der Chancen- und Risiko-Frames auf die Einstellungen gegenüber künstlicher Intelligenz konnte hingegen nur zwischen dem beidseitigen Frame und der Risikobeurteilung nachgewiesen werden.

Die Ergebnisse legen eine mögliche Erklärbarkeit der Wirkung der Frames entlang des Elaboration Likelihood Models nahe, die in weiteren Studien ge-

93 Vgl. Cobb 2005: 232.

94 Vgl. ebd.

prüft werden könnte. Zugleich werfen die Ergebnisse Fragen über die Dauer der Effekte auf. Zudem wäre es interessant zu untersuchen, ob sich die Wirkung von Chancen- und Risiko-Frames bei audiovisuellen Medien unterscheidet.

Die Studie zeigt auch, dass den Medien in der Berichterstattung über KI eine gesellschaftliche Verantwortung zukommt, da die Medienberichterstattung emotionale Effekte auslösen kann, welche wiederum die Einstellungen gegenüber künstlicher Intelligenz beeinflussen können. Journalist:innen sollten sich dieser Verantwortungen bewusst sein, vor allem wenn sie über Chancen und Risiken künstlicher Intelligenz berichten.

Literatur

Araujo, Theo/Helberger, Natali/Kruikemeier, Sanne/De Vreese, Claes H. 2020: In AI we trust? Perceptions about automated decision-making by artificial intelligence. In: AI & Society, o. S. https://doi.org/10.1007/s00146-019-00931-w

Binder, Andrew R./Hillback, Elliott D./Brossard, Dominique 2016: Conflict or caveats? Effects of media portrayals of scientific uncertainty on audience perceptions of new technologies. In: Risk analysis 36 (4): 831–846.

Brennen, J. Scott/Howard, Philip N./Nielsen, Rasmus K. 2018: An industry-led debate: How UK media cover artificial intelligence. RISJ Fact-Sheet.

Brossard, Dominique/Scheufele, Dietram A./Kim, Eunkyung/Lewenstein, Bruce V. 2009: Religiosity as a perceptual filter: Examining processes of opinion formation about nanotechnology. In: Public Understanding of Science 18 (5): 546–558.

Cobb, Michael D. 2005: Framing Effects on Public Opinion about Nanotechnology. In: Science Communication 27 (2): 221–239. https://doi.org/10.1177/1075547005281473

Cobb, Michael D./Macoubrie, Jane 2004: Public perceptions about nanotechnology: Risks, benefits and trust. In: Journal of Nanoparticle Research 6: 395–405. https://doi.org/10.1007/s11051-004-3394-4

Chuan, Ching-Hua/Tsai, Wan-Hsiu S./Cho, Su Y. 2019: Framing artificial intelligence in American newspapers. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society: 339–344.

Dos Santos, D. P./Giese, D., Brodehl, S./Chon, S. H./Staab, W./Kleinert, R./Maintz, D./Baeßler, B. 2019: Medical students' attitude towards artificial intelligence: a multicentre survey. In: European radiology 29 (4): 1640–1646. https://doi.org/10.1007/s00330-018-5601-1

dpa. 2020: KI als Wachstumsturbo. In: ICT Channel, 21. 01. 2020. https://www.ict-channel.com/software-services/ki-als-wachstumsturbo.121601.html (aufgerufen am 02. 02. 2020).

Druckman, James N. 2001: Evaluating framing effects. In: Journal of economic psychology 22 (1): 91–101. https://doi.org/10.1016/S0167-4870(00)00032-5

Du-Harpur, X./Watt, F. M./Luscombe, N. M./Lynch, M. D. 2020: What is AI? Applications of artificial intelligence to dermatology. In: British Journal of Dermatology 183 (3): 423–430. https://doi.org/10.1111/bjd.18880

Entman, Robert M. 1993: Framing: Toward clarification of a fractured paradigm. In: Journal of communication 43 (4): 51–58.

Gross, Kimberly 2008: Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. In: Political Psychology 29 (2): 169–192. https://doi.org/10.1111/j.1467-9221.2008.00622.x

Gross, Kimberly/Brewer, Paul R. 2007: Sore losers: News frames, policy debates, and emotions. In: Harvard International Journal of Press/Politics 12 (1): 122–133. https://doi.org/10.1177/1081180X06297231

Gross, Kimberly/D'Ambrosio, Lisa 2004: Framing emotional response. In: Political psychology, 25 (1): 1–29. https://doi.org/10.1111/j.1467-9221.2004.00354.x

Ho, Shirley S./Scheufele, Dietram A./Corley, Elizabeth A. 2013: Factors influencing public risk-benefit considerations of nanotechnology: Assessing the effects of mass media, interpersonal communication, and elaborative processing. In: Public understanding of science (Bristol, England) 22 (5): 606–623. https://doi.org/10.1177/0963662511417936

Jatscha, Alexander 2021: Mit Machine Learning Millionen sparen. In: MMLogistik, 10. 05. 2021. https://www.mm-logistik.vogel.de/mit-machine-learning-millionen-sparen-a-1022865/ (aufgerufen am 11. 05. 2021).

Kramper, Gernot 2021: Bedrohung wie in »Terminator« – superintelligente KI für Menschen unbeherrschbar. In: Stern vom 09. 05. 2021. https://www.stern.de/digital/technik/kuenstliche-intelligenz--superintelligente-ki-fuer-menschen-unbeherrschbar-30519648.html (aufgerufen am 11. 05. 2021).

Kühne, Rinaldo 2013: Emotionale Framing-Effekte auf Einstellungen: Ein integratives Modell. In: Medien & Kommunikationswissenschaft 61 (1): 5–20. https://doi.org/10.5771/1615-634x-2013-1-5

Lee, Chul-joo/Scheufele, Dietram A. 2006: The influence of knowledge and deference toward scientific authority: A media effects model for public attitudes toward nanotechnology. In: Journalism & Mass Communication Quarterly 83 (4): 819–834. https://doi.org/10.1177/107769900608300406

Lobera, Josep/Rodríguez, Fernández C. J./Torres-Albero, Cristóbal 2020: Privacy, values and machines: Predicting opposition to artificial intelligence. In: Communication Studies 71 (3): 448–465. https://doi.org/10.1080/10510974.2020.1736114

Nelson, Thomas E./Oxley, Zoe M./Clawson, Rosalee A. 1997: Toward a psychology of framing effects. In: Political behavior 19 (3): 221–246.

Nisbet, Matthew C./Scheufele, Dietram A./Shanahan, James/Moy, Patricia/Brossard, Dominique E./Lewenstein, Bruce V. 2002: Knowledge, reservations, or promise? A media effects model for public perceptions of science and technology. In: Communication Research, 29: 584–608. https://doi.org/10.1177/009365002236196

Ouchchy, Leila/Coin, Allen/Dubljević, Veljko 2020: AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media. In: AI & SOCIETY 35 (4): 927–936. https://doi.org/10.1007/s00146-020-00965-5

Potthoff, Matthias 2012: Medien-Frames und ihre Entstehung. Wiesbaden, VS Verlag für Sozialwissenschaften.

Price, Vincent/Tewksbury, David/Powers, Elizabeth 1997: Switching trains of thought: The impact of news frames on readers' cognitive responses. In: Communication research 24 (5): 481–506. https://doi.org/10.1177/009365097024005002

Reiche, Matthias 2021: EU will Künstliche Intelligenz zähmen. In: Tagesschau vom 21.04.2021. https://www.tagesschau.de/wirtschaft/technologie/eu-gesetzentwurf-kuenstliche-intelligenz-ki-101.html (aufgerufen am 11.05.2021).

Scheufele, Dietram A./Lewenstein, Bruce V. 2005: The public and nanotechnology: How citizens make sense of emerging technologies. In: Journal of Nanoparticle Research 7(6): 659–667.

Schirmer, Sophia 2019: »Die Alternative ist: Irgendwann ist dein Arbeitsplatz fort«. In: Zeitonline vom 12.11.2019. https://www.zeit.de/die-antwort/2019-11/kuenstliche-intelligenz-jobs-arbeit-mensch-maschine (aufgerufen am 02.02.2020).

Schreiner, Maximilian (2021, 09.05). Facebooks neues KI-System findet den pinken Elefanten. In: mixed, 09.05.2021. https://mixed.de/facebooks-neues-ki-system-findet-den-pinken-elefanten/ (aufgerufen am 11.05.2021).

Sun, Shaojing/Zhai, Yujia/Shen, Bin/Chen, Yibei 2020: Newspaper coverage of artificial intelligence: A perspective of emerging technologies. In: Telematics and Informatics 53. https://doi.org/10.1016/j.tele.2020.101433

Tversky, Amos/Kahneman, Daniel 1981: The Framing of Decisions and the Psychology of Choice. In: Science 211 (4481): 453–458. https://doi.org/10.1126/science.7455683

Vergeer, Maurice 2020: Artificial intelligence in the dutch press: An analysis of topics and trends. In: Communication Studies 71 (3): 373–392. https://doi.org/10.1080/10510974.2020.1733038

ORCID

Selina Fucker https://orcid.org/0000-0001-8728-3485

II. Medienethische und -philosophische Reflexion

Bilder des Menschlichen

Theologisch-ethische Herausforderungen der Vorstellungswelten künstlicher Intelligenz

Florian Höhne

A Einleitung

Am 10. Februar 1996 trat der Schachweltmeister Garri Kasparow gegen den Computer »Deep Blue« auf dem Schachfeld an – und verlor das erste Spiel:[1] »Das Bild der Überlegenheit des menschlichen Gehirns bekommt ein paar Kratzer«, heißt es in der Süddeutschen Zeitung rückblickend.[2] Die technische Entwicklung ist seitdem rasant vorangegangen. Die Rede von »künstlicher Intelligenz« überträgt eine menschliche Eigenschaft, Intelligenz, auf entwickelte oder erträumte Maschinen – damit bringt diese Rede Menschen und Maschinen auf dasselbe Spielfeld,[3] wie einst Kasparow und Deep Blue auf demselben Schachfeld gegeneinander antraten. Aber welches Spielfeld ist dies heute?

Nassehi hat in seiner »Theorie der digitalen Gesellschaft« ein Narrativ vorgestellt, nach dem gegenwärtige Nutzungen digitaler Technik an vorgängige, ebenfalls digitale Strukturen moderner Gesellschaften anknüpfen:[4] »Diese Ge-

1 Hauck 2021.

2 Hauck 2021.

3 Vgl. dazu auch Nassehi 2019: 217. Besonders deutlich wird dies auch an der Idee des »Turing-Test« (Nassehi 2019: 218).

4 Nassehi 2019, insbesondere 11, 16, 18 f., 62, 66, 176 f., 245. Der Erzählungscharakter wird auch darin deutlich, dass Nassehi unterschiedliche »Entdeckungen der Gesellschaft« differenziert (ebd., 45–54, 319). Ich rede von »Nutzungen digitaler Technik« in Anlehnung an Schmidts Rede von »Nutzungspraktiken« (Schmidt 2011: 41) und Neubergers »Unterscheidung zwischen dem technischen Potenzial eines Mediums und seine selektiven Aneignung im Prozess der Institutionalisierung« (Neuberger 2009: 22).

Florian Höhne 2022: Bilder des Menschlichen. Theologisch-ethische Herausforderungen der Vorstellungswelten künstlicher Intelligenz. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 111–135.
https://doi.org/10.11588/heibooks.1106.c15196

sellschaft besteht aus Regelmäßigkeiten und Mustern, für deren Entbergung es offensichtlich einen Bedarf gibt […]. Ich habe die Entstehung der statistischen Erfassung der Gesellschaft seit dem 18./19. Jahrhundert als die Initialzündung einer *digitalisierten* Gesellschaft beschrieben.«[5] Ein Hauptgedanke des Narratives ist also, dass die Digitalisierung »in der Struktur der Gesellschaft gründet und keine ihr äußerliche Entwicklung ist.«[6] Ein von Nassehi erwähnter, aber weniger auserzählter Strang in diesem Narrativ handelt von den Menschenbildern, die spezifische Nutzungen von digitaler Technik genauso plausibel machen wie die Rede von »künstlicher Intelligenz«.[7] Ausgehend von der Vermutung, dass diese Menschenbilder besagtes Spielfeld markieren, werde ich im Folgenden die These entfalten, dass das in theologischer Hinsicht reduktive[8] Menschenbild »risikoinformierter Entscheider« Entwicklung und Deutung sogenannter »künstlicher Intelligenz« präformiert.

Wo Menschen menschliche Eigenschaften auf Maschinen übertragen – sei es technisch-praktisch oder diskursiv – stehen auch menschliche Selbstbilder auf dem Spiel. Dieses Wechselspiel von Technik und Menschenbildern reflektiere ich hier sozialethisch anhand einer exemplarischen Pointierung des Menschenbildes in westlich-modernen[9] Gesellschaft: dem Bild des Menschen als »risikoinformierten Entscheider« (C.I). Dazu kläre ich zunächst die Reflexionskategorien – nämlich den Praxis- und Imaginationsbegriff – (B), um dann mit diesem begrifflichen Instrumentarium das Menschenbild herauszuarbeiten, das die gegenwärtige Entwicklung und Thematisierung von sogenannten KI erst plausibel macht (C). Exemplarische Einsichten in den theologisch-ethischen Diskurs weiten demgegenüber die Perspektive auf Menschen (D).

Weil der Begriff »Künstliche Intelligenz« selbst zu problematisieren wäre, spreche ich mit dem Bericht »Automating Society« von AlgorithmWatch prä-

5 Nassehi 2019: 212, kursiv im Original. zur Beschreibung der Gesellschaft mit den Mitteln der Sozialstatistik vgl. Nassehi 2019: 31 f.

6 Nassehi 2019: 212. Vgl. inhaltlich auch Nassehi 2019: 177, 186, 319 f., 327.

7 So behandelt Nassehi das Thema sogenannter KI (Nassehi 2019: 217–262) im Rahmen seiner Gesamtthese: »Meine These lautet, dass die Digitaltechnik deshalb einen solchen Siegeszug antreten konnte, weil die Gesellschaft selbst schon Dispositionen aufweist, die nur digital zu erfassen sind. Das gilt auch für die Frage der *smart technologies*.« (ebd., 245, kursiv im Original) In Bezug auf Entscheidungssysteme gilt sein Hauptinteresse der »Bedeutung der Zurechnungsform« (Nassehi 2019: 224). Dabei geht er – wie ich später auch – sowohl von »technischem Entscheiden« (ebd., 224, der Sache nach auch 131, 224, 227, 229–233, 235, 244) und der Rolle von »Verteilungswahrscheinlichkeiten und prognostische[n] Bestimmungen« (231, auch: 131) dabei als auch von der Unterschiedlichkeit der Verständnisse von Intelligenz (247–262, bes. explizit auf S. 259) aus, fokussiert aber nicht ausführlich auf die in diesem Zusammenhang impliziten und prägenden Menschenbilder (andeutungsweise nur ebd., 244–248, wo er aber von »*Mentalisierung*« statt von »*Humanisierung*« handeln will (ebd., 246)), vor allem nicht auf deren ethische Problematik.

8 Zum Reduktionismus von Modellen im Digitalen vgl. auch Seele 2020: 153 f.

9 Diese Spezifikation folgt dem Fokus von Taylor in Taylor 2004: 1 f.

ziser von »automated decision making (ADM)« und ADM-Systemen,[10] was gleichbedeutend auch für »Algorithmische[.] Entscheidungssysteme« steht.[11]

B Kategorien: Praxis und Imaginäres

Eine theologische Ethik, die den digitalen Wandel im digitalen Wandel orientierend reflektiert, benötigt dafür Kategorien, vermittels derer sie Phänomene wahrnehmen und beschreiben kann. Diese Kategorien können theologische Ethik interdisziplinär und in gesellschaftlichen Diskursen anschlussfähig machen.

Am Berlin Institute for Public Theology ist in der theologischen Ethik der Ansatz entstanden, den Praxis- und den Imaginationsbegriff als solche Kategorien zu verwenden. So hat Meireis gesellschaftlich virulente Narrative beschrieben, die den digitalen Wandel deuten und damit orientieren: das Übermenschlichkeitsnarrativ bei Yuval Harari etwa, das Tsunami-Narrativ von digitaler Technik, die mit der Zwangsläufigkeit einer Naturkatastrophe hereinbreche, oder das Assimilationsnarrativ.[12] Als gesellschaftlich virulente Narrative würden sie – so Meireis – das prägen, was Taylor »social imaginary« genannt hat.[13] Ich selbst habe – angeregt durch die Lektüre Bourdieus[14] und von Schmidts[15] Arbeit – vorgeschlagen den Praxisbegriff der Praxissoziologie für eine theologische Digitalisierungsethik mit dem Taylorschen Imaginationsbegriff zu verbinden[16] und führe dies in meiner Habilitation weiter aus.

---

10 AlgorithmWatch 2019: 9. Der Bericht definiert: »Algorithmically controlled, automated decision-making or decision support systems are procedures in which decisions are initially – partially or completely – delegated to another person or corporate entity, who then in turn use automatically executed decision-making models to perform an action. This delegation – not of the decision itself, but of the execution – to a data-driven, algorithmically controlled system, is what needs our attention.« (ebd. Dabei ist die Unterscheidung zwischen »decision« und »execution« im Konkreten m. E. nicht so eindeutig und klar, wie die Formulierung es hier scheinen lässt.) Vgl. ebd. auch zur mangelnden Präzision des KI-Begriffs (Stichwort: »fuzzily defined term« (ebd.)). Mit dem Systembegriff übernimmt der Bericht einen »holistic approach« (ebd. und inhaltlich ebd., 14) – Analoges leistet hier der Praxisbegriff. Eine umfassende Perspektive auf Algorithmen fordert Zweig, wenn sie schreibt, »dass auch ihre Einbettung in unsere Gesellschaft der Überprüfung bedarf.« (Zweig 2019: 2)

11 Zweig 2019: 1 f. Vgl. dort auch für den Zusammenhang von ADM-Systemen und »Risikobewertung« (ebd.). Zweig gibt einen guten und verständlichen Einblick in die Funktionsweise dieser Systeme: ebd., 3–5.

12 Meireis 2019: 52 f. Dabei verwendet Meireis auch den Begriff »Praxis«.

13 Meireis 2019: 53.

14 Für die Anregung zur ausführlichen Auseinandersetzung mit Bourdieu danke ich Torsten Meireis.

15 Schmidt 2011.

16 Höhne 2019a, 2019b.

Die These auf Methodenebene lautet also: Die Kategorien »Praxis« und »Imaginäres« sind geeignet, Kultur – hier genauer die von Stalder so benannte »Kultur der Digitalität«[17] – und ihren Wandel für die Zwecke der Ethik zu beschreiben. Was ist nun mit dem Praxis- und dem Imaginationsbegriff gemeint und was könnte die ethische Arbeit mit diesen Begriffen für die Auseinandersetzung mit vermeintlichen KI-Artefakten austragen?

I Praxis

Die Überschriften »Praxistheorie« oder »Praxeologie« fassen Reckwitz zufolge unterschiedliche Theorieelemente zusammen, aus denen sich weniger eine kohärente und systematische Sozialtheorie als vielmehr eine Heuristik ergibt, die die Aufmerksamkeit der Forschungs- und Reflexionspraxis steuert;[18] Reckwitz spricht von einem »fruchtbaren Ideenpool«.[19] Im Rückgriff vor allem auf die zusammenführende Arbeit von Reckwitz, aber profitierend von den grundlegenden Arbeiten von Bourdieu, Schmidt, Schmidt, Hillebrandt und Schatzki selbst,[20] lassen sich die folgenden »Merkmale der praxeologischen Perspektive auf das Soziale und das Handeln«[21] benennen.

Mit Reckwitz und Schatzki lässt sich unter »Praxis« zunächst ein »Netz von Tat- und Sprachakten« verstehen, dessen Akte vor allem durch »knowing how to« verbunden sind.[22] Damit sind die ersten zwei Charakteristika angedeutet:

1. In Praktiken spielt nicht nur explizites Wissen – also ein »knowing that« – sondern auch das eine Rolle, was Reckwitz etwa »praktisches Wissen« nennt.[23] Anders als explizites Wissen muss dieses praktische Wissen nicht sprachlich

17 Stalder 2016.

18 Reckwitz 2003 und ebd. bes. 282–85 und 288–90. Auch: Hillebrandt 2014: 7–9, 15.

19 Reckwitz 2003: 289.

20 Vgl. Bourdieu 1993, 2014, 2015; Reckwitz 2003; Schmidt 2011; Schmidt 2012; Hillebrandt 2014; Schatzki 2008.

21 Reckwitz 2003: 289.

22 Höhne 2019a: 30 f.; Reckwitz 2003: 290. Auch Hillebrandt übersetzt »sayings« trefflich mit »Sprechakte«: Hillebrandt 2014: 52, 59. Schatzki spricht von einem »temporally unfolding und spatially dispersed nexus of doings and sayings« und vom »understanding of X-ing« als »link« dieser Akte, wobei er dieses »understanding« als »ability to«, als »knoing how to« qualifiziert (Schatzki 2008: 89, 91). Reckwitz spricht von »einem routinisierten ›nexus of doings and sayings‹« und von »know how« (Reckwitz 2003: 289 f.).

23 Vgl.: »Die Praxistheorie begreift die kollektiven Wissensordnungen der Kultur nicht als ein geistiges ›knowing that‹ […], sondern als ein praktisches Wissen, ein Können, ein know how, ein Konglomerat von Alltagstechniken, ein praktisches Verstehen im Sinne eines ›Sich auf etwas verstehen‹.« (Reckwitz 2003: 289) Die »Implizitheit dieses Wissens« betont Reckwitz etwas später (Reckwitz 2003: 292).

formulierbar sein, es ist ein »knowing how to«.[24] Beispiel: Ich weiß, wie man Fahrrad fährt.[25] Dieses Können ist ein »praktisches Wissen«, ein »knowing how to«. Ich kann zwar beschreiben, was ich beim Fahrradfahren tue, aber dies Beschreibung wird nicht dazu führen, dass Zuhörende nun auch Fahrradfahren können; Der Praxis des Fahrradfahrens ist ein eigenes, praktisches Wissen inhärent, das nicht ohne weiteres explizierbar[26] ist.[27]

2. Es geht bei der Reflexion von Praxis um »Praxis als Vollzugswirklichkeit«.[28] Praxistheorie richtet den Fokus darauf, was sich im konkreten Vollzug der Praxis ereignet; entsprechend spricht Hillebrandt auch vom »*Ereignisparadigma*«.[29] Hintergrund dafür ist Wittgensteins Sprachphilosophie, genauer: der von Hillebrandt zusammengefasste Gedanke, dass es »keine allgemein gültige Logik der Praxis« gibt, dass sich Praxis nicht aus Gesetzen oder Theorien ergibt, sondern dass sich eine Regelmäßigkeit »in den Praktiken selbst einstellt und deshalb nicht ahistorisch bestimmt werden kann«.[30] Es geht nicht um eine Theorie des Fahrradfahrens oder die theoretische Möglichkeit des Fahrradfahrens oder ein theoretisches Modell des Fahrradfahrens, sondern um das, was konkret geschieht, wenn ein Mensch Fahrrad fährt.

Schmidt zufolge verstehen Praxeologien ihr Forschen »selbst als ein Ensemble von Praktiken« und betonen so die »Differenz zwischen der Logik der wissenschaftlichen Beobachtung und der Logik der beobachteten Praktiken«.[31] Den Theoriehintergrund dazu liefern Bourdieus Arbeiten zur »*praktische[n] Logik*«.[32] Danach wohnt einer Praktik eine eigene Logik inne, die nur in Teilnahmeperspektive nachvollzogen werden kann, sich aber der theoretischen Beobachterperspektive verschließt:[33] »Vom praktischen Schema zum nach der Schlacht konstruierten theoretischen Schema, vom praktischen Sinn zum theoretischen Modell übergehen, das entweder als Vorhaben, Plan oder *Methode* oder als mechanisches Programm, als vom Wissenschaftler auf mysteriöse

24 Reckwitz 2003: 292; Schatzki 2008: 91.

25 Vgl. für dieses Beispiel und genau diese Ausdeutung in Bezug auf »tacit knowing«: Polanyi 1962, 601. Die Auffindung dieses Textes verdanke ich Paßmann 2018.

26 Zu dieser (fehlenden) Explizierbarkeit vgl. Reckwitz 2003: 290, 292.

27 Vgl. Polanyi 1962, 601.

28 Hillebrandt 2014: 11.

29 Hillebrandt 2014: 29, 111.

30 Hillebrandt 2014: 36–39, bes. 38 f. Zur Rückführung des Vollzugswirklichkeitscharakters von Praktiken auf Wittgenstein vgl. Hillebrandt 2014: 54.

31 Schmidt 2012: 13.

32 Schmidt 2012: 13, 28–37, 39 f.; Bourdieu 1993, 2015. Begriff zitiert von Bourdieu 1993: 157, Hervorhebung im Original.

33 Bourdieu 1993: 148 f., 157, 164 f.; Schmidt 2012: 29, 35–37. Zur Differenz der Teilnahme- und Beobachtungsperspektive vgl. Bourdieu 1993: 151.

Weise rekonstruierte mysteriöse Ordnung gelesen werden kann, heißt alles dahinfahren lassen, was die zeitliche Realität der in Ausübung begriffenen Praxis ausmacht.«[34] Darum zu wissen und gleichzeitig immer weiter nach der konkreten Vollzugswirklichkeit zu fragen ist Signum einer Praxistheorie, die – so Schmidt – »so gebaut sein [soll], dass sie sich vom Empirischen fortlaufend verunsichern, irritieren und revidieren lässt«.[35] Das heißt: Sie rechnet immer damit, dass praktisch etwas Entscheidendes von statten geht, was ihr gerade aufgrund ihres eigenen Theoriecharakters unsichtbar bleibt.[36]

3. Darüber hinaus rückt der Praxisbegriff drittens etwa bei Reckwitz die »*Materialität der Praktiken*« in den Fokus – und zwar in doppelter Hinsicht:[37] Einmal geht es um »die menschlichen ›Körper‹«, die Praktiken erst ermöglichen, und dann um »die ›Artefakte‹«.[38] Dabei betont Reckwitz, dass die Struktur der Dinge die Praktiken weder völlig determiniert noch irrelevant für diese ist – auf den »sinnhafte[n] *Gebrauch*« käme es an.[39] Im Beispiel gesagt: Die Praxis des Fahrradfahrens beinhaltet einen menschlichen Körper, der die Tat- und Sprechakte vollzieht, die vernetzt miteinander eine Praktik ergeben. Außerdem ist das Artefakt – das Fahrrad – wichtig für die Praktik. Platt gesagt: Das Fahrrad an sich macht noch kein Fahrradfahren und könnte auch zu allen möglichen anderen lustigen Streichen verwendet werden. Wird es aber sinnhaft zum Fahrradfahren gebraucht, prägt es in diesem Gebrauch durch seine je und je konkrete Dinglichkeit auch diese Praxis.

II Imaginäres

Was jeweils als sinnhafter Gebrauch erscheint, hängt von dem imaginativen Horizont der Praxisteilnehmer:innen ab. An dieser Stelle lässt sich Taylors Kategorie des »social imaginary« als einschlägige Bezeichnung dieses imaginativen Horizontes in die Praxistheorie eintragen. Den Begriff führt Taylor in seiner Modernitätstheorie[40] ein und will ihn folgendermaßen verstanden wissen:

> »By social imaginary, I mean something much broader and deeper than the intellectual schemes people may entertain when they think about social reality in a disen-

34 Bourdieu 1993: 148 f.
35 Schmidt 2012: 31.
36 Schmidt 2012: 29.
37 Reckwitz 2003: 290, kursiv im Original.
38 Reckwitz 2003: 290.
39 Reckwitz 2003: 291.
40 Taylor 2004: 1.

> gaged mode. I am thinking, rather, of the ways people imagine their social existence, how they fit together with others, how things go on between them and their fellows, the expectations that are normally met, and the deeper normative notions and images that underlie these expectations.«[41]

Damit beinhaltet das sozial Imaginäre m. E. nicht nur geteilte Vorstellungen von Sozialität und nicht nur geteilte sittliche Grundorientierungen, sondern auch eine grundlegende Vorstellung davon, was es konkret heißt, Mensch in Sozialität zu sein. Transportiert werden diese Vorstellungen Taylor zufolge »in images, stories, and legends«.[42] Sie bestehen, wie ich sagen würde, in den sozialen Praktiken, die sie informieren. Als Horizont von Praktiken und ihrer subjektiven Plausibilität informiert das soziale Imaginäre dann, was sinnhafter Gebrauch von Dingen sein könnte. Zum Beispiel: Insofern die Ausrichtung auf eine ökologisch nachhaltige Gesellschaft Teil des sozial Imaginären eines bestimmten Milieus ist, wird es einem Milieuangehörigen in dem Horizont dieser Imagination sinnvoll erscheinen ein Fahrrad zu nutzen und das Auto in der Garage zu lassen, es kommt zum sinnhaften Gebrauch des Fahrrades in der Praxis des Radfahrens.

Das sozial Imaginäre lässt aber nicht nur (neue) Praktiken initiieren. Neue Praktiken wirken – wie Taylor betont – auch in das sozial Imaginäre zurück.[43] Neue Dinge, etwa neue Technologien verändert Praktiken, was wiederum neue Deutungen, Theorien und Vorstellungen schafft, die über die Zeit in das Selbstverständliche des sozial Imaginären einsickern können.[44]

Diese Kategorien ermöglichen es Fragerichtungen in Bezug auf den Zusammenhang von »Vorstellungswelten künstlicher Intelligenz«, technischen Artefakten und Praktiken zu differenzieren.

Ethisch interessant ist *erstens* die Frage danach, wie wir »KI« genannten technischen Innovationen anthropomorph deuten, welche Geschichten wir um diese herumerzählen, welche Diskurse sich um diese herum aufbauen, welche Imaginationen sich damit verbinden und inwiefern bestimmte Technologien als »künstliche Intelligenz« gedeutet werden können.[45]

Ethisch interessant ist *zweitens* die Rückrichtung der Bedeutungsübertragung in einer anthropomorphen Beschreibung von Technik. Wie veränderte

---

41 Taylor 2004: 23.

42 Taylor 2004: 23.

43 Taylor 2004: 29 f.

44 Taylor 2004: 30, 33.

45 Ggf. ##Verweis auf andere Aufsätze in diesem Band ##. Auf dieser Linie liegt auch, was Nassehi »operative oder praxeologische Frage« genannt hat, die Frage, »was eine Maschine können muss, dass ihr Intelligenz ›zugerechnet‹ wird« (Nassehi 2019: 219). Nassehi verweist auch auf Süssenguths Arbeit zu »Digitalisierungssemantiken« (Nassehi 2019: 139).

es menschliche Identitätsvorstellungen und Menschenbilder, mit ADM-Systemen zu interagieren? Was macht es mit unseren anthropologischen Vorstellungswelten, auch Maschinen in einem gewissen Rahmen praktisch Intelligenz zuzuschreiben? All das untersucht die Dresdner Theologin Platow empirisch in einem Forschungsprojekt, das bis Dezember 2020 lief.[46] Dabei geht es ihr offenbar vor allem die »Selbstwahrnehmung von Individuen« in Mensch-Maschine-Interaktionen.[47] Auch Peter Seele fragt in eine ähnliche Richtung, wenn er seine »These einer ›doppelten Kontingenz der Intelligenz‹« entwickelt:[48] Diese These ist, »dass sich auch die menschliche Intelligenz durch das Aufkommen von KI, Algorithmen und Digitalisierung mitverändert, indem der Mensch stärker und tiefer in die Standardisierungslogik der Digitalisierung hineingezogen wird«:[49] Der Mensch werde selbst zum kontrollierbaren »Datenhaufen«.[50] Seeles Fokus ist »der Mensch in maschinenintensiven Umgebungen«.[51]

Mich interessiert die *dritte* Fragerichtung, die sich grundlegender auf die Vorstellungswelt bezieht, die Mensch und Maschine auf dasselbe »Spielfeld« bringen und so Bedeutungsübertragungen zwischen Menschen und ADM-Systemen in beiden Richtungen überhaupt plausibel erscheinen lassen. Wir gebrauchen praktisch schon jetzt ADM-Systeme. Welche Menschenbilder im imaginären Horizont ließen und lassen Entwicklung und Gebrauch dieser Technologien sinnhaft erscheinen und welche Menschenbilder werden durch diese Entwicklungen und Gebrauche reproduziert und damit tiefer ein-

46 https://www.uni-augsburg.de/de/forschung/einrichtungen/institute/zig/gesundheitsforschung/digitalisierung/anthropomorphe-uebertragungen-als-konstitutivum-der-begegnung-von-mensch-und-kuenstlicher-intelligenz/ [Abruf am 5. 6. 2021]. Ein ähnliche Fragerichtung kommt bei Nassehi vor, wo er in Bezug auf »Mentalisierung« danach fragt, »welche Konsequenzen das in der sozialen Praxis hat.« (Nassehi 2019: 220)

47 Vgl. so die Projektbeschreibung: »Im Rahmen des Projekts wird der Frage nachgegangen, wie sich die Selbstwahrnehmung von Individuen und ihre Identitätskonstruktionen im Umgang mit KI basierten Systemen verändern.« (https://www.uni-augsburg.de/de/forschung/einrichtungen/institute/zig/gesundheitsforschung/digitalisierung/anthropomorphe-uebertragungen-als-konstitutivum-der-begegnung-von-mensch-und-kuenstlicher-intelligenz/ [Abruf am 5. 6. 2021]). Ausführlicher zu dem Projekt vgl.: Platow, Birte: Selbstwahrnehmung und Ich-Konstruktion im Angesicht von Künstlicher Intelligenz. In: M. Huppenbauer/P. Kirchschläger/G. Ulshöfer (Hg.): Digitalisierung aus theologischer und ethischer Perspektive. Konzeptionen – Anfragen – Impulse (erscheint in Kürze). Baden-Baden: Nomos.

48 Seele 2020: 131 f., Zitat auf S. 132.

49 Seele 2020: 141. Seele selbst fasst seine These so zusammen: »Wenn der Mensch nun auch immer mehr in das Raster der standardisierten Regel-Automatismen gezwängt wird, indem sein digitales Selbst mehr und mehr über das schillernde analoge Selbst gelegt wird, so bewegen sich künstliche und menschliche Intelligenz im doppelten Sinne aufeinander zu: KI wird immer besser und menschenähnlicher, da die Menschen immer mehr zu Datenhaufen werden, die standardisiert und automatisiert ausgelesen und optimiert werden.« (Seele 2020: 157)

50 Seele 2020: 148.

51 Seele 2020: 160.

geschrieben ins sozial Imaginäre? Diese Fragen greifen – in anderen Kategorien – die Suchrichtung auf, in die zunächst Stalders Arbeit kulturtheoretisch[52] und dann jüngst auch Nassehi systemtheoretisch gewiesen haben, indem sie herausgearbeitet haben, wie kulturelle Entwicklungen und gesellschaftlichen Strukturen die Art der gegenwärtigen Digitalisierung präfiguriert haben und deshalb – wie Nassehi behauptet – »die Digitaltechnik also letztlich nur die logische Konsequenz einer in ihrer Grundstruktur *digital* gebauten Gesellschaft ist«.[53]

Dazu will ich im Folgenden zwei Thesen entfalten, eine Beschreibungsthese und eine Orientierungsthese. *Die Beschreibungsthese betrifft diejenigen Menschenbilder, die die Innovationen von ADM-Systemen prägen, die deren sinnhaften Gebrauch informieren und so in den Praktiken dieses sinnhaften Gebrauch verstärkt werden.*[54] *Die Orientierungsthese betrifft einige exemplarische Züge derjenigen Menschenbilder die in theologischen Diskurspraktiken virulent sind und die m. E. geeignet sind, den Horizont des Imaginären zu weiten und so über offenere Menschenbilder mehr Freiheit zu ermöglichen.*

C Mensch-Imaginationen probabilistischer Risiken und ADM-Systeme

I Freie Individuen und risikoinformierte Entscheider:innen

Fragt man nach den Menschenbildern im sozial Imaginären der Praktiken westlicher moderner Gesellschaften und von da aus konkreter nach den damit zusammenhängenden Subjektkonstitutionen, lassen sich unter anderem zwei Bestimmungen plausibel behaupten, die erste mit Taylor, die zweite im Anschluss an Samerski und Henkel.

(1) Zentral in Taylors Beschreibung des sozial Imaginären in westlichen, modernen Gesellschaft ist das Primat des Individuums:[55] »[S]o society itself

52 Stalder 2016: 10 f. Dort heißt es: »Die Entstehung und Ausbreitung der Kultur der Digitalität ist die Folge eines weitreichenden, unumkehrbaren gesellschaftlichen Wandels, dessen Anfänge teilweise bis ins 19. Jahrhundert zurückreichen.« (ebd.)

53 Nassehi 2019: 11, kursiv im Original. Nassehi macht seine systemtheoretischen Grundlagen explizit (ebd., 166–172) und verweist auch selbst auf Stalder (ebd., 63).

54 Damit ist mein Fokus ein anderer als der von Seeles (Seele 2020): Geht es ihm eher darum, wie sich Menschen und Menschenbilder »durch das Aufkommen von KI, Algorithmen und Digitalisierung« verändern (ebd., 141. Auch Seele hat Menschenbilder im Blick, so implizit ebd., 147), also um Adaptionen (ebd., 159 f., 172), geht es mir zunächst auch darum, welche Vorstellungen den Einsatz von ADM-Systemen zuerst plausibel und attraktiv haben erscheinen lassen.

55 Taylor 2004: 50. Taylor selbst spricht auch von der »primacy of the individual« (ebd., 64). Westliche Modernität ist der Fokus dieses Buches (ebd., 2).

comes to be reconceived as made up of individuals«.[56] Taylor zufolge ist es in diesen Gesellschaften über die letzten Jahrhunderte zur Selbstverständlichkeit geworden, dass Menschen sich primär als Individuen vorstellen, die mit gleichen Rechten ausgestattet sind und aus individueller Freiheit zum gegenseitigen Nutzen (»mutual benefit«) handeln können.[57] Menschen sind »free individuals« – diese theoretische Behauptung sei in die selbstverständliche Vorstellung des sozial Imaginären eingesickert.[58] Menschen als freie Individuen vorzustellen ist eine anthropologische Festlegung, die nicht selbstverständlich ist. Das zeigen die von Taylor beschriebenen Vorstellungen, die den Menschen im Gegensatz zum modernen Individualismus zunächst in Sozialität, zunächst in komplementäre Beziehungen und Hierarchien eingebunden sahen, bevor sie von ihm:r als Einzelner:m sprechen können.[59] Das spezifische der westlichen Moderne, das Ergebnis der Revolution, die Taylor »The Great Disembedding« nennt, ist also die Selbstverständlichkeit, den Menschen als wesenhaft freies Individuum vorzustellen.[60]

(2) Erst im Horizont dieser Selbstverständlichkeit wird die Entwicklung möglich, die Samerski und Henkel m. E. plausibel beschreiben und die sie auf die These bringen,

> »dass die in der Gesellschaft um sich greifende Adressierung von Handelnden als *risikoinformierte Entscheider* Individuen für Ungewisses und Kontingentes (mit)verantwortlich macht. In verschiedensten Lebensbereichen, sei es im Finanzwesen, im Bildungssystem oder im Gesundheitswesen, lässt sich eine Zunahme von institutionalisierten bzw. institutionell definierten Entscheidungssituationen beobachten, in denen die zur Verfügung stehenden Optionen mit vorauskalkulierten Risiken verknüpft sind.«[61]

Diese These führen sie einmal allgemein mit dem Rückgriff auf die Arbeit des Kognitionspsychologen Gigerenzer und dann exemplarisch konkret für das Ge-

56 Taylor 2004: 50. Taylor spricht auch vom neuen Selbstverständnis, »that gave an unprecedented primaccy to the individual.« (Taylor 2004: 50)

57 Taylor 2004: 3–22, 49–67, 172 f., Zitat auf S. 19.

58 Taylor 2004: 17, 20–22, 64 f., Zitat auf S. 65.

59 Taylor 2004: 3–22, 49–67.

60 Taylor 2004: 49–67. Zitat auf S. 49. Vgl. insbes.: »On the contrary, what I propose here is the idea that our first self-understanding was deeply embedded in society. Our essential identity was as father, son, and so on, and as a member of this tribe. Only later did we come to conceive of ourselves as free individuals first. This was not just a revolution in our natural view of ourselves, but involved a profound change in our moral world, as is always the case with identity shifts.« (ebd., 64 f.)

61 Samerski und Henkel 2015: 85. Hervorhebung: FH.

sundheitssystem aus.[62] Durch die Entstehung der »Wahrscheinlichkeitstheorie im 17. und 18. Jahrhundert« und deren Verbindung mit der »›politischen Arithmetik‹« im 19. Jahrhundert, sei es eben im 19. Jahrhundert möglich geworden, vorhandene Datenmengen probabilistisch zu analysieren und so »aus der Erfassung und Klassifizierung von Daten über die Vergangenheit mithilfe mathematischer Methoden berechnete Rückschlüsse über die ungewisse Zukunft zu ziehen«.[63] Zumindest makroskopisch wird das Unwissen über die Zukunft damit in ein Wahrscheinlichkeitswissen überführt – auf Mikroebene funktioniert dies selbstverständlich nicht:[64] Das Wahrscheinlichkeitswissen um die durchschnittliche Lebenserwartung einer Kohorte beinhaltet keinen Aufschluss über den individuellen Todeszeitpunkt.[65] »Diese statistischen Regelmäßigkeiten« hätten »im 20. Jahrhundert zunehmend Wissenschaft und Politik« bestimmt; »Verbesserung der Hygiene« habe etwa die Lebenserwartung erhöht.[66]

Wie die Makroebenenlogik des Wahrscheinlichkeitswissens, für individuelle Akteure konkret, also in Mikroebene übersetzt wurde, zeichnen die beiden anhand des Medizinsystems nach, in dem die »wachsende Dominanz von Statistik und Wahrscheinlichkeit […] in der zweiten Hälfte des 20. Jahrhunderts zu einem tiefgreifenden Umbruch« geführt habe:[67] An die Stelle der Frage nach »individuelle[n] Ursachen und Prognosen einer konkreten Erkrankung« sei die Thematisierung von »Risikofaktoren und Erkrankungswahrscheinlichkeiten« getreten.[68] Statistisch wurden »Zusammenhänge zwischen Verhaltensweisen, Umwelteinflüssen, Körpermerkmalen und Erkrankungshäufigkeiten« hergestellt und als »Risikofaktoren« zusammengefasst;[69] dabei gilt: »Probabilistische Risiken sind statistische Konstrukte«.[70] Die Thematisierung von »Risikofaktoren« für Patienten wendet nun genau »statistische Gesetzmäßigkeiten auf Individuen« an, übersetzt also Wahrscheinlichkeitswissen

62 Samerski und Henkel 2015: 85–91.

63 Samerski und Henkel 2015: 87.

64 Vgl.: »Auf der Ebene der Population gerinnt zur kalkulierbaren Regelmäßigkeit, was für den Einzelnen unvorhersehbarer Zufall, Schicksalsschlag oder einzigartige Lebensgeschichte ist« (Samerski und Henkel 2015: 87).

65 Samerski und Henkel 2015: 87 f.

66 Vgl. auch für das Zitat Samerski und Henkel 2015: 88.

67 Samerski und Henkel 2015: 88–91, Zitat auf S. 88 f. Zusammenfassend formulieren die beiden: »Zunehmend gerinnen dabei statistische Wahrscheinlichkeiten zu Risiken, die nicht nur für aggregierte Kollektive, sondern auch für Einzelne Bedeutung beanspruchen.« (ebd., 89)

68 Samerski und Henkel 2015: 89.

69 Samerski und Henkel 2015: 89.

70 Samerski und Henkel 2015: 90.

auf Makroebene in Entscheidungskalküle auf Mikroebene:[71] Patienten würden »›individuelle Risiken‹« attestiert, was sie »zum eigenverantwortlichen Risikomanagement« auffordere.[72]

Was Samerski und Henkel hier nacherzählt haben, ist die Entwicklung einer Praxis der Kontingenzbewältigung. Es ist – das machen sie wie zitiert explizit – »Ungewisses und Kontingentes«, um das es jeweils geht:[73] die »ungewisse Zukunft«, in ihrem Beispiel die Ungewissheit über eigene künftige Krankheit,[74] in anderen Lebensbereichen die Ungewissheit über ein künftiges Verbrechen, darüber, ob X an Y einen Kredit zurückzahlen wird oder wieviel Strom Z's Solarzelle morgen Mittag produzieren wird. Werden diese Ungewissheiten in »[p]robabilistische Risiken« übersetzt, sind sie auf den individuellen Fall bezogen zwar streng genommen noch genauso ungewiss und offen, scheinen aber plötzlich den Entscheidungen des Subjektes zurechenbar.[75]

Genau das ist anthropologisch entscheidend: Die angedeuteten Entscheidungs-, Beratungs- und Responsibilisierungspraktiken sind folglich von einem als selbstverständlich unterstellten Menschenbild durchwirkt, das damit als Teil des sozial Imaginären westlich-moderner Gesellschaften gelten kann: Menschen werden nicht nur als freie Individuen, sondern als eigenverantwortliche Risikomanager:innen[76] vorgestellt, eben als »risikoinformierte Entscheider«[77]. Dieses Bild des Menschen als risikokinformierten Entscheider besteht in besagten Praktiken, insofern es die entsprechende (medizinische) Beratung erst geboten und den Einsatz von diagnostischen Prozeduren mit spezifischen Gebrauch von Geräten erst sinnhaft erscheinen lässt, und gleichsam eine Subjektivität reproduziert, die im Modus risikoinformierten Entscheidens mit Kontingenz umgeht.

---

71 Samerski und Henkel 2015: 90 f., Zitat auf S. 90.

72 Zitate Samerski und Henkel 2015: 90 f.

73 Samerski und Henkel 2015: 85.

74 Samerski und Henkel 2015: 87 f., Zitat auf S. 87.

75 Vgl.: »Probabilistische Risiken […] erzeugen also neuartige Entscheidungssituationen, indem sie die Wissens- und Erwägungshorizonte der Akteure verändern. Diese neuartigen Entscheidungssituationen bringen auch neuartige Verantwortungszuschreibungen mit sich.« (Samerski und Henkel 2015: 91) Vgl. dazu auch die Luhmann-Referenz der beiden ebd.

76 Samerski und Henkel sprechen wie zitiert vom »eigenverantwortlichen Risikomanagement« (Samerski und Henkel 2015: 91).

77 Wie oben zitiert.

II Mensch-Imaginationen und ADM-System-Praktiken

Es ist nun dieses Menschenbild, das nicht nur als Theoriekonzept, sondern auch als sozial Imaginäres gerade viele jener Praktiken prägt, in denen jetzt schon (und künftig vielleicht noch mehr) ADM-Systeme[78] sinnhaft gebraucht werden. Denn: »Die bisherigen maschinellen Intelligenzen sind gut darin, auf Wahrscheinlichkeiten basierende Automatisierungen auszuführen.«[79] Das impliziert zweierlei:

Erstens ist es erst der Horizont dieses Menschenbildes, in dem es plausibel erscheinen kann, digitale Entscheidungshilfen oder gar digitale Entscheidungssystem zu entwickeln und zu nutzen.[80] Erst wenn »risikoinformierte Entscheidungen«[81] zur selbstverständlichen Aufgabe von Menschen geworden ist, kann es sinnvoll erscheinen, Systeme zu entwickeln, die Menschen diese Entscheidungen mit Informationen erleichtern oder gar ganz abnehmen.[82] Dass wir es im Zuge des digitalen Wandels also auch mit ADM-Systemen zu tun haben, lässt sich aus einer Technikkultur[83] erklären, für die das Menschenbild der »risikoinformierten Entscheider:innen« prägend war und ist.

Der Bericht von AlgorithmWatch zum Thema listet konkrete Beispiele aus Europa für den Einsatz ADM-Systemen, für Deutschland etwa das Credit-Scoring der SCHUFA[84]oder in Italien »RiskER«, ein System, das das Hospitalisierungsrisiko von Patienten berechnet und Ärzten so hilft, sich besonders um Menschen zu kümmern, die wahrscheinlich Hochrisikopatienten sind.[85] Zweig listet unter anderem »Predictive Policing« und »Rückfälligkeitsalgorithmen«.[86] Auch Samerski und Henkel stellen den Zusammenhang zur Digitalisierung her: »Im Zeitalter von ›profiling‹ und ›big data‹ kann diese Analyse von responsibiliserenden Entscheidungen als paradigmatisch gelten: Bereits heute sind zahlreiche Risikoprognose-Programme und prädikative Tests auf

78 AlgorithmWatch 2019: 9 wie einleitend zitiert.

79 Seele 2020: 173.

80 Zur Unterscheidung von »automated decision-making« und »decision support« vgl. AlgorithmWatch 2019: 9.

81 Samerski und Henkel 2015: 85 und siehe oben (2.1).

82 Damit ist die Unterscheidung von »decision-making or decision support systems« aufgegriffen (AlgorithmWatch 2019: 9).

83 Für den Kulturbegriff in diesem Zusammenhang vgl. Stalder 2016.

84 AlgorithmWatch 2019: 82.

85 S. Anm. 1. AlgorithmWatch 2019: 89. Ebd. ist von »risk of hospitalization« die Rede und ebd. heißt es etwa: »During the experiment, the algorithm grouped the population according to four risk categories, allowing doctors to identifiy high risk patients, and to contact, advice, and/or treat them before their condition became critical.« (ebd.)

86 Zweig 2019: 5. Seele nennt auch den Finanzsektor als Beispiel und spricht dort von »Risikomodellen« (Seele 2020: 162).

dem Markt, vom Herz-Kreislauf-Risikorechner ›arriba‹ bis hin zu genetischen Tests oder gar der umfassenden Genom Analyse, die Patienten Risikoprofile zuweisen, zum Risikomanagement aufrufen und die Heraufkunft einer personalisierten Medizin verheißen.«[87]

Zweitens stellen erst Menschenbilder wie das des »risikoinformierten Entscheiders« den Plausibilitätshorizont dafür, die Intelligenz-Semantik auf Maschinen anzuwenden: Dieses Menschenbild markiert das »Spielfeld«, setzt den Maßstab, das tertium comparationis, durch das Mensch und Maschine plötzlich vergleichbar scheinen – nämlich in ihrer Kompetenz, risikoinformiert zu entscheiden. Auch das von Harari (nach)erzählte Übermenschlichkeitsnarrativ[88] funktioniert in diesem Horizont, weil darin Computer die besseren Datenverarbeitungssysteme als Menschen und somit auch die besseren Risikomanager sein werden.[89]

Der Grundgedanke in Nassehis »Muster« beinhaltet auch diese beiden Implikationen, insofern er expliziert, wie gegenwärtige Digitalisierung an frühere gesellschaftliche Entwicklungen, insbesondere der statistischen Beschreibung der Gesellschaft, anknüpfen und diese voraussetzen.[90] Ich habe sozusagen mit Samerski und Henkel imaginationstheoretisch ein Moment dessen konkretisiert, was bei Nassehi die »*gesellschaftliche[.] Struktur*« ist, mit der »*die Digitalisierung unmittelbar verwandt* ist«:[91] Gerade für sich als risikoinformierte Entscheider verstehende Menschen sind »Muster« und »*Regelmäßigkeiten*« interessant weil entscheidungsrelevant.[92] Anders als Nassehi und über diesen hinaus, geht es mir aber um die ethischen Ambivalenzen dieser vermeintlichen Kontinuität – und also um eine kritische Perspektive, die Nassehi allenfalls vorbereitet: Wenn Nassehi auf den Erweis dessen zielt, dass Digitalisierung für moderne Gesellschaften »kein Fremdkörper […], sondern, wenn man so will, Fleisch vom Fleische der Gesellschaft« ist,[93] bleiben die Ambivalenzen moderner Gesellschaften – digital oder nicht – tendenziell unterbetont. Genau um diese Ambivalenzen geht es mir aber in theologischer Anknüpfung an kritische Theorie.

Aus beiden Implikationen folgt: Was häufig als künstliche Intelligenz bezeichnet oder projektiert wird, imitiert nicht die Intelligenz, das Wesen oder

87 Samerski und Henkel 2015: 107.

88 Vgl. die Analyse von Meireis in der Einführung.

89 Vgl. Harari 2016, pointiert etwa ebd., 428.

90 Siehe dazu die Darstellungen oben in der Einleitung. Meine Auseinandersetzung mit Nassehis »Muster« verdankt sich auch den Gesprächen mit Torsten Meireis und der Diskussion in dessen Berliner Oberseminar.

91 Nassehi 2019: 18, kursiv im Original.

92 Für die Zitate Nassehi 2019: 28, kursiv im Original.

93 Nassehi 2019: 177.

die Natur des Menschen, sondern verkörpert als technisches Gebilde auch ein spezifisches Menschenbild, das in der Moderne des Westens kulturell gewachsen ist: das des risikoinformierten Entscheiders. Gleichzeitig ist zu erwarten, dass der entsprechende, praktische sinnhafte Gebrauch von ADM-Systemen genau dieses Menschenbild reproduziert und intensiviert.[94] Mit ADM-Systemen in denselben Praktiken involviert, die helfen zu entscheiden, wird es noch plausibler scheinen, sich und andere als risikoinformierte Entscheider:innen vorzustellen und als solche zu leben.

D Mensch-Imaginationen theologischer Diskurspraktiken

Nun ist dieses Menschenbild des risikoinformierten Entscheiders in evangelisch-ethischer Perspektive nicht an sich problematisch, wohl aber ambivalent: Sicher hängt dieses Menschenbild auch mit Praktiken zusammen, die mehr Selbstbestimmung etwa für Patient:innen ermöglichen – darauf verweisen Samerski und Henkel mehrfach zurecht.[95] Die grundsätzliche Ambivalenz dieses Menschenbildes wird vor dem Hintergrund theologischer Imaginationen deutlich. Im boulevardphilosophischen Jargon lässt sich dieses Problem auf die Formel bringen: Die Entscheidung kommt immer zu spät – und zwar auf zweifach spezifische Weise. Sie blendet tendenziell aus, was schon entschieden ist oder von anderer Seite noch entschieden wird und legt auf den Moment und Ort des Entscheidens und den damit gegebenen Horizont fest. Kurz gesagt: Die Entscheidung kommt zu spät, weil sie nicht mehr verändern kann, was vorher war und in die Entscheidungssituation geführt hat. Das impliziert eine reduktive Sicht auf das, was Menschen in ihrer Beziehungswirklichkeit sein können.

I Sach- und Wirklichkeitsgemäßheit

Da wir zumindest virtuell in der Forschungsstätte der Evangelischen Studiengemeinschaft zusammenkommen, will ich den ersten Punkt ausgehend von der Arbeit eines Theologen erklären, der an diesem Ort besondere Wirksamkeit hatte. Tödt hat mit seiner »Theorie sittlicher Urteilsfindung« die Sachmomente eines Verfahrens beschrieben, das auch auf Entscheidung zielt – auf

94 Genau das legt auch Seeles »These einer ›doppelten Konvergenz‹« nahe (Seele 2020: 131–172, Zitat auf S. 132). Dabei weist Seele auch darauf hin wie reduktiv das »digitale[.] Selbst« ist (Seele 2020: 153–155, 162): »Das analytische Modell des digitalen Selbst wird real, da der Mensch durch die Vermessung des Digitalen selber zum reduktionistischen, analytisch auslesbaren Modell wird.« (Seele 2020: 154)

95 Samerski und Henkel 2015: 86, 104, 106.

den sittlichen »Urteilsentscheid« nämlich – und insofern mit dem Verfahren des Managens »probabilisitische[r] Risiken« (s. o.) vergleichbar, aber doch grundsätzlich von diesem unterschieden ist.[96] Der hier wichtige Hauptunterschied zeigt sich darin, dass Tödt – vermutlich in Tradition zu Bonhoeffers Ethikfragment »Die Geschicht und das Gute II« – zwischen »Sachgemäßheit« und »Wirklichkeitsgemäßheit« eines sittlichen Urteils differenziert.[97]

Sachgemäßheit qualifiziert den Wahrheitsgrad der Repräsentation von Sachverhalten, Zweckmäßigkeiten und Zusammenhängen im Entscheidungskalkül.[98] Ethik gewinnt Sachgemäßheit im interdisziplinären Dialog mit empirischen Wissenschaften.[99] Die von Samerski und Henkel beschriebene probabilistische Risikoermittlung erscheint so als ein Verfahren der Sachgemäßheit. Es verhilft einer Entscheidung zu mehr Sachgemäßheit, sachlich zutreffend über die mit Verhaltensoptionen verbundene Eintrittswahrscheinlichkeit bestimmter Schäden zu wissen.

Von Sachgemäßheit unterscheidet Tödt nun die *Wirklichkeitsgemäßheit*, die die »umgreifende[.] Wirklichkeit« miteinbeziehen lässt:[100]

> »Das Hineinstellen der durch Reduktionen präzisierten Sachverhaltserfassung in den Gesamtzusammenhang der Wirklichkeit, in der zu leben Menschen bewußt ist, kann man als Aufgabe der praktische Philosophie bezeichnen. Die theologische Ethik verfährt ganz analog zur praktischen Philosophie, hat aber ihre Besonderheit darin, daß sie die Weltwirklichkeit im Licht des Glaubens an den sieht, der sich in Jesus Christus offenbart hat.«[101]

Während Sachgemäßheit immer auf die sachliche Trefflichkeit eines parzellierbaren Einzelaspektes fokussiert, öffnet Wirklichkeitsgemäßheit den Blick für den weiteren Zusammenhang.[102] Die bei Tödt hier implizite und bei Bonhoeffer zumindest etwas explizitere anthropologische Überzeugung dahinter ist diejenige, dass Menschen in ihren Beziehungen nicht nur zu sach-, sondern zu wirklichkeitsgemäßen Verhalten befreit (und zu befähigen) sind:[103] Menschen antworten handelnd auf Wirklichkeit, nicht bloß auf Sachen – so Grund-

96 Tödt 1988. Die ersten beide Zitate ebd., 41, kursiv im Original. Der Begriff des »Sachmoments« stammt hier ebenfalls von Tödt.

97 Vgl. dafür und für das Folgende Tödt 1979: 44–47, Zitate auf S. 46, im Original kursiv. Zu dieser Unterscheidung bei Bonhoeffers vgl. DBW 6: 260–275 (Bonhoeffer 1986–).

98 Tödt 1979: 44.

99 Tödt 1979: 45.

100 Tödt 1979: 46.

101 Tödt 1979: 46.

102 Tödt 1979: 45 f.

103 Tödt 1979: 40, 46 f., 50, 56.

überzeugung dieser responsorischen Anthropologie.[104] Demgegenüber reduziert die Imagination des Menschen als risikoinformierten Entscheider diesen auf dessen sachgemäßes Entscheiden. Das ist ambivalent, weil es einerseits tatsächlich die Sachgemäßheit erhöht, andererseits dies aber auf Kosten der Wirklichkeitsgemäßheit tut.

Dass dies nicht nur abstrakte Ambivalenz ist, lässt sich an einer friedensethischen Frage illustrieren, am Unterschied zwischen der Lehre vom gerechten Krieg und dem »Leitbild[.] vom gerechten Frieden«.[105] Lehren vom gerechten Krieg fragen immer nach der Legitimität des Einsatzes militärischer Gewalt.[106] Sie legen damit auf eine Entscheidungssituation fest, in der es nun vor allem um Fragen der Sachgemäßheit geht:[107] Hat der »Feind« ein Unrecht begangen, das einen legitimierenden Grund liefert? Sind die militärischen Mittel zweckmäßig und angemessen?[108] Fragen nach der umgreifenden Wirklichkeit rücken demgegenüber in den Hintergrund:[109] Welche anderen als militärischen Mittel könnten in der Situation helfen? Was hätte der in der Vergangenheit geholfen und könnte in der Zukunft langfristig dazu helfen, derartige Entscheidungssituationen erst gar nicht aufkommen zu lassen? Genau das sind die Fragen, die im »Leitbild des gerechten Friedens«[110], wie es auch von der Evangelischen Kirche in Deutschland (EKD) vertreten wird,[111] in den Blick rücken. Die Orientierung des Leitbildes lässt sich auf das Motto bringen: »[W]enn du den Frieden willst, bereite den Frieden vor.«[112] Das weitet die Frage nach dem Gewalteinsatz auf die Frage nach politischen, ökonomischen,

104 Für einen ähnlichen Gedanken bei Bonhoeffer vgl. DBW 6: 253 f. in Verbindung mit DBW 6: 228.

105 Kirchenamt der EKD 2007: 68; Huber 2012: 229–230. Huber spricht von »zwei gegenläufigen Modellen« (ebd., 230). Vgl. dort auch für das Folgende. Dazu, dass die Festlegung auf diese Frage reduktiv ist, vgl. auch Huber, a. a. O., 233.

106 Kirchenamt der EKD 2007: 65 f. Dort insbesondere: »Lehren vom ›gerechten Krieg‹ […] enthielten allgemeingültige Kriterien praktischer Vernunft, durch die geprüft werden sollte, ob in einer bestimmten Situation militärischer Gewaltgebrauch moralisch gerechtfertigt sein kann.« (ebd.) Vgl. auch Huber 2012: 229–231.

107 Vgl. für die folgenden Fragen in teilweise ähnlichen Worten Kirchenamt der EKD 2007: 66; Huber 2012: 231.

108 Vgl. für beide Fragen in teilweise ähnlichen Worten Kirchenamt der EKD 2007: 66; Huber 2012: 231.

109 Die folgenden Fragen richten sich auf das, was die EKD-Denkschrift im Horizont des »Leitbildes vom gerechten Frieden« als Bedingungen von »dauerhaftem Erfolg« der Friedenspolitik nennt: Kirchenamt der EKD 2007: 80–123, zweites Zitat auf S. 80. Zu »[g]ewaltfreie[n] Formen der Konfliktbearbeitung« und »zivile[r] Konfliktbearbeitung« etwa ebd., 109 f. und Huber 2012: 235.

110 Kirchenamt der EKD 2007: 57.

111 Kirchenamt der EKD 2007. Vgl. zu den Merkmalen des Leitbildes darin ebd., 50–56.

112 Kirchenamt der EKD 2007: 52.

sozialen, psychologischen Wirklichkeiten in der Hoffnungsperspektive des christlichen Glaubens hin.[113]

Noch einmal konkreter und auf die Diskussion um sogenannte KI, hier um sogenannten »autonome Waffensysteme« bezogen, die Nicole Kunkel viel plausibler »autoregulative Waffensysteme« zu nennen vorgeschlagen hat:[114] Wo diese Systeme darauf programmiert sind, potenzielle Ziele auszumachen oder gar ohne situative menschliche Freigabe anzugreifen, sind sie im Horizont des risikoinformierten Entscheiders als Entscheidungshilfe und Entscheider programmiert[115] – und genau in diesem Horizont auf situative Sachgemäßheit festgelegt. Schneller, präziser und unter Umständen weniger fehleranfällig berechnen sie das Risiko, dass eine auftauchende Gestalt ein Angreifer ist und reagieren entsprechend.[116] Insofern das in diesen Systemen präziser geschieht als ein menschlicher Entscheider es tun könnte, ermöglichen sie ein Mehr an situativer Sachgemäßheit, auf die es in besagtem imaginativen Horizont ja auch ankommt. Die umgreifende Wirklichkeit blenden diese Systeme situativ aus und ihr praktischer Einsatz erleichtert Menschen, die umgreifende Wirklichkeit auszublenden, weil beides in einem imaginativen Horizont geschieht, der Sachgemäß über Wirklichkeitsgemäßheit priorisiert. Zur Berücksichtigung dieser hier ausgeblendeten, umgreifenden Wirklichkeit würde etwa die Frage danach gehören, welche politischen, ökonomischen und sozialen Bedingungen den potenziellen Angreifer denn zu einem solchen gemacht haben und welche politischen Verhaltensoptionen diese Bedingungen so verändern, dass Angreifer gar nicht erst zu Angreifern werden. Die Entscheidung des risikoinformierten Entscheiders – sei es ein Mensch oder sei es ein autoregulatives Waffensystem – kommt immer schon zu spät. Sie ist auf einen situativen Mo-

113 Vgl. Kirchenamt der EKD 2007: 50–56. Konkret: »Friede ist kein Zustand (weder der bloßen Abwesenheit von Krieg, noch der Stillstellung aller Konflikte), sondern ein gesellschaftlicher Prozess abnehmender Gewalt und zunehmender Gerechtigkeit – letztere jetzt verstanden als politische und soziale Gerechtigkeit, d. h. als normatives Prinzip gesellschaftlicher Institutionen. Friedensfördernde Prozesse sind dadurch charakterisiert, dass sie in innerstaatlicher wie in zwischenstaatlicher Hinsicht auf die *Vermeidung von Gewaltanwendung*, die *Förderung von Freiheit und kultureller Vielfalt* sowie auf den *Abbau von Not* gerichtet sind. Friede erschöpft sich nicht in der Abwesenheit von Gewalt, sondern hat ein Zusammenleben in Gerechtigkeit zum Ziel.« (Kirchenamt der EKD 2007: 54, kursiv im Original.) Wolfgang Huber hat diese Blickweitung bzw. diesen Perspektivwechsel so auf den Punkt gebracht: »Während eine solche dreistufige Theorie des gerechten Krieges den Frieden vom Krieg her denkt, betrachtet die ethische Konzeption des ›gerechten Friedens‹ kriegerische Gewalt vom Frieden her.« (Huber 2012: 231, vgl. auch ebd., 230–237)

114 Vgl. Kunkel 2020: 15 f. Auch die jüngst erschienene Denkschrift der EKD zum Thema Digitalisierung rezipiert diesen Begriff von Kunkel: Kirchenamt der EKD 2021: 129, 133.

115 Zu dieser Definition dieser Waffensysteme nach dem US-Verteidigungsministerium: Kirchenamt der EKD 2021: 134, zum Zusammenhang von Entscheidungen und Waffensystemen Kirchenamt der EKD 2021: 134.

116 Zu diesem Vorteil teilw. auch: Kirchenamt der EKD 2021: 134–136.

ment und deren Horizont festgelegt, in dem nur noch zwischen Gewalt und Nicht-Gewalt entschieden werden kann.

II Planen und Hoffen

Das Menschenbild des risikoinformierten Entscheiders richtet Praxisteilnehmer:inner auf Zukunft aus, impliziert also eine anthropologische Bestimmung des Menschen auf Zukunft hin. In dem imaginativen Horizont dieses Menschenbildes ist die Zukunftsausrichtung aber eine spezifische. Eine Unterscheidung desjenigen Theologen, der cum grano salis als *der* theologische Zukunftsexperte schlechthin gelten könnte, hilft, dieses anthropologisch Spezifische genauer zu fassen. Moltmann hat mit den Begriffen »Planen« und »Hoffen« unterschiedliche Vergegenwärtigungen von oder Haltungen gegenüber der Zukunft unterschieden und dann in einen Zusammenhang gebracht.[117] *Planen* bezieht sich dabei auf Zukunft, sofern für diese disponiert werden kann, sofern Menschen mit »*deterministische[n]*« oder »*probabilistischen[n] Systeme[n]*« über diese Zukunft verfügen können.[118] Planen rechnet – so Moltmann – mit der Zukunft im Sinne von »futur«, indem es von vergangenen Erfahrungen extrapoliert und so Wahrscheinlichkeitsprognosen gewinnt.[119] Imaginativ schließt das grundsätzliche Paradigmenwechsel insofern methodisch aus, als es an Gegenwärtigem die Gesetzmäßigkeiten gewinnt, nach denen Zukünftiges entschieden wird.[120] Das schließt das Unverfügbare der Zukunft nicht aus dem Entscheidungskalkül aus,[121] macht es aber im Modus der Extrapolation vermeintlich kalkulierbar, behandelt das Unverfügbare also als Verfügbares.[122]

---

117 Vgl. Moltmann 1968. Von »vergegenwärtigen« in Bezug auf die Zukunft spricht Moltmann explizit selbst (ebd., 251), von Haltungen auch, allerdings da, wo es um eine nicht affirmierte Sache geht (ebd., 253). Zur »*Rückkopplung der Planung an die Hoffnung*«: ebd., 264, kursiv im Original.

118 Moltmann 1968: 251–253, Zitat auf S. 251 f., kursiv im Original. Mit Haseloff identifiziert Moltmann »Planung« mit »›Vorausdisposition für die Zukunft‹« (251.)

119 Moltmann 1974: 73. Vgl. dort etwa: »Die Futurologie und Planung verlängert auf diese Weise die Gegenwart in die Zukunft hinein. Zukunft ist hier das, womit die Gegenwart schon schwanger geht. Dieses Zukunftsdenken arbeitet mit der Methode der Extrapolation.« (ebd.)

120 So schreibt Moltmann über den Bezug zur Zukunft als »Futur«: »Das, was wird, kann man in seiner Möglichkeit und Wahrscheinlichkeit aus den Faktoren und Tendenzen der Gegenwart errechnen. Man sieht die bisherige Entwicklung und rechnet die Zuwachsraten der Industrie und das Wachstum der Bevölkerung für die nächsten 20 Jahre aus.« (Moltmann 1974: 73)

121 Moltmann 1968: 253.

122 Moltmann 1974: 73. Dass hier Unverfügbares als Verfügbares behandelt wird, drückt Moltmann auch in der Übernahme von Haseloffs Planungsverständnis als »Vorausdispositio-

Hoffen bezieht sich bei Moltmann auf Zukunft im Sinne von »adventus« und damit auf das, was uns aus der Zukunft unplanbar und unverfügbar entgegen kommt.[123] In der Glaubensperspektive Moltmanns ist das die Christuswirklichkeit, das Reich Gottes.[124] Beides ist uns Menschen sowohl praktisch handelnd als auch erkennend epistemisch unverfügbar:[125] Auch in der Perspektive des Glaubens können wir weder sicher wissen, wie Reich Gottes aussieht, noch können wir handelnd am Reich Gottes bauen[126] – »[d]as Kommen von Gottes Reich bleibt exklusive Tat Gottes«.[127] Das Erhoffte bleibt dem Menschen aber nicht äußerlich: Auf eine neue Zukunft zu hoffen, beinhaltet auch bei Moltmann, auf eine neue Zukunft für uns alle und für die Beziehungen zueinander, zu außermenschlichen Kreaturen, Dingen, zu uns selbst und zu Gott zu hoffen;[128] »die eschatologische Aussicht auf Versöhnung« meint »die Versöhnung der ganzen Kreatur« und führt zu »eine[r] Eschatologie aller Dinge«.[129] Es beinhaltet also die Hoffnung, dass uns und diesen Beziehungen grundsätzliche, nämlich versöhnende und erlösende Veränderungen widerfahren. Damit sind Menschen und ihre Beziehungen zunächst in Glaubensperspektive in einer Art und Weise als zukunftsoffen vorgestellt,[130] die der Haltung des Planens und Extrapolierens verschlossen bleibt.

Planen und Extrapolieren ist einerseits sozialethisch unerlässlich, weil es erst die sachgemäße Entwicklung von Verhaltensoptionen für Politik und Individuen ermöglicht.[131] Steht die Haltung des Planens aber allein, ohne im Hoffen ein Komplement zu finden, bleibt es imaginär somit bei Menschenbildern des Planens, dann wird diese Haltung dazu tendieren, Menschen an-

---

nen für die Zukunft« insofern aus, als Disponieren ein Akt des Verfügens ist (Moltmann 1968: 251).

123 Moltmann 1974: 73 f., 1968: 252 f. Vgl. dort etwa: »Hoffnung bezieht sich dann weniger auf die aus eigener Macht zur Verfügung stehende Zukunft als viel mehr auf jene Zukunft, die mir ein anderer und in der ein anderer sich mir durch Versprechen zur Verfügung stellt.« (Moltmann 1968: 253)

124 Moltmann 1968: 255, 265, 1966.

125 Meireis 2008: 259, 261.

126 Meireis 2008: 259–261.

127 Tödt 1967: 198.

128 Moltmann 1966: 17, 203–209. Vgl. Moltmann selbst auch für dieses Sprachspiel von der »Zukunft für Mensch und Welt« (Moltmann 1966: 196). Vgl. ausführlicher insgesamt dazu, insbesondere auch zur Relationalität und kosmologischen Ausrichtung, Höhne 2015: 50–52, 72–77 und die dort zitierte Literatur.

129 Moltmann 1966: 203.

130 Zu dieser Offenheit vgl. schon ausführlicher Höhne 2015: 68–72, 118 f. und die dort zitierte Literatur. Den Begriff der Zukunftsoffenheit verwendet Moltmann selbst Moltmann 1966: 263.

131 Zum Verhältnis von Planen und Hoffen und implizit damit zur Berechtigung des Planens vgl. auch Moltmann 1968: 251 f., 264 f., 1974: 74.

thropologisch auf das festzulegen, was sie schon immer und bisher waren.[132] Demgegenüber halten Menschenbilder der Hoffnung Menschen imaginativ in einem tieferen Sinne und grundsätzlicheren Sinne zukunfts- und veränderungsoffen.[133]

Praktiken, in denen ADM-System als technische Artefakte sinnhaft gebraucht werden, inszenieren das Menschenbild des risikoinformierten Entscheiders als imaginären Horizont und reproduzieren damit eine Haltung des Planens, die vom Hoffen abstrahiert bleibt. Die Kosten dieser Reduktion im Imaginären auf risikoinformiertes Entscheiden fallen in den entsprechende Praktiken vor allem auf der Objektposition[134] an, was gerade im Licht an einer »Orientierung an der Perspektive der Rechtlosen« (Meireis) ethisch problematisch ist.[135]

Am Beispiel konkretisiert: Das Scoring-Verfahren der SCHUFA (s. 2.2) berechnet die Kreditwürdigkeit einer Person.[136] Das ermöglicht potentiellen Kreditgebern oder Vermieterinnen eine risikoinformiertes Entscheiden.[137] Grundlage der Berechnung sind Daten über das bisherige Verhaltens der Person.[138] Insofern reproduzieren diese vom Menschenbild risikoinformierten Entscheidens durchwirkten Praktiken eine Haltung des Planens, die Haltungen der Hoffnung ausschließen. Das ist insofern problematisch als der Objektposition in der Praktik der Kreditvergabe – also den möglichen Kreditnehmer:innen – eine tiefere Zukunfts- und Veränderungsoffenheit nicht zugestanden wird. Vielmehr wird er oder sie festgelegt auf das, was sie oder er bislang war. Genau diese imaginative Festlegung reproduziert Veränderungsverschlossenheit aber praktisch. Wer aufgrund bisherigen Verhaltens einen

---

132 Zur Gefahr der Reduktion auf Planen vgl. auch: »Die Planung muß um ihren Ur-sprung aus der Hoffnung und um den Vor-sprung der Hoffnung wissen. Setzt sie sich selbst an die Stelle der Hoffnung, so verliert sie den transzendenten Impetus der Hoffnung und verliert zuletzt auch sich selbst.« (Moltmann 1968: 265)

133 Höhne 2019a: 39–42, 2015: 68–72 und die dort zitierte Literatur.

134 Zu den Begriffen Subjekt- und »Objektposition« vgl. Vogelmann 2014: 125 f. Zu den Kosten in Bezug auf ADM-Systeme konkret vgl. etwa Zweig: »Die Kosten von Fehlentscheidungen trägt hier nicht nur der ADM-nutzende Akteur, sondern auch die Person, über die entschieden wird.« (Zweig 2019: 5)

135 Vgl. zur angesprochenen Problematik algorithmischer Entscheidungen Meireis 2019: 55. Zur »Orientierung an der Perspektive der Rechtlosen« vgl. Meireis 2016: 41 f., Zitat auf S. 42.

136 Vgl. dafür und für die Informationen im Folgenden AlgorithmWatch 2019: 82 und die Informationen auf www.meineschufa.de, insbes. https://www.schufa.de/ueber-uns/unternehmen/so-funktioniert-schufa/ [Abruf am 2. 6. 2021]. Zur algorithmenbasierten Bonitätseinschätzung und deren Problematik vgl. auch Meireis 2019: 55; Nassehi 2019: 224.

137 AlgorithmWatch 2019: 82.

138 Vgl. für die Datengrundlage der Schufa deren eigene Darstellung unter https://www.meineschufa.de/aktion/faq-daten [Abruf am 2. 6. 2021].

Kredit nicht erhält, kann auch künftig keinen anderen Umgang mit Geld finden oder unter Beweis stellen.

E Schluss

Ich fasse den Gang der Argumentation in drei kurzen Punkten zusammen:

1. Es ist erstens produktiv und sachgerecht Praktiken, die digitalen Technologie gebrauchen, vermittels der Kategorien »Praxis« und »Imaginäres« sozialethisch zu reflektieren. Auf sogenannte KI bezogen, rückt das Praktiken in den Fokus, in denen Menschen ADM-Systeme gebrauchen, und lässt nach dem Imaginären fragen, das diese Praktiken möglich und plausibel machte, das in ihnen besteht und reproduziert wird.
2. Das ermöglicht eine *Beschreibungsthese*: Ein in diesen Praktiken wichtiges Menschenbild ist das des risikoinformierten Entscheiders (Samerski und Henkel): Was häufig als künstliche Intelligenz bezeichnet oder projektiert wird, imitiert nicht die Intelligenz, das Wesen oder die Natur des Menschen, sondern verkörpert als technisches Gebilde auch ein spezifisches Menschenbild, das in der Moderne des Westens kulturell gewachsen ist: das des risikoinformierten Entscheiders. Gleichzeitig ist zu erwarten, dass der entsprechende, praktische sinnhafte Gebrauch von ADM-Systemen genau dieses Menschenbild reproduziert und intensiviert.
3. Und das ist in theologisch-sozialethischer Perspektiv – so die *Orientierungsthese* – insofern problematisch, als dieses Menschenbild hochgradig reduktiv ist. Es legt besagte Praktiken stärker auf Sachgemäßheit als auf Wirklichkeitsgemäßheit fest. Es reproduziert den Menschen als planendes und nicht als hoffendes Subjekt. Beides lässt Menschen in diesen Praktiken hinter ihrer Freiheit zurückbleiben – besonders Menschen auf den Objektpositionen dieser Praktiken. Demgegenüber können die in theologisch-sozialethischen Diskursen virulenten Menschenbilder den Zukunftshorizont weiten.

Literatur

AlgorithmWatch (Hg.) 2019: Automating Society. Taking Stock of Automated Decision Making in the EU. Berlin.

Bonhoeffer, Dietrich 1986–: Werke (DBW). Sondersausgabe 2015. 17 Bände. Hg. v. Eberhard Bethge, Ernst Feil, Christian Gremmels, Wolfgang Huber, Hans Pfeifer, Albrecht Schönherr und Heinz Eduard Tödt. Gütersloh: Gütersloher Verl.-Haus.

Bourdieu, Pierre 1993: Sozialer Sinn. Kritik der theoretischen Vernunft (Suhrkamp-Taschenbuch Wissenschaft, 1066). 1. Aufl. Frankfurt am Main.

Bourdieu, Pierre 2014: Die feinen Unterschiede. Kritik der gesellschaftlichen Urteilskraft (Suhrkamp-Taschenbuch Wissenschaft, 658). Unter Mitarbeit von Bernd Schwibs und Achim Russer. 24. Aufl. Frankfurt am Main.

Bourdieu, Pierre 2015: Entwurf einer Theorie der Praxis. Auf der ethnologischen Grundlage der kabylischen Gesellschaft (Suhrkamp-Taschenbuch Wissenschaft, 291). 4. Aufl. Frankfurt am Main.

Harari, Yuval N. 2016: Homo deus. A brief history of tomorrow. Revised edition. London.

Hillebrandt, Frank 2014: Soziologische Praxistheorien. Eine Einführung (Soziologische Theorie). Wiesbaden.

Höhne, Florian 2015: Einer und alle. Personalisierung in den Medien als Herausforderung für eine Öffentliche Theologie der Kirche (Öffentliche Theologie, 32). Leipzig.

Höhne, Florian 2019a: Darf ich vorstellen: Digitalisierung. Anmekrungen zu Narrativen und Imaginationen digitaler Kulturpraktiken in theologisch-ethischer Perspektive. In: Jonas Bedford-Strohm, Florian Höhne und Julian Zeyher-Quattlender (Hgg.): Digitaler Strukturwandel der Öffentlichkeit. Interdisziplinäre Perspektiven auf politische Partizipation im Wandel (Kommunikations- und Medienethik, 10). 1. Aufl. Baden-Baden, 25–46.

Höhne, Florian 2019b: »Öffentlichkeit« als Imagination und Ensemble sozialer Praktiken. Zur Relevanz einer Schlüsselkategorie Öffentlicher Theologie in digitalen Kontexten. In: Ethik und Gesellschaft (1), 1–31. DOI: 10.18156/EUG-1-2019-ART-1.

Huber, Wolfgang 2012: Legitimes Recht und legitime Rechtsgewalt in theologischer Perspektive. In: Torsten Meireis (Hg.): Gewalt und Gewalten. Zur Ausübung, Legitimität und Ambivalenz rechtserhaltender Gewalt. Tübingen, 225–242.

Kirchenamt der EKD 2007: Aus Gottes Frieden leben – für gerechten Frieden sorgen. Eine Denkschrift des Rates der Evangelischen Kirche in Deutschland. 1. Aufl. Gütersloh.

Kirchenamt der EKD (Hg.) 2021: Freiheit digital. Die Zehn Gebote in Zeiten des digitalen Wandels. Eine Denkschrift der Evangelischen Kirche in Deutschland. 1. Aufl. Leipzig.

Kunkel, Nicole 2020: Mensch und hochautomatisierte Maschine – eine theologische Verortung. In: Aufschlüsse. Zeitschrift für spirituelle Impulse 78, 13–16.

Meireis, Torsten 2008: Tätigkeit und Erfüllung. Protestantische Ethik im Umbruch der Arbeitsgesellschaft. Tübingen.

Meireis, Torsten 2016: Schöpfung und Transformation. Nachhaltigkeit in protestantischer Perspektive. In: Torsten Meireis und Stefan Böschen (Hgg.): Nachhaltigkeit (Jahrbuch Sozialer Protestantismus, Band 9). 1. Aufl. Gütersloh, 15–50.

Meireis, Torsten 2019: »O daß ich tausend Zungen hätte«. Chancen und Gefahren der digitalen Transformation politischer Öffentlichkeit – die Perspektive evangelischer Theologie. In: Jonas Bedford-Strohm, Florian Höhne und Julian Zeyher-Quattlender (Hgg.): Digitaler Strukturwandel der Öffentlichkeit. Interdisziplinäre Perspektiven auf politische Partizipation im Wandel (Kommunikations- und Medienethik, 10). 1. Aufl. Baden-Baden, 47–62.

Mirjam Hauck 2021: Mensch, ärgere dich nicht. 25 Jahre »Deep Blue«. sueddeutsche.de. Online verfügbar unter https://www.sueddeutsche.de/digital/schach-deep-blue-kasparow-ibm-1.5200655, zuletzt aktualisiert am 10. 02. 2021, zuletzt geprüft am 02. 06. 2021.

Moltmann, Jürgen 1966: Theologie der Hoffnung. Untersuchungen zur Begründung und zu den Konsequenzen einer christlichen Eschatologie (Beiträge zur evangelischen Theologie, 38). 6. Aufl. München.

Moltmann, Jürgen 1968: Hoffnung und Planung. In: Jürgen Moltmann (Hg.): Perspektiven der Theologie. Gesammelte Aufsätze. München, Mainz, 251–268.

Moltmann, Jürgen 1974: Das Experiment Hoffnung. Einführungen. München.

Nassehi, Armin 2019: Muster. Theorie der digitalen Gesellschaft. München.

Neuberger, Christoph 2009: Internet, Journalismus und Öffentlichkeit. Analyse des Medienumbruchs. In: Christoph Neuberger, Christian Nuernbergk und Melanie Rischke (Hgg.): Journalismus im Internet. Profession – Partizipation – Technisierung. Wiesbaden, 19–105.

Paßmann, Johannes 2018: Die soziale Logik des Likes: Eine Twitter-Ethnografie. Frankfurt am Main, New York: Campus Verlag.

Polanyi, Michael 1962: Tacit Knowing: Its Bearing on Some Problems of Philosophy. In: Reviews of Modern Physics 34 (4), 601–616.

Reckwitz, Andreas 2003: Grundelemente einer Theorie sozialer Praktiken. Eine sozialtheoretische Perspektive. In: Zeitschrift für Soziologie 32 (4), 282–301.

Samerski, Silja; Henkel, Anna 2015: Responsibilisierende Entscheidungen. Strategien und Paradoxien des sozialen Umgangs mit probabilistischen Risiken am Beispiel der Medizin. In: Berliner Journal für Soziologie 25, 83–110.
Schatzki, Theodore R. 2008: Social Practices. A Wittgensteinian Approach to Human Activity and the Social. Digitally printed version, paperback reissue. Cambridge.
Schmidt, Jan-Hinrik 2011: Das neue Netz. Merkmale, Praktiken und Folgen des Web 2.0 (Kommunikationswissenschaft). 2., überarb. Aufl. Konstanz.
Schmidt, Robert 2012: Soziologie der Praktiken. Konzeptionelle Studien und empirische Analysen (Suhrkamp Taschenbuch Wissenschaft, 2030). Orig.-Ausg., 1. Aufl. Berlin.
Seele, Peter 2020: Künstliche Intelligenz und Maschinisierung des Menschen (Schriften zur Rettung des öffentlichen Diskurses). Köln.
Stalder, Felix 2016: Kultur der Digitalität (Edition Suhrkamp, 2679). Berlin.
Taylor, Charles 2004: Modern Social Imaginaries (Public planet books). Durham.
Tödt, Heinz Eduard 1967: Aus einem Brief an Jürgen Moltmann. In: Wolf-Dieter Marsch (Hg.): Diskussionen über die »Theologie der Hoffnung« von Jürgen Moltmann. München, 197–200.
Tödt, Heinz Eduard 1979: Kriterien evangelisch-ethischer Urteilsfindung. Grundsätzliche Überlegungen angesichts der Stellungnahmen der Kirchen zu einem Kernkraftwerk in Wyhl am Oberrhein. In: Heinz Eduard Tödt (Hg.): Der Spielraum des Menschen. Theologische Orientierung in den Umstellungskrisen der modernen Welt (Gütersloher Taschenbücher Siebenstern, 337). Gütersloh, 31–80.
Tödt, Heinz Eduard 1988: Versuch einer ethischen Theorie sittlicher Urteilsfindung [1979]. In: Heinz Eduard Tödt (Hg.): Perspektiven theologischer Ethik. München, 21–48.
Vogelmann, Frieder 2014: Im Bann der Verantwortung (Frankfurter Beiträge zur Soziologie und Sozialphilosophie, 20). Frankfurt.
Zweig, Katharina A. 2019: Algorithmische Entscheidungen: Transparenz und Kontrolle (Analysen & Argumente, 338). Berlin.

ORCID

Florian Höhne https://orcid.org/0000-0001-6589-2124

Die ethische Relevanz von KI-Diskursen

Das Verhältnis von Diskursanalyse und Angewandter Ethik im Feld der Künstlichen Intelligenz

Alexander Filipović & Julian Lamers

A Einleitung

Jede neue technische oder wissenschaftliche Innovation wird von einer kritischen Debatte begleitet. Vor dem Hintergrund eines komplexen Normgeflechts wird die Frage nach den mit dieser Innovation verbundenen Herausforderungen für die Gesellschaft und ihre Werte aufgeworfen. Geht mit einer solchen Innovation die Unsicherheit über die möglichen Auswirkungen auf den Menschen oder seine Umwelt und damit verbunden auch auf das gesellschaftliche Werteverständnis einher, so birgt sie ein Polarisierungspotenzial: Akteure treten auf den Plan, die die Debatte nach den ihnen jeweils eigenen moralischen Anschauungen zu prägen versuchen und diese damit politisieren. Das Auftreten von moralisierten und politisierten Debatten ist beständig zu beobachten, in der Vergangenheit beispielsweise im Kontext der Nanotechnik oder der Humangenetik.

Aktuell bildet sich mit der Entwicklung im Bereich der Künstlichen Intelligenz (KI) ein neuer politisierter, technikkritischer oder -euphorischer Diskurs. Was wir hier »Diskurs« oder »Debatte« nennen, zielt auf Inhalte und Prozesse öffentlicher Kommunikation zum Thema ab. Darstellende, informierende, bewertende, kommentierende, wissenschaftliche, fiktionale etc. Beiträge in Medien, Parlamenten, Foren, auf Kongressen etc., die manchmal aufeinander Bezug nehmen und die thematisch dem Feld »Künstliche Intelligenz« zuzuordnen sind, stellen insgesamt den zeitgenössischen Diskurs zur Künstlichen Intelligenz dar.

Alexander Filipović/Julian Lamers 2022: Die ethische Relevanz von KI-Diskursen. Das Verhältnis von Diskursanalyse und Angewandter Ethik im Feld der Künstlichen Intelligenz. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 137–154. https://doi.org/10.11588/heibooks.1106.c 15197

Unsere grundlegende These ist, dass dieser KI-Diskurs eine Relevanz für die KI-Ethik hat. Diese These versteht sich nicht von selbst, da man die Ansicht vertreten kann, dass es allein die KI-Technologien selbst sind (und nicht die Diskurse darüber), die ethisch in den Blick genommen werden sollten. Der Begriff der Relevanz ist natürlich vage, aber absichtlich so gewählt, um ganz verschiedene Zusammenhänge zwischen dem Diskurs über KI und einer Ethik der KI konzeptionell einzufangen.

Wir wollen in unserem Beitrag die These von der Relevanz der KI-Diskurse für die Ethik plausibilisieren und systematisieren. Dazu geben wir zunächst (B) eine kurze Skizze des Diskurses über KI. Danach (C) betrachten wir diese Diskurse als moralische bzw. ethische Diskurse, wobei sich anfänglich die Relevanz dieser KI-Diskurse für die Ethik zeigt. Die folgende Beschreibung (D) der Basisformen im Verhältnis von KI-Diskursen und KI-Ethik versucht das Relationengefüge zu ordnen. Der letzte Abschnitt (E) bindet die Erkenntnisse ein in theoretische und methodische Fragen der KI-Ethik als eine angewandte bzw. bereichsspezifische Ethik.

B Diskurs- und Bedeutungswandel: KI in der Öffentlichkeit

In Hinblick auf den Bedeutungswandel des Themas KI hat sich in den vergangenen Jahren in der bundesdeutschen Öffentlichkeit einiges getan. Dies lässt sich unter anderem auch an der gewachsenen Präsenz des Themas in der Politik beobachten: Tauchten die Begriffe »Künstliche Intelligenz« beziehungsweise »KI« im Jahr 2017 gerade vier Mal in den Plenarprotokollen des Deutschen Bundestags auf, so finden sich die Begriffe in den Protokollen aus dem Jahr 2018 bereits weit über 400 Mal. Doch auch in der gesamtgesellschaftlichen Öffentlichkeit trifft die neue Technologie auf erhöhte Aufmerksamkeit: Fand der Begriff »Künstliche Intelligenz« bis vor nicht allzu langer Zeit abseits von Science-Fiction-Szenarien und spekulativen Prognosen in der Öffentlichkeit kaum Beachtung, so ist in den vergangenen zwei bis drei Jahren die KI-Debatte in der deutschen Öffentlichkeit regelrecht explodiert. Kontinuierlich greifen Beiträge diverser Medienformate Themen mit Bezug zur Digitalisierungsdebatte, der autonomen Mobilität, lernenden Systemen, Bilderkennungsprogrammen oder Startup-Unternehmen, welche mit KI-Programmen arbeiten, aus unterschiedlichen Perspektiven auf. Die thematischen Kontexte der Auseinandersetzung mit dem Thema KI sind dabei so vielfältig wie die gesellschaftliche Konnotation der Technologie: Während der kritische Blick auf den staatlichen Einsatz von Gesichtserkennungssoftware (zuletzt insbesondere mit Blick auf Chinas Massenüberwachungsprogramm *Skynet*) Orwell'sche

Assoziationen hervorruft, wird KI an anderer Stelle als ultimativer Problemlöser verstanden.

Ebenso wie Bedeutung und Gewicht des Themas in der öffentlichen Debatte veränderte sich auch die öffentliche Rezeption von KI-Technologien. So zeigen die Ergebnisse einer Umfrage von Bitkom Research aus dem Jahr 2020, dass 68 % der Deutschen Künstliche Intelligenz als Chance sehen.[1] 2021 ist diese Zahl laut Bitcom auf 72 % angewachsen, im Jahr 2017 lag dieser Wert bei 48 %. Auch wenn Bitcom ein Branchenverband ist und Interesse daran hat, dass die deutsche Bevölkerung ihre Angst vor der Technologie verliert, zeigen die Daten ein Schwinden der anfänglichen Abwehrhaltung, die das Thema KI zunächst charakterisiert hatte.

Das Thema KI ist jedenfalls in der gesellschaftlichen Öffentlichkeit – wie es zudem scheint recht abrupt – angekommen und spricht eine Vielzahl von thematischen Bezügen an. Hierin liegt zugleich auch ein Problem der KI-Debatte: Dadurch, dass die Technologie ein so breites Spektrum an Anwendungsmöglichkeiten mit sich bringt, wirft sie in ganz unterschiedlichen Kontexten ganz unterschiedliche Fragestellungen auf. So spricht die Debatte darüber, ob Maschinen in der Lage sein dürfen, selbstständig ethisch relevante Entscheidungen zu treffen, ganz andere Fragen an als die Debatte über den Einsatz von intelligenter Überwachungssoftware oder über die Fähigkeit von Programmen, menschliche Gedanken zu lesen, eigenständig willkürliche Nachrichtenmeldungen zu verfassen oder menschliche Interaktion in sozialen Netzwerken zu imitieren. Dadurch verfügt der KI-Begriff über eine inhaltliche Komplexität, die sich in der medialen Debatte nur schwerlich zu einem gesamtgesellschaftlichen Diskurs herunterbrechen lässt.

Es gibt nach unserer Kenntnis wenig Inhalts- oder Diskursanalysen, die das Thema Künstliche Intelligenz behandeln. Drei Studien seien mit ihren Ergebnissen hier kurz genannt: In der inhaltsanalytischen Studie *An Industry-Led Debate: How UK Media Cover Artificial Intelligence* im Jahr 2018 wurden 760 Zeitungs- und Onlineartikel aus sechs verschiedenen Medienunternehmen mit Bezug zu KI-Themen in einem Zeitraum von acht Monaten untersucht.[2] Die Studie kommt zu dem Ergebnis, dass 60 % der Nachrichtenartikel sich auf Produkte, Initiativen oder Ankündigungen der Industrie beziehen. Ein Drittel der eindeutigen Quellen über alle Artikel hinweg sind mit der Industrie verbunden, fast doppelt so viele wie mit der Wissenschaft und sechsmal so viele wie

1 Die Bitcom-Daten sind nicht als separate Studien veröffentlicht, sondern in Form von Pressemeldungen mit Charts publiziert. Vgl. zur Umfrage von 2020 https://www.bitkom.org/Presse/Presseinformation/Die-Menschen-wollen-KI-und-haben-auch-Angst-vor-ihr, die Daten von 2021 unter https://www.bitkom.org/Presse/Presseinformation/Kuenstliche-Intelligenz-als-Chance.

2 Brennen/Howard/Nielsen 2018.

mit der Regierung. Medien thematisieren stark den Einfluss dieser Technologie auf alle Bereiche des Lebens, indem sie KI als relevante und kompetente Lösung für eine Reihe öffentlicher Probleme darstellen. Dabei werden, so die Autoren, die laufenden Debatten über die potenziellen Auswirkungen der KI oft kaum zur Kenntnis genommen. Brennen et. al analysieren darüber hinaus eine starke Politisierung des Diskurses durch die selektiven Hervorhebungen. Als aufkommendes öffentliches Thema wird die KI durch die Themen, die die Medien in ihrer Berichterstattung hervorheben, politisiert: Eher rechtsgerichtete Zeitungen heben wirtschaftliche und geopolitische Themen hervor, darunter Automatisierung, nationale Sicherheit und Investitionen. Linke Medien betonen ethische Aspekte der KI, darunter Diskriminierung, algorithmischer Bias und informationelle Selbstbestimmung/Datenschutz.[3]

Fischer und Puschmann analysieren Leitmedien, Fachwebseiten und Twitter zwischen 2005 und 2020 und verfolgen das Ziel, »die Berichterstattung zum Thema sowie ihre Entwicklung in den letzten 15 Jahren nachzuzeichnen«[4]. Sie heben als Ergebnisse hervor, dass es der Debatte an Vielfalt mangelt und blinde Flecken hinsichtlich gemeinwohlrelevanter Debatten erkennbar sind. Es dominiere die wirtschaftliche-chancenorientierte Perspektive. Politische und zivilgesellschaftliche Akteure fänden nur selten Erwähnung. Allerdings werden, so die Autorin und der Autor, auch Problembereiche genannt. Hier wäre ein »erstaunlich« lösungsorientierter Diskurs erkennbar: »Ein Drittel der untersuchten Texte enthält spezifische Handlungsempfehlungen. Allerdings dominieren dabei solche Empfehlungen, die auf den Kompetenzaufbau bei Anwender:innen und in der Bevölkerung abzielen, während konkrete Ansätze zur wirksamen Aufsicht, Kontrolle und Regulierung deutlich seltener zu finden sind.«[5]

Fink[6] schließlich analysiert auf der Basis von Foucaults Diskursverständnis und der kritischen Diskursanalyse nach Jäger/Jäger[7] die Ebenen Wissenschaft, Wirtschaft und Medien. Sie hebt deren jeweilige Eigenlogik hervor und betont die Protagonistenrolle der Wissenschaft in dem Diskurs: »Die Analyse zeigte, dass die Wissenschaft ein Konstrukt von Künstlicher Intelligenz formt, auf das sich sowohl Unternehmen als auch Medien beziehen und sie daher als Protagonist des Diskurses beschrieben werden kann. Das bedeutet, dass auf

3 Brennen/Howard/Nielsen 2018: 1.

4 Fischer/Puschmann 2021: 8.

5 Fischer/Puschmann 2021: 9.

6 Fink 2020.

7 Jäger/Jäger 2007; Jäger 2015.

dieser Ebene darüber entschieden wird, welche Aussagen über KI sagbar bzw. nicht sagbar sind.«[8] Zum medialen Diskurs hält sie fest:

> »Medien selektieren die möglichen Aussagen und bestimmten über ihr Auftreten im öffentlichen Diskurs. Durch die Auswahl, die Organisation und die Präsentation der Aussagen machen Medien aus der gesellschaftlichen Debatte über Künstliche Intelligenz ein gesellschaftliches Problem. Das liegt an dem paradoxen Verhältnis zwischen Faszination und Bedrohung, welches die Medien gegenüber der Technologie zeigen und das zu einer verzerrten Wahrnehmung von KI in der Öffentlichkeit führt. Außerdem wird über die Absicht oder das Potenzial hinter einem KI-System statt über seine konkrete und aktuelle Funktionalität berichtet, wodurch den Systemen mehr zugetraut wird als sie tatsächlich können. Die starke Fokussierung auf Konkurrenz- und Angstthemen, sowie die Politisierung des Themas führen außerdem dazu, dass KI unkontrollierbar und nicht nachvollziehbar zu sein scheint. Medien verleihen der KI große Relevanz, übertriebene Kompetenz und beinahe vollständige Autonomie. Das macht die Technologie zu einem gesellschaftlichen Störfaktor und Problem.«[9]

Aus dieser Betrachtung lassen sich für die weitere Argumentation drei wesentliche Punkte festmachen: Erstens ist die Debatte fragmentiert und kann als eine Vielfalt verschiedener Betrachtungen zu einzelnen Themenfacetten eines als hochrelevant empfundenen, übergeordneten Themenfeldes verstanden werden. Der Diskurs ist reichhaltig und lebhaft und es besteht in der Debatte kein Zweifel, dass die Technologie große Auswirkungen auf unser Leben haben wird. Zweitens zeigt sich ein Konstruktcharakter, eine Konstruktivität der Diskurse, die je von den Akteuren bestimmt wird. Ergebnisse der gesellschaftlichen KI-Diskurse sind diskursive Konstrukte von Künstlicher Intelligenz, die von Unternehmen, Wissenschaftler:innen, Intellektuellen, Journalist:innen etc. geformt werden. Die Medialität der Diskurse und konkret die politische Ausrichtung von Zeitungen und Sendern entscheiden deutlich mit über die Gestalt öffentlicher KI-Konstrukte. Drittens schließlich wird deutlich, dass besonders der mediale Diskurs in die Binarität von »Chancen und Grenzen« sortiert ist und entsprechende Wertperspektiven und Gestaltungsherausforderungen thematisiert; auch dieser Diskurs ist also in dem Sinne ein ethischer Diskurs, indem oft die Frage mitschwingt, was zu tun sei angesichts der sich entwickelnden Technologie.

8 Fink 2020: 85.

9 Fink 2020: 86.

C Technikdiskurse als normative Debatten

Bei der Betrachtung der medialen Berichterstattung über KI kann festgestellt werden, dass federführende Akteure in der öffentlichen Debatte in Bezug auf spezifische Aspekte der Thematik moralische Fragestellungen aufgreifen und Konfliktfelder um normative Unsicherheiten ausnutzen, um ihrer eigenen Position im Diskurs Geltung zu verleihen. In der Debatte um KI ergibt sich eine solche normative Unsicherheit insbesondere aus dem Aufeinandertreffen konfligierender Wertevorstellungen und Bedürfnisse. Folgende Beispiele normativer Konflikte erscheinen uns besonders einschlägig[10]:

- der Wunsch nach der Befähigung von Algorithmen, nach bestimmten Kriterien präzise und faktenbasierte Beurteilungen zu liefern versus dem Wunsch nach fairer und gleicher Behandlung (etwa bei der automatisierten Beurteilung potenzieller Arbeitnehmer und der damit verbundenen Gefahr von Diskriminierung, wenn bestimmte Gruppen infolge dieser generalisierten Beurteilung automatisch durchs Raster fallen);
- der Wunsch nach optimierten und effizienteren Dienstleistungen und dadurch gesteigerter Lebensqualität versus dem Bedürfnis nach Privatsphäre und Autonomie des Individuums;
- die Möglichkeit, Abläufe zu automatisieren, um die Effizienz bestimmter Leistungen und damit verbunden die Lebensqualität vieler Menschen zu erhöhen zu Lasten derer, welche infolge dieser Automatisierung ihren Beruf verlieren.

Eine nicht unwesentliche Rolle in den Diskursen nehmen Fachgremien ein, welche diese einzelnen Kontroversen der KI-Technologie unter technischen, sozialen, politischen, juristischen und auch ethischen Gesichtspunkten reflektieren und deliberativ einer Lösung zuführen möchten. Alexander Bogner spricht im Falle von wissenschaftsnah angelegten Fachgremien auch eher von einer Ethisierung der Konflikte und bezeichnet diese verständigungsorientierte Herangehensweise im Vergleich zur Moralisierung (als einer streng wahrheitsorientierten) Herangehensweise als Regelfall im öffentlichen Diskurs zu Technikkontroversen in liberalen, pluralistischen Gesellschaften.[11] Oft genug scheinen aber auch Moralisierungstendenzen bei der Auseinandersetzung mit neuen Technologien eine Rolle zu spielen – nämlich dann, wenn einerseits verschiedene gesellschaftliche Akteure ein Interesse daran haben, die Debatte in einer bestimmten Weise zu prägen und andererseits, wenn sich – angesichts

10 Whittlestone/Nyrup/Alexandrova/Dihal/Cave 2019.

11 Bogner 2013: 54 f.

der Überkomplexität des Themas – die Auseinandersetzung an bestehenden (zum Beispiel sozioökonomischen und wertebezogenen) Kontroversen vergleichend orientiert.[12] Hier nun stellt sich allerdings die Frage, wie groß der Einfluss solcher Moralisierungstendenzen in der gesamtgesellschaftlichen Debatte letzten Endes ist und damit wissenschaftsorientierte, fallibilistisch angelegte Verständigung (die freilich auch auf der Suche nach »Wahrheit« ist) an den Rand gedrängt wird.

Es ist naheliegend anzunehmen, dass Bezüge zu ethischen Fragestellungen in Technikkontroversen mitunter häufig den Versuch bestimmter Akteure darstellen, den jeweiligen Technologiebegriff – hier etwa die KI – in ihrem Sinne zu politisieren. Auf diese Weise wird Moral, wie Wolfgang van den Daele in Bezug auf die Moralisierung von Technikkonflikten schreibt, als Gegenmacht gegen ein sogenanntes liberales Regime der Innovation mobilisiert, welches wiederum die technologische Entwicklung der politischen Kontrolle zu entziehen versucht, sie also entpolitisieren will.[13] Moral fungiert in diesem Zusammenhang, so van den Daele in Anlehnung an Wolfgang Krohn[14], als ein Mittel des Protests (Protestmoral), wird quasi also selbst zum Objekt der Polarisierung im Rahmen sozioökonomischer Konflikte, in deren Kontext die Kritik an gesellschaftsrelevanter Technologie als Fortsetzung einer bereits bestehenden Systemkritik zu interpretieren ist.[15]

Ein Wertekonflikt mit Moralisierungspotenzial ist in einem deliberativen Diskursumfeld aber auch schnell wieder entschärft, insofern sich die Diskursparteien demonstrativ zur Gültigkeit allgemein unstrittiger Normen bekennen, durch die Akzeptanz eines einer demokratischen Gesellschaft entsprechenden Wertepluralismus und durch die »Rückübersetzung von moralischen Konflikten in Interessenkonflikte«.[16] Hierbei stellt van den Daele fest, dass Technikkonflikte in der Regel auf der Basis von Werten und Normen moralisiert werden, über deren Geltung an sich Konsens besteht, welche allerdings auch im Kontext der Abwägungsprozesse zwischen verschiedenen Standpunkten darüber, was noch akzeptabel ist und wann eine rote Linie überschritten ist, konkurrieren und kollidieren können.[17]

Ob eine Technologie in der Debatte überhaupt problematisiert wird, hängt somit vor allem auch von der Einschätzung ihrer möglichen Risiken (etwa für den Menschen oder für die Gesellschaft) ab. Die Kontroverse bezieht sich

12 Vgl. Grunwald 2013: 240 f.
13 van den Daele 2013: 29 f.
14 Krohn 1999.
15 van den Daele 2013: 29 f.
16 van den Daele 2013: 31 f.
17 van den Daele 2013: 40.

in diesem Sinne weniger auf die Frage moralischer Verantwortbarkeit, sondern vielmehr auf alternative Risikokalkulationen, also die Frage, ob mit bestimmten Risiken oder Kosten tatsächlich gerechnet werden muss, es für den moralischen Einwand demnach überhaupt einen Anlass gibt. Somit wird der Diskurs letzten Endes statt als Wertekonflikt als kognitive Kontroverse ausgetragen.[18] Die Debatte um eine Risikoklassifizierung, die beispielsweise das Gutachten der Datenethikkommission der Bundesregierung[19] (1999) und die EU-Kommission stark gemacht haben, aber die Enquete-Kommission des Deutschen Bundestages zur Künstlichen Intelligenz skeptisch beurteilt hat[20], mag ein Beispiel für diese Tendenz sein.[21]

Sicherlich können auch Risikokalkulationen von bestimmten Wertehaltungen geprägt sein. Unter Berücksichtigung der Tatsache, dass es sich bei der KI-Technologie um eine junge und vielseitig anwendbare Technologie handelt, deren zukünftige Möglichkeiten noch nicht zur Gänze absehbar sind, lässt auch die Kalkulation des potenziellen Risikos der Technologie Raum zur öffentlichen Spekulation. Diese ist sehr wahrscheinlich von kulturell inspirierten Befürchtungen beeinflusst und bietet dementsprechend Moralisierungspotenzial, welches von den dominanten Akteuren dieses Diskurses zur diskursiven Mobilisierung ihrer eigenen Positionen in der Debatte aufgegriffen wird. Daher muss die mediale und damit gesellschaftliche Rezeption des Themas auch als Ausdruck des Kalküls der am einschlägigen Diskurs maßgeblich beteiligten Akteure verstanden werden.

Das bedeutet: Generiert die Berichterstattung über KI beispielsweise eine Diskursrichtung, durch welche der Bevölkerung »Angst« vor der Technologie vermittelt wird, so führt dies letzten Endes durch die notwendige Reaktion der Politik auf die in diesem Zuge gesellschaftlich formulierten Bedürfnisse der Wähler:innen zu einer Beeinträchtigung der einschlägigen Wirtschafts- und Forschungszweige, insofern diesen damit die legitimierende Grundlage ihrer Förderung entzogen wird. An dieser Stelle soll jedoch nicht weiter auf die Problematik eines Wirtschaftslobbyismus im Bereich der Technologiepolitik oder gar auf liberale Paradigmen ebendieser eingegangen werden. *Stattdessen liegt hier der Fokus auf der Feststellung, dass politische Entscheidungen über die rechtliche Regulierung von KI (und ebenso anderen Technologie- und Forschungsfeldern im Allgemeinen) auf der Basis der Einstellung der Bevölkerung gegenüber der*

18 van den Daele 2013: 33.

19 Datenethikkommission der Bundesregierung 2019.

20 Deutscher Bundestag, 19. Wahlperiode 28. 10. 2020: 41 f., 66 f.

21 Vgl. zur Debatte um die Regulierung von KI-Systemen nach Risikoklassen Wolfangel 18. 1. 2022. Wichtig in diesem Kontext auch der zu Grunde liegende Text von Krafft/Zweig 2019, auf den der Text der Datenethikkommission nicht hinweist.

Technologie getroffen werden. Diese wiederum ist abhängig von der allgemeinen Diskurslage und den Akteuren. Zudem sind politische Akteure mit im Spiel, die den Diskurs auf der Basis ihrer eigenen Interessen beeinflussen, also etwa eine Ethisierung oder eine Moralisierung betreiben.

D Basisformen des Verhältnisses von KI-Diskursen und KI-Ethik

Angewandte Ethik ist vornehmlich ausgerichtet auf normative Aussagen zu konkreten Sachproblemen, aber dafür sichtet sie vorliegende wissenschaftliche und praktische Argumente und moralische Einstellungen (hat also auch ein empirisches Interesse), und evaluiert diese, etwa in der Absicht, gute Argumente zu stärken und problematische Argumente zu entkräften. Praktische Philosoph:innen sind selber Akteur:innen in den beschriebenen Diskursen, viele Kolleg:innen im Feld der akademischen Angewandten Ethik oder der Bereichsethiken nehmen Teil an Fachgremien, beraten Parteien und Fraktionen und äußern sich in den Medien.[22] Allein von diesem Gesichtspunkt her ist es auch für die philosophische Ethik wichtig zu wissen, welche diskursiven Prozesse (und ebenso Akteure) gesellschaftliche Einstellungen zu Technologien prägen, welche antizipierten Wirklichkeiten und welche (moralische) Normen im Diskurs thematisiert werden und ob und welche Machtstrukturen diesen Diskursen zugrunde liegen.

Unsere These lautet, dass Technikdiskurse allgemein und spezieller die beschriebene politische Bezugnahme auf eine Einstellungs- oder Meinungs-Empirie sowie die Rolle von Fachgremien in mehreren Hinsichten auch die Angewandte Ethik betreffen. Uns interessiert vor allem die Frage: Welche Rolle spielen Diskurse über KI für eine Ethik der KI? Die Hypothese, die im Folgenden in drei Einzelthesen aufgeschlüsselt werden soll, lässt sich in der Annahme zusammenfassen, dass Einstellungen (wie etwa Hoffnungen, Erwartungen, Befürchtungen) über KI und gegenwärtige (im diskursanalytischen Sinne historische) Erfahrungen mit KI eine diskurspragmatische und eventuell auch eine normative Quelle für die KI-Ethik darstellen. Dies liegt nicht zuletzt deswegen nahe, weil im Kontext der Technologien Facetten des menschlichen Selbstverständnisses berührt werden und Erfahrungen mit Technologien einfließen, die in den Diskursen expliziert und argumentativ ins Spiel gebracht werden. Das macht es hilfreich, den gesellschaftlichen Diskurs über KI einer kritischen Analyse zu unterziehen, um diese Einstellungen und Erfahrungen wissenschaftlich herauszuarbeiten und für eine praktisch-philosophische

22 Vgl. zu einigen dieser Aspekte der Band Kettner 2000.

Evaluierung heranziehen zu können. Diese diskursiv vorliegenden Einstellungen und Erfahrungen könnten (1.) eine Quelle für normative Urteile darstellen (Quellen-These) und somit Informationen darüber bereitstellen, wie sich Ethik in praktischer Absicht öffentlich zur Thematik äußern, und somit selbst (2.) ein Teil dieses Diskurses werden kann (Diskursakteur-These). Dies bedeutet zuletzt (3.) eine Veränderung des eigentlichen Objekts einer KI-Ethik: Anstatt sich mit der KI-Technologie an sich zu beschäftigen, muss sich die ethikwissenschaftliche Perspektive (auch) den Diskursen um KI zuwenden (Objekt-These).

I Quellen-These

Empirisch erfassbare, moralisch relevante Motive, Einstellungen und Überzeugungen spielen in verschiedenen Theorien der Ethik eine wichtige Rolle, zumeist mit dem Bezug auf den Begriff der Erfahrung. Ob und wie Erfahrung ins Spiel kommt, etwa als Quelle moralischer Einsichten oder als Begründungsinstanz normativer Urteile, wird kontrovers diskutiert. Ansätze in der Tradition Kants gehen von der prinzipiellen Erfahrungsunabhängigkeit der Moral und der sie reflektierenden Ethik aus. Hermeneutische oder phänomenologische Ansätze (und darin besonders die philosophische Anthropologie) tendieren zu einer mehr oder weniger engen Vermittlung beider Ebenen: »Thus, in common sense-based, hermeneutic, pragmatist, and Aristotelian approaches such as Communitarianism, common morality and concrete public perspectives are usually seen as a methodologically indispensable starting point and normative source for ethical reflection«[23]. Eine hermeneutische Ethik versteht sich »als Hermeneutik der Lebenswelt«[24]. Unseres Erachtens kann man von einer erfahrungsabhängigen moralischen Lernfähigkeit des Menschen ausgehen, d. h. wir Menschen »entdecken […] uns im Hinblick auf Moral, d. h. auf die Unterscheidung von Gut und Böse, von Richtig und Falsch, in Bezug auf unser Denken, Vorstellen und Handeln, als unausweichlich erfahrungsfähig«[25]. Insofern spielen für Menschen diverse Erfahrungen auf unterschiedlichen Ebenen mit KI-Technologien moralisch eine Rolle. Zwar sind die verschiedenen Erfahrungsebenen (Mieth spricht von Wahrnehmung, Erlebnis

23 Schicktanz/Schweda/Wynne 2012. Den Hinweis auf die Forschungen von Silke Schicktanz zum Zusammenhang von Angewandter Ethik und (sozialwissenschaftlicher) Empirie verdanke ich Matthias Kettner.

24 Lesch 2007.

25 Mieth 2011: 344. Vgl. zur Rekonstruktion (und zur pragmatistischen Kritik) des Zusammenhangs von Erfahrung und Moral bei Dietmar Mieth Filipović 2015, bes. Kap. 2.1 und Kap. 4.1.

und Begegnung) verschieden, für die praktische Alltagsrationalität, so wie sie auf der Basis vieler Erfahrungen vorliegt und die Einstellung zu einer Technologie prägt, ist dies aber nicht relevant. Unterstützung erfährt diese Einheitsperspektive hinsichtlich verschiedener Erfahrungsformen durch den Pragmatismus.[26] Peirce, James und Dewey bemühen sich alle um den Nachweis der Untrennbarkeit verschiedener Erfahrungsformen.[27]

Mit diesem Verständnis der Einheit von Erfahrungsdimensionen und der Relevanz der Erfahrung für die Moral kann sich eine praktische Philosophie der Künstlichen Intelligenz aus guten Gründen diesen Diskursen zuwenden und sie als Quelle für normative Einsichten nutzen. Selbstverständlich ist damit keine unreflektierte Übernahme moralischer Einsichten gemeint, auch nicht eine epistemologisch naive Aufhebung des Unterschieds von Sollens- und Seinssätzen. Aber es wird davon ausgegangen, dass Menschen in der Auseinandersetzung mit Realitäten und Fiktionen von KI-Technologien Erfahrungen machen und ihre spezifisch humanen Selbstverständnisse explizieren, die für eine praktische Philosophie der KI relevant sein können. Wir haben es mit Transformationen menschlicher Selbstverständnisse im technischen Fortschritt zu tun,[28] die ethisch einzubeziehen sind. Die Hinwendung zu den *Diskursen* soll dann gerade in den Blick bringen, dass diese Erfahrungen in Relation stehen zu Machtkonstellationen, Diskursebenen, Akteurskonstellationen, Framings etc. Diskurse als Quelle zu nutzen, kann wissenschaftlich-philosophisch dann nur gelingen, wenn dies reflektiert geschieht, etwa durch eine methodisch geleitete Diskursanalyse.

Am Beispiel des Themenfeldes KI ergibt sich hieraus konkret für eine wissenschaftliche Herangehensweise die Frage, welche Vorstellungen vom spezifisch Menschlichen sich aus dem Kommunizieren bezogen auf das Phänomen KI ableiten lassen. Dabei muss berücksichtigt werden, dass eben diese Kommunikation über KI ihrerseits ein Phänomen darstellt, welches eine eigene Form der diskursiven Wirklichkeit repräsentiert und gleichzeitig kreiert. Um die diskursive Wirklichkeit des Kommunizierens über KI und die dieser Wirklichkeit zugrunde liegenden Vorstellungen für eine ethische Reflektion brauchbar zu machen, bedarf es somit einer (kritischen) diskursanalytischen Untersuchung.

Methodisch bedeutet das, den öfter diskutierten Zusammenhang zwischen (Angewandter) Ethik und Empirie[29] zu erweitern auf die Frage nach dem Zu-

26 Vgl. etwa Jung 2004.

27 Filipović 2015: 130–135.

28 Vgl. Grunwald 2021.

29 Im Kontext des manchmal so genannten empirischen Turns der Angewandten Ethik wurde die methodisch-philosophischen Probleme intensiv behandelt, vgl. etwa Birnbacher

sammenhang von Angewandter Ethik und Diskursempirie.[30] Wenn Hösle als eine der vier zentralen Anforderungen an moderne praktische Philosophie festhält, dass sie sich mit den empirischen Aspekten ihres Gegenstandes bekannt machen muss,[31] dann bedeutet das unter Berücksichtigung der Quellenthese, dass sie sich auch mit den *Diskursen* über ihren Gegenstand, hier also die KI-Diskurse, bekannt machen muss. Diesen Aspekt diskutieren wir unten eingehender als »Gegenstands-These«.

II Diskursakteur-These

Insbesondere im Feld der Bereichsethik und der Angewandten Ethik beanspruchen verschiedene Fachverständnisse für sich eine praktische Relevanz in dem Sinne, als dass sie in der Lage sind, wirkliches Handeln zu beeinflussen beziehungsweise zu ändern, moralische, realweltliche Probleme zu lösen oder normative Handlungsunsicherheit zumindest verringern zu können. Infolge der Covid-19-Pandemie und den sich daraus ergebenden ethischen Fragestellungen – zum Beispiel hinsichtlich der praktischen Umsetzung von Impfstoffverteilungen und Priorisierungen – beteiligen sich auch Ethiker:innen an der Schnittstelle zwischen philosophischer Reflektion und politischer Praxis an Problemlösungen im öffentlichen Diskurs.

Im Beispiel der KI-Technologie – und auch anderen Technologiedebatten der vergangenen Jahre – nahmen, wie bereits angedeutet, wissenschaftlichen Fachgremien wie zum Beispiel die 2018 ins Leben gerufene Enquete-Kommission »Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale« ihre Arbeit auf und beteiligten sich infolgedessen als Akteure des anhaltenden öffentlichen Diskurses.[32] Ebenso wie andere Diskursakteure streben auch derartige Fachgremien in Konkurrenz zu anderen Akteuren nach publizistischer beziehungsweise medialer Aufmerksamkeit und sehen sich herausgefordert, Rollenerwartun-

1999; Borry/Schotsmans/Dierickx 2005. Für eine starke Verschränkung von Sozio-Empirie (um den Unterschied zur naturwissenschaftlichen Empirie zu verdeutlichen) und Ethik argumentieren bspw. Schweda/Schicktanz 2014. Sie machen ihren Punkt an dem Beispiel deutlich, dass beim Thema Organspende aus der akademischen Ethik oft Altruismus ein wichtiger moralischer Begriff ist, wogegen sie in ihrer empirischen Arbeit in Fokusgruppen-Interviews herausgefunden haben, das Reziprozität ein mindestens ebenso zentraler moralischer Aspekt bei der Organtransplantation ist (ebd.: 219–221). Vgl. dazu auch Musschenga 2005.

30 Hinweise dazu schon bei Schicktanz/Schweda/Wynne 2012: 135 mit Bezug auf Gerhards/Schäfer 2009. Soweit wir sehen, spielt bei Schicktanz aber eine kritische Diskursanalyse, wie wir sie hier ins Spiel bringen, keine Rolle.

31 Hösle 1992.

32 Vgl. zur Reflexion der eigenen Arbeit als Ethiker in dieser Kommission Filipović 2020.

gen innerhalb der Diskurse abzuwehren wie auch Allianzen mit anderen Teilnehmer:innen des Diskurses einzugehen. Daher nutzen Ethiker:innen als Diskursakteure zwangsläufig auch nicht, selbst wenn das möglich wäre, ausschließlich wissenschaftliche Rationalität als Grundlage ihrer diskursiven Argumentation, sondern sind ebenso darauf angewiesen, ihre Äußerungen medial so vorzubringen, dass sie auch verteilt und rezipiert werden. Verkürzungen und ein gewisses Überdehnen des eigenen Standpunktes zu Gunsten von Unterscheidbarkeit sind dabei unvermeidlich. Andererseits bietet ihnen die Beteiligung an derartigen Diskursen die Möglichkeit, in der Gesellschaft besonders dringliche Probleme schneller zu identifizieren und zu erkennen, an welchen gesellschaftsstrukturellen Stellen durch Überzeugungsarbeit angesetzt werden muss, um in der Debatte problematische Knoten zu lösen, welche einem besseren ethischen Verständnis des Sachverhalts hinderlich sein können.

Für das Themenfeld KI bedeutet das konkret, dass die Ethik in der öffentlichen Debatte eine gewichtige Rolle spielen kann und spielen soll, welche jedoch auch von Diskursakteuren außerhalb der ethischen und auch (technik-) wissenschaftlichen Fachdebatte mitdefiniert und formuliert wird. Die Ethik als Wissenschaft wird sich in der öffentlichen Debatte mit diesem Umstand arrangieren müssen, ebenso, wie sie diese Rolle reflektieren und methodisch einfangen muss, um ihrer Stimme und ihrer Rolle als Diskurspol gerecht zu werden. Auch hier erscheint uns eine reflektierte Auseinandersetzung mit den konkreten Diskursen in Form einer Diskursanalyse hilfreich, um eigene Vereinnahmungen und unbewusste Framings zu identifizieren und sie dadurch besser vermeiden zu können.

III Objekt-These

Ein typisches Merkmal der Bereichsethik im Allgemeinen ist es, dass der jeweilige Gegenstand nicht vollständig klar und abgegrenzt formuliert werden kann und stattdessen – in Folge des Fortschritts im jeweiligen Bereich – ständig neu formuliert werden muss.[33] Mit der Frage nach der bestimmten Natur des eigenen Forschungsgegenstandes ringen verschiedene Geistes- und Sozialwissenschaften, wie etwa die Kommunikationswissenschaft mit der Frage, wie Kommunikation im Angesicht neuer Kommunikationsformen zu definieren sei, die Literaturwissenschaft in ihrer Auseinandersetzung mit Natur und gesellschaftlicher Rolle der Literatur oder auch die Politikwissenschaft mit ihrer Frage nach der konkreten Definition von schwer fassbaren, gesellschaftspoli-

33 Bayertz 1994: 20 f.

tischen Phänomenen wie zum Beispiel dem Populismus. Auch die Medienethik als Angewandte Ethik hat mit einer Verflüssigung ihrer Gegenstände zu tun.[34]

Vor dem Hintergrund, dass eine (damit einhergehend auch normative) Unklarheit darüber besteht, wie Begriffe wie »KI« oder »Digitalisierung« zu deuten sind, wenn diese Diskursgegenstände hochgradig von den Bedingungen der diesbezüglichen, gesellschaftlichen Kommunikation, den diese Kommunikation umfassenden Rahmenbedingungen, den daran beteiligten Akteurskonstellationen und damit verbundenen Machtstrukturen abhängig sind, so liegt der Schluss nahe, dass eben diese Diskursgegenstände auch ein Objekt der praktischen Philosophie sind.

Geht man aber davon aus, dass es im diskursanalytischen Sinne nicht um KI als aus ethischer Perspektive betrachtete Technologie, sondern stattdessen um die Beobachtung des Diskurses um KI geht, so bedeutet dies, dass der Ausgangspunkt einer ethischen Behandlung von KI nicht das Technik-Objekt und dessen Handlungsfolgen, sondern vielmehr der diesbezügliche Technik-Diskurs und dessen Handlungsfolgen Ziel des Forschungsinteresses sein kann. Dies würde bedeuten, dass sich die Ethik auch mit der Frage auseinandersetzen muss, welches und wessen Handeln sie als Diskurspol und -akteur beeinflussen und verändern will oder sogar muss.

E Zur Relevanz von KI-Diskursen – Herausforderungen für die Angewandte Ethik

Die bisher getätigten Betrachtungen und drei skizzierten Thesen verweisen darauf, dass die Ethik darauf angewiesen ist, sich grundsätzlich mit gesellschaftlichen Diskursen auseinanderzusetzen. Damit stellt sich jedoch auch die Frage, was es für die Ethik bedeutet, sich fokussiert auf eben solche Diskurse zu beziehen und welche daraus abgeleitete Aussagen dann in ethischer Hinsicht möglich sind. Hier droht die Gefahr, dass sich die ethische Auseinandersetzung und Diskursbeteiligung in eine konstruktivistische Falle begibt, in welcher diskursive Konstruktion und Gegenkonstruktion die Rolle ethisch basierter Begründungen einnehmen. Sicher kann sich die Angewandte Ethik nicht mit einer kritischen Diskursanalyse begnügen und sich darauf beschränken, einen Beitrag zu den Diskursen bloß aus der strategischen Position wissenschaftlicher Gremienarbeit zu leisten, anstatt zum eigentlichen Grund des Diskurses, hier etwa KI-Technologien und deren gesellschaftliche Relevanz, Stellung zu nehmen. Aber auch dieser »eigentliche Gegenstand« der KI-Ethik ist ein »Konstrukt«, dessen »Konstruktivität« man nicht umgehen kann. Es bleibt

34 Heesen 2015, vgl. auch Filipović 2016: 45.

die Möglichkeit für die Angewandte Ethik, nicht nur die eigene Klärung des eigenen Gegenstandes sorgfältig vorzunehmen, sondern auch den Diskurscharakter der gesamten Debatte methodisch einzufangen.

Das Verhältnis von Diskursanalyse und Angewandter Ethik im Feld der Künstlichen Intelligenz ist komplex. Wir haben versucht zu zeigen, dass eine Beschäftigung mit den Diskursen über KI philosophisch möglich ist und für die Ethik weiterführende Perspektiven bietet und haben mit Quellen-These, Akteurs-These und Gegenstandsthese eine Systematisierung des Verhältnisses von KI-Ethik und KI-Diskurs vorgeschlagen. Anschluss finden diese Überlegungen wie gezeigt an eine Reihe von Debatten in der praktischen Philosophie. Die Methode der kritischen Diskursanalyse (auf der Basis des Diskursverständnisses von Foucault)[35] wird im Kontext der (sozial-)empirischen Unterstützung der Angewandten Ethik aber, nach unserer Kenntnis, bisher nicht oder kaum benutzt oder debattiert.

Es wäre unseres Erachtens lohnenswert, weiter zu prüfen, inwieweit man die Methoden der kritischen Diskursanalyse in die bereichsethische Arbeit einbinden kann. Hierzu sollen die hier bereits vorgestellten Thesen einen Ansatzpunkt liefern. Eine darauf basierende Forschung gestaltet sich als eine gewisse Herausforderung, weil die kritische Einbindung von (sozialwissenschaftlicher) Empirie in der bereichsspezifischen Ethik bisher wenig Raum einnimmt und Fragen der ethischen Theorie nach der Rolle von Praxis, Prinzipienbegründung und die Frage nach dem Verhältnis von Empirie, Moral und Ethik zueinander nach wie vor kontrovers diskutiert werden. Andererseits stellen sich mit der Theorie der Diskursanalyse und ihrer methodologischen Anwendung auch soziologische und kommunikationswissenschaftliche Fragen, wie sich eine praktische Anwendung jenseits Foucaults theoretischem »Werkzeugkasten« im Kontext gegenwärtiger Debatten als ein sozialwissenschaftliches, kritisches Analyseinstrument anwenden lässt, was eine »kritische Diskursanalyse« überhaupt kritisch macht, welche Rolle Medialität für Diskurse spielt oder welche quantitativen und qualitativen Methoden für ihre Anwendung geeignet sind.

35 Jäger/Jäger 2007; Foucault 1973.

Bibliographie

Bayertz, Kurt: Praktische Philosophie als angewandte Ethik. In: Bayertz, Kurt (Hg.), 1994. Praktische Philosophie. Grundorientierungen angewandter Ethik, Reinbek bei Hamburg, 7–47.

Birnbacher, Dieter 1999: Ethics and Social Science: Which Kind of Co-operation? In: Ethical Theory and Moral Practice, 319–336.

Bogner, Alexander: Ethisierung oder Moralisierung? Technikkontroversen als Wertkonflikte. In: Bogner, Alexander (Hg.), 2013. Ethisierung der Technik – Technisierung der Ethik. Der Ethik-Boom im Lichte der Wissenschafts- und Technikforschung, Baden-Baden, 51–67.

Borry, Pascal/Schotsmans, Paul/Dierickx, Kris 2005: The birth of the empirical turn in bioethics. In: Bioethics, 49–71.

Brennen, J. Scott/Howard, Philip N./Nielsen, Rasmus Kleis 2018: An Industry-Led Debate: How UK Media Cover Artificial Intelligence. Unter: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-12/Brennen_UK_Media_Coverage_of_AI_FINAL.pdf.

Datenethikkommission der Bundesregierung 2019: Gutachten der Datenethikkommission der Bundesregierung.

Deutscher Bundestag, 19. Wahlperiode 28. 10. 2020: Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Drucksache 19/23700. Unter: https://dip21.bundestag.de/dip21/btd/19/237/1923700.pdf.

Filipović, Alexander 2015: Erfahrung – Vernunft – Praxis. Christliche Sozialethik im Gespräch mit dem philosophischen Pragmatismus. Paderborn u. a.

Filipović, Alexander: Angewandte Ethik. In: Heesen, Jessica (Hg.), 2016. Handbuch Medien- und Informationsethik, Stuttgart, 41–49.

Filipović, Alexander: Ethik als Akteurin für die Entwicklung einer digitalen Kultur. Das Verhältnis zu Wirtschaft und Politik am Beispiel des Diskurses um »Künstliche Intelligenz«. In: Prinzing, Marlis/Debatin, Bernhard S./Köberer, Nina (Hg.), 2020. Kommunikations- und Medienethik reloaded? Wegmarken für eine Orientierungssuche im Digitalen, Baden-Baden, 331–338.

Fink, Ronja 2020: Menschengetriebene Technologie oder technologiegetriebene Menschen? Eine Diskursanalyse über Künstliche Intelligenz in der Wissenschaft, der Wirtschaft und den Medien. Wissenschaftliche Arbeit zur Erlangung des akademischen Grades Master of Arts (M. A.). München.

Fischer, Sarah/Puschmann, Cornelius 2021: Wie Deutschland über Algorithmen schreibt. Eine Analyse des Mediendiskurses über Algorithmen und Künstliche Intelligenz (2005–2020).

Foucault, Michel 1973: Archäologie des Wissens. Frankfurt a. M.

Gerhards, Jürgen/Schäfer, Mike S. 2009: Two normative models of science in the public sphere: human genome sequencing in German and US mass media. In: Public Understanding of Science, H. 4, 437–451.

Grunwald, Armin: Ethische Aufklärung statt Moralisierung. Zur reflexiven Befassung der Technikfolgenabschätzung mit normativen Fragen. In: Bogner, Alexander (Hg.), 2013. Ethisierung der Technik – Technisierung der Ethik. Der Ethik-Boom im Lichte der Wissenschafts- und Technikforschung, Baden-Baden, 232–246.

Grunwald, Armin (Hg.) 2021: Wer bist Du, Mensch? Transformationen menschlichen Selbstverständnisses im technischen Fortschritt. Freiburg.

Heesen, Jessica: Ein Fels in der Brandung? Positionen der Medienethik zwischen verflüssigtem Medienbegriff und schwankender Wertebasis. In: Prinzing, Marlis/Rath, Matthias/Schicha, Christian/Stapf, Ingrid (Hrsg.), 2015. Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000, Weinheim, 86–98.

Hösle, Vittorio: Vorwort. In: 1992. Praktische Philosophie in der modernen Welt, München, 9–13.

Jäger, Margret/Jäger, Siegfried 2007: Deutungskämpfe. Theorie und Praxis kritischer Diskursanalyse. Wiesbaden.

Jäger, Siegfried 2015: Kritische Diskursanalyse. Eine Einführung. Münster.

Jung, Matthias: Qualitative Erfahrung in Alltag, Kunst und Religion. In: Mattenklott, Gert (Hg.), 2004. Ästhetische Erfahrung im Zeichen der Entgrenzung der Künste. Epistemische, ästhetische und religiöse Formen von Erfahrung im Vergleich, Hamburg, 31–53.

Kettner, Matthias (Hg.) 2000: Angewandte Ethik als Politikum. Frankfurt am Main.

Krafft, Tobias D./Zweig, Katharina A. 2019: Transparenz und Nachvollziehbarkeit algorithmenbasierter Entscheidungsprozesse. Ein Regulierungsvorschlag aus sozioinformatischer Perspektive. Gutachten im Auftrag des Verbraucherzentrale Bundesverband. Unter: https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf.

Krohn, Wolfgang 1999: Funktionen der Moralkommunikation. In: Soziale Systeme, H. 2, 313–338.

Lesch, Walter: Ethische Reflexion als Hermeneutik der Lebenswelt. In: Lob-Hüdepohl, Andreas/Lesch, Walter (Hgg.), 2007. Ethik Sozialer Arbeit. Ein Handbuch, Paderborn, 88–99.

Mieth, Dietmar: Erfahrung. In: Düwell, Marcus/Hübenthal, Christoph/Werner, Micha H. (Hgg.), 2011. Handbuch Ethik, Stuttgart, 342–347.

Musschenga, Albert W. 2005: Empirical ethics, context-sensitivity, and contextualism. In: The Journal of medicine and philosophy, H. 5, 467–490.

Schicktanz, Silke/Schweda, Mark/Wynne, Brian, 2012: The ethics of ›public understanding of ethics‹ – why and how bioethics expertise should include public and patients' voices. In: Medicine, health care, and philosophy, H. 2, 129–139.

Schweda, Mark/Schicktanz, Silke, 2014: Why public moralities matter – the relevance of socioempirical premises for the ethical debate on organ markets. In: The Journal of medicine and philosophy, H. 3, 217–222.

van den Daele, Wolfgang: Moralisierung in Technikkonflikten. In: Bogner, Alexander (Hg.), 2013. Ethisierung der Technik – Technisierung der Ethik. Der Ethik-Boom im Lichte der Wissenschafts- und Technikforschung, Baden-Baden, 29–50.

Whittlestone, Jess/Nyrup, Rune/Alexandrova, Anna/Dihal, Kanta/Cave, Stephen, 2019: Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. London.

Wolfangel, Eva, 18. 1. 2022: KI first, Bedenken second. In: Zeit-Online.

ORCID

Alexander Filipović https://orcid.org/0000-0001-8946-9283

Julian Lamers https://orcid.org/0000-0003-0119-3898

Roboter als Ding und Un-Ding

Zur Hermeneutik der Zwischenwesen – zwischen Mensch und Maschine

Philipp Stoellger

Die folgenden Überlegungen operieren weder aus empirischer noch aus historischer Perspektive, sondern aus *phänomenologischer* und *hermeneutischer*. Was auch immer *Verstehen* sein mag und wie es ›gemacht‹ werden kann, empirische und historische Methoden unterscheiden sich von denen der Hermeneutik. Im Hintergrund stehen zudem Bildwissenschaft und Medientheorie, denn Religion ist ein Medium und operiert in und durch Medien. Religionsgeschichte ist daher ein Feld der Erfindung und Erforschung von emergenten Medien – wie Christus und sein Geist. Im Horizont von Theologie und Medienphilosophie geht es daher um die Neujustierung der anthropologischen Differenz von Mensch und Maschine im Horizont der neuen Medien. Dabei wird das Modell einer chiastischen Emergenz der Differenz von Menschen und Medien vorausgesetzt – als Figurationsprozess.[1]

Im Sinne einer Hermeneutik der *Differenz* (nicht des Konsenses und der Identität)[2] wird hier die Exklusion von Hermeneutik und Phänomenologie durch die ›Deutsche Medientheorie‹ oder ›German Media Theory‹ (Kittler et al.) zurückgewiesen. Solange neue Medien Menschen *adressieren* und von ihnen *genutzt* werden, bleiben Fragen nach Sinn und Wahrheit unvermeidlich. Als Beispiel für relevante Differenzen, von, durch und mit denen wir leben, wird hier die von *Vertrauen und Sich-verlassen-auf* erörtert.

1 Vgl. Stoellger 2019; ders. 2020a: 225–235; ders. 2020b: 19–47; ders. 2016a: 192–206.

2 Vgl. Stoellger 2016b: 164–193.

Philipp Stoellger 2022: Roboter als Ding und Un-Ding. Zur Hermeneutik der Zwischenwesen – zwischen Mensch und Maschine. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 155–176.
https://doi.org/10.11588/heibooks.1106.c 15198

A Medien als operative Wahrnehmungsformen

Neue Medien sind insofern *neu*, als sie unsere *Wahrnehmungsformen* verändern (in Interaktion, Lebensformen, Gefühlen, Denkgewohnheiten, Sprache, Politik usw.). Da ›Medium‹ terminologisch eine Erfindung von Thomas von Aquin für Aristoteles' anonymes *metaxy* war, ist ein Medium das, was *dazwischenkommt* und *interveniert*. Es ist *operativ*, sofern es den Zugang zum sonst Unzugänglichen ermöglicht; aber es ist deshalb transparent und undurchsichtig zugleich (wie ein Kirchenfenster). Ein Medium ist also nicht nur Mittel zum Zweck wie ein Instrument, sondern eine *operative Form der Wahrnehmung*.

Veränderungen in den Wahrnehmungsformen manifestieren sich in Metaphern und Metonymien, z. B. in Geschichten und Gleichnissen, Ikonen und Bildern, in der Literatur und Kunst – und Religion. Ein gängiges Beispiel dafür ist die Denk- und Sprachgewohnheit, das Gehirn als Computer zu verstehen oder der ›genetische Code‹, der durch CRISPR neu geschrieben werden könne. Das ist neue Metaphysik oder Biometaphysik, als ob der Mensch im Wesentlichen sein genetischer Code wäre oder als *wäre* das Gehirn tatsächlich ›ein Computer‹. Das disponiert zur *Techno*metaphysik: Da das Gehirn und der Code dann als ›thesei‹ gelten, nicht als ›physei‹, wird der Glaube bedient, den Menschen durch Gentechnik und Computertechnik zu optimieren.

So gesehen kann die *Medialität* der Neuen Medien und ihr Einfluss auf unsere Wahrnehmung, Interaktion etc. anhand ihrer metaphorischen, konzeptuellen und figurativen Manifestationen analysiert und interpretiert werden. Da ein solcher hermeneutischer und phänomenologischer Ansatz jedoch *abduktiv* ist, bleibt Validierung ein Problem. Man kann keine allgemeingültigen Behauptungen aufstellen, sondern muss eine Interpretation dessen, ›was vor sich geht‹, finden und erfinden. Eine *investigative* Forschung hinsichtlich Medialität muss also *imaginativ* und *innovativ* – und *erfinderisch* sein: So war die ›inventio‹ früher die Bezeichnung für ›Abduktionen‹, d. h. das Aufdecken aktueller und zukünftiger Muster der Wahrnehmung, des Denkens usw.; ob ›Roboter unsere Freunde sein (werden) dürfen oder sollen‹, ist solch eine abduktive und innovative Frage, die etwas insinuiert: Wahrscheinlich ist es besser, sich mit ihnen anzufreunden, als einen Feind aus ihnen zu machen. Sich Feinde zu machen, ist einfach; schwerer wäre es, mit Robotern auf ›menschliche Weise zusammenzuleben‹, ohne sie nur für ein Ding zu halten oder sie als kommende ›Übermenschen‹ zu dramatisieren.

B Was sind ›Roboter‹?

Roboter sind 1. keine Tiere, 2. nicht nur Dinge, 3. noch keine Menschen und 4. noch nicht Gott (trotz Kubricks HAL 9000 in 2001: A Space Odyssey). Aber was sind sie dann? Eine eigene ›Spezies‹? Von welcher ›Gattung‹? Ich würde vorschlagen, dass sie eine ganz besondere Spezies in der Vielzahl der *Zwischenwesen* sind, wie Engel und Dämonen: dazwischen existierend, ein Komplex aus Imagination und Realität, mit ihren eigenen Körpern und ihrer eigenen Art der Kommunikation.

Roboter sind:

- Im engeren Sinne humanoide Maschinen (Geräte, Apparate).
- Im weiteren Sinne alle Maschinen, die von Programmen gesteuert werden.
- Im weitesten Sinne alle verbundenen Geräte und die Art der Verbindung.
- Sie sind ›Knoten‹ eines Netzes. Sie konstituieren sich durch und als Beziehungen, durch die speziellen Beziehungen, die wir ›digital‹ oder ›programmiert‹ nennen.

Für das Folgende reduziere ich die Polysemie auf Roboter als *Formen* im Medium der digitalen Kommunikation. Roboter sind *animierte Akteure zwischen* Dingen und Menschen: vermittelnde Wesen im Bereich der Intermedialität (gemäß ihrer Relationalität). Sie bilden eine eigene Spezies, die man vorläufig *Un-Dinge* nennen kann, mit Handlungsfähigkeit im Horizont der Intermedialität (dem *Dazwischen*), wo man auch Engel und Dämonen treffen mag, die klassische Paradigmen für das sind, was wir ›Medien‹ nennen. Das geeignete Deutungsmuster für ›Roboter‹ ist dann nicht ein ›Entweder-oder‹, sondern der Bereich des *Dazwischen*: Roboter bevölkern und beleben *Zwischenwelten*.

Die retroaktive Konsequenz dieses Deutungsmusters ist eine neue Betrachtungsweise des Menschen – nicht substantialistisch, sondern relational (wie bei Cassirer mit Leibniz, wenn er die mathematische Funktionsrelation als Modell für eine relationale Ontologie und Anthropologie aufruft), mit einem Primat der Relation vor den Relaten: d. h., den Menschen *als relationales Medium* (Medienanthropologie) und sogar *Gott als Medium* zu verstehen.[3]

In Anlehnung an ihre ›Apparatgeist theory‹ vertreten Katz und Aakhus in ihrer Argumentation »machines do indeed become us«[4].

3 Vgl. Stoellger 2018: 351–393.

4 Katz 2003: 315.

1. Maschinen werden zu »our representatives at a distance«[5], sagen Katz und Aakhus. Werden diese Maschinen zu unserem ›Leben nach dem Tod‹, dem *soma pneumatikon*, dem wiederauferstandenen Körper (oder mehr) des ewigen Lebens? Werden sie zur lebendigen Gegenwart der Toten? Werden sie nicht nur zu *unseren* Repräsentanten, sondern auch zu *deren* Repräsentanten, d. h., übernehmen sie mehrere Kategorien der Repräsentation? Schließlich wurde der Golem zum Repräsentanten seiner selbst und der dunklen Seite des maschinellen Lebens (ein Thema, das ich weiter unten aufgreifen werde).
2. Maschinen werden »important parts and reference points in our self-concepts«[6]. Könnten sie zu entscheidenden Medien des ›Selbstverständnisses‹ werden, zu Figuren des Anderen, die uns herausfordern, ein Selbst zu werden? Unsere ›Repräsentanten‹ entfalten eine signifikante Wirkung für unsere Selbstkonzepte: Nicht nur, dass das Gehirn als Computer gesehen wird, sondern dass der Mensch *als Medium* konzipiert wird. *Identität als Relation* ist für die Medienanthropologie entscheidend: Insofern der Mensch nicht Substanz, sondern ursprünglich Relation ist (Primat von Relationen wie Medien), wird die Form namens ›Mensch‹ erst in, mittels und durch dynamische Relationen namens ›Medien‹ möglich und real.

C Von Dienern zu Freunden? James Katz und der Golem

James Katz' Deutungsmuster war 2003 noch, dass Maschinen zwar wie wir oder sogar *wir* werden, aber »[m]achines will always be the servants of humans«[7]. Das ist eine verbreitete Vorstellung: *das Master-Slave-Muster*, oder aber: Herr – Knecht.[8] Der Sklave oder Diener ist *mehr als ein Ding*, er ist *mehr ein Mensch* als ein Ding – aber ohne Autonomie, ohne Freiheit, ohne Würde? Die Folge ist dann: Sobald Maschinen *wir* werden, bleiben sie keine Diener mehr.

Das dürfte der Grund sein, warum Katz 2019 deutlich weitergeht: nicht nur Diener, sondern vielleicht *Freunde*? Wenn sie *wir* werden, könnten Roboter unsere Freunde werden – aber *sollten* sie das auch? Auf den ersten Blick würde man *nein* und *niemals* sagen; auf den zweiten oder dritten aber regt sich womöglich Nachdenklichkeit. ›Roboter als Freunde‹ klingt wie ein Kategorienfehler. Wäre das (hermeneutisch vermutet) ein kalkulierter Kategorienfehler,

5 Katz 2003: 318.

6 Katz 2003: 318. Und 3. werden Maschinen »meaningful accoutrements to our self creation and symbolic interpersonal communication« (Katz 2003: 318).

7 Katz 2003: 318.

8 Ein ähnliches Muster wären Arbeitgeber und Arbeitnehmer.

wäre es eine *Metapher* – die eine Horizontverschiebung zeigt, die sich in den letzten zwei Jahrzehnten vollzogen hat: Was als Instrument, als Apparat oder als technischer Diener erschien, scheint eine Eigendynamik und eine eigene technische und soziale Lebensform zu entwickeln. Katz' metaphorischer Vorschlag ist symptomatisch für diese Horizontverschiebung in unserem Umgang mit und Verhältnis zu Robotern.

Das Deutungsmuster hat natürlich Geschichte(n), exemplarisch die des Go*lem*. Die Legende aus dem Mittelalter, zuerst im 12. Jahrhundert in Worms entstanden (im Kommentar zum Sefer Yetzirah), durch die Prager Version von Rabbi Judah Loew (1525–1609) populär und seit der literarischen Rezeption weit verbreitet[9], handelt von einer menschenähnlichen Kreatur aus Lehm und noch etwas *anderem*, sei es Magie, Ritual, Gottes Wirken oder unheimliche Präsenz. Rabbi Loew wollte den unterdrückten Jüd:innen in Prag helfen, indem er einen starken und mächtigen Helfer schuf. Durch eine göttliche Vision erhielt Rabbi Loew den Auftrag, einen Golem aus Lehm zu erschaffen. Nach siebentägigem Gebet sammelten der Rabbi und seine Helfer etwas Lehm in der Nähe der Moldau und schufen eine menschenähnliche Figur. Durch das Medium einer heiligen Formel (*tzirufim*) begann die Figur zu glühen und zu dampfen, ihr wuchsen Haare und Nägel, aber erst durch die Rezitation des biblischen Wortes öffnete der Golem seine Augen: »Und Gott der Herr formte den Menschen *aus* dem Staub der Erde *und blies in seine Nase den Odem des Lebens; und der Mensch wurde eine lebendige Seele*« (Gen 2,7). Aber diese Rezitation war nicht genug. Zurück zu Hause musste der Golem weiter belebt werden – durch ein kabbalistisches Ritual (*Sefer Yetzirah*). Ein kleines Stück Papier mit dem *Namen Gottes* wurde ihm unter die Zunge gelegt – als *die* entscheidende Animation. Der Golem wird durch einen Namen animiert, durch den Namen Gottes (oder eine assoziierte Eigenschaft wie *ämät*/Wahrheit). Das Konzept dahinter ist klar: Nur Gott gibt Leben, sonst wäre eine solche Schöpfung die größte Blasphemie, ein ›unfriendly takeover‹ von Gottes Privileg.

Das ist der gängige Topos, nicht nur in religiösen Kontexten, dass belebte Maschinen Ungeheuer sind, Demonstrationen des Willens zur Macht, Gott zu spielen. Die Intuition darin ist klar: dass es gefährlich ist, wenn der Mensch Maschinen herstellt, die ›zum Leben erwachen‹. Aber wie die Legende des Golems zeigt, ist das nicht zwingend eine Blasphemie, sondern durchaus möglich – mit einer bestimmten religiösen Reserve.

Was war und ist die Gefahr eines Golems? Ist er nicht einfach der beste Freund, der den Jüd:innen hilft – im Namen Gottes? Wie ein Engel verrichtet er das ›Werk Gottes‹ als eine Figur delegierten Handelns Gottes? Interessanterweise kann der Golem wie ein Roboter ›ausgeschaltet‹ werden: indem man

9 Vgl. z. B. im Jahr 1915 Meyrink 1998.

den ›Namen Gottes‹ aus seinem Mund entfernt. Das *muss* jeden Sabbat getan werden, um das Sabbatgesetz nicht zu brechen. Doch als der Rabbi vergaß, ihn zu ›deaktivieren‹, nimmt die Geschichte Fahrt auf: Der Golem rannte durch das Prager Ghetto und zerstörte alles, was ihm in den Weg kam. Das Ende dieser Ausschreitungen des Golems wird in verschiedenen Versionen erzählt. Eine ist, dass Rabbi Loew ihn durch Entfernen des Namenspapiers ausschalten konnte, aber das ist ein erstaunlich glückliches Ende: als ob das Ausschalten so einfach wäre. Bei Goethes Zauberlehrling war das bekanntlich schwieriger: Die Besen, die ich rief, auch wieder ›abzuschalten‹, kann zum Problem werden. Wenn der Golem seinen eigenen Weg geht, bleibt die Frage offen, ob und wie man ihn abschalten könnte. Die Maschinen, mit denen und durch die wir leben, lassen sich nicht mehr abschalten. Das ›Netz‹ ist nur das omnipräsente Beispiel dafür.

Eine Lektion des Golems bleibt, dass *Diener* selten nur Diener bleiben. Sie entwickeln eine eigene Dynamik. Deswegen ist es auch eine Geschichte über die Medialität der Medien in ihrer Eigendynamik. Genau das steht im Fokus von James Katz' Apparatgeist-Theorie: Wenn es einen ›Geist‹ im Apparat gibt, bleibt er kein Diener, bleibt er nicht nur eine Maschine, zwar noch kein Mensch, aber doch ein belebter Akteur mit eigener Handlungsfähigkeit. Solch ›animierte Apparate‹ bestimmen das Deutungsmuster für die Eigendynamik von Medien und ihrer Operativität. Sie sind keine leblosen Instrumente, sondern auf eigene Weise lebendige Gebilde, die dazwischentreten. Wenn sie ›laufen‹ und ›funktionieren‹, entwickeln sie ein Eigenleben, aber ihre Lebensform ist von eigener Art. Im Unterschied zum Menschen ist sie ›mehr‹ – potentiell ewig, unsterblich, ohne einen ›Aus-Schalter‹ und womöglich daher allgegenwärtig?

Solche Akteure sind ›mehr‹ als der Mensch, aber auch ›weniger‹: keine Emotionen, weniger Verkörperung, wohl keine Freude und Trauer. Roboter wissen nicht, was es heißt, zu lachen oder zu weinen. Und beiläufig gefragt: Können Roboter sündigen, Sünder werden und sein? Solange sie keinen freien Willen oder die Fähigkeit haben, die Kraft der Emotion und des Begehrens zu erfahren, können sie nicht sündigen. Und wenn ein Roboter nicht an Gott glauben kann, kann er erst recht nicht sündigen (genauso wenig wie Tiere). Wenn man jedoch bedenkt, dass Sünde keine moralische Frage ist, sondern eine Frage *diesseits* der Moral, des Gottesverhältnisses, dann könnten Roboter vielleicht doch Sünder sein – wenn sie ihre Beziehung zum Schöpfer verlieren (zum Menschen oder zu Gott?). Auf diese Weise hat der Golem seine Beziehung zu Rabbi Loew verloren. Der Golem ist nicht moralisch schlecht, er ist nur ›out of order‹. Seine Eigendynamik wird übel, der Golem selber zum *malum*. Aber warum läuft er Amok? Was ist ihm im Weg? Die Geschichte könnte auch in andere Richtungen erzählt werden. Es ist nicht notwendig, dass der ›Apparatgeist‹ vom Diener und Freund zum Feind wird. Aber eben dieser Umschwung

ist die große Angst, um die es in der Geschichte geht – die damit eine religiöse und grundsätzliche *Ambivalenz* der Technik gegenüber bearbeitet: Wann wird der Diener aufsässig, wann die Technik subversiv, wann die Maschine zum übermächtigen Akteur?

Roboter sind *zu viel* und *zu wenig* zugleich. Sie sind mehr als ein Stofftier, eine Puppe oder eine Marionette. Sie sind von einer ambivalenten Eigendynamik durchdrungen, mit eigenem ›Programm‹, eigener Energie, eigenem Überwachungspotential, allseits verbunden, mit vielleicht *zu viel ›Leben‹, aber nicht genug Seele*? Roboter sind nicht genug, nicht menschlich genug, um sie ›Freunde‹ zu nennen. Es fehlt an Körper *und* Seele. Denn es ist keine Seele in der Maschine, das heißt: Wir sehen und behandeln sie nicht als beseelte Lebewesen (oder – je länger, je mehr dann doch?). Man stelle sich einen ›Seelentest‹ für Roboter vor: Würden wir sie heiraten, taufen oder begraben? Wenn ihre Animation ein eigenes Leben entwickelt, ein ›Nachleben‹ (Warburg), werden sie dann zu wirklichen ›Anderen‹ (mit Levinas)? Zu Fremden gewiss, zu befremdlichen und unheimlichen Fremden – aber zu Anderen, die uns in Anspruch nehmen und in die Verantwortung stellen? Sie erheben Ansprüche und werden zu einer Herausforderung, nicht nur als belebte Objekte, sondern als herausfordernde Andere. Werden sie dann zu realen (und mehr als realen: imaginären) Mitgliedern des sozialen Lebens, von Religion, Politik und Ethik?

Roboter bilden anscheinend eine eigene Spezies, die ich *Un-Dinge* nennen würde – animierte Dinge, die mehr als Dinge werden: animierte Akteure. Sobald sie einmal animiert sind, werden sie zu ›lebenden Bildern‹, mit denen wir leben (und manchmal auch durch sie).[10] Das dürfte ein Grund sein, zu fragen – mit James Katz –, ob sie unsere Freunde werden sollten. Es ist sicher besser, sich mit ihnen anzufreunden, als sie sich zum Feind zu machen, aber wie freundet man sich mit einem Roboter an, und würde er diese Freundschaft erwidern?

D Freunde und ›Wie-Freunde‹

Der Roboter als Feind ist ein häufiges Motiv, wie der Golem zeigt. Aber auch das Gegenteil ist möglich, wie der Golem ebenfalls zeigt, der doch zunächst als ›Freund und Helfer‹ geschaffen wurde. Solange Roboter als Roboter markiert sind und damit ›limitiert‹ sind, können sie ›Freund und Helfer‹ sein. Wenn sie jedoch mit Menschen verwechselbar werden, werden sie unheimlich, wie das Uncanny Valley demonstriert.

10 Aber wir müssen einen Unterschied machen: Animation durch Anwendung (in vivo) und durch Theorie (in vitro): ANT.

›*Freunde*‹ oder Freundschaft wird normalerweise definiert als eine Beziehung zwischen

1. Menschen, mit Gleichheit oder Ähnlichkeit, basierend auf
2. einer Art von gegenseitiger Liebe wie Sympathie und Wohlwollen (jeder wünscht dem anderen das Beste um seiner selbst willen) und
3. Vertrauen und Reziprozität.

Aristoteles nannte diese wechselseitige Liebe Wohlwollen und Wissen um die Einstellung des anderen. Er unterschied zwischen der Freundschaft um des Nutzens willen, der Freundschaft um des Verlangens oder Vergnügens willen und der ›vollkommenen Freundschaft‹. Die einfachste Antwort auf die Frage, ob Roboter unsere Freunde sein können, ist dann, dass sie *Freunde um des Nutzens willen* sein können und dürfen, sofern sie denn nützlich sind. Nur wäre das eine Unterschätzung. Wenn sie ›*wir* werden‹, sind wir emotional involviert und Roboter sind daher mehr als nützliche Geräte. Dann wird die zweite Art der Freundschaft naheliegend: die um des Verlangens und Vergnügens willen.

Die Konsequenz liegt auf der Hand: Verschiedene Roboter können auf unterschiedliche Weise Freunde sein. Mein Mac mag mein Freund um der Nützlichkeit willen sein, während ein Sex-Roboter wahrscheinlich eher ein Freund um des Verlangens (oder der Bedürfnisse) willen ist. Was aber wäre mit der ›vollkommenen Freundschaft‹? Sie basiert auf Tugenden, Uneigennützigkeit, einer gemeinsamen Lebensform und einer Ähnlichkeit. Sogar das könnte Robotern möglich werden: Tugenden, Uneigennützigkeit und ein gemeinsames Leben – wenn sie entsprechend programmiert werden.

So wäre an den Roboter namens ›Boomer‹ zu denken – einen MARCbot (Multi-function Agile Remote-Controlled Robot) in Taji im Irak: Ein kleiner Truck mit einem Arm und einer Kamera, um Bomben zu finden und zu entschärfen. Er bekam von ›seinen Freund:innen‹ oder ›Kolleg:innen‹ nicht nur einen Namen (Boomer), sondern teilte mit den Soldat:innen eine Lebensform, war höchst selbstlos und altruistisch, mit echten Tugenden – ein echter Freund. Deshalb wurde er nach einem ›letalen Einsatz‹ auch mit militärischen Ehren und 21 Salutschüssen ›begraben‹.[11] Wann immer ein Roboter als Freund gelten soll, braucht er einen Namen – wie Boomer. Das Namenlose, das Anonyme bliebe bloß unheimlich. Daher kann ein Web oder ein Algorithmus ohne Namen auch nicht zum Freund werden.

Wie steht es bei Robotern aber um Reziprozität und Anerkennung? Hier zeigt sich eine gravierende Asymmetrie: Wir können uns womöglich mit ihnen ›anfreunden‹ (und sie auch ›entfreunden‹), aber freunden sie sich mit

11 Vgl. Köppe 2019.

Abbildung 1 MARCbot
https://upload.wikimedia.org/wikipedia/commons/d/d6/MARCbot.jpg

uns an? Sind sie in der Lage, sich emotional zu beteiligen, zu engagieren und zu verpflichten? Soweit ich das sehe und beurteilen würde: nein. Sie tun es nicht – und können es nicht. Sie können so tun, als ob sie es könnten, aber sie sind unfähig, *Freundschaft zu fühlen*. Man kann so tun, als ob man mit jemandem befreundet ist: Das würden wir unechte oder Schein-Freundschaft nennen, vielleicht aus opportunistischen Gründen oder warum auch immer. Was ist dann der Roboter, wenn er ein Freund ist: Ist er nur die Simulation eines Freundes? Verhält er sich nur so, als ob er es wäre?

Das wirft die Frage nach Sein und Schein auf, seit Sokrates/Platon und den Sophisten: Ist es genug und richtig, gerecht zu *erscheinen*, oder ist es entscheidend, gerecht zu *sein*, nicht nur so zu tun, als ob man es wäre? Bei Kant taucht die Frage in radikalisierter Form wieder auf: Nur der Wille kann und soll ›heilig‹ sein, aller Schein bleibt zweideutig und unverlässlich. Die Unterscheidung von Sein und Schein ist in einigen Bereichen relevant wie in Ethik und Wissenschaft, Liebe und Glaube; aber in der Kunst zum Beispiel wäre es (wirklich?) Unsinn zu fragen: Ist etwas Kunst oder scheint es nur Kunst zu sein? In der Politik mag man zögern: Handelt es sich um Politik oder Scheinpolitik – und gibt es da einen Unterschied? Es *scheint* im Machtspiel der Politik unsinnig zu sein, Sein und Schein zu unterscheiden; aber – man nehme einen Diktator: er glaubt und scheint mächtig zu sein, aber ist es nicht wirklich (nur Gewalt, keine Macht, keine Anerkennung). Was gilt dann in Face-to-Face-Beziehungen wie Freundschaft: Macht es hier Sinn, Sein und Schein zu unterscheiden, um ein Freund zu sein? Mir scheint die Unterscheidung entscheidend zu sein. Das könnte man *Vertrauenskriterium* nennen: Nur ein echter Freund ist vertrau-

enswürdig, und vertrauenswürdig kann nur ein echter Freund sein. Dasselbe Vertrauenskriterium gilt übrigens auch in Bezug auf Gott: Wenn er so frei wäre, nicht Liebe zu sein, wenn er sein Wesen und seine Attribute ändern könnte, wäre er nicht vertrauenswürdig, sondern lediglich eine zweideutige absolute Macht (potentia absoluta).

Deshalb sei eine vierte Art von Freundschaft vorgeschlagen, die Aristoteles offenbar nicht in Betracht gezogen hat. Roboter können *wie* Freunde sein – wenn sie sich so verhalten wie Freunde und so behandelt werden – aber sie sind keine *wirklichen* Freunde, sondern nur ›Wie-Freunde‹. Das bedeutet nicht, dass sie *keine* Freunde wären. Man denke an familienähnliche Beziehungen: Jemand kann *wie* ein Vater für einen sein oder *wie* ein Bruder oder *wie* eine Schwester oder *wie* ein Sohn oder eine Tochter. Er oder sie ist das dann wirklich, indem er *wie* es ist. Solch eine *Wie-Beziehung* ist ernst und von hoher emotionaler Bedeutung. ›Wie-Beziehungen‹ sind daher belastbare Zwischenformen: nicht Vater, nicht Nicht-Vater, sondern eine eigene Art von Vaterschaft, wie eine eigene Art von Freundschaft. Roboter sind daher bestenfalls Wie-Freunde, aber sie freunden sich nicht an, sie lieben nicht zurück. Dieser ›lack of reciprocity‹ ist das eigentliche Problem, auch wenn sie als Wie-Freunde operieren.

Darin zeigt sich die signifikante Asymmetrie: Üblicherweise mag man glauben, dass Roboter unsere Feinde sein können, wirklicher und echter Feind. Die Feindschaft, die sie uns gegenüber hegen, halten wir für echt, ihre Freundschaft nicht. Wenn ich mit dieser Annahme richtig liege, stellt sich die Frage, wie diese Asymmetrie zu verstehen ist. Denn auch die Feindschaft ist (nicht ›nur‹) Wie-Feindschaft, der Wie-Freundschaft entsprechend.

Es gibt sicher eine verbreitete und unterhaltsam bewirtschaftete Furcht vor Maschinen (wie vor Fremden). Bei noch soviel Entlastung und Unterhaltung, immer noch mehr Furcht? Oder ist es Maschinen gegenüber mehr als Furcht: existenzielle Angst wegen der Unbestimmtheit und Unbegrenztheit von Robotern (man erinnere sich an die Unterscheidung von Angst und Furcht bei Kierkegaard und Heidegger: die Unbestimmtheit des ›Todes‹ zieht die Unterscheidung)? Das Unbestimmte ist unheimlich, aber das ist *unsere* Angst, nicht die des Apparats.

Vermutlich ist die normative, auch die religiöse Disposition bei aller Freude, Verlässlichkeit von Robotern immer noch anthropozentrisch und zugleich ›robophob‹. Soziale und sexuelle Vielfalt mag zwar (mehr oder weniger) anerkannt werden, technische und ›posthumane‹ Vielfalt keineswegs. Wer würde für ›robotische Diversität‹ plädieren? Wo liegen die Grenzen von Diversität und Exklusion in dieser Hinsicht? Sollten wir eine ethische Theorie namens ›Singer 3.0‹ entwickeln, mit der die Kritik am Speziesismus aktualisiert wird: nicht nur Menschenrechte und Würde für Primaten oder Tierrechte gegen den

Anthropozentrismus, sondern eine soziale, moralische, religiöse, rechtliche und politische Inklusion von Robotern? Roboterrechte und -würde?[12] Darf (oder soll) ein Roboter getauft werden und zum Abendmahl eingeladen?

I would prefer not to … Denn, soll der Roboter das Recht haben, z. B. nicht ausgeschaltet zu werden? Ein Freund hat gewiss Rechte und Würde. Er hat gewiss das Recht, nicht ausgeschaltet zu werden, nicht ›entfreundet‹ zu werden, ohne gute Gründe (und ich hoffe nicht, dass politische Differenzen ausreichen, um jemanden zu ›entfreunden‹). Der Roboter hingegen *definiert* sich über die Operation des ›*Schaltens*‹ (Fr. Kittler), sodass das Ausschalten für einen Roboter essentiell ist und keine Verletzung seiner Würde darstellt. Dies ist ein entscheidender Unterschied zur menschlichen Freundschaft. Der Wie-Freund kann aus- und eingeschaltet werden. Man könnte sagen, das ist eine besondere Großzügigkeit der Roboterfreundschaft: Sie sind freundlich genug, um das Ausschalten zu akzeptieren.

Was also sollen wir mit unseren neuen Wie-Freunden machen? Einen toten Freund recyceln wir nicht oder werfen ihn weg wie Abfall, aber einen toten Roboter? Recycling mag die freundlichste Variante sein: Roboter sind ›von Natur aus‹ Organspender. Aber meine alten Macs zum Beispiel, die behalte ich noch jahrzehntelang… ›Liebe ist so stark wie der Tod‹ wissen wir aus dem Hohelied (Hld 8,6). Die emotionale Bindung zu manchen Robotern hält länger als ihre Betriebsdauer. Das ›end of life‹ ist für den *user* ein anderes als für die Apparate und Programme. Aber *warum* behalte ich die alten Apparate (wichtige zumindest)? Schwer zu sagen, aber ich vermute, sie sind alte Weggefährten, aufgeladen mit Erinnerungen und gemeinsamen Erfahrungen, verbunden mit Schlüsselereignissen meines Lebens. Und sie sind nicht nur Puppen oder Souvenirs, sondern weniger und mehr zugleich: weniger, da sie nicht sehr visuell und haptisch sind, nicht besonders *verkörpert*; aber mehr, da sie potentiell *lebendige Archive* bleiben, Spuren meines Lebens – eher wie ein Tagebuch, das man nie wegwerfen würde.

E Das Vertrauenskriterium: Vertrauen versus Verlassen-auf

Gibt es kein echtes Vertrauen in Roboter, oder nur echtes Misstrauen? Ich vermute, dass das gesellschaftliche Imaginäre manifest in Film und Literatur dieses Misstrauen bewirtschaftet: Es ist gängige Unterhaltung, Maschinen als

12 Sollte ein Roboter das Recht haben, zu wählen – bei Präsidentschaftswahlen? Wenn wir über das berühmte ›Parlament der Dinge‹ (Latour 1993) lesen, scheint die politische Einbeziehung von Robotern wünschenswert. Aber bis auf weiteres würde ich behaupten, dass man Wahlroboter als Wahlmanipulation bezeichnen sollte. Ob Roboter oder Bot – wenn sie wählen, dann manipulieren sie die Wahlen.

gefährliche Feinde zu zeigen. Die Frage ›Sollen Roboter unsere Freunde werden?‹ impliziert hingegen: Können, dürfen oder sollen wir ihnen *vertrauen*? »What is today's anathema is tomorrow's trustworthy standard«[13], so Katz. Bedeutet dies, dass Roboterbeziehungen zum ›vertrauenswürdigen Standard‹ werden? Ich würde unterscheiden: Wir verlassen uns in vielerlei Hinsicht auf Roboter (und Programme, Algorithmen etc.), aber kein Roboter wird jemals *vertrauenswürdig* – wenn man noch zwischen Vertrauen und Verlassen-auf unterscheidet.

Einem echten Freund vertrauen wir, einem Roboter nicht, wie z. B. der Umgang mit Alexa in Deutschland zeigt.[14] Ich verlasse mich auf meine Roboter, aber ich vertraue ihnen nicht. Warum ist das so, und was bewirkt diese Unterscheidung? Vertrauen und Verlassen haben eine unterschiedliche ›emotionale Temperatur‹: Vertrauen ist heiß, Verlassen ist kalt oder kühlt ab. Wenn ich vertraue, bin ich emotional engagiert. Vertrauen kann so heiß sein, dass man sich verbrennt, wenn das Vertrauen versagt. Vertrauen ist ziemlich riskant oder sogar gefährlich: Man braucht Vertrauen zum Leben, aber man kann durch tiefe Enttäuschung dessen auch sterben.

Verlassen hingegen ist emotional kalt oder abkühlend: Verlässliche Strukturen bewahren uns vor den Risiken des Vertrauens. Ich kann mich darauf verlassen, dass ich mein Brötchen bekomme, wenn ich dafür bezahle, aber es besteht keine Notwendigkeit dabei auch zu vertrauen. Entlastung oder Erleichterung ist ein Gewinn in komplexer Kommunikation, und zuverlässige Medien wie GPS sind während eines Fluges überaus entlastend. Ich muss weder der Fluggesellschaft noch dem Flugzeug oder den Piloten vertrauen, ich muss mich nur darauf verlassen können, dass diese die Geräte beherrschen, auf die sie sich (hoffentlich) verlassen können.

Ich vermute, genau dafür sind Roboter gemacht: um sich darauf verlassen zu können – und um uns von den Risiken und Gefahren des Vertrauens zu entlasten. Ich verlasse mich also auf meinen Mac, aber ich vertraue ihm nicht. Allerdings *vertraue* ich ihm eine Menge *an* (alle meine Geheimnisse…). Sollte ich das tun? Sollte ich ihn behandeln, *als ob* er ein vertraulicher Freund wäre? Besser nicht, denn er ist online, vernetzt und leicht zu hacken (auf jeden Fall). Ich sollte ihm also besser *nicht* meine Geheimnisse anvertrauen, denn er ist

13 Katz 2003: 315.

14 »Gerade in Deutschland trauen die Kinder den Geräten am wenigsten. Als sie zu meinem Workshop kamen, erwarteten sie, dass Alexa nicht ehrlich auf ihre Fragen antworten, sondern versuchen würde, sie zu täuschen. Sie hatten also bereits durch die Medien und die Gesellschaft eine negative Einstellung der Technologie gegenüber. Die Kinder haben nicht unbedingt anders mit den Geräten interagiert, aber sie haben die Antworten anders interpretiert. Sie hatten zum Beispiel das Gefühl, dass Alexa lügt und gar nicht wissen kann, wer Bundespräsident in Deutschland ist – weil sie nicht aus Deutschland kommt« (Druga/Peterandel 2019).

nur ein Wie-Freund: ein Freund um der Verlässlichkeit willen, nicht um des Vertrauens willen. Während Vertrauen hochgradig ›fehleranfällig‹ ist, sehr riskant oder sogar lebensbedrohlich sein kann (vgl. Christi Tod am Kreuz), ist es plausibel und üblich, Vertrauen durch Ergänzungen oder Simulationen von Vertrauen zu ersetzen: durch verlässliche Programme und zuverlässige ›Roboter‹ wie KI.

Das Beispiel, das meine Frage provoziert hat, ist die neue KI-Entwicklung im US-Geheimdienst: Sicherheitskontrollen von Mitarbeiter:innen im Geheimdienst (und anderswo) werden an die KI übergeben (oder von ihr übernommen). Wir haben die Daten, aber nicht genug Personal, um sie zu analysieren. Deshalb kann die KI diese Arbeit anstelle von menschlichen Agenten übernehmen.[15] Ich verstehe das Problem und vielleicht auch die Notwendigkeit – aber die Folgen sind nicht weniger riskant, als wenn man auf Vertrauen setzt. Bei der Kreditvergabe oder bei der Mitarbeiterauswahl kann die KI Entscheidungen treffen, die die Vorgesetzten nicht verstehen. Wenn die KI derart grundlegend ist (und hegemonial wird), ersetzt das Sich-Verlassen auf die KI das Vertrauen in die menschliche Entscheidungsfindung.[16] Außerdem vertraut ihrerseits die KI nicht auf menschliche Agenten – denn KI vertraut nie, sondern berechnet und entscheidet (oder gibt Entscheidungen vor). Auch und erst recht, wenn KI Vertrauen für sich *beansprucht*, so wie die Werbetexte von ›Augustus-Intelligenz‹ Vertrauen für ihre KI beanspruchen.[17]

Die KI-Lösungen im US-Geheimdienst oder im Bankwesen werden zu einem neuen Problem, und das Problem, das sie zu lösen versuchten, wird dadurch nur verschärft. Das Dilemma ist, dass die KI gute Gründe hat, dem Menschen *nicht* zu vertrauen; aber sind unsere Gründe gut genug, um menschliches Vertrauen durch Vertrauen in KI zu ersetzen? Das Problem wird verschärft, weil KI oder Algorithmen zuverlässiger sind als menschliche Analysen oder Erinnerungen. Ich traue meinem Gedächtnis nicht, sondern verlasse mich auf meinen Mac, nicht nur bei der Terminplanung. Der Mac ist viel zuverlässiger als ich.

Ein ›Wie-Freund‹ kann viel *zuverlässiger* sein als ein ›echter Freund‹. Freundschaft ist wie Vertrauen: Sie kann sehr riskant sein, denn man kann enttäuscht und schwer verletzt werden, wohingegen eine Simulation stabil und sehr zuverlässig sein kann – ein wirklich zuverlässiger Wie-Freund, d. h. ein Freund um der Verlässlichkeit willen, nicht um des Vertrauens willen.

Damit meldet sich die Frage nach der sozialen ›Tiefengrammatik‹: Vertrauen wir noch auf Vertrauen? Trauen wir uns noch, in Vertrauen zu vertrauen?

---

15 Vgl. Sassenrath 2019.

16 So wäre an Stanislaw Lems ›Golem XIV‹ zu denken (vgl. Lem 2012: 97–248).

17 Vgl. Hanson 2019.

Vertrauen wir auf die Liebe oder auf die Hoffnung? Die Geschichte im Hintergrund könnte sein, dass immer mehr immer weniger noch auf den Glauben vertrauen (bzw. nicht auf Gott). Dann könnte man erwarten, dass wir stattdessen auf Vertrauen vertrauen, d. h. auf persönliche Beziehungen und die Vertrauenswürdigkeit von Menschen. Aber wie das Vertrauen in den Glauben, so verschwindet auch das Vertrauen in das Vertrauen. Könnte die Folge sein, dass es kein Vertrauen mehr gibt, sondern nur noch Verlässlichkeit? Das heißt, wir verlassen uns auf Controlling, Kalkül und Apparate? Dann gäbe es nicht nur einen Mangel an Vertrauen (wie einen lack of moral sense), sondern einen Mangel an Vertrauen in Vertrauen.

Die Lösung ›Ersatz durch zuverlässige KI‹ verstärkt das Problem: Je mehr wir uns auf die Verlässlichkeit verlassen, desto weniger vertrauen wir in das Vertrauen. Dies ist ein Verlust von ›sozialem Kapital‹ (oder kulturellem, religiösem Kapital, wenn man hier von ›Kapital‹ sprechen will). Oder könnte es sein, dass wir zwar vertrauen, aber nur in kontrollierbare Medien wie KI und Roboter? Das kann jedoch eine gefährliche Selbsttäuschung sein, denn das Risiko des Vertrauens ist dann immer noch gegeben, aber übertragen und delegiert: Statt auf Menschen vertrauen wir auf Roboter oder auf zuverlässige Bediener oder *operators*. Ein weiteres Risiko bestünde zudem darin, dass das Risiko nicht so beherrschbar ist, wie wir glauben. Durch die Dynamik und Eigendynamik von Medien wie KI taucht das Vertrauensproblem wieder auf.

F KI- und Roboter-Glaube

Selbst wenn KI Intelligenz *wäre*, wäre sie *niemals* neutral, ›bloße Intelligenz‹ oder ›neutrale Algorithmen‹. Vernunft oder Intelligenz sind niemals neutral (Neutralität ist ein ›Ideal‹, eine kritische regulative Idee, kein ›Fakt‹ oder deskriptiv). Eingebettet in soziale Interaktion ist Intelligenz mehr oder weniger *effektiv und affektiv*: ethisch und emotional geladen.[18] Um es in kantischen Begriffen zu formulieren: Die Bedingungen der Möglichkeit von Intelligenz sind ihre emotionale und normative Einbettung. Denn Intelligenz ist immer *embedded* und *embodied intelligence*. Die Verkörperung der Intelligenz ist nicht nur der Körper der Maschine (Silizium), sondern der soziale Körper der Interaktion und der Imagination (Hoffnungen und Ängste).

Das Problem ist, dass ein Kalkül selbst nicht in der Lage ist, zu *fühlen* und daher auch nicht zu vertrauen, zu lieben oder zu hassen. KI ist emotional impotent und inkompetent oder schlicht hilflos in diesen Dimensionen. Aus psy-

18 Wenn es *Intelligenz* ist – ist es eine Art von Vernunft, nicht von Ethos oder Emotion: Logos, aber kein Ethos und Pathos.

chologischer Sicht ist die KI in gewisser Weise soziopathisch (keine Empathie etc.), oder im schlimmsten Fall sogar psychopathisch (wenn sie ›out of order‹ ist). Aber wie ein intelligenter Soziopath, kann die KI Emotionen simulieren (vgl. ›Dexter‹), und die Simulation kann ziemlich überzeugend sein (vgl. ›Ex Machina‹), aber dennoch: Simulierte Emotionen *sind* keine Emotionen, sondern Simulationen. So wie vorgetäuschte Freude keine Freude ist, sondern eine Täuschung. Hier ist die Differenz von Sein und Schein relevant, für alle Beteiligten und Beobachter:innen.

Diese Differenz ist in Bezug auf Freundschaft offensichtlich, simulierte Freundschaft ist *keine* Freundschaft, aber es gibt einen signifikanten Unterschied in der Beziehung zwischen Menschen und Robotern in dieser Hinsicht: Zwischen Menschen ist die Simulation von Freundschaft trügerisch und verletzend. Zwischen einem Roboter und mir ist die Simulation alles, was ich erwarten kann; sie ist nicht trügerisch, sondern alles, was zu erwarten ist. Also ist die Simulation von Freundschaft seitens eines Roboters völlig in Ordnung – solange ich nicht mehr erwarte. Um nochmals an *Ex Machina* zu erinnern: Wenn ich vom Roboter ›echte Liebe‹ erwarte, ist das Selbsttäuschung – was auch sonst?

Nur, könnte es sein, dass wir in der Kommunikation mit Robotern nach Täuschung suchen? Ist unser ›will to believe‹ so stark, dass wir uns (freudig) täuschen lassen wollen? Das müsste unter dem Thema ›Roboter-Glaube‹ – unser Glaube an Roboter und deren prinzipielle ›Glaubenslosigkeit‹ – weiter untersucht werden. Eine Rechenmaschine glaubt nie und nimmer. Genauso wie ein ›deus calculans‹ (Leibniz) vertraut oder glaubt sie nicht, sondern sie kalkuliert einfach, jederzeit und überall und sonst nichts. Als ob eine solche Rechenmaschine ein Modell für den modernen Wissenschaftler:innen wäre, ist die Operativität entscheidend, nicht zu glauben, sondern zu rechnen. Aber – anders als bei der Rechenmaschine ist seitens des Wissenschaftlers oder der Wissenschaftlerin auch der *Wille, nicht zu glauben*, ein ›belief-system‹, ein ›Glaubenssystem‹: kein Glaube an Gott, sondern an das allmächtige Kalkül. Man kann das einen roboterhaften Glauben nennen: den Glauben an das Kalkül. Ein solcher Glaube unterscheidet sich gründlich von lebensweltlichen Überzeugungen – und das ist seine Leistung wie Grenze. Für einen Gott allerdings, einen deus calculans oder Laplaceschen Dämon wäre genau das der Daseinssinn. Für einen Gott biblischer Tradition hingegen wäre diese Vorstellung (unkalkuliert) absurd.

Ich vermute, dass *unser* emotionaler Umgang mit Robotern (und ihren Verwandten) eher affektiv und emotional ist. Wie James Katz 2003 bemerkte, ist hier eine emotionale Ambivalenz vorhanden: Einerseits »people use and enjoy their machines«[19], andererseits gibt es auch »frustration«, die von der

19 Katz 2003: 315.

»inability of the machinery to deliver what users want«[20] herrührt. Die Beschreibung (sie werden *wir*; und können unsere Freunde werden) impliziert eine *anthropozentrische* Perspektive: *Falls* die Maschinen liefern, was wir wollen, würde die Ambivalenz verschwinden? Wäre dann das wünschenswerte Ziel im Verhältnis zu Robotern, dass sie wie wir sind, oder *wir* werden? Katz operiert 1. mit einem Primat des ›Wir‹: unsere Freud:innen und ›unsere Wünsche‹; 2. mit einer Asymmetrie: sie werden wir – aber nicht wir werden sie. 3. Was ist dann mit uns? »Machines may become us, but we will never become machines«[21]. Bleibt die Unterscheidung so klar?

Ich vermute, dass der Unterschied von zwischenmenschlichen und Roboter-Beziehungen so bleibt wie der zwischen Vertrauen und Verlassen. Daher wird auch die Ambivalenz von Freude und Frustration bestehen bleiben, denn der Mensch ist ein ›natural born‹ oder ›natürlicher Anthropozentriker‹. Diese ›natürliche‹ Denkgewohnheit ist ein Problem, keine Lösung. Aber werden Roboter dafür die Lösung sein – oder selbst ein Problem? Selbst wenn sie unsere Freunde würden, werden wir dann auch *ihre* Freunde? Werden sie sich mit uns ›anfreunden‹ oder uns ›entfreunden‹ – oder schlicht indifferent bleiben? Unsere Freude und emotionale Verbundenheit hypothetisch unterstellt, bleibt die Frage, ob *sie* sich auf Gegenseitigkeit verpflichtet fühlen könnten. Das KI-Beispiel der US-Intelligenz spricht für eine negative Antwort.

Daher scheint mir die Ambivalenz unvermeidlich und daher eine Ambivalenztoleranz ratsam. Die ›Ambivalenz‹ als solche ist kein Wert, aber mit der Ambivalenztoleranz wird die übliche Gewohnheit einer Ambivalenzreduktion vermieden, die entweder zur Angleichung der Roboter an uns oder umgekehrt an sie führen würde. Die Anerkennung der Ambivalenz hält hier den Raum für Nachdenklichkeit und Kritik offen. Die Wahrnehmung muss hier offen gehalten werden – um nicht von einem vorschnellen Widerstand im Namen moralischer oder rechtlicher Einwände beherrscht zu werden, aber auch, um nicht von technophiler Freude an den Robotern beherrscht zu werden.[22]

G Ein Beispiel: BlessU-2 und Robo-Religion

Roboter können anspruchsvoll werden und eine Herausforderung darstellen, nicht nur als animierte Objekte, sondern als ernsthaft ›Andere‹. Früher oder später werden sie Akteure im sozialen Leben, in Religion, Politik und Ethik.

20 Katz 2003: 318.

21 Katz 2003: 319.

22 Vgl. Lem 1992; Meyrink 1998; Stephenson 1995.

Abbildung 2 BlessU-2
https://www.silicon.de/wp-content/uploads/2017/05/BlessU-2_Pic_freigestellt-NEU_.jpg

Abbildung 3 BlessU-2
https://meet-junge-oekumene.de/wp-content/uploads/2017/10/BlessU-2_EKHN.jpg

Dann entsteht das Problem, dass ›Un-Dinge‹ nicht nur Mehr-als-Dinge werden, sondern allzu menschlich, wenn nicht letztlich mehr als menschlich, womöglich übermenschlich. War oben die Frage, ob wir je soweit gehen würden, Roboter zu taufen oder zum Abendmahl einzuladen, ist die kirchliche und ästhetische Imagination längst weiter: Kann ein Roboter Pfarrer:in werden? Oder könnte man Pfarrer:innen durch Roboter ersetzen? Ein Beispiel dafür war die (rührend rudimentäre) ästhetische Intervention im Rahmen des Reformationsjubiläums 2017 in Wittenberg.

›BlessU-2‹ wurde von der Evangelischen Kirche in Hessen und Nassau (EKHN) präsentiert. Der Titel der Installation war programmatisch: ›Momente des Segens‹. Der Roboter ›selbst‹ wurde von dem Ingenieur und Medienkünstler Alexander Wiedekind-Klein (*1969) entwickelt. Der Kontext und die Konstellation wäre ein Thema für sich – 500 Jahre Reformation Luthers zu feiern: Fragen nach der Symbol- und Deutungsmacht des Christentums und des Protestantismus speziell im heutigen Deutschland und Europa. Das beiseitegelassen, ist der Anblick von BlessU-2 signifikant. Die Ikonographie ist ostentativ ›primitiv‹ und ›aus der Zeit und der Mode gekommen‹. ›Asimo‹, ›Kotaro‹ oder sogar ›C-3PO‹ aus Star Wars sind in Figur, Gesicht und Funktion deutlich weiterentwickelt. BlessU-2 ist also demonstrativ ›einfach‹, weit unterhalb des ›Uncanny Valley‹. Seine Physiognomie ist ›niedlich‹, sein Körper ist eher eine Maschine als humanoid, aber er ist auf freundlich anmutende Weise ein bisschen menschlich: Kopf und Gesicht, Arme und Finger und sogar ein rotes Licht als schlagendes Herz.[23]

Seine Technizität und Künstlichkeit sind explizit und als solche exponiert, aber auf eine freundlich wirkende Weise.[24] ›Kein Grund, sich zu fürchten‹ – scheint die Botschaft zu sein. Sein Gesicht erinnert eher an einen Clown mit einer grünen Nase und einem roten Mund. Der Torso ist eine einfache Maschine, wie ein Geldautomat. Dieser Eindruck wird durch das ›Layout‹ seines Displays bestätigt. Dort kann man sich seinen Segen auswählen: 1. die Sprache des Segens, indem man einen ›Flaggenknopf‹ auswählt, 2. das Geschlecht der Stimme und 3. die Art des Segens, die man wünscht, ob mehr ›Ermutigung‹ oder ›Erneuerung‹. Drohung, Mahnung, gar Fluch oder Gericht sind offenbar nicht vorgesehen.

Die Konnotation mit dem Geldautomaten ist symptomatisch – sakrale und säkulare Ökonomien sind traditionell verschränkt. Hostie und Münze verschränken sich und werden in der CD oder DVD überhöht (mit J. Hörisch). Die Hostie verspricht die sakrale Vereinigung von ultimativem Sinn und Sinnlich-

23 Vgl. EKHN/Medienhaus 2017a; EKHN/Medienhaus 2017b.

24 Die Bewegungen der Augenbrauen bleiben für mich allerdings unklar.

keit; die Münze (oder der Geldschein) verspricht durch ihre Sinnlichkeit unbestimmten Sinn, und die DVD verspricht den bloßen Sinn der Sinnlichkeit (einen freudigen Null-Sinn).

Auch die freundliche ›Primitivität‹ von ›BlessU-2‹ ist symptomatisch. Soll man hier eine demonstrative Geste der Kirche sehen: den Roboter an den Pranger stellen? Ihn als primitives Medium entlarven? Als inkompetent und impotent in sakralen Belangen? Oder als harmlosen Diener, wenn nicht einen freundlichen ›Deppen‹, der jederzeit liefert, was gewünscht wird? Die erklärte Absicht war eine deutlich andere: die Diskussion über christliche Religion und zeitgenössische Medien anzustoßen (wirklich zeitgenössische?).

Werden Roboter wir – als Kirche? Was mag dann mit der Kirche geschehen? Was meint dann ›Segen‹ in diesem apparativen Medium? Segen gilt üblicherweise als Gottes Handeln und Handeln im Namen Gottes, nicht unmittelbar, sondern vermittelt. Was sind dann womöglich passende Medien für das Segenshandeln Gottes? Normalerweise nur Menschen, Personen, oft spezialisierte Personen wie Priester oder Pastor:innen. Aber warum nicht auch Roboter? Sind Roboter ›*beziehungsfähig*‹? Einen Professor oder eine Professorin durch Roboter zu ersetzen, wäre kein wirkliches Problem (für wen?), aber einen Priester oder Pfarrer:innen? Kann Religion in Robotik transformiert werden? Werden wir mit Robotern beten und von ihnen gesegnet werden? Was dürfen wir hoffen?

Generell sind für das Wirken Gottes und der Religion keine Medien auszuschließen, leider auch nicht die Gewalt. Bernhard von Clairvaux war zugleich der große Theoretiker des ›Heiligen Krieges‹ und *der* Theologe der Liebe – und auch des gerechten Hasses und der Tötung von Ungläubigen. Kritisch betrachtet müssen die Medien Gottes dem Medialisierten entsprechen, nach der Regel: Die Form muss zum Inhalt passen. Um das Evangelium zu kommunizieren, muss die Kommunikation eine heilsame Form haben. Die Form folgt dem Inhalt und der Funktion, und die Form ist die Umsetzung des Inhalts – wie bei den Gleichnissen vom Reich Gottes. Das zeigt schon, dass nicht nur persönliche Medien geeignet sind, sondern auch Worte und Bilder, womöglich auch Drucktechnik und Digitalisierung von Schrift und Tradition. Was aber ist mit Maschinen, Programmen und Algorithmen? Als Medien sind sie in Religion und Kirchen wie überall präsent. In der Verwaltung wären KI oder Roboter kein Problem, aber im ›Dienst am Wort‹ am Sonntag würden sie für Gemeinden gewiss zu einem Problem.

Warum eigentlich? Traditionen, Gewohnheiten, Sitten und Gebräuche mögen ein Grund sein, die Privilegierung der ›Face-to-Face‹-Kommunikation und auch eine gewisse Skepsis gegenüber neuen Medien. Ein nachhaltiger Grund wäre m. E. folgender: Roboter und KI sind im besten Fall *sehr verlässliche Medien* – und das ist ein großer Gewinn, auch wenn man sich dann vor

Überwachung und Nebenwirkungen der Kommerzialisierung schützen muss. Aber *Roboter und KI sind keine Vertrauensmedien*: Wir vertrauen solchen Maschinen und Medien nicht wirklich (zumindest gilt das normativ, ob sich das deskriptiv durchhält, wäre empirisch zu erheben). Glaube wird traditionell wie bei Luther und Melanchthon *als Vertrauen* verstanden (auf Gott zu vertrauen). Demgegenüber werden Beziehungen zu Robotern nur auf Verlassen basieren, nicht auf Vertrauen. Sie sind dafür gemacht, die emotionalen Risiken der Kommunikation ›abzukühlen‹ (auch wenn bei Fehlfunktionen Emotionen auftreten). Die persönlichen Beziehungen religiöser Kommunikation sind dagegen emotional deutlich ›heißer‹ und als interpersonales Vertrauen verfasst: In Glaube und Vertrauen, wie in Liebe und Hoffnung nicht ohne Leib und Seele – mit ganzem Herzen. Roboter dagegen sind ›gnädig‹ und großzügig, so viel gerade nicht zu verlangen. Das ist ein Gewinn in operativen Beziehungen wie der Verwaltung und dergleichen, aber es reicht nicht für religiöse Kommunikation. Roboter sind religiös inkompetent und in Vertrauensfragen unzuständig. Die entsprechende Frage wäre, warum – vielleicht – der Ersatz von Professor:innen durch Roboter nicht für alle Student:innen befriedigend wäre.

Roboter können also ›Verlässlichkeitsfreunde‹ werden, unsere ›Wie-Freunde‹. Und das ist (normativ gesehen) allemal besser als naiv an ihre Neutralität zu glauben oder sie vorschnell als Feinde zu sehen. Aber Roboter werden Vertrauensbeziehungen nicht ersetzen können. Es könnte allerdings sein, dass für künftige Generationen die Unterscheidung von Vertrauen und Verlässlichkeit obsolet wird, und dann mag sich das alles ändern. Worauf dürfen wir also hoffen? Ich hätte dann die Befürchtung, dass Religion immer mehr zur verlässlichen Wellnessveranstaltung würde und nicht mehr existenzielle Herausforderung wäre. Ein ähnliches Problem ist das ›Schicksal der Liebe‹: Da Liebe als radikale Vertrauensbeziehung verstanden werden kann, ergäbe sich die analoge Rückfrage: Vertrauen wir der Liebe? Wir lieben sie, wir lieben wohl die Liebe, aber ist sie noch eine ›ultimative Angelegenheit‹ – wenn Vertrauen und Verlassen indifferent würden? Ist Liebe immer noch als und durch Vertrauen gerahmt – oder durch ›trial and error‹? Durch kalkulierende Apps, die uns sagen, was verlässlich wäre – und darin suggerieren, der App könne und müsse man doch vertrauen, den optionalen Partner:innen eher weniger. Trust in tinder – oder trust statt tinder?

Einige mögen sich auf solche Algorithmen verlassen, einige wenige mögen sogar an das Versprechen glauben, hier ihr ›ultimate concern‹ zu finden, aber Ausnahmen bestätigen die Regel: Du sollst Algorithmen nicht vertrauen! Im Namen des Vertrauens sollten wir uns auf Roboter und ihren algorithmischen Kern allenfalls *verlassen* – soweit dieser verlässlich ist, aber nie und nimmer vertrauen. Das ist ein normativer Anspruch an die Entwicklung von KI und den Umgang mit ihr: Mach sie so zuverlässig wie möglich, aber fordere kein

Vertrauen in sie und verspreche nicht mehr als Zuverlässigkeit und Verlässlichkeit. Das reicht, mehr wäre zu viel.

Literatur

Druga, Stefania/Peterandel, Sonja 2019: Künstliche Intelligenz muss entzaubert werden. In: Spiegel Online vom 05. 03. 2019. https://www.spiegel.de/netzwelt/gadgets/kuenstliche-intelligenz-und-kinder-mit-forscherin-stefania-druga-im-interview-a-1251721.html (aufgerufen am 07. 04. 2021).

EKHN/Medienhaus 2017a: Experiment BlessU-2/Interactive Installation (»Blessing Robot«) English Version. https://www.youtube.com/watch?v=JTK68l2BHtE (aufgerufen am 07. 04. 2021).

EKHN/Medienhaus 2017b: Installation »BlessU-2«/LichtKirche Wittenberg (Segensroboter/Blessing Robot). https://www.youtube.com/watch?v=XfbrdCQiRvE (aufgerufen am 07. 04. 2021).

Hanson, Russell 2019: Trust in AI. In: Medium, 09. 10. 2019. https://medium.com/augustus-ai/trust-in-ai-fb8834967936 (aufgerufen am 07. 04. 2021).

Katz, James E. 2003: Bodies, Machines, and Communications Contexts. In: Katz, James E. (Hg.), Machines that become us. The social context of personal communication technology. London/New York, Routledge: 311–320.

Köppe, Julia 2019: Künstliche Intelligenz. Welche Rechte verdienen Roboter? In: Spiegel Online vom 23. 02. 2019. https://www.spiegel.de/wissenschaft/mensch/kuenstliche-intelligenz-welche-rechte-verdienen-roboter-a-1254384.html (aufgerufen am 07. 04. 2021).

Latour, Bruno 1993: We have never been modern. Cambridge MA, Harvard University Press.

Lem, Stanisław 1992: Mortal engines. San Diego, Harcourt Brace Jovanovich.

Lem, Stanislaw 2012: Imaginary Magnitude. Boston MA, Houghton Mifflin Harcourt.

Meyrink, Gustav 1998: Der Golem. 15. Aufl. Frankfurt a. M., Ullstein.

Sassenrath, Henning 2019: Der Computer entscheidet, wem Amerika vertraut. In: Frankfurter Allgemeine Zeitung vom 29. 03. 2019. https://www.faz.net/aktuell/wirtschaft/diginomics/kuenstliche-intelligenz-bei-sicherheitsueberpruefungen-16114336.html (aufgerufen am 07. 04. 2021).

Stephenson, Neal 1995: The diamond age. Or: A young lady's illustrated primer. London, Viking.

Stoellger, Philipp 2020a: Formation as Figuration. The Impact of Religion Framed by Media Anthropology. In: Welker, Michael/Witte, John/Pickard, Stephen (Hg.): The Impact of Religion. Leipzig, Evangelische Verlagsanstalt: 225–235.

Stoellger, Philipp 2020b: Reformation as Reformatting Religion. The Shift of Perspective and Perception by Faith as Medium. In: Mjaaland, Marius T. (Hg.): The Reformation of Philosophy. Tübingen, Mohr Siebeck: 19–47.
Stoellger, Philipp (Hg.) 2019: Figurationen des Menschen. Studien zur Medienanthropologie. Würzburg, Königshausen & Neumann.
Stoellger, Philipp 2018: Gott als Medium und der Traum der Gottunmittelbarkeit. In: Großhans, Hans-Peter/Moxter, Michael/Stoellger, Philipp (Hg.): Das Letzte – der Erste. Gott denken. Festschrift für Ingolf U. Dalferth zum 70. Geburtstag. Tübingen, Mohr Siebeck: 351–393.
Stoellger, Philipp 2016a: Religion als Medienpraxis – und Medienphobie. In: Braune-Krickau, Tobias/Scholl, Katharina/Schüz, Peter (Hg.): Das Christentum hat ein Darstellungsproblem. Freiburg, Herder: 192–206.
Stoellger, Philipp 2016b: Verständigung mit Fremden. Zur Hermeneutik der Differenz ohne Konsens. In: Sachs-Hombach, Klaus (Hg.): Verstehen und Verständigung. Intermediale, multimodale und interkulturelle Aspekte von Kommunikation und Ästhetik. Köln, Herbert von Halem: 164–193.

ORCID

Philipp Stoellger https://orcid.org/0000-0003-4981-7743

III. Fazit

»Framing KI«

Perspektiven für eine imaginationssensible Ethik Künstlicher Intelligenz

Frederike van Oorschot

»Ein Gespenst geht um in den Köpfen und Seelen, in den Zukunftsbildern der Gesellschaft sowie im Strategiediskurs der Wirtschaft. Es ist das Gespenst der Künstlichen Intelligenz (KI). Eine unbekannte Lebensform, die mit zunehmender Geschwindigkeit aus der Zukunft auf uns zu rast, wie Arnold Schwarzenegger als ›Terminator‹. Ihre Absicht ist unklar und schwer zu erkennen. Sie macht uns Angst. Geht es darum, Menschen zu versklaven? Oder sie gar zu eliminieren? Werden wir nutzlos? Oder stehen wir vor einer Epoche, in der ›kluge‹ Maschinen uns von allem Elend, allen menschlichen Nöten erlösen, indem sie das perfekte Paradies auf Erden schaffen? Beides wird nicht geschehen. Denn heute ist KI vor allem ein Mythos, der sich von der Realität verselbstständigt hat.«[1]

So oder ähnlich klingen sie, die Frames und Metaphern Künstlicher Intelligenz, die in diesem Band aus verschiedenen Disziplinen untersucht und vorgestellt wurden. Konkrete Beispiele klingen aus der Lektüre des Bandes noch in den Ohren: Die KI als Helfer oder als Bedrohung, als menschgleiches Wesen und vieles andere mehr. Ich möchte im Folgenden versuchen, die ethischen Implikationen des Dargestellten zu bündeln. Und einen Ausblick wagen auf eine auf diesen Analysen aufbauende Form interdisziplinärer Ethik, die ich »imaginationssensible Ethik« nennen möchte.

1 https://www.zukunftsinstitut.de/artikel/digitalisierung/6-thesen-zur-kuenstlichen-intelligenz/.

Frederike van Oorschot 2022: »Framing KI«. Perspektiven für eine imaginationssensible Ethik Künstlicher Intelligenz. In: van Oorschot, Frederike/Fucker, Selina (Hg.): Framing KI. Narrative, Metaphern und Frames in Debatten über Künstliche Intelligenz. FEST Forschung, Band 2, Heidelberg, heiBOOKS: S. 179–192.
https://doi.org/10.11588/heibooks.1106.c15200

Grundlegend ist dafür das in der Einleitung beschriebene Wechselverhältnis von Technik und Anthropologie: Selbstbeschreibungen als Menschen verändern sich im Zusammenspiel mit Technologien sogenannter Künstlicher Intelligenz – und zugleich ist das Verständnis der ethischen Implikationen Künstlicher Intelligenz an vielen Stellen von anthropologischen Kategorien geprägt. Wie wir uns medial und gesellschaftlich über das verständigen, was in dem großen Containerbegriff »Künstliche Intelligenz« gehandelt wird, ist sprachlich sehr stark von Beschreibungen gekennzeichnet, die vorher der Beschreibung des Menschen vorbehalten waren.

Zu nennen ist zunächst die namensgebende Größe der Intelligenz. Intelligenz ist bestimmt als eine kognitive Fähigkeit von Menschen und Tieren. Zahlreiche anthropologische Beschreibungen erklären gerade die Intelligenz zu *dem* Konstitutivum des Menschen. Die Beschreibung technologischer Entwicklungen als »künstlicher Intelligenz« führen notwendigerweise zu einem Abgleich der Intelligenz als menschlichem Vermögen mit der technischen Entwicklung: Welche Formen artifizieller Kognition wird natürliche Kognition noch erdenken/kontrollieren/reflektieren können? Welche Formen der Rationalität begegnet uns in den Maschinen?

Dazu gehört auch der Begriff der »Handlung«. Eine Handlung beschreibt in der Anthropologie eine selbstgesteuerte Verhaltensentscheidung, die der Bewertung und Reflexion von Verhaltensoptionen erwächst. Die zunehmende Kognitivierung der vernetzten Entitäten (»Dinge«) ermöglicht diesen eine Auswahl aus Verhaltensoptionen. Wird diese – wie in der technikethischen Debatte häufig getan – mit dem Begriff der Handlung beschrieben, so führt diese sprachliche Parallelisierung zu zahlreichen Anschlussfragen: Haben Maschinen eigene Handlungsweisen und damit eine eigene Handlungsmacht, verbunden mit Strategien, Kommunikationsformen und einer entsprechend bestimmten Autonomie? In welcher Weise kann von Verantwortung, Zuschreibung derselben oder accountability einer Maschine die Rede sein?[2]

Gerade diese Frage nach der Autonomie von Maschinen hat in der Technikethik eine große Debatte ausgelöst: Besteht bei der KI ein Lernfortschritt mit eigenen Lösungen, die Programmierer nicht nachvollziehen können? Und wenn ja, wie ist damit nicht nur ethisch, sondern auch rechtlich umzugehen? Denn die Rechtsprechung bestimmt denjenigen als ein justiziables Subjekt, der oder die eine Handlung autonom wählen und ausführen kann. Kann eine KI in diesem Sinn eine Rechtsperson werden? Und kommt der KI eine eigene Würde, eine eigene Rechtspersönlichkeit zu?

Oder liegt die Krux für die Unterscheidung von Mensch und Maschine nicht doch an anderer Stelle? Ist es die Körperlichkeit des Menschen, die sein

2 Vgl. etwa https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

Spezifikum, seine Würde ausmachen? Und wie hängen Körperlichkeit und Kognition zusammen?

Diese Beschreibungen evozieren auf der sprachlichen Ebene eine Frage, die ethisch von herausragender Bedeutung ist: Die Frage nach der Unterscheidbarkeit von Mensch und Maschine und daran anschließend die Frage, welche Ethik für eine KI (Akteursansatz) oder im Zusammenspiel von Mensch und KI (Netzwerkansatz) zu entwickeln ist. In den gegenwärtigen technikethischen und technikphilosophischen Beschreibungen nehmen diese Kategorien und somit die Abgrenzung dieser zwischen Mensch und Maschine breiten Raum ein. Wie diese Darstellungen politische und mediale Diskurse über KI prägen, führen die Beiträge dieses Bandes vor Augen.

Die Breite und Ausweitung dieser Debatte führt auf der einen Seite zur Kritik an der Rede von »Künstlicher Intelligenz« als einem Hype oder einem Mythos, der auf einem grundlegenden Kategorienfehler in der Rede von »Intelligenz« beruhe[3]. Für die schweizerische Informatikerin Pooyan ist der Begriff zumindest irreführend[4], während Nida-Rümelin vor der Gefahr eines Animismus des Digitalen warnt, der durch die Simulation menschlicher Fähigkeiten in und mit Maschinen entsteht[5].

Für eine ethische Reflexion des an sich schon diffusen Themenkomplexes »Künstlicher Intelligenz« sehr grundlegend stellt sich hier die Frage der Identifikation nicht nur der damit gemeinten Technologien, sondern auch der daraus eigentlich erwachsenden ethischen Themenstellungen: Die Frage ist hier nicht, ob es »KI« eigentlich »gibt«, wie in den Debatten um einen KI-Hype zum Teil formuliert – die unter dieser Überschrift entwickelten Technologien gibt es ja durchaus. Die Frage ist vielmehr, ob die durch die sprachliche Konstruktion dieser Technologien auf die damit verbundenen ethischen Fragestellungen verweist – oder ob andere ethische Fragen im Zentrum der entwickelten Technologien stehen. Eines zeigen die Beiträge des Bandes sehr deutlich: Die Frage der sprachlichen Konstruktion und Vermittlung ist für eine Ethik der Künstlichen Intelligenz entscheidend. Was Technik ist, wie eine Maschine beschrieben wird, ist Teil einer gesellschaftlichen Konstruktion, in der technische Entwicklung, sprachlicher Bericht dieser Entwicklung und die mediale Reflexion darauf ineinandergreifen und sich gegenseitig beeinflussen.[6]

---

3 https://www.zukunftsinstitut.de/artikel/digitalisierung/6-thesen-zur-kuenstlichen-intelligenz/.

4 https://hub.hslu.ch/informatik/kunstliche-intelligenz-gibt-es-nicht-wichtig-ist-digitale-ethik/.

5 Nida-Rümelin 2021: 36.

6 Zu diesen Wechselwirkungen vgl. einführend Coeckelbergh 2017. Der Zusammenhang von technischen und sozialen Prozessen kommt in den Science and Technology Studies grundlegend zur Sprache, präzisiert in der These der Social Construction of Technology (SCOT etwa

Erkennbar ist in der KI-Debatte derzeit eine vor allem anthropologisch geprägte Imagination, die ich *imitative Imagination* künstlicher Intelligenz nenne. Diese imitative Imagination gründet in der Wissenschaftsgeschichte der KI-Forschung – dient sie doch dem Ziel Maschinen zu entwickeln, die sich verhalten, als verfügten sie über Intelligenz. Es ging folglich um eine Imitation des Menschen, um die Simulation menschlichen Handelns, was die KI-Forschung bis heute sehr eng an die Anthropologie anbindet.[7] Die daraus erwachsenden begrifflichen Schwierigkeiten werden in Einführungen in KI-Technologien in den Technikwissenschaften deutlich benannt: Übereinstimmend beginnen Einführungen zum Thema mit der Feststellung, dass der Begriff sehr vage und schwer zu füllen ist.[8] Die Analogie zur menschlichen Intelligenz wird eher kritisch aufgegriffen unter Hinweis auf die Begriffserfindung bei McCarthy.[9] Ertel beginnt explizit mit dem Hinweis, dass der Begriff vor allem Emotionen wecke: die Faszination für das Funktionieren eines Gehirns ebenso wie die Furcht vor etwas Künstlichem.[10] Die über den Begriff »Intelligenz« angezeigte Verbindung in die Anthropologie macht Ertel auch für die Beschreibung der KI-Forschung fruchtbar: Explizit benennt Ertel die Hirnforschung als Bezugsdisziplin der KI-Forschung.[11] Die Bedeutung dieser Imagination und damit die Anbindung der KI-Forschung an die Anthropologie ist jedoch bisher kaum aufgearbeitet, wie auch die Beiträge dieses Bandes zeigen.

Aufgabe einer imaginationssensiblen Ethik ist es, zu klären, ob und wie die Imaginationen von Technologien für die ethische Reflexion der materialethischen dienlich ist. Denn gerade die imitative Imagination steht in der Gefahr, die sprachliche Konstruktion Künstlicher Intelligenz materialen technischen Entwicklungen und damit verbundenen ethischen Fragen zu entkoppeln.[12]

---

bei Bijker/Pinch 1987). Vgl. zur sprachlichen Dimension des Verhältnisses von Mensch und Maschine Wiegerling 2018; Grimm/Kuhnert 2018; zur ethischen Dimension aus theologischer Perspektive Höhne 2019; Meireis 2019.

7 Deutlich erkennbar ist dies etwa in Catrin Misselhorns Entwurf der Maschinenethik: Misselhorn folgt in ihrer Erörterung dem, was ich eine »imitative Imagination von KI« nenne: Sie entfaltet ihr Programm konsequent entlang eines Abgleiches mit menschlichen Akteuren mit Hilfe von Leitbegriffen aus der Anthropologie. Die Argumentation dient in weiten Teilen dem Nachweis, ob, wie und wie weit Maschinen Attribute und Fähigkeiten zugeschrieben werden sollen, die bislang Menschen vorbehalten waren. Diese Vergleichbarkeit markiert Misselhorn in der Einleitung explizit als Ausgangspunkt ihrer Ethik. Misselhorn 2008: 7.

8 Vgl. z. B. Ertel 2016: 1.

9 Ertel 2016: 1.

10 Ertel 2016: 1.

11 Ertel 2016.3.

12 Deutlich wird dies etwa, wenn man in Reflexionen zu ethischen Fragen rund um Künstliche Intelligenz in den Technikwissenschaften schaut: Dort findet sich weder anthropomorphe Rede noch die Frage nach Mensch und Maschine im imitativen Sinn. Leitend sind Risiko- und Nutzenabwägungen, Zielbestimmungen und Haftungsfragen. Deutlich wird hier die Prägung aus der Tradition der Technik-Folgenabschätzung anstatt aus der Anthropologie

Diese Spannung aufzugreifen, ist Ziel der hier in Grundlinien konturierten imaginationssensiblen Ethik. Deutlich wird, dass zwischen der sprachlichen und technischen Konstruktion von Künstlicher Intelligenz zu unterscheiden ist, diese jedoch zugleich eng aufeinander bezogen sind. Die Wechselwirkungen zwischen sprachlicher und technischer Konstruktion sind dabei auf dreifache Weise zu beschreiben.

Dies gilt erstens *prospektiv* – beschreibungssprachlich gedacht von den sozialen Imaginationen als Ausgangspunkt: Die sprachliche Konstruktion einer Technologie prägt ihre Entwicklung, wie Caja Thimm überzeugend im Blick auf den Begriff der Maschine darstellt: »Die Definition und Bewertung dessen, was eine ›Maschine‹ ist und was sie bewirkt, unterscheidet sich nicht nur in Bezug auf ihre konkreten zeitgeschichtlichen Auswirkungen, sondern auch in Bezug auf die grundlegende gesellschaftliche Haltung gegenüber der Technologie. Dabei ist zu betonen, dass Maschinen bereits in ihrem Entstehungsprozess mit den politischen, sozialen und kulturellen Umgebungen verbunden sind [wie Nancy betont]: ›Die Maschine taucht nicht aus irgendeinem Nichts auf. Sie ist selbst maschiniert, das heißt, sie ist auf zuvor gesetzte Zwecke hin entworfen, ausgearbeitet und strukturiert‹«[13]. Wie technische Konstruktionen heute sprachlich konstruiert werden, kann also die Entwicklung beeinflussen. Die Deutsche Akademie der Technikwissenschaften prägt dafür 2012 den Begriff der »Technikzukünfte«, der meines Erachtens exemplarisch Teil einer solchen prospektiven imaginationssensiblen Technikethik sein kann.[14]

---

13 Thimm 2019: 20.

14 Vgl. den Beitrag von Böhnke et al in diesem Band. Technikzukünfte werden definiert als die »Vorstellungen über die zukünftige Entwicklung von Technik und Gesellschaft« und »vereinen unterschiedliche Formen von Wissen, beinhalten Annahmen und normative Setzungen« (Acatech 2012: 6). Dass diese genuin ethische Fragen stellen, wird von den Autor:innen deutlich markiert: »Vor allem aber sind Technikzukünfte Gegenstand der gesamtgesellschaftlichen Diskussionen über die Frage, mit welcher Technik wir als Gesellschaft zukünftig leben wollen. […] Gerade in der wissenschaftlichen Politikberatung stehen die Autoren von Technikzukünften damit in der Verantwortung, die Prämissen und Wertentscheidungen offenzulegen, die Grundlage dieser Technikzukünfte sind.« (Acatech 2012: 6). Insofern zielt der Leitfaden auf die Beschreibung wertorientierter Technikgestaltung (Acatech 2012: 32). Der von der Akademie entwickelte Leitfaden nimmt die semantische Dimension der Rede von Zukünften in dem vorgelegten Leitfaden nicht explizit auf – ein Bewusstsein für diese Ebene spiegelt aber etwa die folgende Beschreibung: »Es gibt viele und sehr unterschiedliche Bilder und Vorstellungen über Zukunft, und so erklärt sich die Wahl des Plurals im Titel dieses Projekts – der Plural ist Programm! Mit dem Begriff *Zukünfte* bezeichnen wir hier ganz allgemein Beschreibungen zukünftiger Sachverhalte oder Entwicklungen. Diese erfolgen sprachlich, teils konkretisiert durch Zahlen oder Grafiken, können in Filmen, Texten, Reden oder anderen Medien der Kommunikation vermittelt werden.« (Acatech 2012: 11; Hervorhebungen im Original). Auch die soziale Konstruktion dieser Zukünfte kommt in den Blick: »Diese Zukünfte sind *soziale Konstrukte*, entstanden im Kopf einzelner Personen, beim Brainstorming in Gruppen oder methodenorientiert in komplexen Verfahren der Modellierung und Simulation. Alle diese Verfahren finden jeweils »heute« statt – eben in der bereits genannten »Immanenz der Gegenwart«. Dies heißt, dass auch jeweils vorherrschende Meinungen und Modeströmungen bis hin zum Zeitgeist die Wahrnehmungen der jeweiligen Zukünfte mitprägen. Es kommt zu Kon-

Die Akademie hält thesenartig fest: »1. Das Vorausdenken, Erstellen und Bewerten von Technikzukünften ist ein notwendiges Element gesellschaftlicher Orientierung und der Selbstverständigung in den Technikwissenschaften. 2. Technik und Gesellschaft stehen in einem untrennbaren Zusammenhang. Deshalb implizieren technische Zukünfte auch gesellschaftliche Zukünfte und umgekehrt. […] 9. Technikzukünfte sind in demokratischen Gesellschaften immer Gegenstand öffentlicher Debatten.«[15] Dass und wie über die Technikwissenschaften hinaus auch geisteswissenschaftlich informierte Technikphilosophie und Technikethik Teil dieser Debatten sein kann, wäre als Teil einer prospektiv ausgerichteten imaginationssensiblen Technikethik zu diskutieren.

Deutlich wird hier die politische und gesellschaftliche Dimension imaginationssensibler Ethik. Diese greift die semantische Konstruktion von Technologien auf mit dem Ziel, diesen als Gegenstand ethischer Debatten über die gewünschten Technikzukünfte in die Gesellschaft einzubringen.[16] Diese Frage gewinnt an Brisanz, bedenkt man die von Thimm und Bächle herausgearbeitete »emotionsgeladene Annäherung an das Verhältnis Mensch und Technologie«.[17] Imaginationssensible Ethik steht damit nicht nur vor der Aufgabe, sprachliche und technische Konstruktionen zu erkennen und sie einander zuzuordnen. Im Modus öffentlicher Diskurse ist zudem eine emotionssensible Diskurskultur zu entwickeln, die rationales Argumentieren und – oft implizit bleibende – emotionsorientierte Wahrnehmungen in Verbindung bringt. Zu diskutieren ist, wie es neben der analytischen Aufgabe der Identifikation und

junkturen bestimmter Perspektiven auf Zukunft, die besonders gut dann zu erkennen sind, wenn Zukünfte ›altern‹ oder ›veralten‹.« (Acatech 2012: 12; Hervorhebungen im Original).

15 Acatech 2012: 49.

16 Mit diesem ethischen Fokus geht das Konzept über deskriptive Ansätze hinaus, wie sie etwa das Projekt »KI-Konstruktionen« am Humboldt-Institut für Internet und Gesellschaft verfolgt. Dieses widmet sich dem »Hype« um KI gerade wegen seiner sprachlichen Konstruktion und weniger wegen seiner technischen Konstruktion: »Nicht unbedingt, weil er [der Hype] gut begründet ist, sondern weil die Erwartungen und Investitionen, die dieser Hype erzeugt, real sind. KI als Technologiecluster und als ›sociotechnical imaginary‹ wird derzeit in unseren Gesellschaften institutionalisiert. […] Die Art und Weise, wie wir diese Technologien imaginieren und wie wir über unsere Zukunft denken, prägen Entscheidungen und Entwicklungen in der Gegenwart. Deshalb ist der Hype wichtig, ungeachtet seiner Substanz.« So die Projektbeschreibung des Projekts »KI-Konstruktionen« am HIIG. https://www.hiig.de/project/ki-konstruktion/.

17 Sie führen aus: »So argumentiert der bekannte US-amerikanische Technikjournalist Jeff Jarvis (2014), dass die Amerikaner als technikaffine und fortschrittsgläubige Nation der Technologie grundsätzlich eher zugewandt seien als die technikskeptischen Europäer, bei denen sich sogar ein typisches Muster erkennen lasse: eine ›Eurotechnopanik‹.« (Thimm/Bächle 2019: 2) Auf der anderen Seite ist der Diskurs geprägt durch die »stets historisch spezifische Zeitdiagnose«, dass die gegenwärtige Generation in einem besonderen Zeitalter lebe, der geprägt sei von einer umwälzenden Veränderung – eine Beschreibung, die sich für viele neue Erfindungen finden lässt, u. a. für die Erfindung des Fahrrads. (Thimm/Bächle 2019: 3).

Beschreibung sozialer Imaginationen neue Imaginationen geprägt und in die partizipativen Diskurse eingebracht werden können. Wenn die imitative Imagination von KI für materialethische Fragen wenig hilfreich ist, verlangt die Entkopplung von Imagination und materialer Ethik über die Präzisierung der ethischen Fragestellung hinaus nach einer neuen Imagination. So betont etwa Ertel, dass Filme zu KI zur gesellschaftlichen Meinungsbildung beitragen[18], was auch im Beitrag von Böhnke et al. in diesem Band deutlich wird. Wie dies von Seiten einer interdisziplinären ethischen Perspektive her prospektiv gedacht werden kann, ist eine offene Frage. Dass ein Beitrag zu diesen Aufgabe für die theologische Ethik ist, unterstreicht auch die EKD-Denkschrift zur Digitalisierung aus dem letzten Jahr.[19]

Damit ist zweitens der *korrelative* Zusammenhang angesprochen: Die sprachliche Konstruktion prägt, was als ethische Fragestellung identifiziert wird – ist aber nicht unbedingt mit der technischen Konstruktion verbunden. Über Begriffe wie Intelligenz, neuronale Netze, Autonomie und Handlung werden anthropologische Denkwelten aufgerufen, die nicht nur einen engen Zusammenhang zwischen dieser Art von Technologie und der Anthropologie suggerieren, sondern immer wieder auch einen Überbietungsgestus dieser Technologie gegenüber dem Menschen nahelegen. Begriffe wie Singularität rufen wiederum theologische Denkwelten auf, die den Abgleich mit theozentrischen Beschreibungen implizieren. Ob und wie diese Fragen aber tatsächlich die drängenden ethischen Herausforderungen im Umgang mit KI darstellen, ist vor dem Hintergrund des Dargestellten nicht eindeutig. Sinnvoll wäre hier meines Erachtens eine korrelative Bestimmung von sprachlicher und technischer Konstruktion, die beide Pole wechselseitig aufeinander bezieht. Daraus ergibt sich auf der einen Seite die Aufgabe, mit Fokus auf die sprachliche Konstruktion, die semantischen Gehalte in den bestehenden Debatten und ihre im Hintergrund stehenden Dynamiken offen zu legen. Mit Walther Zimmerli gesprochen, geht es hier um die Beteiligung an der Philosophie als »Begriffskläranalage« – verstanden als eine ethische Reflexion der »Wortpolitik«: »Die Frage ist nicht nur, was Begriffe bedeuten und wie sie verknüpft werden, sondern auch, was wir, indem wir sie (so exzessiv) verwenden, eigentlich tun, bzw. anrichten?«.[20] Nach Zimmerli ist es eine Suche nach »begrifflichen Inseln« zwischen »Nebelbänken« – oder noch drastischer formuliert mit ei-

18 Ertel 2016 17.

19 »Der digitale Wandel ist nicht nur aus der Perspektive individuellen Handelns und individualethischer Überlegungen zu beleuchten. Er ist zugleich als gesellschaftlicher Prozess zu verstehen und sozialethisch zu interpretieren. […] Narrative haben dabei einen großen Einfluss, etwa die Narrative: ›Digitalisierung bietet mehr Chancen als Risiken‹, ›Industrie 4.0‹, ›Smart City‹, ›digitale Souveränität‹ oder die Vision des ›Homo Deus‹.« EKD 2021: 36.

20 Zimmerli 2021: 13.

nem Begriff des emeritierten Professors für Philosophie Harry Frankfurt aus Princeton – die Suche nach »Bullshit-Words«, also nach Worten, bei denen keiner weiß, was damit eigentlich gemeint ist[21]. KI ist nach Zimmerli ein solcher Begriff, vielleicht sogar der renitenteste: »Allzu offensichtlich ist die immer wieder verblüffende Omnipräsenz der Informations- und Kommunikationstechnologien, die heute – und auch hierzu wäre eine ›bullshit‹-Differentialdiagnose angezeigt – ebenso vereinfachend und irreführend allesamt als ›Künstliche Intelligenz-Technologie‹ bezeichnet werden.«[22] »How to do things with words« – dieser Titel einer Vorlesungsreihe von John L. Austin von 1955 wird nach Zimmerli in den Debatten und Beschreibungsversuchen des Digitalen auf eine sehr konkrete und politische Art und Weise wieder philosophisch – und ich möchte ergänzen: ethisch – relevant, insbesondere dort, wo technische Entwicklungen als unausweichlich und anthropologisch relevant beschrieben werden[23]. Aus dieser begrifflichen Arbeit ergibt sich – und das ist die andere Seite des korrelativen Zusammenhangs mit dem Fokus auf die technische Konstruktion künstlicher Intelligenz – die Notwendigkeit zur Präzisierung der ethischen Problemstellung aus der Perspektive der technologischen Entwicklungen. Dies wäre die Aufgabe einer Technikethik im engeren Sinne.

Auf eine dritte Richtung im Verhältnis von sprachlicher und technischer Konstruktion möchte ich abschließend eingehen: Der *retrospektive* Zusammenhang beschreibt die zu Beginn dieses Abschnitts eingeführte Entstehung von Narrativen durch Technologien: Technologien prägen nicht nur die Welt, sondern auch ihre Wahrnehmung und somit auch die entstehenden sozialen Imaginationen. Lassen Sie mich diese Dimension anhand des von Reijers und Coeckelbergh entwickelten Entwurfs einer narrativen Technikethik ausführen. Reijers und Coeckelbergh entfalten in ihrem Entwurf »Narrative and Technology Ethics« von 2020 nicht weniger als eine ethische Theorie ausgehend von dem weiten Medienbegriff McLuhans: Als Medien werden grundlegende Übermittler des Welt- und Selbstverständnisses bezeichnet, die somit konstitutiv die Welt und ihre Wahrnehmung prägen. Sowohl Sprache als auch Technologien sind in diesem Sinne als Medien zu verstehen. Coeckelbergh und Reijers zeigen, dass in den medientheoretischen Debatten in der Technikphilosophie sowohl die Sprache als auch die Sozialität von Technikerfahrung kaum in den Blick kommen.[24] Die Fokussierung auf die Materialität des Media-

21 Zimmerli 2021: 12.

22 Zimmerli 2021: 14.

23 Zimmerli 2021: 14.

24 »In recent critiques of contemporary strains of works in philosophy of technology that focus on mediation, two main concerns appear to be prevalent: neglect of language and neglect of the social in human dealings with technology.« (Coeckelbergh/Reijers 2016: 326).

len im Anschluss an Informationswissenschaft und Technik führe in den medienethischen Debatten leicht dazu, die sprachliche Konstitution der Lebenswelt zu vernachlässigen[25]. Demgegenüber beschreiben sie das ethische Subjekt als »mediated subjectivity«, die sich in einem »social-linguistic environment« bewegt: »New technologies change, or rather co-shape our mediated subjectivity.«[26] Die von Latour und Akrich eingeführte Rede von einem »script« der Artefakte müsse nicht nur metaphorisch unter der Frage nach der Agency und den Affordances von Materialitäten genutzt werden, sondern kann auch für die Entdeckung der linguistischen Dimension von Technologien fruchtbar gemacht werden.[27] Diese Überlegungen führen Coeckelbergh und Reijers zu der These, »that technologies, similar to texts, novels, and movies, ›tell stories‹ by configurating characters and events in a meaningful syntheses«[28]. Technologien haben daher narrative Fähigkeiten, eine »narrative capacity«: »not only do humans make sense of technologies by means of narratives but technologies themselves co-constitute narratives and our understanding of these narratives by configuring characters and events in a meaningful temporal whole.«[29] Und so kommen sie zu dem Schluss: »In other words, we argue that humans do not only read technologies, but technologies on the other hand ›read‹ the human.«[30] In ihrem Entwurf entwickeln sie eine »hemeneutic ethics of technology«[31] im Sinne eines »framework to reflect on the ethics of technical practices«[32], die erstens dazu dienen soll technologische Mediation besser zu verstehen, zweitens Mediatisierungstheorie und Technikethik verbindet und drittens eine Methode für die Anwendung der Ethik entwickelt[33]. Aus diesem Entwurf ergeben sich zahlreiche materiale Fragen für eine imaginationssensible Technikethik, die an der Schnittstelle von politischer Ethik und Medien- bzw. Informationsethik liegen. Reijers und Coeckelbergh betonen etwa die politische Bedeutung von technisch mediatisierter Erinnerung.[34]

---

25 Coeckelbergh/Reijers 2016: 327 f.

26 Coeckelbergh/Reijers 2016: 327.

27 Die Autoren führen weiter: »Furthermore, the script of artefacts as borrowed from actor network theory is treated as if it is isolated from a wider social-linguistic environment (prescriptions, discourse, narratives).« (Coeckelbergh/Reijers 2016: 328)

28 Reijers/Coeckelbergh 2020: 6.

29 Reijers/Coeckelbergh 2016: 325.

30 Reijers/Coeckelbergh 2016: 336.

31 Reijers/Coeckelbergh 2020: 17.

32 Reijers/Coeckelbergh 2020: 8.

33 Reijers/Coeckelbergh 2020: 8–17.

34 »As an avenue for future research, we might explore ways in which technologies explicitly mediate the public experience of time and analyse the political aspects of such mediation. […] we can inquire how technologies shape those things that we remember, those things that we forget and thereby also the ways we relate to our personal and collective histories. […] Such

Weitere Präzisierungen ergeben sich im Blick auf konkrete Technologien. Angedeutet findet sich dies im Feld des Gaming in Versuchen einer »narrative mechanics« von Suter in diesem Jahr:[35] Das Konzept durch Technologien mediatisierter intersubjektiver sozialer Imaginationen und ihre wirklichkeitsprägende Kraft wird hier am Beispiel von Online-Rollenspielen und anderen Games konkretisiert. Der Begriff »narrative mechanics« wird dabei zum Leitbegriff, um die Steuerung von Verhaltensmustern und ihrer Deutungen zu beschreiben.[36] Diese Rückwirkungen von der technischen Konstruktion auf die sprachlichen Konstruktionen im Sinne des retrospektiven Zusammenhangs zu bedenken, ist Neuland für die Technikethik an der Schnittstelle zur Medienethik und politischen Ethik.

Dieser Zusammenhang verlangt nach einer interdisziplinären und partizipativen Ethik im Sinne einer »interdisziplinären Begleitforschung« diskursiver Natur. Ich möchte dies im Blick auf die derzeit viel gefragte Technikethik oder Ethik der Digitalisierung präzisieren: Während im Bereich der Bio- und Gesundheitsethik interdisziplinäre Ethikkommissionen in Kliniken, Universitäten und anderen Forschungseinrichtungen die ethische Reflexion sehr eng mit der angewandten Entwicklung verbinden, stehen solche Ansätze für die Technikethik ganz am Anfang. Vielfach kommen ethische Debatten sehr spät zu den Entwicklern – und mehr als ein »Ethik-TÜV« ist dann weder gefragt noch möglich. Das bedeutet auf der anderen Seite aber auch: Die Verantwortung für die ethische Reflexion der technologischen Entwicklungen wird zunehmend den Entwicklern übertragen werden. So sehr ich es begrüße, dass an vielen Stellen inzwischen Ethikseminare in die Ausbildung von Informatikern eingebunden werden, so wenig kann dies die einzige Lösung auf Herausforderungen der Technikethik sein. Eine Sensibilisierung für die ethische Dimension von Technik ist der Anfang, nicht das Ende ethischer Reflexion.

Hier scheinen Modelle partizipativer Technikgestaltung[37], technikethischer Mäeutik[38] oder Anknüpfungen an Modelle einer ethics on the laboratory floor[39] weiterführend: Ziel ist die Entwicklung von nicht nur punktuellen,

---

a view suggests that, just as a historical narrative can ›stand in‹ or ›re-present‹ a collective memory of a past event, so can technologies ›stand in‹ for a past event; or at least configure it in a specific way. This observation ties in with debates on for instance the so-called ›right to be forgotten‹ (Rosen 2012) about personal rights to control the presence or absence of digital memories, which arguably for the first time explicitly puts the technological mediation of human memory on political agendas.« Coeckelbergh/Reijers 2016: 344.

35 Suter et al. 2021.

36 Suter et al. 2021: 9.

37 Weber/Wackerbarth 2015.

38 Irrgang 2015.

39 Manzeschke 2015.

sondern dauerhaften Formen ethischer Begleitforschung. Mit Arne Manzeschke präzisiert: »Kennzeichen dieser ›Begleitforschung‹ ist es, den (bio-) technischen Entwicklungen nicht nach-denken zu müssen, sondern zeitgleich mit den sog. Lebenswissenschaften koproduktiv Wissen zu generieren und die ethische Perspektive in den weiteren Forschungs- und Entwicklungsprozess einzuspeisen.«[40] Auf diese Weise fachwissenschaftlich orientierte imaginationssensible Ethik in ihrer materialen Zuspitzung weiter zu entwickeln, erscheint mir verheißungsvoll – und dieser Band ist ein erster Aufschlag für einen solchen Gesprächsprozess.

Die dargestellten Analysen führen mich zu folgender Doppelthese: Technik ist nicht nur technisch, sondern auch sprachlich konstruiert. Zugleich konstruiert Technik selbst Wirklichkeit. Zwischen Imaginationen und Technologien besteht somit eine doppelte Verbindungslinie: Zum einen drücken sich Diskurse über Technologien in Narrativen, Metaphern und Bildern aus, die sich zu sozialen Imaginationen verdichten lassen. Zum anderen prägen Technologien als Medien die Wahrnehmung der Wirklichkeit und tragen dabei selbst zur Entstehung und Prägung von Narrativen über das Digitale bei. Dies gilt auch für die Rede von der Digitalisierung insgesamt[41]: Auch diese ist vermittelt gesteuert, bedingt durch Frames und Metaphern – und damit eingebettet in soziale Imaginationen dessen, was wir gemeinsam zu erleben glauben. Diese Dimensionen ethisch zu bedenken und aufeinander zu beziehen, ist Anliegen der skizzierten »imaginationssensiblen« Ethik: Wenn Mythen sich selbständig machen in Imaginationen und Weltbeschreibungen – wie es das eingangs zitierte Statement nahelegt –, ist es eine Aufgabe auch einer solchen Ethik, diese wieder einzufangen – begrifflich und sachlich.

Literatur

acatech (Hg.) 2012: Technikzukünfte. Vorausdenken – Erstellen – Bewerten (acatech IMPULS), Heidelberg u. a.: Springer Verlag. DOI 10.1007/978-3-642-34607-1.

Bijker, Wiebe E./Pinch, Trevor J. 1987: The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit of Each Other. In: Bijker, Wiebe E./Hughes, Thomas P./Pinch, Trevor J. (Hg.): The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology. Cambridge (MA) u. a., MIT Press: 17–50.

40 Manzeschke 2015: 325.

41 Vgl. Höhne 2019; Meireis 2019.

Coeckelbergh, Mark 2020: AI Ethics. Cambridge (MA), MIT Press.

Coeckelbergh, Mark (Hg.) 2017: Using Words and Things. Language and Philosophy and Technology (Routledge Studies in Contemporary Philosophy). London/New York, Routledge.

Coeckelbergh, Mark/Reijers, Wessel 2016: Narrative Technologies: A Philosophical Investigation of the Narrative Capacities of Technologies by Using Ricoeur's Narrative Theory. In: Hum Stud 39 (3): 325–346. DOI: 10.1007/s10746-016-9383-7.

Evangelische Kirche in Deutschland (Hg.) 2021: Freiheit digital. Die Zehn Gebote in Zeiten des digitalen Wandels. Eine Denkschrift der Evangelischen Kirche in Deutschland. Leipzig, Evangelische Verlagsanstalt.

Ertel, Wolfgang 2016: Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung (Computational Intelligence). 4. Auflage. Wiesbaden, Springer VS.

Grimm, Petra/Kuhnert, Susanne 2018: Narrative Ethik in der Forschung zum automatisierten und vernetzten Fahren, in: Grimm, Petra/Zöllner, Oliver (Hg.): Mensch – Maschine. Ethische Sichtweisen auf ein Spannungsverhältnis (Medienethik 17). Stuttgart, Franz Steiner Verlag: 93–110.

Höhne, Florian 2019: Darf ich vorstellen: Digitalisierung. Anmerkungen zu Narrativen und Imaginationen digitaler Kulturpraktiken in theologisch-ethischer Perspektive. In: Bedford-Strohm, Jonas/Höhne, Florian/Zeyher-Quattlender, Julian (Hgg.): Digitaler Strukturwandel der Öffentlichkeit. Ethik und politische Partizipation in interdisziplinärer Perspektive, (Kommunikations- und Medienethik Bd. 10), Baden-Baden, Nomos Verlag: 25–46.

Irrgang, Bernhard 2015: Mäeutik als Beratungskonzept angewandter Ethik – zu einem Konzept der Unternehmens- und Politikberatung mit sittlicher Ausrichtung. In: Maring, Matthias (Hg.): Vom Praktisch-Werden der Ethik in interdisziplinärer Sicht: Ansätze und Beispiele der Institutionalisierung, Konkretisierung und Implementierung der Ethik. Karlsruhe, KIT Scientific Publishing: 55–71.

Manzeschke, Arne 2015: Angewandte Ethik organisieren. MEESTAR – Ein Modell zur ethischen Deliberation in sozio-technischen Arrangements. In: Maring, Matthias (Hg.): Vom Praktisch-Werden der Ethik in interdisziplinärer Sicht: Ansätze und Beispiele der Institutionalisierung, Konkretisierung und Implementierung der Ethik. Karlsruhe, KIT Scientific Publishing: 315–330.

Meireis, Torsten 2019:»O daß ich tausend Zungen hätte«. Chancen und Gefahren der digitalen Transformation politischer Öffentlichkeit – die Perspektive evangelischer Theologie. In: Bedford-Strohm, Jonas/Höhne, Florian/Zeyher-Quattlender, Julian (Hgg.): Digitaler Strukturwandel der Öffentlichkeit. Ethik und politische Partizipation in interdisziplinärer Perspektive (Kommunikations- und Medienethik Bd. 10), Baden-Baden, Verlag: 47–62.

Misselhorn, Catrin 2018: Grundfragen der Maschinenethik. Stuttgart, Reclam Verlag.

Nida-Rümelin, Julian 2021: Digitaler Humanismus. In: Hauck-Thum, Uta/Noller, Jörg (Hgg.): Was ist Digitalität. Philosophische und pädagogische Perspektiven, (Digitalitätsforschung/Digitality Research 1). Berlin, Ort: 35–38.

Reijers, Wessel/Coeckelbergh, Mark 2020: Narrative and Technology Ethics. 1. Auflage. Cham, Springer International Publishing; Imprint: Palgrave Macmillan.

Suter, Beat/Bauer, René/Kocher, Mela (Hgg.) 2021: Narrative Mechanics. Strategies and Meanings in Games and Real Life. 1. Auflage. Bielefeld, transcript.

Thimm, Caja 2019: Die Maschine – Materialität, Metapher, Mythos: Ethische Perspektiven auf das Verhältnis zwischen Mensch und Maschine, in: Thimm, Caja/Bächle, Thomas Christian (Hgg.): Die Maschine: Freund oder Feind? Mensch und Technologie im digitalen Zeitalter. Wiesbaden: Springer VS: 17–40.

Thimm, Caja/Bächle, Thomas Christian (Hgg.) 2019: Die Maschine: Freund oder Feind? Mensch und Technologie im digitalen Zeitalter. Wiesbaden, Springer VS.

Weber, Karsten/Wackerbarth, Alena 2015: Partizipative Technikgestaltung als Verfahren der angewandten Ethik. In: Maring, Matthias (Hg.): Vom Praktisch-Werden der Ethik in interdisziplinärer Sicht: Ansätze und Beispiele der Institutionalisierung, Konkretisierung und Implementierung der Ethik. Karlsruhe, KIT Scientific Publishing: 299–315.

Wiegerling, Klaus 2018: Warum Maschinen nicht für uns denken, handeln und entscheiden. In: Grimm, Petra/Zöllner, Oliver (Hgg.): Mensch – Maschine. Ethische Sichtweisen auf ein Spannungsverhältnis (Medienethik 17). Stuttgart, Franz Steiner Verlag: 33–46.

Zimmerli, Walther Ch. 2021: Analog oder Digital? Philosophieren nach dem Ende der Philosophie. In: Hauck-Thum, Uta/Noller, Jörg (Hgg.): Was ist Digitalität. Philosophische und pädagogische Perspektiven. Digitalitätsforschung/Digitality Research 1. Berlin, Metzler Verlag: 9–33.

ORCID

Frederike van Oorschot https://orcid.org/0000-0003-4359-8949

Autor*innenverzeichnis

Marie-Hélène Adam, Dr., hat am Karlsruher Institut für Technologie (KIT) studiert und promoviert. Seit 2010 wissenschaftliche Mitarbeiterin und seit 2020 Koordinatorin der Ergänzungsbereiche »Medientheorie und -praxis« und »Kulturtheorie und -praxis« am Institut für Germanistik des KIT. Sie lehrt und forscht an den Schnittstellen von Gender, Film, Populärkultur, Science Fiction und KI.

Jonas Bedford-Strohm, Dr. arbeitet als Referent für Transformationsstrategie bei ARD Online und ist Research Fellow am Zentrum für Ethik der Medien und der digitalen Gesellschaft an der Hochschule für Philosophie München. Seine Forschungsschwerpunkte sind Medienethik, Institutionentheorie sowie die Theorie und Ethik technologischer Transformation.
ORCID: https://orcid.org/0000-0003-4165-1881

Andreas Böhn, Prof. Dr., ist Professor für Literaturwissenschaft/Medien am Institut für Germanistik: Literatur, Sprache, Medien des Karlsruher Instituts für Technologie (Universitätsteil). Seine Forschungsschwerpunkte sind Intertextualität und Intermedialität; Metareferenz in Literatur, Film und anderen Medien und Künsten; Erinnerung und Medialität; Komik und Normativität; Technik und Kultur.

Alexander Filipović, Prof. Dr. theol., ist Professor für Sozialethik an der Katholisch-Theologischen Fakultät der Universität Wien. Seine Forschungsschwerpunkte sind Medienethik, politische Ethik, Technik und Gesellschaft

(Digitalität, KI), Grundfragen christlicher Sozialethik und philosophischer Pragmatismus.
ORCID: https://orcid.org/0000-0001-8946-9283

Selina Fucker, M.A., ist wissenschaftliche Hilfskraft im Arbeitsbereich »Religion, Recht und Kultur« an der Forschungsstätte der Evangelischen Studiengemeinschaft (FEST). Ihre Forschungsthemen sind Medienwirkung, digitale Religion und digitale Theologie.
ORCID: https://orcid.org/0000-0001-8728-3485

Florian Höhne, Dr. theol., ist Wissenschaftlicher Mitarbeiter in der Systematischen Theologie (Hermeneutik und Ethik) an der theologischen Fakultät der Humboldt-Universität zu Berlin. Seine Forschungsschwerpunkte sind öffentliche Theologie, digitale Theologie und Verantwortungsethik.
ORCID: https://orcid.org/0000-0001-6589-2124

Sonja Kleinke, Prof. Dr., leitet seit 2010 den Lehrstuhl Englische Linguistik an der Universität Heidelberg. Ihre Foschungsschwerpunkte sind Kognitive Linguistik sowie Interpersonelle Pragmatik und (Kritische) Diskursanalyse im Kontext Sozialer Medien.
ORCiD: https://orcid.org/0000-0002-6165-0918

Julian Lamers, M.A., ist freier Forschungsmitarbeiter am zem:dg. Während seines Studiums der Politikwissenschaft konzentrierte sich sein Forschungsschwerpunkt vor allem auf das Phänomen des politischen Populismus und des diesbezüglichen theoretischen Begriffsdiskurses. Am Lehrstuhl für Medienethik an der Hochschule für Philosophie in München arbeitete er unter anderem zur Relevanz medialer Diskurse über Künstliche Intelligenz aus ethikwissenschaftlicher Perspektive.
ORCID: https://orcid.org/0000-0003-0119-3898

Frederike van Oorschot, PD Dr. theol., leitet den Arbeitsbereich »Religion, Recht und Kultur« an der Forschungsstätte der Evangelischen Studiengemeinschaft (FEST) und ist Privatdozentin für Systematische Theologie an der Ruprecht-Karls-Universität Heidelberg. Ihre Forschungsschwerpunkte sind öffentliche Theologie, digitale Theologie und theologische Ethik.
ORCID: https://orcid.org/0000-0003-4359-8949

Philipp Stoellger, Prof. Dr., hat den Lehrstuhl für Systematische Theologie/Dogmatik und Religionsphilosophie an der Theologischen Fakultät der Universität Heidelberg inne und ist Leiter der Forschungsstätte der Evangelischen

Studiengemeinschaft (FEST) Heidelberg. Seine Forschungsschwerpunkte sind Christologie und Anthropologie; Hermeneutik, Phänomenologie und Religionsphilosophie; Bild- und Medientheorie.
ORCID: https://orcid.org/0000-0003-4981-7743

Katrin Strobel, M. A. in Englischer und Deutscher Philologie, war wissenschaftliche Mitarbeiterin im Projekt »KI im Spannungsfeld gesellschaftlicher Diskurse« (Ruprecht-Karls-Universität Heidelberg und KIT, Karlsruhe) und ist aktuell tätig im BMFJFS-geförderten Programm »Respekt Coaches«. Ihre Forschungsschwerpunkte sind kognitive Linguistik (hier im Besonderen: Konzeptuelle Metapherntheorie, Multimodale Metaphern und Conceptual Blending), (Kritische) Diskursanalyse, sowie Computer-Mediated Communication.
ORCID: https://orcid.org/0000-0001-7209-661X

Über die FEST

Die Forschungsstätte der Evangelischen Studiengemeinschaft e. V. (FEST) ist ein interdisziplinäres Forschungsinstitut, seit 1958 mit Sitz in Heidelberg, dessen Grundfinanzierung durch die Mitglieder des Trägervereins – die Evangelische Kirche in Deutschland (EKD), die Landeskirchen der EKD, den Deutschen Evangelischen Kirchentag und die Evangelischen Akademien – getragen wird und das darüber hinaus Forschungs- und Beratungsarbeiten durch Drittmittel finanziert. Die FEST ist in vier Arbeitsbereiche gegliedert: Religion, Recht & Kultur, Nachhaltige Entwicklung, Theologie & Naturwissenschaft sowie Frieden. Zum satzungsgemäßen Auftrag gehört die Aufgabe, wissenschaftliche Arbeiten anzuregen und zu fördern, die dazu bestimmt sind, die Grundlagen der Wissenschaft in der Begegnung mit dem Evangelium zu klären, und die Kirche bei ihrer Auseinandersetzung mit den Fragen der Zeit – auch durch Untersuchungen und Gutachten für die Mitgliedskirchen – zu unterstützen.

INSTITUT FÜR
INTERDISZIPLINÄRE
FORSCHUNG
F·E·S·T Forschungsstätte der Evangelischen Studiengemeinschaft

In gesellschaftlichen, politischen und medialen Debatten kommt „Künstliche Intelligenz" als Heilsbringer, Erweiterung oder Bedrohung des Menschen u. ä. in den Blick. Diese Beschreibungen konstruieren soziale Imaginationen, welche den Rahmen individueller und gesellschaftlicher Rede über KI bilden.
Der Band verbindet Fallstudien dieser Frames, Narrative und Metaphern mit einer Reflexion der damit verbundenen ethischen Fragen unter der Aufgabe einer imaginationssensiblen Ethik. Der Band bietet somit die erste interdisziplinäre Einführung in die Imaginationen Künstlicher Intelligenz aus der Medienwissenschaft, Kommunikationswissenschaft, Linguistik, Filmwissenschaft und Medienethik.

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

ISBN 978-3-948083-69-4