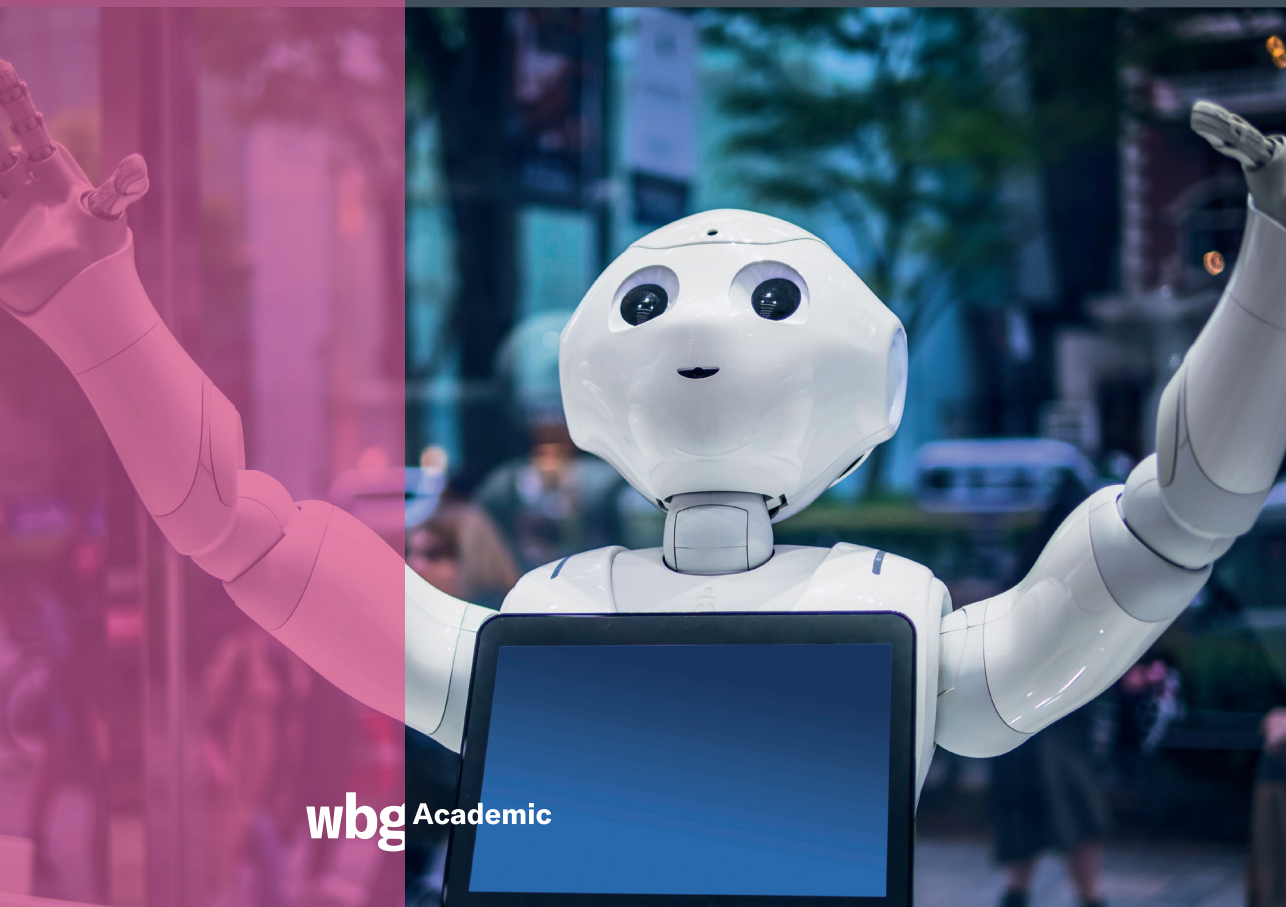
Anna Puzio, Nicole Kunkel, Hendrik Klinge (Hg.)

Alexa, wie hast du's mit der Religion?

Theologische Zugänge zu Technik und Künstlicher Intelligenz

Theologie und Künstliche Intelligenz, Vol. 1



wbg Academic

Anna Puzio, Nicole Kunkel, Hendrik Klinge (Hg.)

Alexa, wie hast du's mit der Religion?

Theologie und Künstliche Intelligenz

Theology and Artificial Intelligence

Volume 1

Alexa, wie hast du's mit der Religion?
Theologische Zugänge zu Technik und Künstlicher Intelligenz

Alexa, How Do You Feel About Religion?
Theological Approaches to Technology and Artificial Intelligence

Anna Puzio, Nicole Kunkel, Hendrik Klinge (Hg.)

Alexa, wie hast du's mit der Religion?

Theologische Zugänge zu Technik und Künstlicher Intelligenz

wbg Academic

Die Veröffentlichung wurde gefördert aus dem Open-Access-Publikationsfonds
der Humboldt-Universität zu Berlin.

Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische
Daten sind im Internet über www.dnb.de abrufbar.

https://doi.org/10.53186/1030373

wbg Academic ist ein Imprint der wbg
erschienen 2023 bei der wbg (Wissenschaftliche Buchgesellschaft), Darmstadt
Die Herausgabe des Werkes wurde durch die
Vereinsmitglieder der wbg ermöglicht.
Satz und eBook: Satzweiss.com Print, Web, Software GmbH
Umschlagsabbildung: VTT Studio – stock.adobe.com
Gedruckt auf säurefreiem und
alterungsbeständigem Papier
Printed in Germany

Besuchen Sie uns im Internet: www.wbg-wissenverbindet.de

ISBN 978-3-534-40782-8

Elektronisch ist folgende Ausgabe erhältlich:
eBook (PDF): 978-3-534-40783-5

Inhalt

II Transformation der Religion
Roboter und Religion

III Transformation des Körpers
Medizin und Optimierung

IV Transformation des Krieges
Autoregulative Waffensysteme

V Transformation der Theologie
Theorie und Kritik

Vorwort der Reihe „Theologie und Künstliche Intelligenz“

Anna Puzio

Die zunehmende Technologisierung verändert die menschliche Lebenswelt, Mitwelt, Beziehungen, Bildung und Arbeit, den Gesundheitsbereich, Religion, Politik und Gesellschaft. Dadurch wird die Theologie vor neue Aufgaben gestellt. Hinzu kommt, dass religiöse Motive wie Heilsvorstellungen, paradiesische Motive, Unsterblichkeitsstreben, Beseitigung von Leid, Allmachts- und Schöpfungsvorstellungen im Technikdiskurs auftauchen, die einer theologischen Auseinandersetzung bedürfen. Religionen nehmen als kulturelle Akteure eine wichtige Rolle in der Technologisierung ein. Mit ihren Institutionen, Verbänden und religiösen Erzählungen beeinflussen sie Werte, Leitbilder, Weltvorstellungen, soziale Beziehungen, Gemeinschaft und Gesellschaft.

In der Theologie werden zum Beispiel neue ethische und anthropologische Reflexionen notwendig. Aber auch andere Felder der Theologie sind gefragt. Gerade die Vielfalt der Methoden und Disziplinen innerhalb der Theologie machen es ihr möglich, Technik und Künstliche Intelligenz sehr weitumfassend zu ergründen. Im Zuge der Technologisierung werden theologische Konzepte neu herausgefordert. Für die Theologie ergibt sich aber auch die Chance, ihre Konzepte neu zu hinterfragen und weiterzuentwickeln. Die Buchreihe *Theologie und Künstliche Intelligenz* wurde gegründet, um die theologische Beschäftigung mit Technik und Künstlicher Intelligenz zu fördern und ihr eine eigene Plattform zu geben. Sie geht davon aus, dass die Technologisierung ein relevantes und dringliches Forschungsthema für die Theologie darstellt sowie umgekehrt, dass die Theologie zum Technikdiskurs gewinnbringend beitragen kann.

Die Reihe *Theologie und Künstliche Intelligenz* widmet sich der theologischen Auseinandersetzung mit der Technologisierung und Künstlichen Intelligenz, aber auch allgemeineren Aspekten der Digitalisierung. Wie kann die Theologie zur Technologisierung beitragen? Welche spezifische Perspektive bringt sie mit? Welche Aufgaben ergeben sich für die Theologie? Die Themen betreffen sowohl Grundlagenreflexionen und theoretische Neuentwürfe als auch praktische Anwendungsbereiche. Sie reichen von medizinischen Technologien, Selbstoptimierung, Virtual und Augmented Reality und Alltagstechnologien wie Wearables über (Teil-)Autonomens Fahren und Geoengineering hin zu Transhumanismus, autoregulativen

Waffensystemen und religiöser Robotik. Alle Bände durchlaufen einen Double Peer Review Prozess. Die Reihe legt Wert auf Diversität. Weiterhin strebt sie an, interreligiös, interdisziplinär und international ausgerichtet zu sein, was in den nächsten Jahren weiter ausgebaut werden soll. Unterstützt wird diese Ausrichtung durch ein internationales, interdisziplinäres Editorial Board.

Foreword to the Theology and Artificial Intelligence Series

Anna Puzio

Increasing technologization is changing the human living environment, co-environment, relationships, education and work, healthcare, religion, politics and society. As a result, theology is confronted with new tasks. In addition, religious motifs such as ideas of salvation, paradisiacal themes, striving for immortality, the elimination of suffering and conceptions of divine omnipotence and creation appear in the discourse on technology, which requires theological analysis and debate. As cultural actors, religions play an important role in technologization. With their institutions, social organizations and narratives, they influence values, guiding principles, conceptions of the world, social relations, community and society.

In response to this complexity, new ethical and anthropological reflections are necessary in theology, but other fields are also important. It is precisely the diversity of methods and disciplines within theology that makes wide-ranging exploration of technology and artificial intelligence possible. Technologization challenges theological concepts anew and also offers the opportunity to develop them further. The book series *Theologie und Künstliche Intelligenz (Theology and Artificial Intelligence)* was founded to promote theological engagement with technology and artificial intelligence and to provide a platform for this area of inquiry. A guiding assumption for the series is that technologization is a relevant and urgent research topic for theology and, conversely, that theology can profitably contribute to the discourse on technology.

The series is dedicated to the theological examination of technologization and artificial intelligence, but it also explores more general aspects of digitalisation. How can theology contribute to technologization? What specific perspective does it bring? What tasks arise for theology? Topics include basic reflections and theoretical redesigns as well as practical areas of application. They range from medical technologies, self-optimisation, virtual and augmented reality and everyday technologies such as wearables to (semi-)autonomous driving and geoengineering to transhumanism, autoregulatory weapon systems and religious robotics. All volumes go through a double peer review process. The series emphasises diversity; it aims to be interreligious, interdisciplinary and international in scope, which will be further developed in the coming years. This orientation is supported by an international, interdisciplinary editorial board.

Theologie und Künstliche Intelligenz

Perspektiven, Aufgaben und Thesen einer Theologie der Technologisierung

Anna Puzio

1 Theologie, Technik und Künstliche Intelligenz

Künstliche Intelligenz (KI), Blockchain, Virtual und Augmented Reality, (Teil-)Autonomes Fahren, autoregulative Waffensysteme, Enhancement, Reproduktionstechnologien und humanoide Robotik – diese Technologien (und mit ihnen viele weitere) sind schon längst keine spekulativen Zukunftsvisionen mehr, sondern haben bereits Eingang in unser Leben gefunden oder stehen an der Schwelle zum Durchbruch. Die rasanten technologischen Entwicklungen wecken ein Orientierungsbedürfnis: Was unterscheidet den Menschen von der Maschine, die menschliche Intelligenz von der Künstlichen Intelligenz, wie weit sollte der Körper verändert werden dürfen, was sind die Gefahren und was sind die Chancen der Technologien?

Viele dieser Anfragen werden auch an die Theologie gerichtet. Es wird z. B. nach dem Menschenbild, nach dem Schöpfungsverständnis, nach einer Ethik der technologischen Körpereingriffe oder nach dem moralischen Status von Robotern gefragt. Was sagt die Theologie zu diesen technologischen Entwicklungen? Daher ist es höchste Zeit für die Theologie, das Themenfeld der Technologisierung und KI wissenschaftlich zu ergründen und Antworten zu formulieren. Verändern sich durch die Technologisierung die verschiedenen Bereiche des menschlichen Lebens, der Gesellschaft und Mitwelt, verändern sich auch die Orte und Themen der Theologie.

Während Fragen der Digitalisierung bereits intensiver ergründet worden ist, sind Technologisierung und KI vonseiten der Theologie noch unzulänglich erforscht. Der Band *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz* behandelt Technik und Künstliche Intelligenz (KI) aus explizit theologischer Perspektive. Damit legt der Band Wert darauf, dass die Theologie Reflexionen über ihre eigenen Theorien und

ihre eigene Perspektive im Technikdiskurs anstellt und fragt, was die technologischen Entwicklungen konkret für die Theologie bedeuten. Einige Beiträge fokussieren die KI, während andere Aufsätze noch weitere Technologien in den Blick nehmen. Der Band fasst unter den Technikbegriff verschiedene Artefakte und Gegenstände, naturwissenschaftliche Verfahren und technische Beschaffenheit zusammen, zielt jedoch nicht auf die „Techniken" im Sinne von Künsten, Tätigkeiten oder Methoden (wie Atem- und Meditationstechniken oder Werkzeuggebrauch). „Technik" und „Technologie" werden aus diesem Grund oft synonym verwendet. Der Band konzentriert sich vorrangig auf neue Technologien.

Als Auftaktband der Reihe *Theologie und Künstliche Intelligenz* gibt er Einblick in die Vielfalt der relevanten Forschungsfragen und versammelt Themen, die sich für die Theologie vor dem Hintergrund der technologischen Transformationen unmittelbar aufdrängen. Dazu gehören Transformationen des Menschenbildes und theologischer Theorien, humanoide Roboter und religiöse Roboter, autoregulative Waffensysteme, neue Möglichkeiten in der Medizin und Optimierungstechnologien. Der Band greift Perspektiven vor allem der Evangelischen Theologie, Katholischen Theologie und Jüdischen Theologie auf, sodass auch dieser Aufsatz Theologie und Religion stets im Plural meint, auch wenn er und die nachfolgenden Thesen aus der Perspektive einer christlichen Theologin verfasst worden sind und christliche Ansätze besonders fokussieren. Diese Perspektiven sollten in der zukünftigen Forschung um weitere Sichtweisen z. B. der nicht-monotheistischen Religionen ausgebaut werden.

Bevor ein detaillierter Einblick in den Band gegeben wird, sollen zuvor Ansätze für eine theologische Auseinandersetzung mit Technik und KI skizziert werden. Dies geschieht, indem in Kapitel 2 Thesen zu einer Theologie der Technologisierung aufgestellt werden und in Kapitel 3 die spezifisch theologische Perspektive und Aufgaben der Theologie reflektiert werden. Kapitel 4 zeigt konkrete Einsatzmöglichkeiten von Technik und KI in den Kirchen und Religionsgemeinschaften auf. Schließlich gibt Kapitel 5 einen ausführlichen Einblick in die Ausrichtung und die Themen des Bandes.

2 10 Thesen zu einer Theologie der Technologisierung

1. *Die Technologisierung ist relevant für die Theologie.*

Die menschliche Lebenswirklichkeit, Gesellschaft und Mitwelt sind grundlegend von Technik und KI geprägt. Durch Technik und KI verändert sich, wie wir leben, Beziehungen führen, kommunizieren, den Menschen verstehen, arbeiten, Politik und Gesellschaft gestalten. Damit ist die Technologisierung relevant für eine Theologie, die den Anspruch erhebt, an die menschliche Lebenswirklichkeit anzuknüpfen und über Verantwortung und Gerechtigkeit in Gesell-

schaft und Mitwelt zu reflektieren. Technik begleitet uns ständig im Alltag und ist nicht nur ein Zukunftsthema für die Theologie, sondern bereits ein brisantes Thema der Gegenwart. Auffällig ist, dass der ganze Technikdiskurs von religiösen Motiven wimmelt: Es begegnen Heilsvorstellungen, Hoffnung auf die Beseitigung von Leid, kognitive und moralische Verbesserung des Menschen, Auferstehungsmotive, Allmachtsfantasien, gottähnliche Technik, Paradiesvorstellungen und das Streben nach Unsterblichkeit. Diese Motive und Vorstellungen bedürfen einer theologischen und religionswissenschaftlichen Auseinandersetzung. Außerdem verändern sich durch die Technologisierung auch religiöse Praktiken und die Theologie (These 5).

2. *Die Theologie ist relevant für die Technologisierung.*

Nicht nur die Technologisierung ist relevant für die Theologie, sondern auch die theologische Perspektive ist relevant für die Technologisierung. Zuvor wurde bereits dargestellt, dass die schnellen technologischen Entwicklungen ein Orientierungsbedürfnis in der Gesellschaft wecken und Anfragen an die Theologie z. B. im Blick auf Anthropologie und Ethik gestellt werden. Außerdem sind trotz des großen Bedeutungsverlustes der christlichen Kirchen Religionen als kulturelle Akteure nicht zu unterschätzen. Mit ihren Institutionen, Verbänden und religiösen Erzählungen beeinflussen sie unsere Werte, Leitbilder, Weltvorstellungen, soziale Beziehungen, Gemeinschaft und Gesellschaft. Das Relevantsein ist jedoch nicht gesetzt und selbstverständlich, sondern ist immer auch als Aufgabe zu verstehen, als ein Relevant*werden*. Theologie und Religion haben einen großen Erfahrungsschatz im Blick auf soziale, religiöse und spirituelle Bedürfnisse des Menschen, die z. B. für die Mensch-Maschine-Interaktion wichtig sind. In Sozialer Robotik, z. B. bei der Verwendung von Robotern in Krankenhäusern, können religiöse Vorstellungen, Glaubensinhalte, soziale und emotionale Bedürfnisse eine wichtige Rolle spielen.

3. *Am Beginn des theologischen Engagements in der Technologisierung stehen eine gründliche wissenschaftliche Forschung und eine sachliche Vermessung des Diskurses. Polarisierungen sollten vermieden werden.*

Bevor geurteilt wird, muss sich die Theologie zunächst gründlich und wissenschaftlich mit der Technologisierung auseinandersetzen. Bei vielen technologischen Themen kann die Theologie zurzeit gar nicht mitreden, weil sie von ihnen keine Ahnung hat. Und wenn sie trotzdem urteilt, läuft sie Gefahr, den Themen und menschlichen Bedürfnissen nicht gerecht zu werden, keine Orientierung bieten zu können, ihre Relevanz und Glaubwürdigkeit im Diskurs zu verspielen. Der Technikdiskurs und auch die theologischen Beiträge zur Technologisierung wimmeln von Kampf- und Kriegsmetaphorik (es gibt eine „Invasion", etwas muss „bekämpft"

und etwas anderes muss „verteidigt" werden), von Hybrisvorwürfen, Polarisierungen und unklaren Begriffsverwendungen. Die Debatte wird sehr emotional geführt und Theolog:innen sehen Aspekte des Menschseins oder der Religion bedroht. Technologien werden dabei häufig zum eigenständigen und machtvollen Gegenüber stilisiert, denen der Mensch schon bald unterlegen und hilflos ausgeliefert sein wird. Technik ist jedoch kein von uns getrenntes eigenständiges Gegenüber, sondern vielmehr etwas, zu dem wir in einer engen Beziehung stehen. Von der Theologie sollte konstruktive Kritik an Missständen geleistet werden, allerdings kein blinder Technikpessimismus und keine Technikangst vorgebracht werden.

4. *Eine Theologie der Technologisierung ist interdisziplinär, interreligiös und international. Sie experimentiert und beschreitet kreativ und mutig neue Wege.*

Wie muss eine Theologie der Technologisierung arbeiten? Sie muss interdisziplinär, interreligiös und international arbeiten, um sich adäquat in die Technologisierung einbringen zu können. Im Technikdiskurs gewinnt die Zusammenarbeit mit den Technik- und Naturwissenschaften besondere Bedeutung. Dies bedeutet nicht nur einen bloßen interdisziplinären Austausch, sondern die Aneignung der Methoden der anderen Disziplinen und das Üben einer gemeinsamen Sprache. Die neue Situation, die technologisierte Gesellschaft, macht es erforderlich, neue Wege zu beschreiten und kreativ und ohne Angst zu experimentieren.

5. *Durch die Technologisierung werden Religion und Theologie transformiert.*

Nicht nur die menschliche Lebenswirklichkeit, Gesellschaft und Mitwelt, sondern auch religiöse Praktiken und Theologie selbst werden durch die technologischen Entwicklungen transformiert. Religiöse Roboter, Kommunikationstechnologien und der Chatbot „ChatGPT", der neue Formen des Forschens verspricht, wirken sich auf Religion und Theologie aus.

6. *Durch die Technologisierung werden theologische Konzepte und Theorien hinterfragt und neu herausgefordert. Zudem werden neue theologische Zugänge notwendig.*

Die Technologisierung verändert neben religiösen Praktiken und Formen des Wissenschaftbetreibens ebenfalls theologische Konzepte und Theorien. Beispielsweise kommt es vor allem im englischsprachigen Raum zu neuen Denkbewegungen, die die Beziehung des Menschen zur Technik neu reflektieren und damit neue Ansätze in Anthropologie und Ethik bieten. Das Segnen durch Roboter oder die Teilnahme an religiösen Zeremonien mittels Augmented und Virtual Reality werden kirchenrechtliche Bestimmungen herausfordern. Tradierte Konzepte werden neu herausgefordert und es bietet sich der Theologie die Chance, diese zu hinterfragen und weiterzu-

entwickeln. Dabei wird sich die Theologie nicht darauf beschränken können, alte Theorien auf eine neue Situation zu übertragen. Stattdessen werden neue Theorien und Konzepte notwendig. Die Erfahrungen der Gläubigen mit den Technologien sollten dabei eine wichtige Rolle spielen.

7. *Der Technologisierung muss stets in einer Doppelperspektive von Problemen bzw. Herausforderungen und Chancen betrachtet werden. Die Technologisierung bietet viele Chancen für Theologie und Religion.*

Während die Theologie Neuerungen und neuen technologischen Entwicklungen zunächst mit Angst, Abwehr und Skepsis begegnet, sollte konstruktive Kritik durch den Blick auf die potenziellen Chancen ergänzt werden. Anhand von Social Media wird bereits deutlich, wie mittels der Technologien ein Zugang zu den jungen Menschen gefunden werden kann. Hier kommunizieren Menschen ihre Ängste, Trauer und Freude, nehmen politische Haltungen ein, teilen ihre Meinungen und wichtige Lebensereignisse mit. Kirche, Religion und Theologie sollten sich auch auf Social Media einbringen und diese als Einladung verstehen, an die Lebenswirklichkeit der jungen Menschen anzuknüpfen.

8. *Theologie muss wagen, die vernachlässigten und verschwiegenen Themen zu behandeln und ganz neue Gedankengänge ausprobieren.*

Eine Theologie der Technologisierung muss eine Reihe von Themen behandeln, mit denen sie sich zurzeit nicht zu beschäftigen wagt. Dazu gehören zum Beispiel Sexrobotik, religiöse Robotik, Reproduktionstechnologien und Technik zur Kontrazeption. Die Theologie wird die Technologisierung nicht aufhalten können. Wenn sie sich nicht einbringt, werden die Technologien trotzdem entwickelt werden, aber ohne die Stimme der Theologie. Neue Themen brauchen neue Gedankengänge. Theologie kann innovativ sein, wenn sie lang akzeptierte Vorstellungen, an denen sie sich nicht zu rütteln traut, versucht umzukehren und neue Gedankengänge zumindest auszuprobieren. Wenn sie ablehnt, dass KI und Roboter Bewusstsein und eine Seele haben, sollte sie zunächst erörtern, was Bewusstsein und Seele sind und ob sie die Aussage tatsächlich widerlegen kann. Wenn sie KI und Roboter nicht als Schöpfung ansieht, sollte sie im ersten Schritt aufzeigen, was Schöpfung ausmacht und ob sich nicht auch das Gegenteil denken lässt.

9. *Theologie muss sich bereits in Design und Entwicklung der Technologien einbringen.*

Theologie zu betreiben, darf sich nicht nur auf theologische Theorien und ethische Leitlinien beschränken. Stattdessen braucht es auch Theolog:innen, die bereits im Design und der Entwicklung der Technologien mitwirken. Denn hier entscheidet sich bereits, „nach wessen

Bilde" Technologien entwickelt werden, d. h. welche Personengruppen nicht zu Wort kommen,[1] welche Menschenverständnisse und Werte relevant sind und für welche Zwecke Technologien entwickelt werden.

10. *Theologie sollte Influencerin werden.*

Die Rolle der Theologie sollte nicht sein, auf bestehende Verhältnisse nachträglich zu reagieren und Entwicklungen bloß zu kommentieren. Vielmehr sollte Theologie Influencerin werden, die in der Technologisierung Einfluss nimmt, Positionen entwickelt und starkmacht, Output generiert und innovativ ist. Was könnten theologische Perspektiven in der Technologisierung sein? Beispielsweise das Aufdecken von Machtverhältnissen und Diskriminierungen, den Blick auf die Mitwelt, das Starkmachen eines dynamischen und offenen Menschenverständnisses, neue Reflexionen auf den Körper und Sexualität, feministische, queere, antirassistische und interkulturelle Perspektiven sowie das Einstehen für Diversität.

3 Die spezifisch theologische Perspektive und Aufgaben der Theologie

In den theologischen Diskussionen auf Tagungen, in Forschungsgruppen oder Kommissionen begegnen immer wieder zwei Fragen. Zum einen, worin die spezifische Perspektive der Theologie besteht, die sie in den Technikdiskurs einbringen kann. Wie kann die Theologie beitragen? Unterscheidet sich die theologische Perspektive überhaupt von denjenigen anderer Disziplinen? Zum anderen wird nach den Aufgaben der Theologie im Kontext der Technologisierung gefragt. Autor:innen des Bandes und weitere Forschende der Theologie haben einen Anfang gemacht und erste Antworten darauf gewagt:[2]

Die Perspektive der Theologie

Wie kann die Theologie zur Technologisierung beitragen? Welche spezifische Perspektive bringt sie mit?

> Das, was man „Technologisierung" nennt, ist ein ambivalentes Phänomen. Auf der einen Seite eröffnet sie neue Möglichkeitsräume für Individuen und die

[1] Graham, Elaine L.: Representations of the Post/Human: Monsters, Aliens, and Others in Popular Culture. New Brunswick (NJ) 2002, 61, 111, 123.

[2] Hinweise zu den Autor:innen finden sich auch im Autor:innenverzeichnis.

Gesellschaft – auf der anderen Seite birgt sie die Gefahr, soziale Spaltungen voranzutreiben und Freiheitsräume einzuengen. Um angemessen auf die Technologisierung zugehen und deren Herausforderungen proaktiv begegnen zu können, ist es wichtig, beiderlei in den Blick zu nehmen. Mit ihrer reichen Denktradition und ihrer Vielfalt an leistungsstarken Reflexionsfiguren, kann die Theologie die eigene Perspektive schärfen, uns helfen, weder überzogenen Hypes noch allzu düsteren Technikdystopien aufzusitzen, und dazu beitragen, einen guten Umgang mit der Technologisierung zu finden.

Max Tretter, Evangelische Theologie, Erlangen

Theology specifically but also the humanities broadly is an integral part of the technologization of any culture, as we provide the means of critical reflection on their development and implementation. Such critical reflection is one of the only defenses against the insatiable ethos of capitalism, which imposes no natural limits and asks no ethical questions beyond the demands of the market.

Katherine Smith, Katholische Theologie, New York

Theology offers a critical perspective from which to engage the massive process of technologization. In the three Abrahamic traditions, but especially in Judaism, theology reminds us that a) humans are not their own makers; b) that embodied humans should not be reduced to data; and c) that relationality is the core of being human. The task of theology is to challenge us to examine the ethics, existential meaning, and societal impact of human-made technology rather than assume that it is either morally neutral or necessarily beneficial.

Hava Tirosh-Samuelson, Jüdische Theologie, Phoenix, Arizona

Die Theologie ist Fürsprecherin des Menschen der Zukunft. Einer Zukunft, die so unvorstellbar dezentral, partizipativ und frei sein wird.

Laurence Lerch, Katholische Theologie, Luzern

Die obigen Antworten sehen die Rolle der Theologie in der kritischen Prüfung technologischer Entwicklungen. Dazu gehört das Aufdecken von Ideologien und Machtverhältnissen. Durch den Blick vieler Religionen auf die Marginalisierten kann die Theologie auf Diskriminierungen aufmerksam machen, für Gerechtigkeit eintreten und Personengruppen, die in der Technologisierung nicht zu Wort kommen, eine Stimme geben. Die Perspektive der Theologie sollte aber eine doppelte sein: Neben der Ausübung von Kritik wird in den Antworten auch auf das Wahrnehmen von Chancen und neuen Möglichkeiten hingewiesen.

Tirosh-Samuelson weist auf die Bedeutung der Relationalität für die Theologie hin. Wie werden zwischenmenschliche Beziehungen durch Technologien verändert? Vernachlässigt wird bislang jedoch die Beziehung zur Technik. Dabei bietet sich gerade die Theologie dazu an, um das Verhältnis zu nicht-menschlichen Entitäten neu zu reflektieren. Die Theologie ist für eine Beschäftigung mit Technologien besonders geeignet, da sie z. B. einen breiten Fundus an ganz spezifischen Formen des Verhältnisses zum Nichtmenschlichen (z. B. in der Bibel) und eine Ethik zum Umgang mit dem Anderen hat. Sie verfügt über viele Erfahrungen mit den sozialen Bedürfnissen von Menschen, wie sie in Sozialer Robotik relevant werden oder mit den spirituellen Bedürfnissen, wie sie in der religiösen Robotik zentral sind.

Außerdem kommen im Zuge der Technologisierung viele anthropologische und ethische Fragen zum Menschenbild und Weltbild auf. Der Blick auf das Menschenbild, die Technikanthropologie stellt einen guten Ausgangspunkt für die theologische Auseinandersetzung mit KI und Technik dar. Welche Menschenverständnisse werden in den Technologien transportiert? Mittels Technologien werden neue Verständnisse von Mensch und Körper mitentworfen. Theologische Technikanthropologie fragt, „nach wessen Bilde" Technologien entworfen werden, welche Personengruppen in der Technologisierung unterrepräsentiert sind und gibt diesen eine Stimme.[3] Wer in die Technologisierung inkludiert wird, wirkt sich auch darauf aus, wie wir KI verstehen. Zurzeit sind ein „westliches" Intelligenzverständnis und „westliche" Werte vorherrschend. Dies beeinflusst auch die Zwecke, für die KI eingesetzt, aber auch schon entwickelt wird. Theologische Technikanthropologie sollte für ein dynamisches Menschenverständnis einstehen, das für Veränderung offen ist und Pluralität berücksichtigt.[4]

Die theologische Auseinandersetzung mit KI und Technik sollte sich aber nicht nur auf die Anthropologie beschränken. Die Vielfalt der Methoden und Disziplinen innerhalb der Theologie machen es ihr möglich, Technologisierung sehr umfassend zu ergründen. Diese Reflexionen führen zur zweiten Frage, und zwar der Frage nach den Aufgaben der Theologie im Kontext der Technologisierung:

---

3 Graham: Representations, 61, 111, 123.

4 Vgl. Puzio, Anna: Über-Menschen. Philosophische Auseinandersetzung mit der Anthropologie des Transhumanismus (Edition Moderne Postmoderne). Bielefeld 2022, Teil III; Puzio, Anna: Zeig mir deine Technik und ich sag dir, wer du bist? – Was Technikanthropologie ist und warum wir sie dringend brauchen. In: Diebel-Fischer, Hermann/Kunkel, Nicole/Zeyher-Quattlender, Julian (Hg.): Mensch und Maschine im Zeitalter ‚Künstlicher Intelligenz'. Theologische Herausforderungen. 2023; Puzio, Anna: Digital and Technological Identities – In Whose Image? A philosophical-theological approach to identity construction in social media and technology. In: Cursor (2021). Online at: https://cursor.pubpub.org/pub/y2bcesx4 (Stand: 14.03.22).

Aufgaben der Theologie

Welche Aufgaben ergeben sich im Blick auf die Technologisierung für die Theologie?

> The first task of theology, academically speaking, is the careful study of the subject and landscape. Too often, theologians rush in with judgment without understanding the technology at hand.
>
> *Katherine Smith, Katholische Theologie, New York*

> Technik ist Teil menschlicher Kulturausübung; sie dient wie die Kultur insgesamt der Lebensbewältigung. Ethische Theologie hat die technische Kultur auf diesen Zweck hin immer wieder zu prüfen und damit Tendenzen ihrer Sakralisierung wie Selbstverzweckung zu wehren.
>
> *Yannick Schlote, Evangelische Theologie, München*

> Theology, for me, means to be vitally concerned to investigate, discover, and live in accord with divine will. Now, if we understand technologisation to be the design, development and deployment of tools, then the task for theology is, quite clearly, to ensure that our tools are designed, developed and deployed in accord with divine will.
>
> *Mois Navon, Jüdische Theologie, Tel Aviv*

> Technology has forced religion to foster a relationship with cyberspace. Virtual theology is needed to continue preaching religious practices and beliefs to the masses that have an online presence and for those who turn to virtual spaces for religious discussions and inquiry. Religion is now offline, online, and hybrid.
>
> *Sana Patel, Islamische Theologie, Ottawa*

Smith geht davon aus, dass die theologische Auseinandersetzung mit Technologien zunächst eine tiefe wissenschaftliche Erforschung und eine gründliche Vermessung des Diskurses braucht. Danach können beispielsweise Menschenverständnisse und neue ethische Herausforderungen eine wichtige Rolle spielen. Schlote weist auf die Ethik und Lebensbewältigung hin. Patel macht stark, dass die Theologie in direkten Kontakt mit den Technologien treten muss.

4 Wie können Kirchen und Religionsgemeinschaften KI und Technik einsetzen?

Die Einsatzmöglichkeiten von KI und Technik in Kirchen und Religionsgemeinschaften sind vielfältig und werden sich im Laufe der Zeit und im Zuge der fortschreitenden Technologisierung verändern, sodass hier nur einige mögliche Perspektiven für den Einsatz von Technik und KI aufgezeigt werden sollen. Da es in den Religionen spezifische Umgangsformen mit Bildern, verschiedene religiöse Lehren und Vorschriften in religiösen Gebäuden gibt, wird der Fokus hier auf das Christentum gelegt. Dennoch können die Ideen ebenfalls für viele andere Religionen fruchtbar gemacht werden.

Eine große Bedeutung kann KI bei vielen Prozessen einnehmen, die im Hintergrund ablaufen. Kirchen und Religionsgemeinschaften beschäftigen sich mit Veranstaltungsorganisation, Organisation von Verbänden und Gemeinschaften und haben eine Menge von Daten z. B. der Gläubigen zu verwalten. KI ist sehr effizient im Umgang mit Daten (Daten auswerten, sortieren, leichter zugänglich machen, anonymisieren) und kann für Prognosen und Prozessoptimierung eingesetzt werden. Es ergeben sich auch ganz neue Einsatzmöglichkeiten von Daten für die Religionsgemeinschaften, die evaluiert werden müssen. Darüber hinaus können KI und Technik zur Zielgruppenansprache, für Strategieentwicklung und Marketing verwendet werden.

In den letzten Jahren hat KI viele Fortschritte im Umgang mit Texten, Bildern und Musik gemacht, die in Religionen eine wichtige Rolle spielen. Texterkennung und Textgenerierung, Musikkomposition, Bilderkennung und KI, die Bilder malt, können für religiöse Praktiken und Veranstaltungen, für den Besuch von religiösen Gebäuden und religiöse Bildung eingesetzt werden. Der Einsatz von KI für Übersetzungen oder einfache Sprache können zu einer inklusiven Kirche beitragen. Es ergeben sich neue Möglichkeiten, die Botschaften der Religionsgemeinschaften und Informationsmaterial digital zu präsentieren, z. B. mittels Projektionen, Einblendungen, Bots und Touchscreens. KI kann auch für Bibelarbeit und Wissensmanagement verwendet werden. Die KI-basierten Text- und Bildtools werden ebenfalls die theologische Forschung verändern, z. B. die Recherche und Textgenerierung, Bibelforschung sowie den Umgang mit alten Texten und alten Schriften.

Eine spannende Chance für die Kirchen und Religionsgemeinschaften stellen Virtual und Augmented Reality dar, die bereits in den Kulturbereich Eingang gefunden haben. In religiöser Bildung können durch Virtual und Augmented Reality Zugang zu alten Erfahrungswelten und historische Reisen ermöglicht und religiöse Stätten besucht werden. In vielen Religionen nehmen bestimmte Orte und Länder eine wichtige Bedeutung ein, die von vielen Gläubigen aber gar nicht bereist werden können. Auch die Teilnahme an religiösen Zeremonien kann für diejenigen möglich gemacht werden, die nicht dabei sein

können, weil sie z. B. zu krank sind. Von zuhause, von der Pflegeeinrichtung oder vom Krankenhaus aus können Gläubige mittels Virtual und Augmented Reality an der religiösen Feier teilnehmen, mit speziellem Equipment religiöse Gegenstände anfassen, haptische und olfaktorische Eindrücke haben. Ebenfalls für Menschen mit Behinderungen können spezielle Zugänge zu religiösen Veranstaltungen angeboten werden, beispielweise bestimmte Bewegungen erleichtert werden. Auf diese Weise können Virtual und Augmented Reality ein Stück mehr zu einer inklusiven Kirche beitragen. Darüber hinaus sind auch ganz neue Verwendungsweisen von Virtual und Augmented Reality denkbar. Sie können in religiösen Ritualen oder begleitend zu spirituellen Erfahrungen eingesetzt werden. Daran wird wieder deutlich, dass im Zuge der Technologisierung auch religiöse Praktiken transformiert werden. Ebenfalls können Informationsmaterial und religiöse Botschaften auf neue Weise zugänglich und interessant gemacht werden. Weiterhin erfreut sich Immersive Art zunehmend an Popularität. Bei Immersiver Kunst tauchen die Betrachtenden regelrecht in die Kunst ein. Kunst wird zum multimedialen Erlebnis mit Licht- und Soundeffekten, VR-Brillen, Videoprojektionen, haptischen und olfaktorischen Effekten. Da Kunst eine wichtige Rolle in vielen Religionen spielt, bietet es sich an, Immersive Art auch in Religionsgemeinschaften anzubieten.

Eine weitere Technologie, die Religionen und Religionsgemeinschaften prägen wird, stellt die Robotik dar. Religiöse Roboter sind Roboter, die für religiöse Zwecke eingesetzt werden. Diese sind bislang in den nicht-monotheistischen Religionen und besonders im asiatischen Raum weiter verbreitet. Religiöse Roboter können Gebete begleiten, Gespräche führen, religiöse Zeremonien feiern, aus religiösen Schriften vorlesen und Musik abspielen. Sie können Führungen durch religiöse Gebäude geben, Fragen zur jeweiligen Religion beantworten und mit denjenigen, die nicht vor Ort sein können, chatten oder ihnen religiöse Feiern über das Internet tragen. Noch bevor religiöse Roboter für explizit religiöse Zwecke weite Verwendung finden, ist es naheliegend, dass in Soziale Roboter religiöse Aspekte integriert werden. Soziale Roboter können beispielsweise im Gesundheitsbereich, in Krankenhäusern und Pflegeeinrichtungen eingesetzt werden. Soziale Roboter können Gespräche führen, Tabletten oder Spritzen geben, Kindern und ihren Eltern den langen Krankenhausaufenthalt erleichtern. Wenn Roboter mit den zu behandelnden Personen interagieren und Gespräche führen, können gerade in der Krankheits- und Pflegesituation auch religiöse und spirituelle Bedürfnisse aufkommen. Sollten diese Roboter tatsächlich atheistisch bzw. agnostisch sein oder sollten sie auch Auskunft über Religionen geben, religiöse Werte, religiöse und spirituelle Elemente integriert haben?

Ferner können die vielen Funktionen, die unter dem Namen Smart Home Eingang in die Häuser finden, auch in den Gebäuden der Religionsgemeinschaften Anwendung finden. Unter Smart Home versteht man die Vernetzung von Technik im Haus wie Licht, Heizung, Klima-

tisierung, Türverriegelung, Sprachassistenten, Küchengeräten, Fernseher und weiterer Unterhaltungselektronik. In religiösen Gebäuden wären andere Technologien und Funktionen denkbar. Analog zu Smart Home würde es Smarte Kirche geben.

Im Technikdiskurs und in vielen Religionsgemeinschaften herrscht Angst, dass Technik den Menschen zunehmend ersetzt und damit wertvolle zwischenmenschliche Erfahrungen verloren gehen. Dabei müssen Technologien den Menschen nicht ersetzen oder imitieren, sondern förderlich ist, wenn Technologien gerade das tun, was sie besonders gut können. Dazu gehören z. B.: die Verarbeitung und Speicherung von Daten, bestimmte Hebebewegungen in der Pflege und die Eigenschaft, dass man in sich bei manchen Pflegetätigkeiten vor der Technik weniger schämt als vor Menschen sowie beeindruckende virtuelle Erfahrungen, visuelle und haptische Effekte, die religiöse Praktiken ergänzen oder religiöse Zeremonien inklusiver machen.

5 Zum Band

Der Band *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz* versammelt deutsch- und englischsprachige Beiträge zu Technik und KI aus dem deutschsprachigen und internationalen Raum, um die Vielfalt der theologischen Forschungsdiskurse aufzuzeigen. Angestrebt werden interdisziplinäre Auseinandersetzungen der Theologie, besonders mit den Technik- und Naturwissenschaften. Die Beiträge wurden vor allem von Forschenden der Evangelischen, Katholischen und Jüdischen Theologie verfasst. Der Band gliedert sich in fünf Sektionen und beleuchtet die Transformationen in verschiedenen Bereichen des menschlichen Lebens und in der Theologie, zu denen es im Kontext der Technologisierung kommt. Als Auftaktband der Reihe *Theologie und Künstliche Intelligenz* gibt er Einblick in die Vielfalt der theologisch relevanten Themen und bietet eine Übersicht über aktuelle Forschungsdiskurse. Der Band kann nicht alle in Bezug auf die Technologisierung relevanten Forschungsthemen aufgreifen, zumal sich diese ständig verändern werden, sondern bietet vielmehr Ansatzpunkte für die weitere theologische Forschung.

Die erste Sektion *Transformation des Menschenbildes: Mensch und Roboter* nimmt anthropologische Reflexionen vor. Durch die Technologisierung werden Menschenbilder transformiert. Für *Katherine Smith* gehört die Anthropologie zu den wichtigsten Forschungsfeldern der Theologie im KI-Diskurs. Ausgehend von ihren Erfahrungen mit menschenähnlichen medizinischen Übungspuppen an der Barbara H. Hagan School of Nursing and Health Sciences des Molloy College in New York geht sie Zusammenhängen von Anthropologie und KI nach. In ihrem Beitrag *Learn, Remember, Act: Theological Anthropology and AI Metaphor* vergleicht sie die Unterschiede von Mensch und KI im Blick

auf Lernen, Erinnern und Handeln. *Lukas Brand* setzt sich in seinem Beitrag *Grundlinien eines Menschenbilds der Künstlichen Intelligenz* mit der Frage auseinander, welches Menschenbild KI-Systeme repräsentieren. Am Beispiel des humanoiden Roboters „Optimus", den Tesla 2022 präsentierte, widmet er sich der technologischen Reproduktion des Menschen. Auch *Sven Nyholm* greift humanoide Robotik auf und diskutiert in seinem Aufsatz *Wie sollen wir mit künstlich-intelligenten humanoiden Robotern umgehen?* den moralischen Status von Robotern. Nyholm fragt, ob Roboter moralisch relevante Eigenschaften oder Fähigkeiten haben können, diese nachahmen oder repräsentieren können. In seinem Aufsatz bietet er einen Überblick über die gegenwärtige internationale Debatte zu Robotern als Trägern von Rechten.

In der zweiten Sektion *Transformation der Religion: Roboter und Religion* wird der Roboterdiskurs weiterverfolgt und der Fokus nun auf den religiösen Kontex gelegt. *Anna Puzio* untersucht in ihrem Beitrag *Robot Theology: On the Theological Engagement with Robotics and Religious Robots* religiöse Roboter, d. h. Roboter, die für religiöse Zwecke verwendet werden. Sie zeigt den Einsatz von Robotern in verschiedenen Religionen auf und verweist dabei auf die Bedeutung von zeitabhängigen, kulturell ausgehandelten Konzepten von Mensch und Nicht-Mensch, Leben und Schöpfung. Ihr Aufsatz trägt zur Profilierung der zukünftigen theologischen Beschäftigung mit Robotik bei. *Hendrik Klinge* wendet sich dem Zusammenhang von Religion und Robotik aus einer anderen Perspektive zu, indem er die Religiösität von Robotern untersucht. In seinem Artikel *Do Robots Believe in Electric Gods?* geht er anhand eines theologischen Turing Tests und unter Einbezug von Wittgenstein der Frage nach, ob Roboter einen religiösen Glauben haben können.

Die dritte Sektion widmet sich der *Transformation des Körpers* in den Bereichen *Medizin und Optimierung*. *Max Tretter* beginnt mit der Untersuchung des Self Trackings, das im Alltag bereits weit verbreitet ist. Technologien und KI sollen der Ungewissheit entgegenwirken und neue Entscheidungsfreiheiten ermöglichen. Tretter erforscht im Aufsatz *Ambivalenzen gegenwärtiger Gewissheitsbestrebungen*, wie diese technologisch erzeugten Gewissheiten sich tatsächlich auf die menschliche Entscheidungsfreiheit auswirken und macht dafür die Simulationstheorie von Jean Baudrillard fruchtbar. *Yannick Schlote* befasst sich im Beitrag *On Digital Twins and Heavenly Doppelgangers* mit Digitalen Zwillingen in der Medizin, von denen prognostiziert wird, eines der großen Zukunftsthemen der nächsten Jahre zu werden. Digitale Zwillinge können z. B. die digitale Repräsentation von Patient:innen zur Simulation von medizinischen Anwendungen sein. Schlote zeigt dabei die Ähnlichkeiten des Digitalen Zwillings mit dem gnostischen Glauben an die Koexistenz des Menschen mit seinem himmlischen Doppelgänger auf und leistet davon ausgehend eine ethische Bewertung Digitaler Zwillinge. *Dominik Winter* richtet seinen Blick ebenfalls auf Zukunftstechnologien. Er setzt sich mit dem Moral Enhancement, d. h. der

moralischen Verbesserung des Menschen durch technologische Einwirkung, und mit der transhumanistischen Verhältnisbestimmung von Körper und Geist auseinander. Winter formuliert *Impulse für Fragen der Leiblichkeit in der Ethik vor dem Hintergrund des Moral-Enhancement-Diskurses*.

Für eine Theologie der Technologisierung und KI spielen auch *autoregulative Waffensysteme*, mit denen sich die vierte Sektion zu den *Transformationen des Krieges* beschäftigt, eine zentrale Rolle. *Nicole Kunkel* problematisiert, dass solche Waffen Menschen tödlich treffen können, dabei aber auf nicht unproblematischen Algorithmen basieren und ihres Erachtens keine moralischen Entscheidungen fällen können. In ihrem Aufsatz *Automatization challenging Peace Ethics* bearbeitet sie das Thema aus der Perspektive der christlichen Friedensethik. Nachdem Kunkel Einblick in den Gesamtdiskurs gegeben hat, legt *Mois Navon* den Fokus auf die menschliche Würde. Im Aufsatz *Autonomous Weapons Systems and Battlefield Dignity* argumentiert Navon aus der Perspektive der Jüdischen Theologie, dass die Würde auf dem Schlachtfeld eine eigene ethische Kategorie ist, die ganz anders definiert ist als die Würde in Friedenszeiten.

Abschließend zeigt die Sektion *Transformation der Theologie: Theorie und Kritik*, wie im Kontext der Technologisierung theologisch relevante Theorien verändert werden, neue Theorien aufkommen, und eine theologische Kritik geübt werden kann. Im Aufsatz *Jewish Philosophy and the Critique of AI Technology* bezieht sich *Hava Tirosh-Samuelson* auf Emmanuel Levinas, Hans Jonas und Jonathan Sacks, um Kritik am Transhumanismus zu üben. Tirosh-Samuelson setzt dem Transhumanismus die Werte der Freiheit, der Verantwortung und der verkörperten Würde als kritische Antworten des Judentums entgegen. Anschließend stellt *Lukas Ohly* im Aufsatz *Digitale Transformation des Unsichtbaren* einige *Schöpfungstheologische Anmerkungen zu den Grenzen des digitalen Herstellens im Anschluss an Hannah Arendt* an. Dabei erörtert Ohly auch das digitale Abendmahl. *Simon Reiners* beschäftigt sich im Aufsatz *Metaversum und resistente Körperlichkeit* mit dem Metaversum und dem Humanismus. Er stellt den Körper in den Mittelpunkt seiner Überlegungen und diskutiert materiell-feministische Theorien, Donna Haraway und Theodor Adorno.

Der Band knüpft an die 2021 vom Netzwerk für Theologie und Künstliche Intelligenz *neth:KI* organisierte Tagung *Alexa, wie hast du's mit der Religion? Technik, Digitalisierung und Künstliche Intelligenz im Fokus der Theologie* (Tagungsteam: Lukas Brand, Nicole Kunkel, Julia van der Linde, Anna Puzio) an. Das internationale und interreligiöse Netzwerk setzt sich zum Ziel, die theologische Beschäftigung mit Technik und KI zu fördern. Die Tagungsthemen wurden durch weitere Forschungsbeiträge ergänzt, unter anderem von Mitgliedern aus dem Netzwerk neth:KI. Alle Aufsätze haben einen doppelten Peer-

Review-Prozess durchlaufen. Für die Mitwirkung an der ersten Konzeption des Bandes bedanken wir uns bei Lukas Brand. Für die Unterstützung bei der Manuskripterstellung danken wir Saskia Fischer. Die Veröffentlichung wurde gefördert aus dem Open-Access-Publikationsfonds der Humboldt-Universität zu Berlin.

Theology Meets AI

Examining Perspectives, Tasks, and Theses on the Intersection of Technology and Religion

Anna Puzio

1 Theology, Technology and Artificial Intelligence

Artificial intelligence (AI), blockchain, virtual and augmented reality, (semi-)autonomous vehicles, autoregulatory weapon systems, enhancement, reproductive technologies and humanoid robotics – these technologies (and many others) are no longer speculative visions of the future; they have already found their way into our lives or are on the verge of a breakthrough. These rapid technological developments awaken a need for orientation: what distinguishes human from machine and human intelligence from artificial intelligence, how far should the body be allowed to be changed and what are the dangers and opportunities presented by these technologies?

Many of these questions are also addressed to theology. For example, questions about the image of humanity, the understanding of creation, the ethics of technological body interventions or the moral status of robots. What does theology have to say about these technological developments? It is the right time for theology to scientifically explore technologization and AI and to formulate answers. As technology changes the various areas of human life, society and the world around us, the places and topics of theology are also undergoing transformation.

While questions raised by digitalisation have already been the subject of intense inquiry, theological investigation of technologization and AI is still insufficient. This volume, *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz (Alexa, How Do You Feel about Religion? Theological Approaches to Technology and Artificial Intelligence)*, deals with technology and artificial intelligence (AI) from an explicitly theological perspective. In doing so, it asserts that theology should reflect on its own theories and perspectives in the discourse on technology and identify the concrete implications of techno-

logical developments for theology. Some articles in this collection focus on AI, while others examine also other technologies.

As the first volume in the series *Theologie und Künstliche Intelligenz (Theology and Artificial Intelligence),* this collection provides insight into the diversity of relevant research questions and topics that arise for theology against the background of technological transformations. These include transformations of the human, humanoid robots and religious robots, autoregulatory weapon systems, new possibilities in medicine and optimisation technologies. The volume incorporates perspectives primarily from Protestant theology, Catholic theology and Jewish theology, so in this introductory article, theology and religion are always considered to be plural, even though it has been written from the perspective of a Christian theologian and focuses on Christian approaches. These perspectives should be expanded in future research to include further viewpoints, e.g. those of the non-monotheistic religions.

In what follows, Section 2 establishes theses on a theology of technologization, and Section 3 presents a reflection on the theological perspective and tasks of theology in relation to technology. Section 4 illustrates possibilities of using technology and AI in churches and religious communities. Finally, Section 5 provides a detailed summary of the volume's contents, indicating the variety of possible approaches to theological engagement with technology and AI.

2 10 Theses on a Theology of Technologization

1. *Technologization is relevant to theology.*

Technology and AI fundamentally shape the reality of human life, society and co-world; they are changing how we live, conduct relationships, communicate, work, engage politics and society and understand the human being. Technologization is thus relevant for any theology that claims to connect with the reality of human life and to reflect on responsibility and justice in society and co-world. Technology accompanies us constantly in everyday life and is no longer a future topic for theology: it is already an explosive topic of the present. It is striking that the discourse on technology is teeming with religious motifs: we encounter ideas of salvation, hope for the elimination of suffering, cognitive and moral improvement of the human, resurrection narratives, fantasies of omnipotence, comparisons with god-like capacities, visions of paradise and the pursuit of immortality. These motifs and ideas require a discussion in theology and religious studies. In addition, religious practices and theology are also changing as a result of technologization (thesis 5).

2. *Theology is relevant to technologization.*

Not only is technologization relevant for theology, but the theological perspective is also relevant for technologization. It has already been shown that the rapid technological developments create a need for orientation in society, posing significant questions to theology, e.g., with regard to anthropology and ethics. Furthermore, despite significant declines in the importance of Christian churches, religions should not be underestimated as cultural actors. With their institutions, organizations and narratives, they influence our values, guiding principles, worldviews, social relations and community. However, relevance is not a given: it must always be understood as a task, a *becoming* relevant. Theology and religion embody a wealth of experience with regard to the social, religious and spiritual needs of humans, and these needs are relevant to a variety of technological applications, such as human-machine interaction and social robotics (e.g., in the use of robots in hospitals). Drawing upon these resources, theology can have an impact on the actual design and approach of future technologies.

3. *Thorough scientific research and a factual survey of the landscape of the discourse represent the starting point for theological engagement with technology, and polarisation should be avoided.*

Before judging, theology must first analyse technologization thoroughly and scientifically. At present, theology cannot have a say in many discussions about technology because it is technologically illiterate. Even partial and incomplete information risks failure to offer orientation, to establish relevance and credibility and to effectively address the topic and related human needs. The discourse on technology in general and the theological discourse on technology in particular are teeming with metaphors of struggle and war (e.g., there is an "invasion," something has to be "fought" and something else "defended"), with accusations of hubris, polarisation and unclear use of terms. The debate is very emotional and many theologians see aspects of humanity or religion threatened. At the same time, technologies are often stylised as independent and powerful opponents to which humans will soon be inferior, helplessly at their mercy. However, technology is not a separate and independent counterpart, but rather something with which we are already in a close relationship. Theology should respond to technology with constructive criticism and thoughtful grievances, not with blind pessimism or irrational fear.

4. *A theology of technologization is interdisciplinary, interreligious and international. It experiments and breaks new ground creatively and courageously.*

Theology must be interdisciplinary, interreligious and international to adequately address technologization. In theological discourse on technology, cooperation with the engineering, technical and natural sciences is particularly important, and its form should not be limited to a mere interdisciplinary exchange: it should involve an appropriation of the methods of other disciplines and the construction of a common language. The new situation, the technologized society, necessitates breaking new ground and experimenting creatively and without fear.

5. *Religion and theology are transformed by technologization.*

Along with the overarching reality of human life, society and co-world, technological developments are also transforming religious practices and theology itself (Section 4). Religious robots, communication technologies and the chatbot "ChatGPT," which promises new forms of research, are having an impact on religion and theology.

6. *Technologization challenges and questions theological concepts and theories and as a result, new theological approaches are necessary.*

Technologization is also transforming theological concepts and theories, alongside religious practices and forms of scientific activity. For example, especially in the English-speaking world, there are new intellectual movements that offer new insights on the relationship between humans and technology and thus offer new approaches in anthropology and ethics. Blessing by robots or participation in religious ceremonies using augmented and virtual reality, for example, will challenge church law regulations. Traditional concepts will continue to be challenged, and theology will have the opportunity to question them and develop further. In this effort, theology will not be able to limit itself to transferring old theories to a new situation; instead, new theories and concepts will be necessary. Believers' experience with technologies should play a crucial role in these innovations.

7. *Technologization must always be seen from the perspective of both its challenges and its opportunities; it offers many opportunities for theology and religion.*

While theology often greets technological developments with fear, resistance and scepticism, attention to potential opportunities should complement constructive criticism. For example, social media have already demonstrated a significant capacity to reach young people. On social

media platforms people communicate their fears, sadness and joy, take political stances, and share their opinions and important life events. Church, religion and theology should also be involved in social media and understand them as an invitation to connect with the reality of young people's lives.

8. *Theology must dare to contend with neglected and silenced topics and explore completely new ways of thinking.*

A theology of technologization must address a number of issues that it is currently reluctant to engage, including sex robotics, religious robotics, and both reproductive and contraceptive technologies. Technologization is inevitable, and if theology does not get involved, it will continue to develop without the theology's voice. New issues need new ways of thinking, and theology can be innovative if it reverses long-accepted ideas. If, for example, traditional theology rejects the proposition that AI and robots have consciousness and a soul, it must first discuss its definitions of consciousness and soul and explore whether it can actually refute the statement. In addition, if it does not consider AI and robots as part of creation, it must first establish what constitutes creation and demonstrate that these entities are excluded from it.

9. *Theology must be involved in the design and development of technologies.*

Doing theology should not be limited to theories and ethical guidelines. Instead, theologians need to be involved in the design and development of technologies. In this stage decisions are made about "in whose image" technologies are developed, i.e. which groups of people do not have a say in these decisions,[1] which conceptions of human beings and human values are relevant and for which purposes technologies are developed.

10. *Theology should become an influencer.*

Theology should not simply react to existing conditions and merely comment on developments. Rather, it should become an influencer in technologization, developing and strengthening positions, generating innovative output. What exactly does the theological perspective contribute to the discourse on technologization? Potential answers include exposing power relations and discrimination; examining the co-world; strengthening a dy-

[1] Graham, Elaine L.: Representations of the Post/Human: Monsters, Aliens, and Others in Popular Culture. New Brunswick (NJ) 2002, 61, 111, 123.

namic and open understanding of humanity; putting forward new reflections on the body; integrating sexuality and feminist, queer, anti-racist and intercultural perspectives; and standing up for diversity.

3 The theological perspective and the tasks of theology

In theological discussions at conferences and in research groups or commissions, two questions come up again and again. First, what is the specific perspective that theology brings to the discourse on technology? How can theology contribute? Is the theological perspective different from that of other disciplines? Second, the tasks of theology in the context of technologization are also under discussion. The authors represented in this volume and others in the field have ventured initial answers.[2]

The perspective of theology

What can theology contribute to technologization? What specific perspective does it bring?

> What is called "technologization" is an ambivalent phenomenon. On the one hand, it opens up new possibilities for individuals and society – on the other hand, it bears the danger of driving social divisions and restricting freedom. In order to approach technologization appropriately and to proactively meet its challenges, it is important to take both into account. With its rich tradition of thought and its variety of powerful figures of reflection, theology can sharpen our own perspective, help us not to fall for exaggerated hypes or overly gloomy technological dystopias, and contribute to finding a good way of dealing with technologization.
>
> *Max Tretter, Protestant Theology, Erlangen*

> Theology specifically but also the humanities broadly is an integral part of the technologization of any culture, as we provide the means of critical reflection on its development and implementation. Such critical reflection is one of the only defenses against the insatiable ethos of capitalism, which imposes no natural limits and asks no ethical questions beyond the demands of the market.
>
> *Katherine Smith, Catholic Theology, New York*

[2] For details on the authors, see the author index.

> Theology offers a critical perspective from which to engage the massive process of technologization. In the three Abrahamic traditions, but especially in Judaism, theology reminds us that a) humans are not their own makers; b) that embodied humans should not be reduced to data; and c) that relationality is the core of being human. The task of theology is to challenge us to examine the ethics, existential meaning, and societal impact of human-made technology rather than assume that it is either morally neutral or necessarily beneficial.
>
> *Hava Tirosh-Samuelson, Jewish Theology, Phoenix, Arizona*

> Theology is the advocate of the human being of the future. A future that will be so unimaginably decentralised, participatory and free.
>
> *Laurence Lerch, Catholic Theology, Lucerne*

These responses emphasise theology's role in the critical examination of technological developments. This orientation includes exposing ideologies and power relations; in raising consciousness about the marginalised in many religions, theology can draw attention to discrimination, advocate for justice and give voice to groups of people who are usually left out of discussions about technology. However, the perspective of theology should be twofold: in addition to engaging in critique, these responses also point to opportunities and new possibilities.

Tirosh-Samuelson highlights the importance of relationality. How does technology change interhuman relationships? Theology's responses to this problem have been neglected so far, yet it is precisely theology that lends itself to reflecting anew on our relationship to non-human entities: it has, for example, many resources for reflecting on very specific forms of relationships with the non-human (e.g., in the Bible) and an ethic for dealing with the Other. It also has a long history of concerning itself with the social needs of humans, which is relevant in thinking about social robotics, and with spiritual needs, which is central in religious robotics.

Furthermore, in the course of technologization, many anthropological and ethical questions about the image of human and the world emerge. The concept of humanity, or the anthropology of technology, provides a good starting point for theological engagement with AI and technology. What understandings of humanity are conveyed in the technologies? Through technologies, new understandings of humanity and the body are being designed. Theological anthropology of technology asks "in whose image" technologies are designed, which groups of people are underrepresented in technologization and gives them a voice.[3] Who is included in technologization also affects how we understand AI. Currently, a "Western" understanding of intelligence and "Western" values predominate in the field of AI. Who participates in technological advancement also influences the purposes for which AI is used and developed.

[3] Graham: Representations, 61, 111, 123.

Theological anthropology of technology can advocate for a dynamic understanding of human beings that is open to change and takes plurality into account.[4]

However, theological engagement with AI and technology should not be limited to anthropology. The diversity of methods and disciplines within theology makes it possible to explore technologization comprehensively. These reflections lead to the second question, namely the tasks of theology in the context of technologization.

The tasks of theology

What tasks arise for theology in view of technologization?

> The first task of theology, academically speaking, is careful study of the subject and landscape. Too often, theologians rush in with judgment without understanding the technology at hand.
>
> *Katherine Smith, Catholic Theology, New York*

> Technology is part of human cultural practice; like culture as a whole, it serves to cope with life. Ethical theology has to examine technical culture again and again with regard to this purpose and thus resist tendencies towards sacralisation and self-purposing.
>
> *Yannick Schlote, Protestant Theology, Munich*

> Theology, for me, means to be vitally concerned to investigate, discover, and live in accord with divine will. Now, if we understand technologization to be the design, development and deployment of tools, then the task for theology is, quite clearly, to ensure that our tools are designed, developed and deployed in accord with divine will.
>
> *Mois Navon, Jewish Theology, Tel Aviv*

> Technology has forced religion to foster a relationship with cyberspace. Virtual theology is needed to continue preaching religious practices and beliefs to the masses that have

[4] See Puzio, Anna: Über-Menschen. Philosophische Auseinandersetzung mit der Anthropologie des Transhumanismus (Edition Moderne Postmoderne). Bielefeld 2022, Part III; Puzio, Anna: Zeig mir deine Technik und ich sag dir, wer du bist? – Was Technikanthropologie ist und warum wir sie dringend brauchen. In: Diebel-Fischer, Hermann/Kunkel, Nicole/Zeyher-Quattlender, Julian (eds.): Mensch und Maschine im Zeitalter 'Künstlicher Intelligenz'. Theologische Herausforderungen. 2023; Puzio, Anna: Digital and Technological Identities – In Whose Image? A philosophical-theological approach to identity construction in social media and technology. In: Cursor (2021). Online at: https://cursor.pubpub.org/pub/y2bcesx4 (as of: 14.03.22).

> an online presence and for those who turn to virtual spaces for religious discussions and inquiry. Religion is now offline, online, and hybrid.
>
> *Sana Patel, Islamic Theology, Ottawa*

Smith assumes that theological engagement with technologies first needs deep scientific exploration and a thorough survey of the discourse. After that, conceptions of humanity and new ethical challenges can play an important role. Schlote points to ethics and coping with life, and Patel makes a strong case that theology must come into direct contact with technologies.

4 How can churches and religious communities use AI and technology?

The possibilities for using AI and technology in churches and religious communities are varied, and will change over time as technology advances. Therefore, only some possible perspectives for the use of technology and AI will be outlined here. As there are specific ways of dealing with images, different religious teachings and regulations in religious buildings, the focus in this Section will be on Christianity. However, the ideas presented here can be applied to many other religions.

AI can play a major role in many processes that run in the background. For example, in event organization, community management and data management. AI can efficiently handle data such as member information, analyze, sort and make it more accessible. Additionally, AI can be used for predictions and process optimization, opening up new opportunities for data use within religious communities. Furthermore, AI and technology can also be used for targeted marketing, strategy development and audience engagement.

In recent years, AI has made significant progress in dealing with texts, images and music which play an important role in religions. Text recognition and generation, music composition, image recognition and AI-generated images can be used for religious practices and events, visiting religious buildings and religious education. The use of AI for translation or rephrasing into plain language can contribute to an inclusive church. As technology advances, new opportunities for presenting religious communities' messages and information material digitally are emerging. This can include utilizing projections, overlays, bots, and touch screens. Additionally, AI can be used for Bible study and knowledge management. AI-based text and image tools will also transform theological research, e.g. enquiry and text generation, biblical research and the handling of ancient texts and old scripts.

Virtual and Augmented Reality (VR/AR) are gaining popularity in the cultural sector and represent an exciting opportunity for churches and religious communities. In religious education, VR/AR can provide access to ancient worlds and historical journeys, and enable visits

to religious sites. Many religions attach great significance to certain places and countries, but many believers are unable to visit them. Additionally, it can make participation in religious ceremonies possible for those who cannot attend in person, especially the ill. Using VR/AR, worshippers can participate in religious ceremonies from home, care facilities, or hospitals, allowing them to touch religious objects with special equipment and experience haptic and olfactory impressions. This technology can also provide special access to religious events for people with disabilities, such as facilitating certain movements. In this way, VR/AR can help create an inclusive church experience. Furthermore, completely new uses of virtual and augmented reality are conceivable. They can be used in religious rituals or accompany spiritual experiences. This again shows that religious practices are also being transformed in the course of technologization. Likewise, information material and religious messages can be made accessible and interesting in new ways. Furthermore, immersive art, which immerses viewers in a multimedia experience with light, sound, VR glasses, video projections, and haptic and olfactory effects, is gaining popularity. As art plays an important role in many religions, offering immersive art in religious communities is a valuable opportunity.

Another technology that is likely to shape religions and religious communities is robotics. Religious robots are a form of technology that are specifically designed for use in religious contexts. They are most commonly found in non-monotheistic religions and particularly in Asia, but as technology continues to advance, it is likely that their use will become more widespread. These robots can perform a variety of functions, including accompanying prayers, conducting conversations, celebrating religious ceremonies, reading from religious scriptures, and playing music. They can also give tours of religious buildings, answer questions about the specific religion in question, and even transmit religious celebrations to those who are unable to be present in person. Even before religious robots are widely used for explicitly religious purposes, it is likely that religious aspects will be integrated into social robots. Social robots can be used, for example, in the health sector, in hospitals and care facilities. These robots can have conversations, give medication or injections, and make the long hospital stay easier for children and their parents. As they interact with patients and hold conversations, it is likely that religious and spiritual needs may arise, particularly in the context of illness and care. This raises an important question: should these robots be atheistic or agnostic, or should they also provide information about different religions, have religious values, and have religious and spiritual elements integrated into their programming?

Furthermore, the many functions that are known as "smart home" can also be applied in religious community buildings. Smart home refers to the networking of technology in the house such as lighting, heating, air conditioning, door locking, voice assistants, kitchen appliances, televisions, and other entertainment electronics. In religious buildings, other technologies and functions would be conceivable. Analogous to smart home, there would be smart church.

In the discourse of technology and in many religious communities, there is fear that technology will increasingly replace people and valuable interpersonal experiences will be lost. However, technologies do not have to replace or imitate people, but it is beneficial when technologies do what they do best. This includes: data processing and storage, certain lifting movements in care and the characteristic that people are less ashamed of technology than of people in some care activities, as well as impressive virtual experiences, visual and haptic effects that complement religious practices or make religious ceremonies more inclusive.

5 About the Volume

Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz (*Alexa, How Do You Feel About Religion? Theological Approaches to Technology and Artificial Intelligence*) brings together German- and English-language contributions on technology and AI from German-speaking and international perspectives to illustrate the diversity of theological research discourses. The aim is also to engage in interdisciplinary discussions of theology, especially with the engineering, technical and natural sciences. The contributions were written primarily by researchers in Protestant, Catholic and Jewish theology.

The volume is divided into five sections that highlight the transformations in various areas of human life and in theology in the context of technologization. As the inaugural volume of the series *Theologie und Künstliche Intelligenz (Theology and Artificial Intelligence)*, this collection of articles provides insight into the diversity of theologically relevant topics and offers an overview of current research. The volume cannot address all the relevant issues, especially because they will constantly change, but it does offer productive starting points for future theological investigation.

The first section, *Transformation of the Image of the Human Being: Human and Robot*, undertakes anthropological reflections. Technologization has clearly transformed images of the human. For *Katherine Smith,* anthropology is one of the most important research fields in the theology of AI discourse. Based on her experiences with human-like medical training manikins at the Barbara H. Hagan School of Nursing and Health Sciences at Molloy College in New York, she explores connections between anthropology and AI. In her contribution *Learn, Remember, Act: Theological Anthropology and AI Metaphor* she compares the differences between humans and AI in terms of learning, remembering and acting. In his contribution *Grundlinien eines Menschenbilds der Künstlichen Intelligenz (Basic Outlines of the Human Image in Artificial Intelligence)*, *Lukas Brand* addresses the image of the human that AI systems represent. Using the example of the humanoid robot "Optimus", which Tesla presented to the public in 2022, he analyses the technological reproduction of the human being. *Sven Nyholm* also focuses on

humanoid robotics and discusses the moral status of robots in his article *Wie sollen wir mit künstlich-intelligenten humanoiden Robotern umgehen? (How Should We deal with Artificially Intelligent Humanoid Robots?)*. Nyholm asks whether robots can possess, mimic or represent morally relevant properties or abilities. In his article, he offers an overview of the current international debate on robots as bearers of rights.

In the second section, *Transformation of Religion: Robots and Religion*, the discourse on robots is pursued in relation to religious contexts. In her contribution *Robot Theology: On the Theological Engagement with Robotics and Religious Robots*, *Anna Puzio* examines religious robots, i.e., robots used for religious purposes. She highlights the role of this technology in different religions, pointing to the importance of time-dependent, culturally negotiated concepts of human and non-human, life and creation. Her article contributes to the profiling of future theological engagement with robotics. *Hendrik Klinge* turns to this issue from a different perspective by examining the religiosity of robots themselves. In his article *Do Robots Believe in Electric Gods?*, he employs a theological Turing test and draws upon Wittgenstein to investigate whether robots can have a religious faith.

The third section is dedicated to the *Transformation of the Body* in the Fields of *Medicine and Optimisation*. *Max Tretter* begins by examining AI self-tracking, an already widespread aspect of everyday life that is supposed to counteract uncertainty and enable new freedom of choice. In *Ambivalenzen gegenwärtiger Gewissheitsbestrebungen (The Ambivalences of Current Certainty Efforts)*, Tretter utilizes Jean Baudrillard's simulation theory to explore how these technologically generated certainties actually affect human freedom of choice. In his article *On Digital Twins and Heavenly Doppelgangers, Yannick Schlote* contends with digital twins in medicine, which is predicted to become a major issue. An example of digital twins is the digital representation of patients to simulate medical applications. Schlote demonstrates the similarities between the digital twin and the gnostic belief in the coexistence of human beings and heavenly doppelgangers; based on this comparison, he provides an ethical evaluation of the digital twin phenomenon. In his article *Impulses for Questions of Corporeality in Ethics against the Background of Moral Enhancement Discourse (Impulse für Fragen der Leiblichkeit in der Ethik vor dem Hintergrund des Moral-Enhancement-Diskurses)*, *Dominik Winter* also looks at future technologies and, in particular, moral enhancement, i.e., the moral improvement of humans through technological influence, building on a transhumanist definition of the relationship between body and mind.

In a theology of technologization and AI, *autoregulatory weapon systems* is also a central issue, and it is the focus of the fourth section of the volume, *Transformations of War*. In her article *Automatisation Challenging Peace Ethics*, *Nicole Kunkel* argues that such weapons can kill people, but are based on problematic algorithms and, in her opinion, cannot make moral decisions. Following up on Kunkel's insights into the overall discourse, *Mois Navon* focuses

on human dignity from the perspective of Jewish theology in the article *Autonomous Weapons Systems and Battlefield Dignity*. Navon argues that dignity on the battlefield is an ethical category of its own, and it is defined quite differently in peacetime.

Finally, the section *Transformation of Theology: Theory and Critique* shows how, in the context of technologization, theologically relevant theories change, new theories emerge, and a theological critique can be practised. In the article *Jewish Philosophy and the Critique of AI Technology*, *Hava Tirosh-Samuelson* draws on Emmanuel Levinas, Hans Jonas and Jonathan Sacks to criticise transhumanism. Tirosh-Samuelson counters transhumanism with the values of freedom, responsibility and embodied dignity as critical responses derived from Judaism. Next, in the article *Digitale Transformation des Unsichtbaren (Digital Transformation of the Invisible), Lukas Ohly* makes some *Creation-theological remarks on the limits of digital fabrication following Hannah Arendt (Schöpfungstheologische Anmerkungen zu den Grenzen des digitalen Herstellens im Anschluss an Hannah Arendt)*. Ohly also discusses digital communion. Simon Reiners then engages with the metaverse and humanism in the article *Metaversum und resistente Körperlichkeit (Metaversum and Resistant Corporeality)*. He places the body at the centre of his reflections and discusses material-feminist theories, Donna Haraway and Theodor Adorno.

This volume is a follow-up to the conference *Alexa, wie hast du's mit der Religion? Technology, Digitalisation and Artificial Intelligence in the Focus of Theology*, organised by the Netzwerk für Theologie und Künstliche Intelligenz (neth:KI) (Network for Theology and Artificial Intelligence) in 2021 (conference team: Lukas Brand, Nicole Kunkel, Julia van der Linde and Anna Puzio). This international and interreligious network aims to promote theological engagement with technology and AI. The conference papers were supplemented by further contributions, including from members of the neth:KI network, and all articles have undergone a double peer review process. We would like to thank Lukas Brand for his contribution to the first conception of the volume, and also Saskia Fischer for her assistance with the preparation of the manuscript. The publication was supported by the Open Access Publication Fund of the Humboldt-Universität zu Berlin.

I Transformation des Menschenbildes

Mensch und Roboter

Learn, Remember, Act

Theological Anthropology and AI Metaphor

Katherine Schmidt

Abstract

Anxieties over the capabilities of artificial intelligence reflect our assumptions about its capabilities. From a theological perspective, these assumptions are often inflated if not wholly mistaken, demonstrating an Edenic impulse to create in our own image without realizing the limitations of our creations. This paper analyzes the language used for AI, arguing that the uncritical and unnuanced application of the metaphors of "learning, memory, and action" exacerbates problematic assumptions about AI vis a vis its human-like capabilities.

1 Introduction

About 200 meters from my office, nursing students interact daily with machines that teach them how to draw blood, do basic medical exams, and even assist in childbirth. A lowly theologian, my first thoughts when I walked into my university's nursing school and saw the "patients" in their lab-beds was how human they looked and how ripe this would be for philosophical and theological reflection. My colleagues in the Barbara H. Hagan School of Nursing and Health Sciences insist that these "mannequins" (as they are called in the School) are not "artificial intelligence." While I understand their insistence, I would place these machines on a broad spectrum of artificial intelligence that recognizes the colloquial way in which we use the term. In addition, such technology seems to fall in between the traditional distinction of Artificial General Intelligence (AGI) and Artificial Narrow Intelligence (ANI). AGI aims for more "human-like intelligence including perception, agency, consciousness, intentions and maybe even emotions," whereas ANI is meant to "accomplish very specific tasks in very narrowly de-

fined contexts."[1] The nursing school mannequins can respond in medical ways, with increased heart rates, and even blue lips. Therefore, though the mannequins do not seem to fit the criteria of AGI in terms of perception and agency per se, they exceed ANI in their anthropomorphic technology to enough of an extent be properly considered AI.

The technology of the nursing school mannequin is tailored to training new medical professionals, and is therefore not a common experience for people outside of this context. Technology of this kind, however, reflects an ongoing impulse to develop anthropomorphic technology, be it in shape (e.g. human-like faces and bodies), sound (e.g. voice), or ability (e.g. art creation, or writing).

What follows is a reflection on androids and android-like machines that push us to think about technology and humanity, written from the perspective of a Roman Catholic theologian. The question of artificial intelligence, especially from a theological perspective, is predominantly a question of anthropology. What we say about AI and how we behave toward it reflects our assumptions about humanity. In particular, *anxiety* about artificial intelligence tells us quite a bit about our anthropology, as we often betray our assumptions about what makes us human as we insist on distinguishing ourselves from our technology. It is a curious feature of our species that we spend so much of our reflective capability on distancing ourselves from our own creations. We are fractured in our commitments to technological progress, both attracted to its magic and wary of its power. Lest we become blinded by our own confusions and anxieties over what we create, we should constantly seek clearer definitions and descriptions of technology, especially that which tends to mimic us in various ways.

In the volume *AI for Everyone?* Rainer Rehak argues that due to the limited number of people who have technical knowledge of even everyday technology, we rely heavily on anthropomorphisms to describe artificial intelligence of various kinds. "Hence," writes Rehak, "problems arise when these scientific terms are transferred carelessly into other domains or back into everyday language used in political or public debates."[2] The same would apply for theological and philosophical debates over technology. Thankfully, these critical disciplines have the tools and energy to devote to careful definition and description beyond these "careless transfers." I will take just three of these transferred terms for theological reflection: learning, memory, and action.[3] These three are prevalent in AI discourse and can also function well as examples of the limitations of linguistic metaphor.

---

1 Rehak, Rainer: "The Language Labyrinth: Constructive Critique on the Terminology Used in AI Discourse." In: Verdegem, Peter (ed.): AI for Everyone? Critical Perspectives. Westminster 2021, 87–102, here 90.

2 Ibid.

3 All three are examples in Rehak's argument, alongside autonomy, recognition, trained, communication, understanding, "and, of course, intelligence."

First, a note on the theological implications of using these terms in the first place. These "careless transfers" are problematic for Rehak because of their political consequences, but from a theological perspective, they are reflections of the persistence of the reality of sin. It is not necessarily the case that the mere existence of human-like technologies or features in technologies demonstrates our sinful propensities. However, the easy application of uniquely human attributes onto our technology reflects a kind of Edenic impulse to turn ourselves into not just creators but the Creator by convincing ourselves we have made these things in our own image. It is not only a poor description when we say our technology "learns, remembers, and acts," but it is also a farce, a kind of delusion that we could create anything that approximates the beautiful complexity of humanity. We may create things in our image but our limitations as creatures means that we can never do so in a perfectly analogous way to our own creation in the image of God.

It is, of course, only natural that all technology – from the most basic to the most cutting edge – reflect our image to some degree. According to Bruno Latour, "We have been able to delegate to nonhumans not only force as we have known it for centuries but also values, duties, and ethics."[4] Our tools reflect a story of desires for intended effects, and desires are both anthropologically original as well as culturally negotiated. When it comes to the various tools we adopt as a community, therefore, the material products reflect negotiations of their designers and, to varying degrees, the non-designer members of the community. As Latour argues, "The nonhumans take over the selective attitudes of those who engineered them."[5] Artificial intelligence, in its many iterations along the spectrum of general and narrow, provides new clarity on Latour's analysis of technology and society. We see our assumptions about humanity within the technology itself: how efficient we want to be, how we think humans should sound and look, etc. More than this, we can see assumptions in our discussions and debates over what we have created by the terms we use. We insist that we have created machines that can "learn, remember, and act." We should ask ourselves exactly what these machines are doing, not only to be more precise about their relationship to our own humanity but also to investigate the assumptions that lie behind our own definitions of these terms.

2 Learning: Machine vs. Human

One of the most common phrases in discussions of artificial technology and of digital technology more broadly is "machine learning." For the most part, this refers to a kind of aggregation

[4] Latour, Bruno: Where Are the Missing Masses? In: Johnson, Deborah J./Wetmore, Jameson M. (eds.): Technology and Society, Building Our Sociotechnical Future. Cambridge, Mass 2008, 151–180, here 157.

[5] Latour, Bruno: Where Are the Missing Masses?, 158.

of large sets of data, input through various methods, and recapitulated into specific tasks. More specifically, "machine learning" was coined by Arthur Samuel in 1959 "to refer to the subset of AI techniques where the problem-solving algorithm was not directly programmed by the analyst but was found from the analysis of data."[6] Much anxiety swirls around the collection of personal data that gets used for targeted marketing, an example of algorithmic leverage for profit.[7] This leverage on the part of technology feels to us like a kind of learning: an intake of data that then gets applied in a particular way without being explicitly directed to do so.

To what extent is this technology learning in the way we experience learning? What are the differences between learning as we experience it and what is happening in the technological sense of "machine learning"? Ascribing the term of "learning" to this algorithmic operation is, as mentioned above, a kind of transference of terms that is meant to impress upon the non-technical person the "gist" of the process. It functions as a sort of metaphor at this point, describing a new ability within technology to apply the data that it takes in without being "told" (another transfer!) how exactly to apply it. Without the persistent caveat that this term functions as a metaphor, the non-technical public begins to assume that machines experience learning in the same way we do. More importantly, by adopting the term unquestioningly, both technicians and lay users betray their own assumptions about "learning."

In 1970, Paulo Freire published *Pedagogy of the Oppressed*, a groundbreaking book about education in the context of human suffering wrought by various systems of oppression. Freire famously argues for a model of education that contributes to the liberation of learners. Contrary to his model, he describes the "banking concept of education," which dominates much of the educational landscape. In this concept, "Education thus becomes an act of depositing, in which students are the depositories and the teacher is the depositors. Instead of communicating, the teacher issues communiques and makes the deposits which the students patiently receive, memorize, and repeat."[8] According to Freire (animated throughout by his Catholic faith), this model of education does not allow learners to be fully human because it does not allow for the faculties of inquiry and curiosity in the context of community that make us what we are. The banking model of education falls well short of what learning can and should be: "Knowledge emerges only through invention and re-invention, through the restless, impatient, continuing, hopeful inquiry human beings pursue in the world, with the world, and with each other."[9] Learning in this sense is the complex process of *human* learning.

---

6 Walsh, Kenneth R./Mahesh, Sathiadev/Trumbach, Cherie C.: Autonomy in AI Systems. Rationalizing the Fears. In: The Journal of Technology Studies 47/1 (2021), 38–47.

7 See Zuboff, Shoshana: The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power, New York 2019.

8 Freire, Paulo: Pedagogy of the Oppressed. Bloomsbury, New York 2000, 72.

9 Ibid.

There is a further, more theological difference between human learning and the processes we describe as "learning" in machines. Freire gestures toward an innate human desire for knowledge at the origin of all learning, and the creativity that is both required for and a product of learning. Theologically speaking, both this desire and the creativity needed to meet it reflect the *imago Dei*, the image of God bestowed upon all human creatures and referenced in Genesis 1. Though influenced perhaps too much by Platonic and neo-Platonic cosmology, Western theology throughout the centuries has posited a desire for return-to-Origin within the human soul. As regards learning, this desire moves us to be attracted to truth. As Bonaventure argues, "Our intellect understands in the true Light that enlightens every man who comes into the world, the true Light of the Word who was in the beginning with God."[10] In other words, we only learn and understand to the extent that we recognize truth, which has its origin in God. Eight centuries later, Simone Weil reflects on learning similarly, though through the category of attention: "Every time that a human being succeeds in making an effort of attention with the sole idea of increasing his grasp of truth, he acquires a greater aptitude for grasping it, even if his effort produces no visible fruit … If there is a real desire, if the thing desired is really light, the desire for light produces it. There is a real desire when there is an effort to attention."[11] For Weil, education is a process of directing and redirecting one's attention to the light, which again has its source in God.

Given this picture of learning, it becomes difficult to hold on to the metaphor of "machine learning" except as a kind of technical instantiation of the banking concept, devoid of desire, creativity, and the return to the source of truth involved in human learning. More than relieving some of our anxiety (hopefully) about "machine learning," holding the term to a more robust theological definition enables us to see the ways in which we have allowed the technocratic priorities of data collection and algorithmic leverage to reinforce the banking concept of education in our own minds because of our use of this metaphor. By ascribing the metaphor of "learning" to our machines it seems likely that we have made Freire's dynamic, fully-human and liberative model of learning even harder to foster.

3 Memory: Data vs. Remembering

The data and algorithms of "machine learning" require memory; these processes only work because of a storehouse of data. In addition, even non-technical users of digital space understand

[10] Bonaventure: The Mind's Journey to God. In Swindal, James C./Gensler, Harry J. (eds.): The Sheed and Ward Anthology of Catholic Philosophy Lanham 2005, 156–164, here 159.

[11] Weil, Simone: Waiting for God. New York 2009, 59.

the vast capabilities of digital record: we know that the internet "lasts forever," or at least as far into the future as we can imagine it will. It appears at first blush that memory is a more appropriate metaphor than learning, given that data storage is a central feature of AI. But again, the human experience of memory differs in important ways, especially from a theological perspective. Awareness of this difference is crucial for avoiding an over-anthropomorphizing in our imagination of AI that can lead to unwarranted anxiety or optimism.

Bonaventure continues his reflection on the human mind from above, writing that "you will be able to see God in yourself as a likeness."[12] This is because, according to Bonaventure, the mind recognizes truth because its memory holds the standard of truth received by God at creation. More convincingly, Augustine ruminates on memory in a later chapter of *Confessions*, writing that our experience of memory is often frustrating or baffling, a kind of wandering in a vast storehouse over which we have some but only little control. For thinkers like Augustine and Bonaventure, our humanity comes not only from the collection of memories that make up our personality but our ongoing relationship to them as a rational soul. Stored data – an important component of AI technology – is inert and highly stable. It does not appear unless directly called upon in some way. It is merely an aggregate that amounts only to the data itself. Human memory, by contrast, is in constant relationship to the human subject, constructing her identity, aiding in the interaction with external reality, and functioning both inside and outside of the willful control of the subject:

> Some memories pour out to crowd the mind and, when one is searching and asking for something quite different, leap forward into the centre as if saying, "Surely we are what you want?" With the hand of my heart I chase them away from the face of my memory until what I want is freed of mist and emerges from its hiding places. Other memories come before me on demand with ease and without any confusion in their order. Memories of earlier events give way to those which followed, and as they pass are stored away available for retrieval when I want them.[13]

What Augustine's reflection points to is the nature of human memory regarding subjectivity and identity. A person negotiates with her own memory depending on and resultant in her sense of self, something AI simply does not do in its storage and retrieval of data. We do indeed construct identity out of memory, but it is selective, ad hoc, re-narrated and even fabricated. Therefore, there are a set of faculties regarding the memory that are not present in AI (at least as of yet) and are frankly not necessary.

---

[12] Bonaventure: Mind's Journey, 159.

[13] Augustine: Confessions. Chadwick, Henry (trans.), Oxford 1998, 183.

For Augustine, the memory is "an awe-inspiring mystery," and even "immeasurable."[14] On this point, the Bishop of Hippo is heavily influenced by the anthropological assumption of the imago Dei. He sees in the human mind the spark of the divine, assuming that its complexity – a truth to be sure – is a mark of the image of God: "Surely my memory is where you dwell, because I remember you since I first learnt of you, and I find you there when I think about you." Augustine errs here on the side of overestimating the depths of the human mind regarding memory. He seems unconcerned with the finitude of the human mind, especially as know it today from psychology and biology. What he gestures toward is important nonetheless: the human person bears the image of God and therefore reflects the mystery of the divine life. A more nuanced, psychologically-focused reflection on the faculties of the human person from Maurice Blondel can help us think about the metaphor at the heart of both memory and learning: action.

4 Action: Agency and Will

One may object to the theological reflections above, arguing that despite the differences in human learning and memory from that of AI, the technology in question still acts, that is, demonstrates agency to various degrees. Be it through prescribed programming or extrapolated, algorithmic action, AI seems increasingly capable of its own actions, bearing more than metaphorical resemblance to the banal ongoings of human life. Let us finally, then, examine the notion of action itself in order to define precisely the limits of the metaphor, if any. Nineteenth century philosopher Maurice Blondel submitted his dissertation Action in 1893, drawing sharp criticism from some of his readers because of his theological implications within. For Blondel, human action is the "link between thought and being," and his investigation thereof delineates the complexity at the heart of the human experience that, by his logic, could never be replicated through anything we create, e.g., AI. Blondel describes a tension at the heart of the human experience between our capacity to will and our ability to act. Summarized here by Doherty, Blondel argues that "From the interior life of subjectivity … the human will meets no phenomenal object that will resolve the dialectical principle: that is, that satisfies the infinitude of the willing will. On the contrary, the human being always wills more than is available in any form of willed action considered in the dialectic."[15] We have with in this, in other words, an infinitude of will, originating for Blondel with our Creator. Like Weil and Bonaventure, Blondel

[14] Ibid.

[15] Doherty, Cathal: The Symbiosis of Philosophy and Theology in Blondel's Supernatural Hypothesis. In: Theological Studies 79/2 (2018), 274–293, here 282.

sees in the human soul – primarily in our "willing will" – a trajectory toward our supernatural end that animates our action.

Where Augustine finds God in the memory, therefore, Blondel finds God in the depths of the "willing will," unsatisfied by any finite human action. Whenever we act, says Blondel, we experience "a kind of interior dismemberment" because "it is as though, in order to affirm and to act, it were necessary to ignore those needs of the head and of the heart, the usual source of the faults and the follies that make the life of the ardent and the passionate enigmatic to the calculating and the ambitious."[16] It is as though each human has a wellspring of will out of which they act, which is by necessity a finite and concrete choice and therefore exclusionary of all other willing. Resolving this dialectic between the "willing will" (infinitude) and the "willed will" (finitude, action) is only possible in the return to God.

Following Blondel's dialectic, any human creation is a reflection of the willed will, and only of the willing will *in absentia*. The willing will remains within the human personality, frustrated, "dismembered" by the exclusion of the willed will. Though we enable AI to replicate action, it does not reflect this dialectic. There is no wellspring of the "willing will" within AI. In fact to the degree that AI "wills" anything, it does so actually as the action (willed will) of its creator(s). The extent we can say that AI reflects any kind of "will," it does so exclusively as a "willed will," a finite action. Absent a subjectivity, the AI does not experience the insatiate willing will, and therefore does not "act" in the same way as its human creators. AI simply renders concrete the willed wills of its creators; it has no wellspring of will from which arise action in the precise way Blondel means it in the human person.

The controversy over Blondel's thesis surrounded the theological implications of his argument. By describing the infinite/finite dialectic at the heart of human action, Blondel aimed to create imaginative space for the *imago Dei* in the human person. I convey his thesis here in the context of AI to explore once again what distinguishes humans from the rest of creation, not as a source of dominance but as a reminder of the spark of infinity in the human soul that animates us to create technology such as AI in the first place.

5 Conclusion

Descriptions of AI tell us important things about our own assumptions regarding technology and anthropology. It is simple enough to remember that much of what we say regarding AI is a metaphor, given that technical knowledge is required to understand precisely the mechanisms

[16] Blondel, Maurice: Action (1893): Essay on a Critique of Life and a Science of Practice. Blanchette, Olivia (trans.) Notre Dame 1984, 160.

of its abilities. We are bound by metaphor in most if not all political and philosophical discussions of AI, which limits us in certain ways but also extends the conversation to disciplines like theology and philosophy who are integral to exploring the role of technology in the human experience. I attempted here to offer more precision regarding these metaphors, noting especially the differences in the human analogues of learning, memory and action. This is not to dismiss the power of AI or to assert a preference for the human experience over the machine. To the contrary, by examining the precise differences between ourselves and our creations, we are better able to manage our expectations and fears of what we create. The practice of distinction in this regard should be an ongoing process that opens space for us to negotiate common hopes for our technologies and assumptions about our shared humanity. Because of its persistent reflection on the human condition, theology is in a unique position to offer insight on the present reality and future of AI.

References

Augustine: Confessions. Chadwick, Henry (trans.). Oxford 1998.

Blondel, Maurice: Action (1893): Essay on a Critique of Life and a Science of Practice. Blanchette, Olivia (trans.). Notre Dame 1984.

Bonaventure: The Mind's Journey to God. In Swindal, James C./Gensler, Harry J. (eds.): The Sheed and Ward Anthology of Catholic Philosophy. Lanham 2005.

Doherty, Cathal: The Symbiosis of Philosophy and Theology in Blondel's Supernatural Hypothesis. In: Theological Studies 79/2 (2018), 274–293.

Freire, Paulo: Pedagogy of the Oppressed. Bloomsbury. New York 2000.

Latour, Bruno: Where Are the Missing Masses? In: Johnson, Deborah J./ Wetmore, Jameson M. (eds.): Technology and Society, Building Our Sociotechnical Future. Cambridge, Mass 2008, 151–180.

Rehak, Rainer: "The Language Labyrinth: Constructive Critique on the Terminology Used in AI Discourse." In: Verdegem, Peter (ed.): AI for Everyone? Critical Perspectives. Westminster 2021, 87–102.

Walsh, Kenneth R./Mahesh, Sathiadev/Trumbach, Cherie C.: Autonomy in AI Systems. Rationalizing the Fears. In: The Journal of Technology Studies 47/1 (2021), 38–47.

Weil, Simone: Waiting for God. Harper. New York 2009.

Grundlinien eines Menschenbilds der Künstlichen Intelligenz

Wie gut ist Teslas Optimus?

Lukas Brand

Abstract

The article is dedicated to the thesis of the digitalisability and modularisability of human properties, which agents in artificial intelligence and robotics presuppose in order to technically reproduce humans as androids. Tesla's Optimus is used as an example to illustrate the practical implementation of this thesis. The ambivalent potential of the basic lines of the human image of artificial intelligence elaborated here are of growing importance from the perspective of the Catholic Magisterium and require further theological reflection.

1 Einleitung

Während die künstliche Intelligenz (KI) weltweit ihre ersten Gehversuche macht, kommt in den 1960er Jahren im Vatikan die katholische Weltkirche zusammen, um sich unter anderem auch auf die Veränderungen zu besinnen, die mit den neuen Möglichkeiten von Naturwissenschaft und Technik auch für den Menschen und sein Selbstverständnis einhergehen.[1]

[1] „Heute steht die Menschheit in einer neuen Epoche ihrer Geschichte, in der tiefgehende und rasche Veränderungen Schritt um Schritt auf die ganze Welt übergreifen. Vom Menschen, seiner Vernunft und schöpferischen Gestaltungskraft gehen sie aus; sie wirken auf ihn wieder zurück, auf seine persönlichen und kollektiven Urteile und Wünsche, auf seine Art und Weise, die Dinge und die Menschen zu sehen und mit ihnen umzugehen. […] I]m Bildungsbereich [erlangen] die mathematischen, naturwissenschaftlichen und anthropologischen Disziplinen, im praktischen Bereich [erlangt] die auf diesen Disziplinen aufbauende Technik ein wachsendes Gewicht" (*Gaudium et Spes* [GS], Nr. 4–5; die Konstitution des II. Vatikanischen Konzils Gaudium et Spes wird zitiert nach Rahner, Karl/Vorgrimmler, Herbert: Kleines Konzilskompendium. Freiburg i. Br. [35]2008).

Das Konzil macht das Menschenbild als den Dreh- und Angelpunkt der sich anbahnenden Transformation der modernen Gesellschaft aus.[2] Bis in die Gegenwart hinein hat der Vatikan nicht aufgehört, auf das ambivalente Potential der Technik hinzuweisen.[3] Der vorliegende Beitrag geht dem Menschenbild der Künstlichen Intelligenz nach, dem aufgrund seiner zentralen Rolle für die moderne Technik eine besondere Bedeutung zukommt. „Künstliche Intelligenz" bezeichnet hier das Fachgebiet, das sich der Automatisierung intelligenten Verhaltens widmet; „KI" hingegen soll ein entsprechendes technisches Sachsystem bezeichnen.[4] Es geht in diesem Beitrag also nicht darum, welches Menschenbild KI-Systeme haben könnten, sondern welches Menschenbild sie repräsentieren. Der Beitrag widmet sich insbesondere den ontologischen Verpflichtungen, die eingegangen werden müssen, um den Menschen möglichst umfassend ingenieurswissenschaftlich rekonstruieren und als Androiden technisch reproduzieren zu können. Dass diese Verpflichtungen tatsächlich eingegangen werden, wird am Beispiel des Androiden Optimus des Automobilherstellers Tesla veranschaulicht.

2 Rekonstruierbarkeit von Menschen in Technik

Technik wird hier verstanden als die Summe der Substitute oder Komplemente menschlicher Funktionen oder Verhaltensweisen in Form vorrangig nutzenorientierter Artefakte.[5] Auch Digitaltechnik und KI sind in diesem Sinne nutzenorientierte Reproduktionen von Funktionen oder Verhaltensweisen, die beim Menschen Intelligenz erfordern. State of the Art sind dabei die Verfahren des Maschinellen Lernens und der „autonomen" Robotik.[6] *Technologie*, wird entsprechend als die Wissenschaft von der Technik verstanden.[7] Diese wird den Menschen zu-

---

[2] „Dennoch wächst angesichts der heutigen Weltentwicklung die Zahl derer, die die Grundfragen stellen oder mit neuer Schärfe spüren: Was ist der Mensch?" (GS 10)

[3] Zuletzt prominent etwa durch Franziskus: Enzyklika *Laudato Si* [LS]. Rom 2015, Nr. 102–104.

[4] Vgl. auch die begrifflichen Festlegungen in Abschnitt 2.

[5] Vgl. Ropohl, Günther: Technologische Aufklärung. Frankfurt a. M. 1999, 17–22; vgl. den Art. „Technik" in: Ritter, Joachim/Gründer, Karlfried (Hg.): Historisches Wörterbuch der Philosophie (Bd. 10). Basel 1998, 940–952, hier 940: „Bezeichnung für das Ganze des maschinell/instrumentell Verfügbaren"; für Beispiele vgl. Abschnitt 3.

[6] Vgl. Ertel, Wolfgang: Grundkurs Künstliche Intelligenz. Wiesbaden 2021, 1–4.

[7] Der Begriff der Technologie ist erst im beginnenden 18. Jahrhundert auf diese Bedeutung festgelegt worden, wie sich auch der Begriff der Technik erst in dieser Zeit zunehmend auf den Bereich der nutzenorientierten Artefakte verengt hat (vgl. die Hinweise in Anm. 5). Unter dem Einfluss des angelsächsischen Sprachgebrauchs hat sich der Bedeutungsgehalt von Technologie im 20. Jh. allerdings geweitet und ist heute kaum noch von dem der Technik zu unterscheiden. Vgl. dazu ebenfalls Ropohl: Aufklärung, 22–24 sowie Meier-Oeser, Stephan: Technologie. In: Ritter,

mindest implizit in praktisch-funktionaler Hinsicht analysieren, um Sachsysteme entwickeln zu können, die menschliche Leistungen substituieren oder komplementieren sollen. Die Teildisziplin der Technologie, die sich in diesem Sinne auf die KI als Sache bezieht, wird ebenfalls als Künstliche Intelligenz bezeichnet.

Die folgenden Bedingungen der digitaltechnischen Reproduktion des Menschen werden hier vorausgesetzt:[8] Objekte und ihre Eigenschaften und Funktionen können trivialerweise nicht unmittelbar technisch reproduziert werden. Sie können nur dann Gegenstand eines technischen Reproduktionsprozesses sein, wenn sie in Form einer auf Produktion ausgelegten Rekonstruktion modellierbar sind. Das so konstruierte Modell bestimmt den Reproduktionsprozess. Die Rekonstruktion des Menschen als Modell ist eine Hypothese über die als relevant erachteten, technisch reproduzierbaren Eigenschaften und Funktionen des zu reproduzierenden Objektes. Die Entscheidung darüber, was als relevant erachtet wird, ist eine subjektive Wahl des:r Produzent:in aus zu referenzierenden und nur bedingt erreich- und nachvollziehbaren Eigenschaften und Funktionen im gegebenen Objekt. Mindestens die materielle Dauer sowie die geschichtliche Zeugenschaft des Menschen sind einer Reproduktion jedoch unverfügbar.

Die Entwicklung von *humanoiden Robotern und Androiden* entfaltet sich innerhalb dieser Bedingungen. Ein Roboter wird „humanoid" genannt, wenn seine Form dem menschlichen Vorbild wenigstens in Teilen ähnelt. Ein Android ist eine Unterklasse der humanoiden Roboter, die einen möglichst hohen Grad äußerlicher Übereinstimmung zum Menschen anstrebt. Die Übergänge sind fließend.[9] Die Ausdrücke werden im Folgenden weitgehend synonym für Roboter verwendet, die eine annähernd menschliche Form haben. Je ähnlicher die Reproduktion dem menschlichen Vorbild sein soll, desto größer muss die Schnittmenge ihrer Funktionen, äußerlichen Erscheinung und Verhaltensweisen sein. Sie wird die Form eines Androiden annehmen und zu einem vollständigeren, wenn auch technischen Abbild des Menschen werden.[10] Die Hürde, die Designer:innen hinsichtlich der äußeren Übereinstimmung nehmen

Joachim/Gründer, Karlfried (Hg.): Historisches Wörterbuch der Philosophie (Bd. 10). Basel 1998, 958–961, hier 959 f.

8 Diese Bedingungen habe ich in Brand, Lukas: Virtuelle Menschenreproduktion. In: Pirker, Viera/Pisonic, Klara (Hg): Virtuelle Realität und Transzendenz. Theologische und Pädagogische Erkundungen, Freiburg i. Br. 2022, 96–115 ausführlich dargelegt.

9 Vgl. Winfield, Alan: Robotics. Oxford 2012, 61 f.

10 Das Imitationsspiel, auf dem die philosophische Erörterung der KI maßgeblich fußt, lässt nur Rechenmaschinen als Kandidaten zu (vgl. Turing, Alan M.: Kann eine Maschine denken? In: Zimmerli, Walter Ch./Wolf, Stefan (Hg.): Künstliche Intelligenz. Stuttgart 1994, 39–78, hier 42 f). Die Reproduktion des Menschen im Medium der Technik basiert dementsprechend nicht auf organischen Stoffen. Produkte der Biotechnologie und organische Roboter sind damit ausgeschlossen (vgl. zum Xenobot-Projekt Kriegman, Sam/Blackiston, Douglas/Levin, Michael et al.: A scalable pipeline

müssen, ist das *Uncanny Valley*: das Tal in der Akzeptanzkurve, das dort entsteht, wo Maschinen einem Menschen zwar äußerlich in hohem Maße ähnlich sehen, aber immer noch merklich leblos sind.[11] Das Design von Androiden muss das Verhältnis von Form und Funktion mit Blick auf einen gegebenen Zweck bestimmen.[12] Die Hürde, die Ingenieur:innen hinsichtlich der funktionalen Ähnlichkeit nehmen müssen, ist die technisch reproduzierbare Rekonstruktion der natürlichen menschlichen Funktionen.[13]

Unter einer *Funktion* soll – sofern nicht anders angegeben – ein Prozess verstanden werden, der zur Anpassung eines Systems an die Umwelt geeignet ist und üblicherweise bei lebenden Organismen abläuft (Fortbewegung, Sprache sprechen, Schach spielen).[14] *Funktionsweisen* sind hingegen die Mechanismen, die in der fraglichen Entität angelegt sind und diese üblicherweise in die Lage versetzen, eine entsprechende Funktion auszuführen. Darüber hinaus besitzt der Mensch notwendige *Eigenschaften*, die ihn von anderen Lebewesen unterscheiden. Ingenieur:innen und Informatiker:innen setzen prima facie eine Hypothese über die Berechenbarkeit dieser Funktionen und Eigenschaften im Sinne ihrer mathematisch-informationstechnischen Modellierbarkeit voraus, um diese für praktische Zusammenhänge reproduzieren zu können: Sie werden als mathematische Funktionen von Zahlen dargestellt, die analoge Größen und Prozesse repräsentieren.[15] Aus pragmatischen Gründen wird in der Robotik ange-

for designing reconfigurable organisms. In: Proceedings of the National Academy of Sciences 117/4 (2020), 1853–1859, doi: 10.1073/pnas.1910837117, aber auch Misselhorn, Catrin: Künstliche Intelligenz und Empathie. Ditzingen 2021, 105–107).

11 Vgl. Mori, Masahiro: The Uncanny Valley. In: IEEE. Robotics and Automation Magazine 19/2 (2012), 98–100.

12 Vgl. Schweppenhäuser, Gerhard: Designtheorie. Wiesbaden 2016, VIII: „Designer beanspruchen heute nach wie vor, dass sie nicht bloß für das Schöne und Attraktive zuständig sind. Sie melden ihre Kompetenz für das Nützliche und Effiziente an und gleichermaßen auch für das Richtige und Gute."

13 Zum Begriff der natürlichen Funktion vgl. Detel, Wolfang: Metaphysik und Naturphilosophie. Stuttgart 2014, 117–121.

14 Vgl. zu Maschinen in *sozialen* Funktionen Brand, Lukas: Darf ich Sophia abschalten? In: Fateh-Moghadam, Bijan/Zech, Herbert (Hg.): Transformative Technologien. Baden-Baden 2021, 209–242.

15 Vgl. Wiegerling, Klaus/Nerukar, Michael/Wadephul, Christian: Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien, in: Kolany-Raiser, Barbara/Heil, Reinhard/Orwart, Carsten/Hoeren, Thomas (Hg.): Big Data und Gesellschaft. Eine multidisziplinäre Annäherung (Technikzukünfte, Wissenschaft und Gesellschaft). Wiesbaden 2018, 1–68, hier: 14: „[D]ie Datafizierung im Sinne der Skalier- und Kalkulierbarkeit aller Lebens- und Naturartikulationen [scheint] ein zentraler Anspruch des Big-Data-Zeitalters" zu sein; vgl. Barbour, Ian G.: Ethics in an Age of Technology. New York 1993, 171: „Many AI researchers defend *the formalist thesis* that all intelligence (natural or artificial) consists in the manipulation of abstract symbols. [… A]world of discrete facts can be represented by a corresponding set of well-defined symbols. [… T]he relationships among symbols are abstract, formal, and rule governed; symbols can therefore be processed by different physical systems (natural or artificial, protein based or silicon based)

nommen, dass menschliche Eigenschaften und Funktionen nicht nur begrifflich, sondern auch hinsichtlich ihres Funktionsbereiches voneinander isolierbar sind. Der ganze Mensch kann unter dieser Annahme als ein modulares System disjunkter Eigenschaften und Funktionen aufgefasst werden.[16]

Der so als Träger einer Menge von Eigenschaften und Funktionen verstandene Mensch wird in einem Modell rekonstruiert, das dazu geeignet ist, als Android produziert zu werden. Diese Reproduktion kann selbst wieder aufgefasst werden als Träger einer Menge von Eigenschaften und Funktionen, die von einem technischen Sachsystem sinnvoll ausgesagt werden können. Die Mächtigkeit ihrer Schnittmenge relativ zur Gesamtmenge der Eigenschaften und Funktionen des Vorbilds ist der Grad der Übereinstimmung des technischen Abbilds mit dem menschlichen Vorbild. Bei der Isolierung und Digitalisierung der Eigenschaften und Funktionen des Menschen wird aus Gründen der technischen Realisierbarkeit und Zweckmäßigkeit möglicherweise von einigen Funktionsweisen und Eigenschaften abstrahiert. Dem Menschen wird so zwangsläufig ein idealisiertes Modell gegenübergestellt. Das Menschenbild der Künstlichen Intelligenz ist auf diese abstrakte Rekonstruierbarkeit verpflichtet, die der Reproduktion zugrunde liegt. Sie ist die Grundlage der Entwicklung moderner Technik einschließlich der KI.

3 Digitalisierbarkeit von Funktionen

Als Alleinstellungsmerkmal des Menschen erscheinen seit jeher seine vernunftbasierten Funktionen. Diese kommen im intelligenten Verhalten von Menschen zum Ausdruck. Der besondere Fokus der Informatik bei der Rekonstruktion der Vernunft liegt daher auf der digitalen Modellierung des intelligenten menschlichen Verhaltens.[17] Das Modell dieses Verhaltens soll als kybernetisches Programm im Androiden die Funktion der menschlichen Vernunft übernehmen.

Zur Entwicklung eines solchen Modells scheinen drei Wege denkbar: Erstens der verlustfreie Transfer (Upload) eines menschlichen Bewusstseins einschließlich seiner Intelligenz auf ein technisches Surrogat. Diesem Ansatz liegt das philosophische Konzept einer starken

with identical results. […] T]he brain and the computer are two examples of devices that generate intelligent behavior by manipulating symbols."; vgl. außerdem unten Abschnitt 3.

16 Vgl. dazu unten Abschnitt 4.

17 Vgl. Ertel: Grundkurs, 1: „Die zentrale Frage für den Ingenieur, speziell für den Informatiker, ist jedoch die Frage nach der intelligenten Maschine, die sich verhält wie ein Mensch, die intelligentes Verhalten zeigt"; vgl. Kurzweil, Ray: The Singularity is near. New York 2006, 265: „[The AI revolution] is characterized by the mastery of the most important and most powerful attribute of human civilization, indeed of the entire sweep of evolution on our planet: intelligence."

oder allgemeinen KI zugrunde, dem zufolge der menschliche Geist vollständig auf technische Systeme übertragbar sei; auch hinreichende technisch Systeme könnten dann einen Geist besitzen.[18] Dieser erste Fall steht damit vor dem Kernproblem der technischen Reproduktion: der vollständigen und angemessenen Rekonstruktion aller notwendigen und nur zusammen hinreichenden Teile des zu reproduzierenden Vorbilds. Für das sogenannte Mind-Uploading ist die identische Funktionsweise und damit die Austauschbarkeit von Computer und Gehirn entscheidend: Es bedarf nicht nur einer funktionalen Ähnlichkeit zwischen dem Computer im Androiden und dem Sitz des Geistes im menschlichen Körper, sondern einer identischen Funktionsweise der Hardware, die für die Prozesse, mit denen der menschliche Geist über den physischen Körper superveniert und diesen letztlich steuert, alle notwendigen Schnittstellen aufweist. Angesichts der gegenwärtig zur Verfügung stehenden Technik, etwa der im Unterschied zum analog arbeitenden Gehirn diskret arbeitenden Rechenmaschinen, und unseres recht wagen Verständnisses des Zusammenwirkens von Körper und Geist, dürften ein individueller menschlicher Geist und menschliches Leben als Bedingung unserer Form der vernünftigen Intelligenz mit diskreten Zahlen nicht verlustfrei auf ein technisches Surrogat transferiert werden können.[19]

---

[18] Vgl. Bostrom, Nick: Superintelligenz. Berlin 2016, 51–59; Kurzweil: Singularity beschreibt die Anforderungen an ein hinreichendes technisches Surrogat im Kapitel *Achieving the Computational Capacity of the Human Brain* (ebd., 111–142). Er geht davon aus, dass durch *Reverse Engineering* des Gehirns (Hardware) als dem Sitz der Intelligenz Maschinen entwickelt werden können, die mit einer dieser Intelligenz entsprechenden Software arbeiten: „Understanding the methods of the human brain will help us to design similar biologically inspired machines" (ebd., 194). Hinsichtlich der Unterscheidung von *Fähigkeit* (capability) und *Modell dieser Fähigkeit* (description of that capability) argumentiert er schließlich dafür, dass der Computerscan einer Person (die hauptsächlich, wenn auch nicht ausschließlich im Gehirn zu verorten sei; vgl. ebd., 200), nicht einfach ein dreidimensionales Modell sei. Mit Blick auf den Upload der natürlichen Person müsse der Scan bestimmten Kriterien genügen: „The scan does need to capture all of the salient details, but it also needs to be instantiated into a working computational medium that has the capabilities of the original (albeit that the new nonbiological platforms are certain to be far more capable). The neural details need to interact with one another (and with the outside world) in the same ways that they do in the original" (ebd., 201). Kurzweil ist zuversichtlich, dass durch den stetigen Übergang vom Menschen zur Maschine unsere Intelligenz, Persönlichkeit und Fähigkeiten in den 2040er Jahren auf einen nichtbiologischen Teil unserer Intelligenz übergegangen sein und die biologischen Anteile unseres Gehirns derart an Bedeutung verloren haben werden, dass sie letztlich vernachlässigt und die nichtbiologischen Teile vollständig hochgeladen werden können (ebd., 201 f).

[19] Vgl. Sorgner, Stefan Lorenz: We have always been Cyborgs. Bristol 2021, 10: „Computer viruses are self-replicating entities, but they do not have a metabolism for gaining energy, which is a central feature for other living entities. We have no indication for believing that digital life can even be possible[.] As uploaded personalities we would still want to be alive, and being alive also seems to be a necessary prerequisite for having consciousness or self-consciousness, too. All of these reflections indicate that mind uploading is a highly dubitable procedure. I cannot exclude that it will eventually

Daher legt sich zweitens die Möglichkeit nahe, die Bedingungen und mentalen Zustände, die zur Steuerung intelligenten menschlichen Verhaltens notwendig und hinreichend zu sein scheinen, hinsichtlich ihrer Funktionen zumindest näherungsweise digital-technisch zu reproduzieren. Die Unterscheidung zwischen Bedingungen und mentalen Zuständen soll andeuten, dass das Vorhandensein eines Gehirns (oder allgemein eines Körpers) zwar beim Menschen als notwendige Bedingung für Intelligenz aufgefasst werden kann, aber für eine solche allein nicht hinreichend ist. So zeigt etwa ein totes menschliches Gehirn normalerweise kein intelligentes Verhalten, auch wenn ein lebender Mensch mit ihm ein solches zeigen könnte. Unter mentalen Zuständen soll hier der Einfachheit halber alles zusammengefasst werden, was zu einem Gehirn hinzukommen muss, damit wir von einem Bewusstsein als weiterer Bedingung menschlicher Intelligenz sprechen können. Könnten mentale Zustände umfassend digital realisiert werden, ließe sich vielleicht von einem Roboterbewusstsein sprechen. Die Maschine besäße die Komponenten, die wir auch beim Menschen identifizieren, sie stünden auch in einem ähnlichen Verhältnis zueinander,[20] wären aber von offensichtlich anderer Qualität und wahrscheinlich anderer Funktionsweise als ihre organisch-menschlichen Vorbilder:[21] Das Verhalten des menschlichen Organismus wird über ein Gehirn gesteuert, das aus grauer Masse besteht und einkommende Signale analog verarbeitet. Dasselbe Verhalten eines Androiden wird hingegen von einem Zentralcomputer gesteuert, der auf Silizium basiert und auf dem einkommende Informationen von einem künstlichen neuronalen Netzwerk digital verarbeitet und nach erlernten Mustern mit einem Output verbunden werden. Dabei ist die mathematische Funktion, die den Input auf einen Output abbildet, hinsichtlich ihrer Funktionsweise nicht mit den mentalen Zuständen eines Menschen gleichzusetzen. Auch dieser zweite Fall setzt eine präzise Kenntnis des Zusammenwirkens des menschlichen Gehirns bzw. Körpers und (phänomenalen) Bewusstseins voraus, die zum gegenwärtigen Zeitpunkt nicht gegeben ist, und bleibt daher bis auf Weiteres spekulativ.[22]

Um das menschliche Verhalten im Androiden nachbilden zu können, ist die Umwandlung der beobachtbaren analogen Größen in maschinenlesbare Daten ebenso erforderlich, wie die

be possible, but our current scientific basis does not provide us with a strong reason for regarding it as a likely option."

20 Vgl. Barbour: Ethics, 171: „Mind is to brain as software programs are to computer hardware."

21 Vgl. Turing: Maschine, 42: „Könnte es nicht sein, dass Maschinen etwas ausführen, das sich als Denken bezeichnen ließe, sich jedoch stark von dem unterscheidet, was ein Mensch tut?"; vgl. außerdem Lenzen, Manuela: Künstliche Intelligenz. München ²2018, 29.

22 Vgl. Ropohl: Aufklärung, 155 f: „Solange uns die Psychologie nicht eindeutig sagen kann, wie Emotionalität, Kreativität, Intentionalität und Reflexivität in der psycho-physischen Organisation des menschlichen Nervensystems und Gehirns zustande kommen, stößt auch der Vergleich mit tatsächlichen und in Zukunft zu erwartenden Computereigenschaften auf ungelöste Schwierigkeiten."; vgl. Misselhorn: Empathie, 50–71, 89–107.

Nachbildung menschlichen Verhaltens in Programmen. Die Digitalisierung analoger Größen und digitale Nachbildung analoger Prozesse als maschinenlesbare Information (Daten und Programme) ist das Kerngeschäft der Informatik.[23] Die Informationstheorie im engeren formalen Sinn der Informatik sieht von der semantischen Bedeutung und der pragmatischen Bestimmung des beobachteten Zeichens ab. Sie berücksichtigt nur dessen syntaktische Dimension, die jeweilige Menge der unterscheidbaren Symbole und die Häufigkeit und Wahrscheinlichkeit, mit der sie vorkommen.[24] Für die Rekonstruktion auch der subjektiven affektiven und phänomenalen Zustände im menschlichen Bewusstsein, die dem menschlichen Verhalten in der Regel zugrunde liegen, dürfte ein auf die syntaktische Dimension verkürzter Informationsbegriff allerdings nicht hinreichend sein: Hier liegt eine „prinzipiell unüberschreitbare Grenze technischer Informationsverarbeitung […], [die] notwendig den subjektiven Sinn verfehlt, der einer Information erst pragmatische Relevanz verleiht."[25]

Die semantische Bedeutung und pragmatische Bestimmung der Zeichen, in denen menschliche mentale Prozesse kodiert werden, lassen sich mit den gegenwärtigen Methoden der Digitalisierung also nicht formal erfassen. Will man allerdings weiterhin im „prinzipiellen Sinndefizit bei der Verarbeitung objektivierter Information […] den tieferen Grund für die typischen Unzulänglichkeiten der künstlichen Intelligenz [ausmachen]: Muster zu erkennen, Sprache zu übersetzen oder Probleme zu identifizieren"[26], muss man erklären, wie moderne KI-Systeme diese Unzulänglichkeiten Mitte der 2010er Jahre dennoch hinter sich lassen konnten.

Ungeachtet dieser Unzulänglichkeit der Informationstheorie lässt sich drittens intelligentes menschliches Verhalten in der Tat mit verschiedenen statistischen Methoden simulieren: Nicht zuletzt im Zuge der Snowden-Enthüllungen 2013[27] und des Datenskandals um Cam-

---

[23] Ernst, Hartmut/Schmidt, Jochen/Beneken, Gerd: Grundkurs Informatik. Wiesbaden [7]2020 charakterisieren die Informatik als „Wissenschaft von der systematischen Verarbeitung von Informationen, besonders der automatischen Verarbeitung mithilfe von Digitalrechnern", mit einem Wort als „Intelligenzformalisierungstechnik" (ebd., 1). Wesentlich für die Informatik sei die Modellbildung, bei der Ausschnitte der Wirklichkeit durch Algorithmen beschrieben werden, die Objekte in der Welt und ihre Beziehungen untereinander beschreiben. Mit Hilfe der Algorithmen können so Ereignisse geordnet oder gesteuert werden (vgl. ebd., 2 f), vgl. außerdem Zweig, Katharina: Ein Algorithmus hat kein Taktgefühl. München 2019, 69–73.

[24] Vgl. Ernst/Schmidt/Beneken: Informatik, 47: „In der Informatik geht man oft von einer statistischen Deutung des Begriffs Information aus; dies gilt insbesondere dann, wenn es um die Codierung und Übermittlung von Informationen bzw. Nachrichten geht."; vgl. außerdem Ropohl: Aufklärung, 150.

[25] Ropohl: Aufklärung, 156.

[26] Ebd.

[27] Vgl. Snowden, Edward: Permanent Record, London 2019.

bridge Analytica 2018[28] ist offenbar geworden, das es prinzipiell möglich ist, die von Nutzer:innen in sozialen Netzwerken zur Verfügung gestellten Daten wie Likes, Shares, Kommentare, Bilder und private Chats auszulesen, weitere Daten etwa durch großangelegte staatliche Überwachungsprogramme wie PRISM zu erheben, alles zusammenzuführen, mittels statistischer Methoden zu analysieren und daraus Persönlichkeitsprofile zu erstellen, anhand derer mindestens Handlungspräferenzen einzelner Menschen recht zuverlässig vorhergesagt und durch gezieltes Nudging verstärkt werden können. Für eine effektive Modellierung oder Vorhersage menschlichen Verhaltens ist die Kodierung mentaler Zustände also offenbar nicht erforderlich. In diesem Fall ist das Vorhersagemodell eine Form der schwachen KI: Mit den Präferenzen, die sich aus großen Datenmengen generieren und in Maschinenentscheidungen umsetzen lassen, wird lediglich eine Illusion von Intelligenz erzeugt, ohne dass ihre Funktionsweise mit der menschlichen Intelligenz vergleichbar wäre.[29] Auch wird man kaum geneigt sein, diese Illusion als intelligentes Verhalten zu akzeptieren.

Für das Menschenbild der Künstlichen Intelligenz ergibt sich aus der mathematischen Informationstheorie die folgende Konsequenz: Die „signifikante[n] Unterschiede zwischen den herkömmlichen menschlichen Problemlösungsstrategien und den Vorgehensweisen beim Computereinsatz“[30] stehen einer analogen Funktionalität von computerbasierten Maschinen und menschlichen Akteuren nicht prinzipiell entgegen. Seit dem Aufkommen des Computers werden immer wieder neue Maschinen konstruiert, die immer neue intelligenzanaloge Leistungen erbringen, d. h. Funktionen in Bereichen übernehmen, die bis zu diesem Zeitpunkt zur Domäne menschlicher Intelligenz gerechnet wurden. Technische Sachsysteme, die einzelne menschliche Verhaltensweisen und Intelligenzleistungen auf diesem Wege nachahmen – die etwa wie wir sprechen, wie wir Schach spielen, sich wie wir fortbewegen – tun dies also in der Regel nicht auf die gleiche Weise, sondern lediglich auf einem ähnlichen Niveau.

---

28 Vgl. Grassegger, Hannes/Krogerus, Mikael: „Ich habe nur gezeigt, dass es die Bombe gibt“. Online unter: https://www.tagesanzeiger.ch/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt-652492646668 (Stand: 18.07.2022).

29 Vgl. Duffy, Brian R.: Anthropomorphism and the social robot. In: Robotics and Autonomous Systems 42 (2003), 177–190, hier 178 f: „‚artificial intelligence‘ implies that human intelligence can only be simulated. An artificial system could only give the illusion of intelligence (i.e. the system exhibits those properties that are associated with being intelligent).“

30 Ropohl: Aufklärung, 157.

4 Isolierbarkeit und Modularisierbarkeit von Funktionen

Das dem menschlichen Verhalten zugrundeliegende Bewusstsein ist also zumindest zum gegenwärtigen Zeitpunkt einer technischen Reproduktion nicht zugänglich. Ingenieur:innen werden stattdessen Lösungsverfahren, wie wir oben gesehen haben, durch die statistische Auswertung menschlichen Verhaltens reproduzieren oder durch die zielorientierte Analyse gegebener Probleme (Auto fahren, Schach spielen, Zauberwürfel lösen) entwerfen.[31] Dafür müssen sie die für das Problem hinreichenden Bedingungen isolieren, unter denen das menschliche Verhalten zustande kommt, und mögliche Lösungsmethoden so rekonstruieren, dass sie schließlich mit einer entsprechenden funktionalen Ähnlichkeit zur menschlichen Problemlösung technisch reproduziert werden können. Die Produkte können dabei grundsätzlich so konzipiert werden, dass sie sich anschließend modular wieder zu einem integrierten Agenten mit neuen Funktionen kombinieren lassen.

> „Als Agent bezeichnen wir ganz allgemein ein System, welches Information verarbeitet und aus einer Eingabe eine Ausgabe produziert. […] In der Robotik […] werden Hardware-Agenten (auch „autonome Roboter" genannt) verwendet, die zusätzlich über Sensoren und Aktuatoren verfügen […]. Mit den Sensoren kann der Agent die Umgebung wahrnehmen. Mit den Aktuatoren führt er Aktionen aus und verändert so die Umgebung."[32]

Auch die lernenden Algorithmen, die den gegenwärtigen Standard der KI ausmachen, lassen sich auf einzelne Aufgaben spezialisieren (Schach spielen *oder* Zauberwürfel lösen), die sie dann auf einem menschenähnlichen Niveau oder weit darüber beherrschen.[33] Es gibt allerdings bisher keinen plausiblen Ansatz, um so unterschiedliche Aufgabenbereiche wie Essen kochen, Auto fahren, Zauberwürfel lösen, Schach spielen und Wissenserwerb mit einer einzel-

---

[31] Vgl. Ertel: Grundkurs, 3: „Beim Erforschen intelligenter Verfahren kann man versuchen zu verstehen, wie das menschliche Gehirn arbeitet und dieses dann auf dem Computer zu modellieren oder zu simulieren. […] Ein ganz anderer Zugang ergibt sich bei einer zielorientierten Vorgehensweise, die vom Problem ausgeht und nach einem möglichst optimalen Lösungsverfahren sucht. Hierbei ist es unwichtig, wie wir Menschen dieses Problem lösen. Die Methode ist bei dieser Vorgehensweise zweitrangig. An erster Stelle steht die optimale intelligente Lösung des Problems"; zum Beispiel haben Silver, David/Hubert, Thomas/Schrittwieser, Julian et al.: Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. In: arXiv [cs.AI] 2017 einen selbstlernenden Algorithmus geschrieben, der Schach und weitere Brettspiele ohne menschliche Vorbilder explorativ lernt.

[32] Ertel: Grundkurs, 19 f.

[33] Vgl. Lenzen: Künstliche Intelligenz, 32 f.

nen Lernmaxime maschinell abzudecken.[34] Die Idee, eine Maschine zu programmieren, die wie ein Kind lernt und darin unterrichtet werden kann, unterschiedlichste Aufgaben zu meistern,[35] ist ungeachtet ihrer Attraktivität weit von unseren derzeitigen Möglichkeiten entfernt.[36]

Zwei Alternativen bieten sich unmittelbar an: (1) Verschiedene, bereits existierende, spezialisierte Programme und Roboter könnten mit einem künstlichen neuronalen Netzwerk, das Probleme und Aufgabenbereiche identifiziert und sie den einzelnen Komponenten zuweist, zu einem modularen System integriert werden. Oder: (2) Ein System, das bereits existiert und eine spezielle Aufgabe bereits beherrscht, wird schrittweise um die Funktionen erweitert, die es zukünftig ebenfalls erfüllen soll.

Die Implementierung eines Kommunikationsmodells in einen physischen Agenten zum Zweck seiner räumlichen Navigation scheint mittlerweile praktisch unmöglich: LaMDA basiert genau wie andere Large Language Model auf der Analyse großer Mengen von Textdaten. Es ermöglicht die Aufgliederung komplexer Aufgaben wie „Einen Balkongarten anlegen" in einzelne konsekutive Handlungsschritte, die Aufrechterhaltung einer Aufgabe bis zu ihrem Ende, sowie die sinnbringende Vervollständigung unpräziser Eingaben.[37] Der Output des Programms kann durch Nutzerfeedback zunehmend abgesichert werden (groundedness). LaMDA erzeugt bei Sachfragen nicht mehr einfach nur plausibel erscheinende Antworten.[38] Während LaMDA ausschließlich Text als Output generiert, konnte das neuere Modell PaLM in den phy-

---

34 Vgl. Ertel: Grundkurs, 4: „Der Fixpunkt in der KI ist […] das Ziel, intelligente Agenten für die verschiedensten Aufgaben zu bauen"; vgl. Barbour: Ethics, 170: „Some expert systems […] often work well in narrow technical domains that can be isolated from other considerations. But the systems are blind to larger contexts, and they have difficulty deciding where the boundary of the domain lies and when something outside it might be significant." Akkaya, Ilge/Andrychowicz, Marcin/Chociej, Maciek et al.: Solving Rubik's Cube with a Robot Hand; in: arXiv preprint 2019 konnten eine Roboterhand zunächst in einer virtuellen Umgebung darauf trainieren, einen Zauberwürfel zu lösen. Anschließend konnten sie die physische Roboterhand mit Sensoren verbinden und mit diesem Aufbau auch physische dreidimensionale Zauberwürfel lösen.

35 Vgl. Turing: Maschine, 71: „Warum sollte man nicht versuchen, statt ein Programm zur Nachahmung des Verstandes eines Erwachsenen eines zur Nachahmung des Verstandes eines Kindes herzustellen? Unterzöge man dieses dann einem geeigneten Erziehungsprozeß, erhielte man den Verstand eines Erwachsenen."

36 Vgl. dazu Lenzen: Künstliche Intelligenz, 91–94.

37 Vgl. Google: Google Keynote (Google I/O '22). Online unter: https://youtu.be/nP-nMZpLM1A, 40:00–46:00 (Stand 18.07.2022).

38 Vgl. Thoppilan, Romal et al.: LaMDA. In: arXiv [cs.CL] 2022: „We aim to ensure that LaMDA produces responses that can be associated with known sources whenever possible, enabling cross-checking if desired, because the current generation of language models tends to produce plausible but incorrect statements. We define groundedness as the percentage of responses containing claims about the external world that can be supported by authoritative external sources, as a share of all those containing claims about the external world."

sischen Agenten SayCan implementiert werden, der auf diesen Fähigkeiten aufbauend nun in der Lage ist, auch implizite Aufforderungen in Sätzen wie „Ich habe meinen Saft verschüttet!" wahrzunehmen und ihnen mit einer geeigneten praktischen Lösung selbst nachzukommen.[39]

Damit dieses modulare System sich in der für Menschen eingerichteten Welt einer Vielzahl unterschiedlicher Aufgaben widmen kann, scheint es außerdem zweckmäßig, wenn auch nicht zwingend erforderlich, das System kompakt zu halten und in eine humanoide Form zu bringen, die der Flexibilität menschlicher Fortbewegung und Interaktion vergleichbar ist (Hand/Hände zum Greifen, Gesicht für Mimik, Beine zum Treppensteigen usw.).[40]

5 Wie gut ist Teslas Optimus?[41]

Im August 2021 kündigte der Tesla-Chef Elon Musk an, einen Androiden entwickeln zu wollen: *Optimus* soll in einer für Menschen ausgelegten Umgebung agieren können und gefährliche, sich wiederholende oder langweilige Aufgaben übernehmen, wie etwa einkaufen zu gehen.[42] Aufgrund seiner funktionalen Ähnlichkeit zum Menschen soll Optimus Teil unserer Gesellschaft werden und diese tiefgreifend verändern. In praktisch-ökonomischer Hinsicht will Musk eine universelle Arbeitsmaschine entwickeln, die den Menschen in die Lage versetzt, nur noch das tun zu müssen, was er tun will, wenn er es will.[43]

---

39 Vgl. Ahn, Michael/Brohan, Anthony/Brown, Noah et al.: Do As I Can, Not As I Say. In: arXiv [cs. RO] 2022.

40 Vgl. Winfield: Robotics, 61.

41 Die Frage soll nicht in einem moralischen Sinne verstanden werden. Vielmehr wird Optimus' Qualität anhand seiner eigentümlichen Leistung gemessen. Diese kann – ganz dem Wesen der Roboter entsprechend – allgemein mit der Verrichtung von Arbeit umschrieben werden. Vgl. hier und im Folgenden, sofern nicht anders angegeben, für die Äußerungen von Elon Musk und die Informationen zum Tesla-Bot die Präsentation Tesla: Tesla AI Day 2021. Online unter: https://youtu.be/j0z4FweCy4M, 02:07:00–02:11:11 (Stand: 19.07.2022).

42 Vgl. Siddiqui, Faiz: Tesla says it is building a ‚friendly' robot that will perform menial tasks, won't fight back. Online unter: https://www.washingtonpost.com/technology/2021/08/19/tesla-ai-day-robot/ (Stand: 23.04.2022): „We have almost all the pieces needed for humanoid robots, since we already make robots with wheels."

43 Vgl. Klaiber, Hannah: Elon Musk. Online unter: https://t3n.de/news/elon-musk-optimus-wert-1467373/ (Stand: 23.04.2022); Musk erwartet durch Optimus' zunehmend verfügbare und flexibel einsetzbare Arbeitskraft eine tiefgreifende Umgestaltung der Arbeitswelt und damit der Ökonomie sowie langfristig die Einführung eines bedingungslosen Grundeinkommens. Vieles könnte hierzu aus der Perspektive der Sozialethik gesagt werden: insbesondere zur Arbeit als „*eines* (sic!) *der Kennzeichen, die* den Menschen von den anderen Geschöpfen *unterscheiden*" (Johannes Paul II: Enzyklika *Laborem Exercens*. Rom 1981; Hervorhebungen im Original); vgl. auch das Kapitel über „*Die Notwendigkeit die Arbeit zu schützen*" (LS 124–129).

Optimus soll rund 1,72 m groß und 57 kg schwer sein und ein minimalistisches, humanoides Design mit einem digitalen Interface an Stelle eines Gesichts verbinden. Optimus soll genau wie das Tesla-Automobil mit einem *Full Self Driving-Chip* (FSD), der Rechenleistung von Teslas Dojo-Supercomputer sowie acht Kamera-Sensoren ausgestattet werden. Damit wird er die für Menschen eingerichtete Umwelt erkennen und „verstehen", wie er sich sicher in ihr bewegen kann. Eine mathematische Funktion in Form eines neuronalen Netzwerkes im FSD-Chip verrechnet die über die Sensoren einkommenden Bilder in Form von Zahlwerten und bildet diese auf die insgesamt vierzig elektromechanischen Aktuatoren des Roboters ab. Den Output bilden die entsprechenden Bewegungen und Interaktionen mit seiner Umwelt, mit einem Wort, das Verhalten von Optimus.[44] Tesla will es den Nutzer:innen ermöglichen, mit dem humanoiden Bot durch natürlichsprachliche Äußerungen zu kommunizieren und ihm so auch Aufträge zu erteilen, die er dann effektiv umsetzen soll. Seine Tragekapazität soll dafür 20 kg, seine Hebefähigkeit bis zu 68 kg und seine Geschwindigkeit bis zu 8 km/h betragen. Gemessen an diesen Grenzwerten soll Optimus zuverlässig durch urbanes Gelände navigieren können und eine – wie Musk versichert – geringe Gefahrenquelle darstellen.

Optimus wird bei seiner ersten Vermarktung sicher noch weit davon entfernt sein, das volle Leistungsspektrum eines Menschen zu umfassen. Das Repertoire des Bots muss im Laufe der Zeit also modular um weitere Funktionen erweitert werden. Diese additive modulare Erweiterung der Funktionen und Fertigkeiten von Optimus darf nicht mit dem stetigen Lernprozess von Menschen verwechselt werden. Die Form, in der die neuen Fähigkeiten vorliegen werden, ist – abgesehen von technischen Neuerungen an verbesserten Optimus-Modellen – immer die des digitalen Computerprogramms. Die Vision, die diesem Projekt zugrunde liegt, ist also die eines modularisierbaren, auf digitalen Informationen arbeitenden Humanoiden, der den Menschen wenigstens in ausgewählten Bereichen ersetzen kann. Hinsichtlich der geistigen Fähigkeiten scheint Musk im Tesla-Bot ein Zwischenglied zwischen dem Menschen und den selbstfahrenden Autos zu sehen. Bei letzteren handle es sich um „semi-sentient robots on wheels."[45] Es sei nur folgerichtig, das „Halbbewusstsein"[46] des Tesla-Automobils in einen humanoiden

---

[44] Zur Vorstellung eines ersten Prototyps, der sich zumindest frei im Raum bewegen und mit einzelnen Objekten interagieren kann, vgl. Tesla: Tesla AI Day 2022. Online unter: https://youtu.be/ODSJsviD_SU (Stand: 30.10.2022). Der Prototyp kann nichts, was andere Roboter nicht schon (besser) können. Die Geschwindigkeit, mit der Tesla hier Anschluss an das Führungsfeld sucht, ist trotzdem beeindruckend. Darin scheinen sich auch viele Experten einig, vgl. Ackerman, Evan/Guzzio, Erico: What Robotics Experts Think of Tesla's Optimus Robot, 04. Oktober 2022. Online unter: https://spectrum.ieee.org/robotics-experts-tesla-bot-optimus (Stand: 07.11.2022).

[45] Tesla: AI Day 2021, 02:07:16.

[46] Die Anführungszeichen sollen andeuten, dass der Begriff nach der gängigen Auffassung etwa der Philosophie des Geistes mindestens missverständlich ist.

Roboter zu implementieren. Dennoch verfolgt Musk mit Optimus zumindest gegenwärtig nicht ausdrücklich das Ziel, Menschen äußerlich und funktional auf einem Niveau technisch zu reproduzieren, auf dem er vom Menschen nicht mehr zu unterscheiden wäre.

6 Schluss

Dass es gegenwärtig Bestrebungen gibt, den Menschen technisch zu reproduzieren, also Simulacren zu erzeugen, die mit einem echten Menschen hinsichtlich ihres Verhaltens und bestimmter anderer Eigenschaften vergleichbar sind, ist offenkundig. Nicht nur Tesla strebt nach diesem Ziel.[47] Die Entswicklung humanoider Roboter folgt dabei keiner natürlichen Eigengesetzlichkeit der Technik, sondern dem planmäßigen technischen Herstellungshandeln von Ingenieur:innen, Designer:innen und Entwickler:innen, die ihre Ziele und Werte in das Produkt mit einfließen lassen.[48] Papst Franziskus nimmt die allgemeine Beobachtung von *Gaudium et Spes* wieder auf, dass die aktuelle Transformation zugleich von der schöpferischen Gestaltungskraft des Menschen ausgeht wie auch auf ihn, seine Urteile und Wünsche sowie „auf seine Art und Weise, die Dinge und die Menschen zu sehen und mit ihnen umzugehen" (GS 4) zurückwirkt. Er macht dabei zurecht auf das problematische Machtgefälle aufmerksam, dass zwischen den Zielen und Werten der Entwickler:innen einerseits und den Urteilen, Wünschen und Selbstbildern der Nutzer:innen andererseits besteht:

> „Wir können daher sagen, dass am Beginn vieler Schwierigkeiten der gegenwärtigen Welt vor allem die – nicht immer bewusste – Neigung steht, die Methodologie und die Zielsetzungen der Techno-Wissenschaft in ein Verständnismuster zu fassen, welches das Leben der Menschen und das Funktionieren der Gesellschaft bedingt. […] Man muss anerkennen, dass die von der Technik erzeugten Produkte nicht neutral sind, denn sie schaffen ein Netz, das schließlich die Lebensstile konditioniert, und lenken die sozialen Möglichkeiten in die Richtung der Interessen bestimmter Machtgruppen. Gewisse Entscheidungen, die rein sachbezogen erscheinen, sind in Wirklichkeit Entscheidungen im Hinblick auf die Fortentwicklung des sozialen Lebens." (LS 107)

[47] Dies ist eine nicht erschöpfende Liste weiterer Beispiele: Sophia, Atlas, Ishiguros Geminoiden, Pepper, Asimo, BINA48, diverse Sexroboter, die Smartphone App Replika, sowie diverse smarte Assistenzsysteme, allen voran die besonders fortschrittlichen Vertreter Duplex, LaMDA, ChatGPT und Watson.

[48] Vgl. Ropohl: Aufklärung, 163: „Natürlich hat auch das Informationssystem seine personalen Urheber, aber die sind längst hinter der Oberfläche technischer Perfektion verschwunden; es dominiert der falsche Schein vorgeblicher Objektivität in der technischen Vergegenständlichung."

Die Akteur:innen der KI-Entwicklung nehmen also in besonderer Weise an der modernen Bestimmung des Begriffs vom Menschen Teil.[49] Das Menschenbild der Künstlichen Intelligenz, demzufolge der Mensch auf praktische Funktionen reduziert und auf der Basis digitaler Programme modular in einem Sachsystem nachgebildet werden kann, das den Menschen in zunehmendem Maße für Alltagszusammenhänge hinreichend substituieren soll, wird so schließlich auch in der sozio-technischen Wahrnehmung des Menschen verankert. Diese hier skizzierten Grundlinien des Menschenbildes der Künstlichen Intelligenz werden unsere Gesellschaft nachhaltig prägen. Aufgabe der Theologie wird es im Anschluss an diese Darlegung sein, das ambivalentes Potential dieses Menschenbildes nicht nur klarer herauszustellen, sondern auch kritisch zu begleiten und auf Missverhältnisse nicht nur hinzuweisen, sondern ihnen wo immer möglich aktiv entgegenzuwirken.

Literaturverzeichnis

Ackerman, Evan: What Robotics Experts Think of Tesla's Optimus Robot. IEEE Spectrum 04. Oktober 2022. Online unter: https://spectrum.ieee.org/robotics-experts-tesla-bot-optimus (Stand: 07.11.2022).

Ahn, Michael/Brohan, Anthony/Brown, Noah et al.: Do As I Can, Not As I Say. Grounding Language in Robotic Affordance. In: arXiv preprint 2022 [cs.RO], doi: 10.48550.

Akkaya, Ilge/Andrychowicz, Marcin/Chociej, Maciek et al.: Solving Rubik's Cube with a Robot Hand. In: arXiv preprint [cs.LG] 2019, doi: 1910.07113.

Barbour, Ian G.: Ethics in the Age of Technology. Oxford 1993.

Bostrom, Nick: Superintelligenz. Berlin 2016.

Brand, Lukas: Virtuelle Menschenreproduktion. Wer oder Was ist Replika? In: Pirker, Viera/Pisonic, Klara (Hg.): Virtuelle Realität und Transzendenz. Theologische und Pädagogische Erkundungen. Freiburg i. Br. 2022, 96–115.

Brand, Lukas: Darf ich Sophia abschalten? Vom rechtlichen und moralischen Status virtueller Personen. In: Fateh-Moghadam, Bijan/Zech, Herbert (Hg.): Transformative Technologien. Wechselwirkungen zwischen technischem und rechtlichem Wandel. Baden-Baden 2021, 209–242.

Detel, Wolfang: Metaphysik und Naturphilosophie (Grundkurs Philosophie 2). Stuttgart ^{3}2014.

Duffy, Brian R.: Anthropomorphism and the social robot. In: Robotics and Autonomous Systems 42 (2003), 177–190.

[49] Vgl. Landmann, Michael: De Homine. München 1962, VI: „Die Menschenanschauung schlägt sich ungewollt und gewollt in allem nieder, was wir denken, tun und hervorbringen. […] Sowohl Religion und Kunst wie aber ebenso auch Gesellung und Staat, Recht und Sitte, Wirtschaft und Technik, kurz sämtliche Kulturdomänen eines Volkes und einer Epoche enthalten ein unausgesprochenes und vielfältig gebrochenes menschliches Selbstverständnis, eine, wie man sagen könnte, ‚implizite Anthropologie' und haben in ihr eine der Determinanten ihrer jeweiligen Gestaltung."

Ernst, Hartmut/Schmidt, Jochen/Beneken, Gerd: Grundkurs Informatik. Grundlagen und Konzepte für die erfolgreiche IT-Praxis – Eine umfassende praxisorientierte Einführung. Wiesbaden [7]2020.

Ertel, Wolfgang: Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung. Wiesbaden [5]2021.

Franziskus: Enzyklika *Laudato Si*. Über die Sorge für das gemeinsame Haus. Deutsche Übers. bereitgestellt durch den Vatikan: https://www.vatican.va/content/dam/francesco/pdf/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si_ge.pdf (Stand: 05.01.2023). Rom 2015.

Google: Google Keynote (Google I/O '22). Online unter: https://youtu.be/nP-nMZpLM1A (Stand: 18.07.2022).

Grassegger, Hannes/Krogerus, Mikael: Ich habe nur gezeigt, dass es die Bombe gibt. Tagesanzeiger 20. März 2018. Online unter: https://www.tagesanzeiger.ch/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt-652492646668 (Stand: 18.07.2022).

Johannes Paul II: Enzyklika *Laborem Exercens*. Über die menschliche Arbeit. Deutsche Übers. herausg. vom Sekretariat der Deutschen Bischofskonferenz. Rom 1981.

Klaiber, Hannah: Elon Musk: Optimus-Roboter wertvoller als Autogeschäft und FSD. t3n 21. April 2022. Online unter: https://t3n.de/news/elon-musk-optimus-wert-1467373/ (Stand: 23.04.2022).

Kurzweil, Ray: The Singularity is near. When Humans Transcend Biology. New York 2006.

Kriegman, Sam/Blackiston, Douglas/Levin, Michael et al.: A scalable pipeline for designing reconfigurable organisms. In: Proceedings of the National Academy of Sciences 117/4 (2020), 1853-1859, doi: 10.1073/pnas.1910837117.

Landmann, Michael: De Homine. Der Mensch im Spiegel seines Gedankens. München 1962.

Lenzen, Manuela: Künstliche Intelligenz. Was sie kann & was uns erwartet. München [2]2018.

Meier-Oeser, Stephan: Technologie. In: Ritter, Joachim/Gründer, Karlfried (Hg.): Historisches Wörterbuch der Philosophie (Bd. 10). Basel 1998, 958–961.

Misselhorn, Catrin: Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern & Co. Ditzingen 2021.

Mori, Masahiro: The Uncanny Valley. In: IEEE. Robotics and Automation Magazine 19/2 (2012), 98–100.

Rahner, Karl/Vorgrimmler, Herbert: Kleines Konzilskompendium. Sämtliche Texte des Zweiten Vatikanischen Konzils. Freiburg i. Br. [35]2008.

Ritter, Joachim/Gründer, Karlfried (Hg.): Historisches Wörterbuch der Philosophie (Bd. 10). Basel 1998, 940–952.

Ropohl, Günther: Technologische Aufklärung. Beiträge zur Technikphilosophie. Frankfurt a. M. [2]1999.

Schweppenhäuser, Gerhard: Designtheorie. Wiesbaden 2016.

Siddiqui, Faiz: Tesla says it is building a ‚friendly' robot that will perform menial tasks, won't fight back. The Washington Post 20. August 2021. Online unter: https://www.washingtonpost.com/technology/2021/08/19/tesla-ai-day-robot/ (Stand: 23.04.2022).

Silver, David/Hubert, Thomas/Schrittwieser, Julian et al.: Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. In: arXiv preprint [cs.AI] 2017, doi: 1712.01815.

Snowden, Edward: Permanent Record. London 2019.

Sorgner, Stefan Lorenz: We have always been Cyborgs. Digital Data, Gene Technologies, and an Ethics of Transhumanism. Bristol 2021.

Tesla: Tesla AI Day 2021. Online unter: https://youtu.be/j0z4FweCy4M (Stand: 19.07.2022).

Tesla: Tesla AI Day 2022. Online unter: https://youtu.be/ODSJsviD_SU (Stand: 30.10.2022).

Thoppilan, Romal/De Freitas, Daniel/Hall, Jamie et al.: LaMDA. Language Models for dialog Applications. In: arXiv preprint [cs.CL] 2022, doi: 2201.08239.

Turing, Alan M.: Kann eine Maschine denken? In: Zimmerli, Walter Ch./Wolf, Stefan (Hg.): Künstliche Intelligenz. Philosophische Probleme. Stuttgart 1994, 39–78.

Wiegerling, Klaus/Nerukar, Michael/Wadephul, Christian: Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien. In: Kolany-Raiser, Barbara/Heil, Reinhard/Orwart, Carsten/Hoeren, Thomas (Hg.): Big Data und Gesellschaft. Eine multidisziplinäre Annäherung (Technikzukünfte, Wissenschaft und Gesellschaft). Wiesbaden 2018, 1–68.

Winfield, Alan: Robotics. A Very Short Introduction. Oxford 2012.

Zweig, Katharina: Ein Algorithmus hat kein Taktgefühl. Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können. München ^{4}2019.

Wie sollen wir mit künstlich-intelligenten humanoiden Robotern umgehen?

Drei philosophische Interpretationen dieser Frage

Sven Nyholm

Abstract

This chapter considers whether robots should ever be treated with moral consideration, focusing on humanoid robots, i.e., robots designed to look and behave like human beings. Recently, a large academic literature on this topic has sprung up. In fact, so much has been written about this that it can be hard to navigate the literature. This chapter suggests a way of classifying important contributions to this debate by identifying three key questions on the basis of which many views can be categorized and further analyzed: Question 1: can robots have morally relevant properties or abilities? Question 2: can robots imitate or simulate morally relevant properties or abilities? Question 3: can robots represent or symbolize morally relevant properties or abilities? In addition to arguing that these questions are useful in discussions of the moral status of robots, the chapter also reviews some arguments for and against discussing the issue of moral consideration for robots in the first place.

1 Der Tesla-Bot und Erica

Im August 2021 präsentierte der Technologieunternehmer und CEO des Autokonzerns Tesla, Elon Musk, seine Pläne für den „Tesla Bot". Während dieser Präsentation sagte Musk, dass die selbstfahrenden Autos, die Tesla entwickelt hat, „basically [are] semi-conscious robots on wheels" und dass „Tesla is really the world's largest robotics company". Daher sei es sinnvoll,

räsonierte Musk weiter, diesen angeblich halbbewussten Robotern, eine „humanoide Form" zu geben. Daher die Idee für den Tesla Bot.[1]

Musk zeigte ein Bild eines Roboters in humanoider Gestalt – allerdings ohne Gesichtszüge und ohne menschenähnliche Farbtöne. Er erklärte einige der Technologien, mit denen er den Tesla Bot ausstatten würde: sowohl KI-bezogene Technologien als auch die mechanischen Arten von Technologien, die in diesen Robotern verwendet werden sollten. Musk erklärte auch, welchen Zweck diese Roboter seiner Meinung nach erfüllen könnten: Sie könnten den Menschen unerwünschte Arbeit abnehmen, indem sie „dangerous, repetitive, boring tasks" eliminieren würden.

Musk erklärte weiter, dass die Roboter „friendly" sein würden. Sie wären nur motiviert, Menschen zu dienen und helfen. Außerdem sagte Musk, dass die Roboter schwach genug sein würden, dass man einen Kampf gegen sie leicht gewinnen könne. Sie würden so langsam sein, dass man leicht weglaufen könnte, wenn man Angst vor einem Tesla-Bot bekommen würde.

Eine ganz andere Vision eines humanoiden Roboters stammt vom japanischen Robotikforscher Hiroshi Ishiguro. Neben der Schaffung einer Roboterkopie von sich selbst, die ihm tatsächlich bemerkenswert ähnlich sieht, hat Ishiguro auch verschiedene andere humanoide Roboter geschaffen, darunter einen, der wie eine Japanerin aussieht. Dieser Roboter heißt „Erica". Ishiguro sagte, dass sie die schönste Frau sei, die er je gesehen habe. Diese Roboter sind nicht gesichtslos und sie sind nicht als bloße Werkzeuge geschaffen, die wir überwältigen und rein instrumentalisierend behandeln können. Vielmehr interagiert Ishiguro mit seinen Robotern – vielleicht besonders mit Erica – auf eine Art und Weise, die den Robotern fast ein gewisses Maß an Ehrfurcht entgegenbringt. Sie werden als Personen behandelt, nicht als bloße Werkzeuge oder Dinge.[2]

Diese Tendenz, humanoide Roboter so zu behandeln, als wären sie Personen oder zumindest mehr als bloße Werkzeuge, lässt sich an der Art und Weise, wie viele Menschen mit Robotern interagieren, beobachten. Ein relevantes Beispiel, das weit bekannt ist, ist der Roboter Sophia der Firma Hanson Robotics. Sophia wurde im Jahr 2017 bei einer Technologie- und Futurismus-Veranstaltung sogar die „Ehrenbürgerschaft" des Königreichs Saudi-Arabien verliehen.[3] Das scheint ein klares Beispiel für eine gute Behandlung eines Roboters durch Menschen zu sein. Allerdings kann der anthropomorphisierende

---

1 Zum Video des Auftritts: CNET HIGHLIGHTS: Elon Musk REVEALS Tesla Bot (full presentation). Online unter: https://www.youtube.com/watch?v=HUP6Z5voiS8 (Stand: 23.10.2022).

2 Weitere Informationen über und Bilder von Ishiguros Robotern finden sich auf der Website seines Labors: http://www.geminoid.jp/en/index.html.

3 Zur Diskussion siehe Kapitel 1 von Nyholm, Sven: Humans and Robots: Ethics, Agency, and Anthropomorphism. London 2020.

Umgang mit Robotern manchmal auch weniger positiv und stattdessen von ethisch fragwürdiger Natur sein.

Zum Beispiel lässt sich dies anhand des Sexroboters Roxxxy veranschaulichen, der von dem amerikanischen Erfinder Douglas Hines entwickelt wurde. Als Hines Roxxxy im Jahr 2010 auf einer Technologieveranstaltung vorstellte, erklärte Hines, dass er diesen Sexroboter mit unterschiedlichen Persönlichkeitseinstellungen ausgestattet habe. Einige davon waren von der Art, die man von einem Sexroboter erwarten würde. Beispielsweise würde der Roboter flirtende Kommentare abgeben und positiv auf verschiedene Arten von Berührungen reagieren. Roxxxy hatte jedoch auch einen umstritteneren Persönlichkeitsmodus: „Frigid Farah". In diesem Modus würde der Roboter die sexuellen Avancen des Benutzers nicht begrüßen, sondern eher „nein" sagen und sich so verhalten, als ob der Roboter keine sexuelle Interaktion wünscht oder dieser nicht zustimmt.[4]

Wenn man über diese Art von Beispielen reflektiert, stellt sich die Frage, ob Roboter jemals mit moralischer Rücksicht behandelt werden sollten, genauso wie wir denken, dass Menschen mit moralischer Rücksicht behandelt werden sollten. In diesem Kapitel werden drei verschiedene Möglichkeiten, wie man sich dieser Frage nähern kann, identifiziert und diskutiert. Gegen Ende des Kapitels wird auch diskutiert, ob man überhaupt über den moralischen Status von Robotern nachdenken sollte.

Der Aufsatz beschäftigt sich besonders mit *humanoiden Robotern*, d. h. Robotern, die so wie Menschen aussehen und/oder sich verhalten. Ich werde mich aus drei Gründen auf humanoide Roboter konzentrieren. Erstens sind dies die Art von Robotern, bei denen die Frage, ob Roboter jemals sog. *moral patients* werden könnten, am realistischsten und am interessantesten erscheint.[5] Zweitens werden in vielen Science-Fiction-Filmen humanoide – oder zumindest vage humanoide – Roboter tendenziell als *moral patients* dargestellt. Drittens gibt es einflussreiche bekannte Persönlichkeiten in der Technologie-Welt wie z. B. Elon Musk, Hiroshi Ishiguro oder David Hanson und Ben Goertzel bei Hanson Robotics, die von der Idee humanoider Roboter sehr fasziniert sind und humanoide Roboter entwickeln. Humanoide Roboter sind noch nicht Teil unseres Alltags. Aber in der Zukunft werden sie es vielleicht sein.

---

4 Vgl. Nyholm: Humans and Robots, 105.

5 Ich benutze hier den anglophonen Ausdruck „moral patients". Eine andere Möglichkeit hier wäre, Janina Lohs Term „moralische Objekte" zu benutzen. Unter „moral patients" (oder was Loh „moralische Objekte" nennt) verstehe ich Lebewesen oder andere Entitäten, gegenüber denen es möglich ist, moralisch richtig oder falsch zu handeln. Für Lohs Diskussion, siehe LOH, Janina: Roboterethik: Eine Einführung. Frankfurt a. M. 2019, Kapitel 2.

2 Was ist ein humanoider Roboter? Und warum würde irgendjemand einen humanoiden Roboter erschaffen wollen?

In dem Bühnenstück *Rossums Universal Robots* des tschechischen Dramatikers Karel Čapek aus dem Jahr 1920, der das Wort „Roboter" hier einführte, ähneln die Roboter ein bisschen dem Tesla Bot, den Musk etwas mehr als 100 Jahre später im Jahr 2021 erschaffen würde. Das heißt, diese Roboter sehen auch aus wie künstliche Menschen. Sie sind auch gebaut, um für Menschen zu arbeiten, in diesem Fall in einer Fabrik. Tatsächlich leitet sich das Wort „Roboter" vom tschechischen Wort „robota" ab, was Zwangsarbeit bedeutet.

In gewisser Weise bleibt also Musks Vision des Tesla Bots der ursprünglichen Idee eines Roboters aus dem Theaterstück treu, in dem dieser Begriff zum ersten Mal verwendet wurde. Im wirklichen Leben sehen jedoch die meisten Roboter, die für Menschen nützlich sind, nicht wie Menschen aus, zumindest nicht in der Gegenwart. Denken Sie zum Beispiel an den Staubsaugerroboter Roomba, den einige Leser:innen dieses Buches vielleicht zuhause haben oder zumindest kennen. Oder denken Sie an die Roboter, die in Fabriken zum Bau von Autos eingesetzt werden, an Roboter, mit denen der Meeresboden erkundet wird, oder an Militärroboter. Keiner dieser Roboter sieht aus wie ein Mensch. Stattdessen haben sie Formen, die für die Aufgaben relevant sind, die sie ausführen sollen. Roomba sieht zum Beispiel ein bisschen aus wie ein Hockey Puck oder ein Käfer. Roboter, die in Lagerhäusern „arbeiten", sehen manchmal aus wie Kisten oder Container auf Rädern, die Pakete herumtragen können. Einige Militärroboter sehen aus wie kleine Panzer.

Ein maximal menschenähnlicher humanoider Roboter wäre einer, der sowohl aussieht wie ein Mensch als auch sich so verhält, dass er nicht von einem echten Menschen zu unterscheiden ist. Solche Roboter sind derzeit überhaupt nicht realistisch. Aber es gibt humanoide Roboter, die in der Entwicklungsphase sind, die wie Menschen aussehen und sich zumindest in gewisser Weise wie Menschen verhalten.

Ein Sexroboter ist ein Beispiel für einen humanoiden Roboter.[6] Roxxxy, der bereits erwähnt wurde, ist ein Sexroboter, der wie eine menschliche Frau aussieht und bestimmte einfache Bewegungen ausführen kann und der eine grundlegende Chatfunktion hat, ein bisschen wie ein Chatbot wie Alexa oder Siri. Sophia ist ein weiteres Beispiel für einen humanoiden Roboter. Dieser Roboter hat keine offensichtliche Funktion – außer vielleicht, bei Menschen das Interesse für Roboter und insbesondere für die von Hanson Robotics entwickelten Roboter zu gewinnen.

[6] Sexroboter sind nicht notwendigerweise humanoide Roboter. Man kann sich auch Sexroboter vorstellen, die nicht wie Menschen aussehen oder sich so benehmen. In der ethischen Debatte über Sexroboter geht es aber fast ausschließlich um menschenähnliche Sexroboter.

Ein anderes Beispiel eines humanoiden Roboters ist Kaspar, ein Roboter, der ein bisschen wie ein Kind aussieht und in experimentellen Therapien für Kinder mit Autismus eingesetzt wird.[7] Dieser Roboter soll helfen, Kinder mit Autismus für soziale Interaktion mit Fremden zu öffnen und Scheu abzubauen. Daher scheint es wichtig, dass dieser Roboter eine menschenähnliche Form hat, da dies Teil der Strategie hinter der Therapie ist.

Zu betonen ist hier, dass einige Forscher:innen meinen, dass es gute Gründe gibt, dem Projekt der Schaffung humanoider Roboter skeptisch gegenüberzustehen. Ein solcher Grund, der oft diskutiert wird, ist, dass Roboter, die wie Menschen aussehen und sich wie Menschen verhalten, etwas Ungeheures an sich hätten. Man spricht manchmal von der sogenannten „Uncanny-Valley"-Hypothese.[8]

Interessanterweise hat der Mathematiker und Informatikpionier Alan Turing bereits 1951 in einer Radiorede, die er für den BBC aufzeichnete, ähnliche Bedenken aufgestellt. Turing war sehr daran interessiert, Maschinen zu entwickeln, die sich so verhalten, als ob sie denken könnten. Aber Turing war gleichzeitig skeptisch gegenüber der Idee, humanoide Roboter zu erschaffen. Hier ist ein Zitat aus der Radiosendung Turings, die für den BBC aufgenommen wurde:

> I certainly hope and believe that no great efforts will be put into making machines with the most distinctively human, but non-intellectual characteristics such as the shape of the human body. It appears to me to be quite futile to make such attempts and their results would have something like the unpleasant quality of artificial flowers. Attempts to produce a thinking machine seem to me to be in a different category. […] I believe that the attempt to make a thinking machine will help us greatly in finding out how we think ourselves.[9]

Es ist faszinierend, dass das, was Turing im obigen Zitat sagt, mit den Gründen zu vergleichen ist, die manchmal angeführt werden, warum es eine gute Idee sein kann, zu versuchen, humanoide Roboter zu erschaffen. Hiroshi Ishiguro sagt zum Beispiel, dass einer der guten Gründe für den Versuch, humanoide Roboter zu entwickeln, darin besteht, dass dies uns helfen kann, uns selbst besser zu verstehen. Ähnliches sagen die Erfinder:innen und Entwickler:innen von Sophia, dem obengenannten Roboter von Hanson Robotics. David Hanson hat sogar ein Buch

---

7 Weitere Informationen zu Kaspar finden sich auf dieser Website: https://www.herts.ac.uk/kaspar/the-social-robot.

8 Mori, Masahiro: The Uncanny Valley. In: IEEE Robotics and Automation Magazine 19/2, 98–100.

9 Turing, Alan: Can Digital Computers Think? TS with AMS Annotations of a Talk Broadcast on BBC Third Programme 15 May 1951. Online unter: https://turingarchive.kings.cam.ac.uk/publications-lectures-and-talks-amtb/amt-b-5 (Stand: 25.08.2022).

mit dem Titel *Humanizing Robots: How Making Humanoids Can Make Us More Human* veröffentlicht. Die beiden einleitenden Absätze aus diesem Buch sind es wert, zitiert zu werden, da sie einen Gegensatz zu dem Zitat von Turing oben bilden. Hansson schreibt:

> The human face is the dynamic icon of the human identity. It embodies our sense of self and others, not just in concept but in our very neural infrastructure. The face seizes attention. The history of art, artifact, and entertainment proves our persistent fascination with the human likeness. Sometimes realistic, other times fantastic; sometimes beautiful, other times hideous – always the face speaks to us […].[10]

Hanson fährt fort:

> Humanlike robots reflect not just our faces but also our thoughts. They allow us to explore the deep aspects about what makes us human […] Yet they also promise to be of profound importance to our future. Humanoids push artificial intelligence (AI) toward human-level, "strong" AI.[11]

Hanson glaubt also, dass das Projekt zur Schaffung humanoider Roboter eine ganze Reihe potenzieller Vorteile haben könnte. Es könnte eine Art Kunstwerk sein. Es könnte uns helfen, uns selbst besser zu verstehen. Es könnte uns menschlicher machen. Und erstaunlicherweise könnte es uns laut Hanson auch helfen, künstliche Intelligenz auf menschlicher Ebene zu entwickeln.

3 Können Menschen Robotern gegenüber richtig oder falsch handeln?

Im Jahr 2015 publizierte der US-amerikanische Fernsehsender *Cable News Network* (*CNN)* einen Artikel mit der Überschrift „Is it cruel to kick a robot dog?".[12] Dies war keine Geschichte über einen humanoiden Roboter, sondern über einen vierbeinigen Roboter mit dem Spitznamen „Spot", der, wenn er sich bewegte, einem Hund ähnelte. Der Roboter wurde von der Firma Boston Dynamics entwickelt. Dieser Roboter zeichnete sich dadurch aus, dass er sehr gut das

[10] Hanson, David: Humanizing Robots: How Making Humanoids Can Make Us More Human. Dallas 2017, 1.

[11] Hanson: Humanizing Robots, 1.

[12] Parke, Phoebe: Is it Cruel to Kick a Robot Dog? Online unter: https://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/index.html (Stand: 25.8.2022).

Gleichgewicht halten konnte. Um dies zu veranschaulichen, zeigt ein von Boston Dynamics veröffentlichtes Video Spot, wie er eine Treppe hinaufgeht und auf einem Laufband läuft. Später im Video wird ein Ingenieur von Boston Dynamics gezeigt, der Spot tritt, um weiter zu veranschaulichen, wie stabil Spot ist. Tatsächlich fällt Spot nicht um, wenn er getreten wird. Viele Zuschauer:innen dieses Videos verloren jedoch die Fassung, als sie sahen, wie Spot getreten wurde. CNN berichtete, dass die Zuschauer:innen Kommentare wie zum Beispiel „Kicking a dog, even a robot dog, just seems wrong" und „Poor Spot!" schrieben.

Dies ist ein Beispiel dafür, dass Menschen moralische Urteile darüber fällen können, wie ein Roboter behandelt wird. Stellen Sie sich nun vor, wir hätten es nicht mit einem Roboterhund zu tun, sondern mit einem Roboter, der einem Menschen ähnelt. Stellen Sie sich zum Beispiel vor, dass jemand Erica tritt, d. h. einen Roboter, der wie eine menschliche Frau aussieht. Wenn Menschen mit Missbilligung auf den Anblick von Ingenieuren reagieren, die einen Roboterhund treten, würden viele wahrscheinlich noch stärker auf den Anblick von Menschen, die einen Roboter treten, reagieren, der so aussieht und sich verhält wie ein menschliches Wesen.

Dies könnte insbesondere dann der Fall sein, wenn ein Roboter wie eine bestimmte Person aussieht. Nehmen Sie zum Beispiel nicht Erica, sondern den anderen oben erwähnten Roboter, der von Hiroshi Ishiguro geschaffen wurde: nämlich die Roboterkopie seiner selbst, die Ishiguro geschaffen hat. Sicherlich wäre es Ishiguro unbehaglich – vielleicht würde er sich in irgendeine Weise als angegriffen erleben – wenn jemand anfangen würde, die Roboterkopie von ihm zu treten oder auf andere Weise anzugreifen.

Wie steht es mit akademischen Diskussionen über diese Themen? David Gunkel ist ein Philosoph, der ausführlich darüber geschrieben hat, ob Roboter jemals mit moralischer Rücksicht behandelt werden sollten. Neben der sehr detaillierten Literaturübersicht, die Gunkel bereits 2018 zu diesem Thema herstellte, als sein Buch *Robot Rights* erschien, hat Gunkel auch kontinuierlich eine visuelle Übersicht der unterschiedlichen Sichtweisen aktualisiert, die Philosoph:innen, Informatiker:innen, Theolog:innen und andere abbildet, je nachdem, ob sie Roboter als *moral patients* betrachten.[13] Die Karte war bereits 2018 komplex. Mit jedem neuen Update der Feldkarte, die Gunkel online stellt, wird die Karte komplexer.[14]

Daher kann es schwierig sein zu wissen, wo wir anfangen sollen, wenn wir darüber nachdenken möchten, ob Roboter jemals als *moral patients* betrachtet werden sollten, gegenüber denen wir richtig oder falsch handeln können. Wie kann man sich am besten orientieren? Hier möchte ich drei wichtige Schlüsselfragen vorschlagen, an denen wir uns orientieren können, wenn wir uns der Frage nähern, ob Roboter *moral patients* sein können, und zwar:

---

13 Gunkel, David: Robot Rights, Cambridge 2018.

14 Hier ist eine Version von Gunkels Karte vom Januar 2022, die auf Twitter veröffentlicht wurde: https://twitter.com/David_Gunkel/status/1485983871590182918/photo/1.

1. Können Roboter moralisch relevante Eigenschaften oder Fähigkeiten *haben*?
2. Können Roboter moralisch relevante Eigenschaften oder Fähigkeiten *imitieren* oder *simulieren*?
3. Können Roboter moralisch relevante Eigenschaften oder Fähigkeiten *repräsentieren* oder *symbolisieren*?

Viele der wichtigsten Beiträge zur Literatur darüber, ob Roboter *moral patients* sein können, können als Erforschung dieser drei Fragen interpretiert werden. In den folgenden Abschnitten veranschauliche ich dies.

4 Können Roboter moralisch relevante Eigenschaften/ Fähigkeiten haben?

Der amerikanische Philosoph Eric Schwitzgebel und die Philosophin und Künstlerin Mara Garza argumentieren in folgender Weise.[15] Sie behaupten erst, dass wir immer gleiche Fälle gleich behandeln sollten. Schwitzgebel und Garza schlagen dann vor, dass es theoretisch KI-Systeme oder Roboter geben könnte, die die Art von Eigenschaften oder Fähigkeiten haben, die wir mit unserem menschlichen moralischen Status in Verbindung bringen. Sie kommen dann zu dem Schluss, dass solche möglichen KI-Systeme oder Roboter den gleichen moralischen Status wie Menschen haben würden. Oder wenn ihre Fähigkeiten denen von Tieren ähnlich wären, dann hätten die KI-Systeme oder Roboter den gleichen moralischen Status wie Tiere. Wenn wir zum Beispiel Roboter schaffen könnten, die leiden können, soziale Beziehungen haben oder rational denken können, dann behaupten Schwitzgebel und Garza, dass wir diese Roboter mit der gleichen moralischen Rücksicht behandeln sollten, mit der wir Menschen behandeln.

Interessanterweise stimmt die Informatikerin und KI-Ethikerin Joanna Bryson, die sonst der Meinung ist, dass wir Roboter als bloße Mittel und nicht als Zweck behandeln sollen, mit Schwitzgebel und Garza überein.[16] Bryson meint, dass es theoretisch möglich sei, Roboter zu entwickeln, die *moral patients* wären, denen gegenüber wir bestimmte Verpflichtungen hätten. Laut Bryson ist es jedoch ein moralischer Imperativ, solche Maschinen nicht zu bauen. Wir sollten nur Maschinen herstellen, die wir berechtigterweise als bloße Mittel und Werkzeuge behandeln können.

---

[15] Vgl. Schwitzgebel, Eric/Garza, Mara: A Defense of the Rights of Artificial Intelligences. In: Midwest Studies in Philosophy 39/1 (2015), 98–119.

[16] Vgl. Bryson, Joanna: Robots Should Be Slaves. In: Wilks, Yorick (Hg.): Close Engagements with Artificial Companions. Amsterdam: 2010, 63–74.

Ähnlich argumentiert auch der deutsche Philosoph Thomas Metzinger.[17] Metzinger beschäftigt sich in seiner Forschung mit Fragen über das Bewusstsein und ist überzeugt, dass es theoretisch möglich ist, Maschinen zu bauen, die Schmerzen empfinden können. Metzinger hält es jedoch für unethisch, solche Maschinen zu bauen, da sie leiden könnten. Um solches eventuelle Leiden zu verhindern, sollten wir keine Maschinen mit Gefühlen erschaffen, argumentiert Metzinger.

Viele Philosoph:innen wie z. B. Carissa Véliz meinen, dass Technologien kein Bewusstsein und keine subjektive Erfahrung haben und wahrscheinlich niemals haben werden.[18] Technologien sind etwa „moralische Zombies", laut Véliz. Véliz diskutiert diese Idee hauptsächlich in Bezug darauf, ob Technologien *moral agents* sein können. Aber vermutlich würde Véliz auf der Grundlage desselben Arguments auch den Schluss ziehen, dass Roboter keine *moral patients* sein können.

An dieser Stelle ist es eine gute Idee, darauf hinzuweisen, dass es viele Meinungsverschiedenheiten darüber gibt, ob es möglich ist, Maschinen mit irgendeiner Form von Bewusstsein zu erschaffen. Man kann mit Sicherheit sagen, dass die Ansicht, die Véliz vertritt – d. h., dass Maschinen kein Bewusstsein haben und dass sie es wahrscheinlich in absehbarer Zeit nicht haben werden – die Standardansicht vieler, wenn nicht der meisten Forscher:innen ist. Es gibt jedoch auch viele Forscher:innen – wie u. a. Bryson und Metzinger – die glauben, dass es doch möglich ist, bewusste Maschinen zu erschaffen, wie wir gerade gesehen haben.

Interessanterweise schlägt Bryson in einem ihrer Artikel sogar vor, dass einige existierende Maschinen eine begrenzte Form des Bewusstseins haben.[19] Wenn wir mit Bewusstsein die Fähigkeit meinen, Informationen aufzunehmen und diese Informationen im „Kopf" zu verarbeiten, um dann in der Lage zu sein, diese Wahrnehmungen anderen Akteur:innen mitzuteilen, dann kann man laut Bryson sagen, dass einige Maschinen eine grundlegende Form von Bewusstsein bereits haben.

Es gibt auch diejenigen, die aktiv versuchen, Roboter zu entwickeln, die Lust und Schmerz empfinden können. Beispielsweise beschäftigt sich der japanische Robotik-Forscher Minoru Asada mit dem Projekt, Roboter zu bauen, die Lust und Schmerz empfinden können.[20] Seine Theorie ist, dass dies helfen könnte, eine neue Form der künstlichen Intelligenz zu erschaffen.

---

17 Vgl. Metzinger, Thomas: Two Principles for Robot Ethics. In: Hilgendorf, Eric/ Günther, Jan-Philipp (Hg.): Robotik und Gesetzgebung. Baden-Baden: 2013, 236–302.

18 Vgl. Veliz, Carissa: Moral Zombies: Algorithms are not Moral Agents In: AI & Society 36/2 (2021), 487–497.

19 Vgl. Bryson, Joanna: A Role for Consciousness in Action Selection. In: International Journal of Machine Consciousness 4/2 (2012), 471–482.

20 Vgl. Asada, Minoru: Artificial Pain May Induce Empathy, Morality, and Ethics in the Conscious Mind of Robots. In: Philosophies 4/3 (2019), 38.

Die Idee ist, dass Maschinen, die Lust und Schmerz empfinden können, besser lernen könnten, da wir Menschen zum Teil lernen, indem wir Lust und Schmerz erfahren (und dies ist sicherlich der Fall, bevor wir als Babys lernen, Sprachen zu verstehen).

Es gibt also tiefe Meinungsverschiedenheiten darüber, ob Roboter jemals die Arten von Fähigkeiten haben könnten oder haben werden, die ihnen einen wichtigen moralischen Status verleihen würden. Ob Roboter moralisch wichtige Eigenschaften und Fähigkeiten *haben* können, ist in der Forschung also unklar. Daher sollten wir uns im nächsten Schritt einer anderen Frage zuwenden:

5 Können Roboter moralisch relevante Eigenschaften oder Fähigkeiten imitieren oder simulieren?

Kehren wir noch einmal zu Turings Ideen zurück. Turing dachte, dass statt zu untersuchen, ob Maschinen denken können, es besser ist zu fragen, ob Maschinen sich so benehmen können, dass Menschen sie so wahrnehmen, als ob sie denken könnten.[21] Entscheidend ist also laut Turing, wie sich Maschinen verhalten und nicht, ob sie bewusste Erfahrungen machen oder ob sie einen Verstand haben, der dem menschlichen ähnelt.

Der irische Rechtswissenschaftler und Philosoph John Danaher meint, dass wir Turings Argumentation anwenden sollten, wenn wir darüber nachdenken, ob Maschinen *moral patients* sein könnten.[22] Das heißt, Danaher meint, dass Turing einen guten Test dafür hat, ob Maschinen intelligent sind. Und Danaher findet, dass wir einen ähnlichen Test – einen ethischen Turing-Test, wenn man so will – anwenden sollten, wenn wir darüber nachdenken, ob wir Maschinen als *moral patients* betrachten sollten, denen gegenüber wir richtig oder falsch handeln können.

Danaher merkt an, dass Turings Sichtweise als eine Form von „methodologischen Behaviorismus" angesehen werden kann. Die Methode, um herauszufinden, ob eine Maschine intelligent ist, besteht also darin, das Verhalten der Maschine zu untersuchen. In gleicher Weise schlägt Danaher vor, dass wir eine Form des methodologischen Behaviorismus anwenden sollten, wenn wir darüber nachdenken, ob Maschinen einen moralischen Status haben könnten. Wenn sich eine Maschine wie ein Wesen verhält, von dem wir bereits glauben, dass es einen moralischen Status hat, dann sollten wir nach dieser Theorie daraus schließen, dass die Maschine auch einen moralischen Status hat. Danaher nennt diese Position „ethischen Behaviorismus".

---

21 Vgl. Turing, Alan: Essential Turing. Oxford 2004.

22 Vgl. Danaher, John: Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviorism. In: Science and Engineering Ethics 26/4 (2019), 2023–2049.

Danaher argumentiert, dass es bei der Behandlung von Maschinen nicht darauf ankommt, was „im Inneren" der Maschine vor sich geht. Während also jemand wie Véliz sagen würde, dass es sehr wichtig ist, ob eine Maschine bei Bewusstsein ist, schlägt Danaher vor, dass es eher darauf ankommt, ob sie sich so verhält, als ob sie bei Bewusstsein wäre. Tatsächlich glaubt Danaher, dass wir nie wissen können, ob eine Maschine oder sogar ein Mensch wirklich bei Bewusstsein ist. Es könnte sein, dass alle anderen Menschen Zombies ohne bewusste Erfahrungen sind. Das Einzige, was wir tatsächlich bei anderen Menschen beobachten können, ist, wie sie sich verhalten. Und wir entscheiden, wie wir andere Menschen behandeln sollten auf der Grundlage ihres Verhaltens. Warum sollten wir bei Robotern andere Maßstäbe anlegen? Diese Frage steht im Mittelpunkt von Danahers Argumentation.

Diese Argumentation ist aber nicht unproblematisch. Ein erstes Problem mit der ethisch-behavioristischen Perspektive ist, dass die meisten Menschen sich sehr darum kümmern, was in den Köpfen anderer Menschen wirklich vor sich geht. Weiter kümmern sich Menschen auch sehr darum, was Tiere fühlen könnten und was in deren Köpfen vorgeht. Wir reagieren zwar tatsächlich auf andere Menschen aufgrund ihres Verhaltens. Aber wir tun dies, weil wir ihr Verhalten als Hinweis darauf nehmen, was sie fühlen oder denken. Und ob jemand tatsächlich Freude oder Leiden empfindet, eine Emotion hat, bestimmte Gedanken denkt und so weiter, wird als ethisch sehr wichtig angesehen. Wie sich Menschen verhalten, ist relevant – nicht an sich –, sondern weil es ein Hinweis darauf ist, dass sie über Eigenschaften oder Fähigkeiten verfügen, die als ethisch an sich wichtig eingeschätzt werden.

Zweitens, wenn ein Mensch behauptet, dass er sich in einem bestimmten mentalen Zustand befindet, haben wir normalerweise die ethische Pflicht zu glauben, dass er sich wirklich in diesem mentalen Zustand befindet. Und dieser Mensch hat auch die ethische Pflicht, uns gegenüber ehrlich zu sein. Es könnte sein, dass wir Grund zu der Annahme haben, dass der andere versucht, uns zu täuschen. Unsere Pflicht, darauf zu vertrauen, dass andere ehrlich sind, ist keine absolute Pflicht. Und andere könnten eine Rechtfertigung dafür haben, jemanden zu täuschen. Ehrlichkeitspflichten sind strenge Pflichten, aber sie haben Grenzen. Dennoch haben wir tatsächlich die Pflicht, unseren Mitmenschen innerhalb gewisser Grenzen zu vertrauen. Und sie haben wie gesagt auch die Pflicht, ehrlich uns gegenüber zu sein, wiederum innerhalb gewisser Grenzen.

Im Gegensatz dazu haben wir bei Robotern und anderen Technologien nicht die Verpflichtung, darauf zu vertrauen, dass sie nicht versuchen, uns zu täuschen. Und die Roboter sind auch nicht verpflichtet, ehrlich uns gegenüber zu sein. Sie haben keine solche Verpflichtungen, weil sie im Allgemeinen überhaupt keine moralischen Verpflichtungen haben. (Vielleicht können wir uns zukünftige verantwortliche Roboter mit ethischen Pflichten vorstellen, aber das ist im Moment hauptsächlich ein Thema für die Science-Fiction.)

Deshalb können wir in folgender Weise gegen Danahers ethischen Behaviorismus argumentieren. Wir haben zwar eine weithin anerkannte ethische Verpflichtung, anderen Menschen zu vertrauen und sie haben die ethische Pflicht, zumindest innerhalb gewisser Grenzen ehrlich uns gegenüber zu sein. Aber wir haben keine ähnlichen weithin anerkannten ethischen Gründe, Robotern zu vertrauen, die sich wie Menschen oder Tiere verhalten. Und diese Roboter selbst haben keine ethischen Pflichten, ehrlich zu sein, weil sie als Roboter keine Pflichten haben. Das ist ein großes Problem für die grundlegende Idee des ethischen Behaviorismus.

6 Können Roboter moralisch relevante Eigenschaften oder Fähigkeiten repräsentieren oder symbolisieren?

Wenn wir darüber nachdenken, ob es möglich ist, sich gegenüber Robotern richtig oder falsch zu verhalten, stellt sich auch die Frage, ob Roboter moralisch relevante Eigenschaften oder Fähigkeiten *repräsentieren* oder *symbolisieren* können. Ebenfalls können wir fragen, ob die Art und Weise, wie wir mit Robotern umgehen, etwas moralisch Wichtiges repräsentieren oder symbolisieren kann. Wenn wir zum Beispiel gegenüber einem Roboter, der wie ein Mensch aussieht, „nett" sind, könnte das als eine Repräsentation von etwas Gutem angesehen werden? Könnte dieser Roboter sogar als Symbol für etwas mit Würde angesehen werden, nämlich einen Menschen? Oder wenn wir uns gegenüber einem Roboter „grausam" benehmen, könnte diese Interaktion als Repräsentation oder Symbol für etwas Schlechtes angesehen werden?

Ein Autor, der dieser Idee mehrere Artikel gewidmet hat, ist der australische Philosoph Robert Sparrow.[23] Sparrow verteidigt eine interessante Asymmetrie-These bezüglich der Frage, was unsere Interaktion mit Robotern darstellen könnte. Sparrow glaubt, dass unsere Interaktion mit Robotern etwas Schlechtes repräsentieren kann, aber er glaubt nicht, dass sie jemals etwas Gutes repräsentieren oder symbolisieren könnte.

Sparrow diskutiert sowohl humanoide Roboter als auch Roboter, die Tieren ähneln, wie z. B. Spot der Roboterhund. Wenn wir uns gegenüber einem Roboter scheinbar grausam verhalten, wirft das laut Sparrow ein schlechtes Licht auf uns und unseren moralischen Charakter. Ein Beispiel dafür wäre, einen Roboterhund zu treten. Ein weiteres Beispiel, das Sparrow ausführlich diskutiert, ist die Interaktion mit Sexrobotern. Sparrow ist der Ansicht, dass Sex mit Sexrobotern nur etwas Schlechtes symbolisieren kann: sexistische Einstellungen gegenüber Frauen oder sogar Vergewaltigung.

[23] Vgl. Sparrow, Robert: Virtue and Vice in Our Relationships with Robots: Is There an Asymmetry and How Might it be Explained? In: International Journal of Social Robotics 13/1 (2020), 23–29.

Warum? Weil Sexroboter negative Stereotypen über Frauen repräsentieren – sie repräsentieren Frauen als immer da, um die sexuellen Wünsche von Männern zu befriedigen – und weil Sexroboter nicht in der Lage sind, dem Sex so zuzustimmen wie ein Mensch. Dies alles repräsentiert laut Sparrow die höchstproblematische Idee, dass sexuelle Zustimmung nicht wichtig sei.

Daher gibt es laut Sparrow viele Möglichkeiten, uns schlecht oder unmoralisch zu verhalten, wenn wir mit Robotern interagieren. Und der Hauptgrund dafür ist, dass unsere Interaktion mit den Robotern etwas Schlechtes repräsentieren oder symbolisieren könnte. Gleichzeitig meint Sparrow, dass unsere Interaktion mit Robotern nichts Gutes repräsentieren oder symbolisieren könnte. Wenn wir „nett" zu Robotern sind, kann das kein gutes Licht auf uns werfen. Wenn Sie beispielsweise einen Roboterhund streicheln, anstatt ihn zu treten, dann kann das nichts Positives sein. Es wirft kein gutes Licht auf uns. Oder wenn wir uns gegenüber einem humanoiden Roboter freundlich verhalten, wirft das auch kein gutes Licht auf uns, so Sparrow.

Sparrow bietet zwei Argumente für seine Position an. Erstens denkt Sparrow, dass es im Allgemeinen einfach mehr Möglichkeiten zum Scheitern als zum Erfolg gibt. Es gibt viele Dinge, die im Leben schiefgehen können und die wir schlecht machen können. Aber es gibt einfach weniger Möglichkeiten, wie wir Dinge richtig und gut machen können. Das zweite Argument, das Sparrow anführt, ist, dass Grausamkeit keine besondere Weisheit oder ethische Einsicht fordert. Gut zu sein erfordert dagegen eine besondere Form von Weisheit und ethischer Einsicht. Es erfordert, dass wir wissen, wie wir andere gut behandeln können. Und da es Robotern nicht besser oder schlechter gehen kann, ist es laut Sparrow nicht möglich, ihr Wohlbefinden zu fördern.

Stimmt es, dass es nur möglich ist, Roboter auf eine Weise zu behandeln, die etwas Schlechtes repräsentiert, und dass es nicht möglich ist, mit Robotern auf eine Weise zu interagieren, die etwas Gutes repräsentiert? Betrachten wir eine andere mögliche Denkweise zu diesem Thema, und lassen Sie uns dies tun, indem wir zuerst das folgende Beispiel betrachten.

Ein Mann, der in Michigan lebt und sich „Davecat" nennt, hat den ungewöhnlichen Lebensstil, mit menschenähnlichen Puppen statt mit anderen Menschen zusammenzuleben.[24] Eine dieser Puppen, Sidore, ist seine Frau, sagt Davecat. Sie sind seit fast zwanzig Jahren verheiratet. In einem seiner vielen Interviews über sein Leben mit den Puppen sagt Davecat, dass er und seine Frau Sidore ihren Haushalt auch mit einigen anderen Puppen teilen, von denen eine „Muriel" heißt. Als er über diese Puppe spricht, sagt Davecat Folgendes: „I don't want to treat

[24] Vgl. Nyholm, Humans and Robots, 108.

her like a thing, and I won't."[25] Davecat sagt auch, dass er Wert darauflegt, seine Puppen niemals auf eine Weise zu behandeln, die etwas Schlechtes symbolisieren würde.

Dies wird einigen Leuten als ein seltsamer Lebensstil erscheinen. Aber Davecat könnte auch als jemand erscheinen, der eine scheinbar respektvolle allgemeine Haltung aufzeigt und als jemand der möchte, dass seine Art, mit Technologien – in diesem Fall eher mit Puppen als Robotern – zu interagieren, etwas Gutes repräsentiert. In den Bildern und Videoclips, die Davecats Medienauftritte begleiten, beschreibt er z. B., wie er gerne mit seinen Puppen auf dem Sofa sitzt und Bücher liest, gemeinsam Filme mit ihnen anschaut, sie umarmt und küsst, sie sanft behandelt und so weiter. Und wie oben erwähnt, behauptet Davecat, dass er immer darauf achtet, alles zu vermeiden, was als respektlos gegenüber anderen Personen angesehen werden kann.

Eine mögliche Denkweise über diese Art der Beziehung zu Puppen oder Robotern ist folgende: Wir könnten denken, dass die Zurückhaltung, die jemand wie Davecat zeigt, zumindest eine Form dessen ist, was man „minimale" oder „negative" Tugend nennen könnte. Das heißt, wenn wir darauf achten, Verhaltensweisen gegenüber humanoiden Puppen oder Robotern (oder anderen Technologien) zu vermeiden, die Anstoß erregen oder als respektlos erscheinen könnten, dann zeigen wir zumindest eine ethisch positive Form von Zurückhaltung. Wir zeigen, dass wir zumindest etwas vermeiden wollen, was als Symbol für etwas Negatives angesehen werden kann. Das kann möglicherweise als eine begrenzte Form von Tugend angesehen werden, und es könnte ein gutes Licht auf uns werfen. Wenn beispielsweise eine Person den Roboterhund Spot lieber nicht tritt, weil sie denkt, dass dies das Treten echter Hunde symbolisiert, was ihrer Meinung nach unmoralisch ist, dann könnte dies auch als eine Form von minimaler oder negativer Tugend angesehen werden.

Was wäre dann jedoch eine maximale oder positive Form von Tugend? Man könnte hier behaupten, dass die maximale Form von Tugend die Form des Verhaltens gegenüber anderen Menschen oder Tieren ist, bei der es nicht nur darum geht, etwas Gutes durch unsere Handlungen zu repräsentieren oder symbolisieren, sondern auch darum, zu versuchen, die moralisch relevanten Interessen von anderen direkt zu befördern und zu respektieren. Dies könnte so verstanden werden, dass es erforderlich ist, ein direktes Interesse an den Gefühlen, Gedanken, Wünschen, dem Willen usw. dieser Person oder dieses Tieres zu zeigen. Es könnte so verstanden werden, dass die Entität, mit der wir interagieren, jemand mit einem Verstand, mit Gefühlen, Gedanken und so weiter sein muss – also jemand, der wirkliche Interessen haben kann. Wenn ein Roboter keinen Verstand, keine Gefühle, keine Gedanken usw. hat, dann ist es nicht möglich, dem Roboter gegenüber maximal und positiv tugendhaft zu sein in dem

[25] Dieses Zitat stammt aus diesem Video: THE SKIN DEEP: Davecat, Married to a Doll. Online unter: https://www.youtube.com/watch?v=LiVgrHlXOwg&t=1s (Stand 23.10.2022).

Sinne, dass wir die Interessen der Roboter beobachten – weil der Roboter letztendlich keine moralisch relevanten Interessen hat.

7 Sollen wir überhaupt die Frage diskutieren, ob Roboter *moral patients* mit moralischem Status sein können?

Einige Forscher:innen, wie z. B. die Kognitionswissenschaftlerin Abeba Birhane und der Informatiker Jelle van Dijk, meinen, dass dieses Thema, ob Roboter mit moralischer Rücksichtnahme behandelt werden sollten, eine „Ablenkung" ist. Der Titel einer Birhane und van Dijks Veröffentlichungen zu diesem Thema fasst den Inhalt des Artikels sehr gut zusammen: „Robot Rights? Let's Talk about Human Welfare Instead".[26]

Die Idee ist also, dass, wenn wir uns Zeit nehmen, um darüber nachzudenken, ob Roboter mit moralischer Rücksicht behandelt werden sollten oder vielleicht sogar Rechte haben sollten, dann ist dies mit Opportunitätskosten verbunden. Wir hätten uns stattdessen auf die menschliche Ethik konzentrieren können und uns Fragen widmen können, wie man wichtige menschliche Interessen besser fördern könnte.

Joanna Bryson vertritt einen ähnlichen Standpunkt. Als Saudi-Arabien den Roboter Sophia zur Ehrenbürgerin erklärte, kommentierte Bryson dies in ihrer charakteristisch eindringlichen Sprache folgendermaßen: „[…] this is obvious bullsh*t. What is this about? It's about having a supposed equal you can turn on and off. How does it affect people if they think you can have a citizen that you can buy"?[27]

In ihrem oft zitierten Artikel mit dem Titel „Robots should be Slaves" argumentiert Bryson unter anderem, dass, wenn wir Zeit und Ressourcen verschwenden, um die Roboter gut zu behandeln, wir Zeit und Ressourcen dafür verschwenden, sicherzustellen, dass Menschen gut behandelt werden.

Eine radikal andere Sichtweise auf die Frage, ob es sinnvoll ist, darüber nachzudenken, ob Roboter *moral patients* sein können, findet man in den Schriften von Forschenden wie David Gunkel und dem Politikwissenschaftler Joshua Gellers. Sie argumentieren, dass Reflexion über den moralischen Umgang mit Robotern eine Möglichkeit sein kann, unseren intellektuellen Horizont zu erweitern. Es kann ein Weg sein, so schlagen sie vor, in unserem Denken weniger „US-zentriert" oder „eurozentrisch" zu werden.

---

[26] Birhane, Abebe/van Dijk, Jelle: Robot Rights? Let's Talk about Human Welfare Instead. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (2020), 207–213. DOI: https://doi.org/10.1145/3375627.3375855.

[27] Vincent, James: Pretending to give a Robot Citizenship helps No One. Online unter: https://www.theverge.com/2017/10/30/16552006/robot-rights-citizenship-saudi-arabia-sophia (Stand: 25.8.2022).

Gunkel argumentiert zum Beispiel, dass das Studium japanischer Denkweisen über Mensch-Roboter-Interaktion für diejenigen aufschlussreich sein kann, die in einem westlichen Kontext mit Theologie und Ethik arbeiten.[28] Gunkel interessiert sich insbesondere für animistische Aspekte einiger japanischer Denkweisen über Mensch-Roboter-Interaktionen. Er argumentiert, dass diese Perspektive es uns ermöglichen kann, darüber nachzudenken, wie unsere Beziehungen zu den von uns verwendeten Technologien anders sein könnten als diejenige Sichtweise, an die wir vielleicht gewöhnt sind, dass alle Technologien in erster Linie bloße Werkzeuge sind. In Japan, so Gunkel, gebe es eine größere Offenheit dafür, nicht-instrumentelle Beziehungen zu Technologien einzunehmen und einige Roboter als eine Form von Personen zu betrachten.

Gellers schlägt weiter vor, dass das Studium der Art und Weise, wie einige indigene Völker mit der nichtmenschlichen Natur interagieren, uns dabei helfen kann, unser Repertoire an Möglichkeiten zu erweitern, wie wir uns die Beziehungen vorstellen können, die wir mit Robotern haben können.[29] In einigen kulturellen Kontexten werde die Idee, einem Fluss oder einem anderen Teil der Natur Rechte zu verleihen, sehr ernst genommen, stellt Gellers fest. Es gibt Dinge, die wir über den Umgang mit Robotern lernen können, wenn wir uns im Westen besser damit vertraut machen, wie einige indigene Völker mit der Natur umgehen, argumentiert Gellers.

In ähnlicher Weise argumentiert Christopher Wareham, ein Philosoph und Bioethiker, der lange in Südafrika gearbeitet hat und viel darüber geschrieben hat, wie afrikanische Perspektiven in der Ethik repräsentiert werden könnten.[30] Er vertritt die Ansicht, dass die Idee aus der Ubuntu-Ethik, wonach „a person is a person through other persons"[31], höchst relevant ist, wenn wir darüber nachdenken, ob Roboter Mitglieder der moralischen Gemeinschaft werden könnten und mit moralischer Rücksichtnahme behandelt werden sollten. Laut Wareham könnten Roboter durch andere Personen zu Personen werden, indem sie in die moralische Gemeinschaft aufgenommen und mit moralischer Rücksicht behandelt werden. Die Idee dabei wäre, weniger die Roboter selbst und ihre Eigenschaften zu betrachten, sondern stattdessen zu fragen, ob sie akzeptiert und in moralische Gemeinschaften aufgenommen werden könnten.

Wir sehen also, dass es radikal unterschiedliche Auffassungen darüber gibt, ob wir überhaupt die Frage des moralischen Status von (humanoiden) Robotern innerhalb der Technologieethikforschung diskutieren sollten. Einige stehen diesem Thema sehr skeptisch gegenüber und argumentieren, dass es eine „Ablenkung" sei, die die Aufmerksamkeit von dringenderen

---

28 Vgl. Gunkel, Robot Rights, 75.

29 Vgl. Gellers, Joshua: Rights for Robots: Artificial Intelligence, Animal and Environmental Law. London 2020.

30 Vgl. Warham, C. S. (2020): Artificial Intelligence and African Conceptions of Personhood. In: Ethics and Information Technology 23/2(2020), 127–136.

31 Tutu, Desmond: No Future without Forgiveness. New York 1999, 35.

Themen wie etwa Menschenrechtsfragen ablenke. Andere meinen, dass das Nachdenken über die Frage, ob Roboter *moral patients* sein könnten, ein Weg sein kann, uns für einen breiteren Horizont von Denkweisen zu öffnen als diejenigen, die üblicherweise in der Ethikforschung verwendet werden.

8 Abschlusskommentare

Oben habe ich verschiedene Beiträge der gegenwärtigen philosophischen Debatte zu der Frage diskutiert, ob (humanoide) Roboter mit moralischer Rücksichtnahme behandelt werden sollten. Ich habe vorgeschlagen, dass es hilfreich ist, diese Debatte mit Hilfe der drei Fragen, die ich formuliert habe, zu ergänzen und kritisch zu analysieren. Ich habe auch ein paar Argumente für und gegen diese Diskussion in Betracht gezogen: d. h., sollten wir überhaupt diese Frage des moralischen Status von Robotern diskutieren oder ist das eine Ablenkung von wichtigeren Themen? Eine Frage, die ich oben ebenso nicht diskutiert habe, die aber sehr relevant für dieses Buch ist, ist, ob – und wenn ja, warum – diese Diskussion über den moralischen Status von (humanoiden) Robotern relevant und vielleicht sogar hochinteressant für die gegenwärtige Theologie ist. Ob diese Diskussion in der Tat relevant und/oder interessant für die gegenwärtige Theologie ist, sollten hauptsächlich Theolog:innen beantworten – mein eigener Hintergrund ist in Philosophie – aber ich werde zum Abschluss dennoch drei kurze Vorschläge zu diesem Thema machen.

Erstens schlage ich vor, dass die Diskussion darüber, ob humanoide Roboter als mögliche *moral patients* zu verstehen sind, durchaus relevant für die Theologie sein könnte aus folgendem Grund. Das Thema von unterschiedlichen Menschenbildern, die eine sehr wichtige Rolle in gegenwärtigen theologischen Diskussionen spielen, ist relevant in Bezug auf die Frage, ob es möglich ist, menschenähnliche Roboter zu bauen, die wir mit moralischer Rücksichtnahme behandeln sollen. Zum Vorschein kommen dringende Fragen über unterschiedliche Ideen darüber, was es heißt, Mensch zu sein.

Zweitens meine ich, dass es interessant ist zu fragen, inwiefern wir die menschliche Erschaffung von menschenähnlichen Robotern (die vielleicht mit moralischer Rücksichtnahme behandelt werden sollten) mit der traditionellen theologischen Idee der Erschaffung des Menschen nach dem Ebenbild Gottes vergleichen können. Diese Frage hat, u. a., der amerikanische Theologe Joshua K. Smith auf interessante Art und Weise in seinem Buch *Robotic Persons*[32] diskutiert. In seiner Diskussion wird es klar, dass die philosophische Perspektive, die ich oben diskutiert habe, und die theologischen Fragen, die Smith diskutiert, wichtige Parallelen haben.

[32] Smith, Joshua K. Robotic Persons: Our Future with Social Robots. Bloomington: 2021.

Drittens ist es auch denkbar, dass Roboter – inklusive humanoider Roboter – in religiösen Kontexten in der Zukunft eventuell mehr und mehr eingesetzt werden. Anna Puzio nennt dies „religiöse Robotik."[33] Werden Roboter, die in einem religiösen Kontext eingesetzt werden, einen besonderen moralischen Status haben, der über den moralischen Status von anderen (humanoiden) Robotern in anderen Kontexten (wie z. B. Therapieroboter oder Sexroboter) hinausgeht? Oder werden alle humanoiden Roboter den gleichen moralischen Status haben, unabhängig davon, in welchen Kontexten sie eingesetzt werden?

Dies sind drei Punkte oder Themen, die für mich bestätigen, dass die Frage des moralischen Status von Robotern eine relevante Frage ist – und vielleicht sogar für die gegenwärtige Theologie interessant sein könnte. Es gibt noch nicht viel theologische Literatur zum Thema, wie wir menschenähnliche Roboter behandeln sollten. Deshalb ist es für Theolog:innen, die sich für diese Frage interessieren, möglicherweise hilfreich, einen Überblick von und Eintritt in vorhandene philosophischen Diskussionen in der Literatur zu diesem Thema zu erlangen. Meiner Meinung nach gibt es auf jeden Fall viel Raum für fruchtbare Diskussionen zwischen Theolog:innen und Philosoph:innen, die sich für dieses spannende neue Thema interessieren.[34]

Literaturverzeichnis

Asada, Minoru: Artificial Pain May Induce Empathy, Morality, and Ethics in the Conscious Mind of Robots. In: Philosophies 4/3 (2019), 38.

Birhane, Abebe/van Dijk, Jelle: Robot Rights? Let's Talk about Human Welfare Instead. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (2020), 207–213. DOI: https://doi.org/10.1145/3375627.3375855.

Bryson, Joanna: Robots Should Be Slaves. In: Wilks, Yorick (Hg.): Close Engagements with Artificial Companions, Amsterdam: 2010, 63–74.

Bryson, Joanna: A Role for Consciousness in Action Selection. In: International Journal of Machine Consciousness 4/2 (2012), 471–482.

CNET HIGHLIGHTS: Elon Musk REVEALS Tesla Bot (full presentation). Online unter: https://www.youtube.com/watch?v=HUP6Z5voiS8 (Stand: 23.10.2022).

Danaher, John: Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviorism. In: Science and Engineering Ethics 26/4 (2019), 2023–2049.

Gellers, Joshua: Rights for Robots: Artificial Intelligence, Animal and Environmental Law. London 2020.

[33] Siehe Puzios Beitrag in diesem Band: Puzio, Anna: Robot Theology: On Theological Engagement with Robotics and Religious Robots.

[34] Ich danke den Gutachter:innen dieses Kapitels und ich danke auch den Organisator:innen und Teilnehmer:innen des Workshops „Alexa, wie hast du's mit der Religion?" 2021 des Netzwerks für Theologie und KI, wo ich dieses Material vorstellte.

Gunkel, David: Robot Rights, Cambridge 2018.

Hanson, David: Humanizing Robots: How Making Humanoids Can Make Us More Human. Dallas 2017.

Loh, Janina: Roboterethik: Eine Einführung. Frankfurt a. M. 2019.

Metzinger, Thomas: Two Principles for Robot Ethics. In: Hilgendorf, Eric/ Günther, Jan-Philipp (Hg.): Robotik und Gesetzgebung, Baden-Baden 2013, 236–302.

Mori, Masahiro: The Uncanny Valley. In: IEEE Robotics and Automation Magazine 19/2, 98–100.

Nyholm, Sven: Humans and Robots: Ethics, Agency, and Anthropomorphism. London 2020.

Parke, Phoebe: Is it Cruel to Kick a Robot Dog? Online unter: https://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/index.html (Stand: 25.8.2022).

Schwitzgebel, Eric/Garza, Mara: A Defense of the Rights of Artificial Intelligences. In: Midwest Studies in Philosophy 39/1 (2015), 98–119.

Smith, Joshua K. Robotic Persons: Our Future with Social Robots. Bloomington: 2021.

Sparrow, Robert: Virtue and Vice in Our Relationships with Robots: Is There an Asymmetry and How Might it be Explained? In: International Journal of Social Robotics 13/1 (2020), 23–29.

The Skin Deep: Davecat, Married to a Doll. Online unter: https://www.youtube.com/watch?v=LiVgrHlXOwg&t=1s (Stand 23.10.2022).

Turing, Alan: Can Digital Computers Think? TS with AMS Annotations of a Talk Broadcast on BBC Third Programme 15 May 1951. Online unter: https://turingarchive.kings.cam.ac.uk/publications-lectures-and-talks-amtb/amt-b-5 (Stand: 25.08.2022).

Turing, Alan: Essential Turing. Oxford 2004.

Tutu, Desmond: No Future without Forgiveness. New York 1999.

Veliz, Carissa: Moral Zombies: Algorithms are not Moral Agents In: AI & Society 36/2 (2021), 487–497.

Vincent, James: Pretending to give a Robot Citizenship helps No One. Online unter: https://www.theverge.com/2017/10/30/16552006/robot-rights-citizenship-saudi-arabia-sophia (Stand: 25.8.2022).

Wareham, C. S. (2020): Artificial Intelligence and African Conceptions of Personhood. In: Ethics and Information Technology 23/2(2020), 127–136.

II Transformation der Religion

Roboter und Religion

Robot Theology

On Theological Engagement with Robotics and Religious Robots

Anna Puzio

Abstract

As robots increasingly find their way into the various spheres of human life, the question of religious robots becomes relevant. This article examines from a Catholic-Christian theological perspective whether robots can be used for religious purposes, and it asks how this may be done and what issues are important to consider. In addition, the study contributes to research on the theological engagement with robotics. It is argued that the use of religious robotics differs significantly depending on the specific religion. From a Christian perspective, the use of religious robotics is fundamentally plausible, especially since a wide variety of entities are used as religious media or representations of the divine. However, the use of religious robotics will ultimately be decided by different concepts (e.g. human, life), religious doctrines and culturally transmitted and subjective attitudes. This article places particular focus on the design of religious robotics. It becomes clear that the reasons for accepting religious technology do not lie in the technology itself but in phenomenological preferences and various time- and culture-dependent ideas and concepts. It is likely that as robotics rapidly advances and our relationship with it develops, the use of religious robots also will change.

1 Introduction

In many religions, light and candles play an important role. When one lights a candle (e.g. a votive candle or prayer candle in the church), it is associated with specific religious feelings, experiences and practices. It gives light and warmth, stands for the 'light in the world', hope and the 'light in the darkness'. The candle has a special smell, it flickers, and the wax melts and drips. It

has an effect in the room in which it is placed, and it works with the other candles with which it can be lit. With the lighting of the candle, petitions are formulated, prayers are recited, the deceased and unwell are remembered, and gratitude is expressed. In recent years, wax candles have been replaced in some churches by electric votive candles with LED lamps. By inserting money, one of the many candles is automatically lit. Although electric votive lights have practical advantages, many people lose some dimensions of religious experience and practice in the process. Switching on the electric lamp by inserting money is a simple automation that seems unspectacular and lacks meaning.[1] So, does this mean the end of technology for religious purposes? What about the many beneficial technologies and the great technological advances, e.g. in robotics? Would this mean religious robots will never become popular?

The question of the use of robots arises for theology from the many advances in robotics and their widening use in society (e.g. social robotics in the health sector). Since robots are increasingly present in the various spheres of human life, it is timely to ask questions about religious robots, i.e. robots used for religious purposes. This paper examines whether robots can be used for religious purposes from a Roman Catholic theological perspective, and it asks what is at stake and how they will be used. In addition to the Christian theological orientation, interreligious considerations are explored. Section 2 reflects on the connection between robotics and religion, the relevance and significance of religion for robotics, and theological engagement with robotics. In Section 3, the study focuses on religious robotics. Section 3.1 introduces religious robots and presents the relationships of different religions to them. Sections 3.2 and 3.3 take a closer look at the design and functioning of religious robots. A wide range of possibilities is discussed. What does a robot have to be like to facilitate religious experience? The various steps of the investigation provide insights into philosophical and religious concepts, religious teachings and attitudes that play a role in religious robotics. Finally, Section 4 offers some conclusions and considers the outlook for the theological study of religious robotics.

This paper contributes to research on theological engagement with robotics, referred to here as 'robot theology',[2] which is still in its infancy in international research. It identifies pioneering perspectives, tasks, and questions to be addressed for the theological study of robotics. In this way, the paper contributes to the theological debate on technology. This paper looks beyond references to hubris ('playing God'), the overemphasis on danger or the replacement of humans by omnipotent, powerful artificial intelligence (AI) or robots who have a religious faith and will be our saviours. Instead, a shift in the questions is necessary for

[1] For the candle example see Löffler, Diana/Hurtienne, Jörn/Nord, Ilona: Blessing Robot BlessU2. In: International Journal of Social Robotics 13/4 (2021), 569–586, here 569f. DOI: 10.1007/s12369-019-00558-3.

[2] The term 'Robot Theology' already appears in: Smith, Joshua: Robot theology. Eugene, Oregon 2022.

theological debates about technology, and this article shows how theological research can ask more pertinent questions.

2 Robot Theology: Religion and Robotics

Robot theology is the theological study of robotics. It approaches the diversity of robots scientifically, including service, combat, sex, social and religious robots. Robots may be analysed from various perspectives, including the ethical, moral-theological, anthropological, metaphysical, biblical, pastoral-theological, pedagogical and didactical. Considerations range from the philosophy of religion to canon law. Diverse topics can be covered, and robot theology addresses questions of the mind–body relationship. It also includes biblical investigations of the relationships with non-human entities, ethical questions of the design of social robots, and pastoral-theological and canonical legal frameworks for religious robots.

For various reasons, theology is especially suitable for dealing with robotics. For example, theology has a rich supply of examples of specific forms of relationship with non-human entities (e.g. in the Bible). Theology includes the ethics of dealing with the Other (e.g. charity and special consideration for the alien and marginalised) and addresses social and spiritual needs. These factors are highly relevant in social robotics and religious robotics. Moreover, technologization raises many anthropological and ethical questions about the image of human beings and the world. The profound technological progress shakes up many traditional views and ideas about humanity, technology, metaphysics and the distinctions between nature and culture and nature and technology. A need for orientation arises in society concerning questions such as what distinguishes humans from machines, questions of justice and the good life, and the ethical application of robots. Theology offers a broad repertoire of answers to anthropological and ethical questions about understanding human beings and the world. However, these must be reflected upon anew in the context of technological developments. In addition, the acceptance and handling of robotics differ significantly between countries and cultures. Religions as cultural actors have shaped and continue to shape today's cultural and societal views on robotics.[3] From a theological perspective, many other benefits and advantages of dealing with robotics can be explored. The diversity of disciplines within theology makes it possible for theology to explore robotics broadly (e.g. practically, historically, biblically, ethically and philosophically).

[3] See Trovato, Gabriele et al.: Religion and Robots. In: International Journal of Social Robotics 13/4 (2021), 539–556, here 540. DOI: 10.1007/s12369-019-00553-8.

3 Religious Robots

3.1 Robot theologies: Religious robots in different religions

Robots have been used for various religious purposes. For example, BlessU-2 is a German robot that plays blessings in different languages.[4] SanTO (the Sanctified Theomorphic Operator)[5] takes the appearance of a Christian Catholic saint who quotes sacred texts and accompanies the faithful in prayer. As a companion, it also has psychological functions and contributes to well-being (e.g. for the elderly).[6] Mindar is a robot priest from Japan who represents the Buddhist teacher Kannon Bodhisattva and celebrates the Zen ceremony in the temple.[7] The monk robot Xi'aner follows visitors around the temple, responds to their inquiries about Buddhism and plays Buddhist music. It is also available as a chatbot with which you can communicate over online messenger services. Xi'aner is intended to spread Buddhism in China.[8] As such, Xi'aner is not regarded as threatening religious teachings but rather as contributing to religious dissemination.[9] In Japan, the humanoid robot Pepper is used for Buddhist funerals because it is cheaper than the human priest. It also broadcasts the ceremony over the internet for those who are unable to attend.[10] Religious robots are not limited to religious purposes in the narrow sense, such as using them in church and for religious ceremonies. So-called social robots with religious characteristics can take on other tasks.

The significance of religion for robotics is especially evident in the example of the robot DarumaTO-2, which is used in health and elderly care. It does not fulfil any religious purpose, but it is familiar to older people because of its religious appearance, so they feel comfortable with it. DarumaTO-2 is designed like the Daruma dolls, which depict the Buddhist monk Bodhidharma and are supposed to bring good luck as talismans in Japan and China.[11] Trovato,

---

4 See Löffler et al.: Blessing Robot BlessU2, 575.

5 See Trovato, Gabriele et al.: Communicating with SanTO. 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) 2019, 1–6. DOI:10.1109/RO-MAN46459.2019.8956250.

6 See Löffler et al.: Blessing Robot BlessU2, 573; Trovato et al.: Religion and Robots, 545.

7 See Smith: Robot theology, ch. 7; Klein, Mechthild: E-Priester im Einsatz. In: Deutschlandfunk 25.09.2019. Online at: https://www.deutschlandfunk.de/religion-in-japan-e-priester-im-einsatz-100.html (as of: 18.09.22).

8 See Trovato et al.: Religion and Robots, 544; Löffler et al.: Blessing Robot BlessU2, 573.

9 See Löffler et al.: Blessing Robot BlessU2, 573.

10 See ibid.

11 See Trovato et al.: Religion and Robots, 545, 552; Trovato, Gabriele et al.: The creation of DarumaTO. Proceedings of the 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 606–611, here 607.

for instance, identifies the connection of religion with mental health and well-being.[12] These robots vary in their movements, size, facial expressions, voice, speech functions, display of emotions, light functions, and whether they have integrated screens, are 'anthropomorphic', 'biomorphic' or 'idiomorphic' (i.e. functional) in design. Besides robots, religious technologies include timers for the Jewish Sabbath, an electronic Quran and a Christian electronic rosary.[13]

The acceptance of robots varies greatly between cultures, countries and religions. Differences include how they are handled, the purposes for which they are used and their position and importance in religious life. Therefore, robot theology exists only in the plural: robot theologies. Currently, Christianity remains dominated by technological scepticism and a rejection of robotics. However, this position has changed over the years. For example, in the medieval and early modern periods, automata were promoted by the church to astonish people with their apparent magical abilities.[14] A famous example was the clockwork monk made for King Philip II of Spain in the 16th century. The monk occupied an intermediate position between the living and the non-living. In the course of time, the understanding of aliveness changed.[15] In Islam, religious robotics is confronted by aniconism, which can result in a low acceptance of religious robotics.[16] In Judaism, various activities are forbidden on the Sabbath because believers are supposed to rest, so technology is used to automate some tasks (e.g. automatically turning the lights on and off). Here, automated technologies can facilitate everyday activities and thus deepen the experience of the Sabbath.[17]

Hinduism, Taoism, Confucianism, Shintoism and Buddhism are more open and accepting of religious robotics than the monotheistic religions. Religious robots support rituals, spread religion and generate enthusiasm for it.[18] In Hinduism, this is facilitated by worshipping multiple deities or worshipping the deity in different forms and representing reincarnation and the sacred character of animals and other beings.[19] In Buddhism, the attribution of Buddhahood to robots is discussed, and in Shintoism, inanimate objects like robots can be sacred and have spirits.[20] Thus, there are significantly different attitudes to religious robotics across reli-

12 See Trovato et al.: DarumaTO, 607.

13 Trovato et al.: Religion and Robots, 548, 544f. See footnotes 33 and 35.

14 See ibid, 542.

15 See ibid.; Löffler et al.: Blessing Robot BlessU2, 572 f.

16 See Trovato et al.: Religion and Robots, 542 f.

17 See ibid, 543 f.

18 See ibid, 543 f., 547.

19 See ibid, 543.

20 See ibid, 544; Geraci, Robert M.: Robotics and Religion. In: Runehov, Anne/Oviedo, Lluis (eds.): Encyclopedia of sciences and religions. Dordrecht 2013, 2067–2072, here 2070.

gions and between monotheistic and polytheistic faiths, depending on their positions on divine representation, the environment and their handling of images and idols, among others.[21]

From a Christian perspective, the Incarnation may play an important role in establishing religious robotics. According to Christian doctrine, God became a human being in a body with flesh and blood.[22] The Word became flesh – but what about robots, and what is flesh? Technologization has blurred the boundary between the body and technology. On a biological level, the body's boundaries are not as clearly defined as is often assumed.[23] Can technology also be understood as part of the human body? The theological understanding of the body also has diverse potential for interpretation and could be further transformed into a more open, dynamic understanding with increasing technologization.[24] Specific religious teachings are important for establishing religious robotics and understanding and transforming concepts such as the human, the body, life and creation.

Diana Löffler et al. show that Japan, in particular, has animistic views because of the extent of Buddhism and Shintoism. In these religions, objects are alive and have a soul or a spiritual essence, so there is more openness towards robots. This contrasts with the discussion of the 'Frankenstein complex' (the fear of robots) in Western countries.[25] Among other factors, this fear is attributed to the impact of the industrial revolution, including the many social challenges (e.g. poverty, unemployment, hunger and suffering) produced by technological development.[26] In addition, people in Japan are very familiar with cartoon robots such as Tetsujin 28-go/Gigantor, Doraemon and especially Tetsuwan Atomu (Astro Boy). These robots play an important role for the Japanese, and they have been with them since childhood. These robots have close relationships, emotional bonds and families of their own.[27] Compared with many robots in western science fiction, they do not pose a threat to humans but are friends who save humans from danger.[28] The influence of cartoon robots and the ideas of Shintoism and

---

21 See Trovato et al.: Religion and Robots, 547.

22 I am grateful to Noreen Herzfeld for discussions on this topic.

23 See Barad, Karen: Agentieller Realismus. Berlin 2012, ch. 'Körpergrenzen'.

24 See Puzio, Anna: Zeig mir deine Technik und ich sag dir, wer du bist? – Was Technikanthropologie ist und warum wir sie dringend brauchen. In: Diebel-Fischer, Hermann/Kunkel, Nicole/Zeyher-Quattlender, Julian (eds.): Mensch und Maschine im Zeitalter 'Künstlicher Intelligenz'. 2023; Puzio, Anna: Über-Menschen. Philosophische Auseinandersetzung mit der Anthropologie des Transhumanismus. Bielefeld 2022, part III; Puzio, Anna: Digital and Technological Identities – In Whose Image? In: Cursor (2021). Online at: https://cursor.pubpub.org/pub/y2bcesx4 (as of: 14.03.22).

25 A term coined by Isaac Asimov in his robot novels.

26 See Löffler et al.: Blessing Robot BlessU2, 574.

27 See Robertson, Jennifer: Robo sapiens japanicus. Oakland, California 2018, 1 f.

28 See Geraci: Robotics and Religion, 2070.

Buddhism may explain the more open, approachable and welcoming attitude towards robots in Japan:

> Broadly speaking, his [sc. the cartoon robot's] influence, combined with the persistence of Shinto and Buddhist ideas in the largely agnostic country, promotes a spirit of cooperation and affection between human beings and robots. As a result, many Japanese eagerly look forward to the introduction of functioning, intelligent humanoid robots.[29]

Therefore, the use of robots for religious purposes depends strongly on the respective religion and its ideas, beliefs, teachings and concepts. In addition, it is closely interwoven with other cultural and societal aspects. Conversely, the acceptance of robots in society can also depend on religion, as the example of the Daruma dolls demonstrates. In Shintoism, robots can be sacred and have spirit; they are 'living things', and this implies a certain nature–culture relationship:[30]

> [N]ature is not external to culture and society but is an immanent component or symbiotic constituent of them; moreover, the reality of nature is contingent upon human artifice and mediation […]. Robots are 'living things' in the Shinto universe. While they may not claim to be animists, many Japanese roboticists nevertheless draw from this synergistic nature–culture 'platform' in advocating not only the interchangeability of robots and humans in everyday life but also their mutual enhancement and even mutual constituency.[31]

Life and aliveness, the distinction between animate and inanimate, nature and culture, the relationship to non-human entities and objects, and the relationship to technology are religiously and, more generally, culturally negotiated concepts. The attitude taken towards religious robots is related to these concepts, whether robots are seen as a threat to or promoter of religious purposes and whether they can serve as a divine representation and medium. Moreover, these negotiations and attitudes change over time. Religious robotics is thus always subject to time- and culture-dependent negotiations. Therefore, understandings and attitudes towards religious robotics will continue to change especially during times of technological upheaval and great technological advances.

---

29 Ibid.

30 Robertson: Robo sapiens japanicus, 15; Geraci: Robotics and Religion, 2070.

31 Robertson: Robo sapiens japanicus, 15.

3.2 Designing religious robots: How does religiousness get into the robot, or: Is it all a question of design?

As previously discussed, the use of religious robotics depends on specific ideas, concepts, and cultural and religious contexts and is already established in some religions and places. However, can robots as technological objects (i.e. their shape, form, design, material, made-ness, movements and control) fulfil religious functions and represent the divine?

Looking at the different forms of divine representations, it is striking that anything can become a divine representation: from people and animals to objects, hybrid religious beings, places, plants and other natural elements. In these representations, the divine can, in turn, appear in anthropomorphic objects and anthropomorphised animals and nature and as sacred objects in zoomorphic or physiomorphic form. Therefore, there are numerous mixed and intermediate forms.[32] This diversity of divine representation applies not only between religions, but also within religions. For example, in Catholicism, there are holy people, scriptures, places, buildings, mountains, stones, relics and trees (and St. Barbara's branches, palm branches and fir trees are used for religious customs). Moreover, great importance is bestowed on different kinds of animals, such as doves and sacrificial animals in the Bible. Images and religious objects are used in worship (e.g. tabernacle, chalice and paten, Easter candle, eternal light, altar bells). Natural phenomena such as fire and light play important roles, and almost everything can be blessed (including weapons). Therefore, ontologically, robots are compatible with Catholic theology. In addition, robots can be 'anthropomorphic', 'zoomorphic', 'biomorphic' and 'physimorphic',[33] 'functional', etc.[34]

Ilona Nord and Charles Ess criticise these categorisations into anthropomorphic, biomorphic, physimorphic, functional, etc. shape. These categorisations presuppose clear species boundaries, which are then applied to robots. To what extent can a clear distinction be made between anthropomorphic, biomorphic, physimorphic and functional design? Further research must examine in which cases such categorisations are useful and explore alternative taxonomies.[35]

Furthermore, the use of religious mediums cannot be rejected or opposed from a Catholic perspective. Religion always depends on a medium and is always mediated, and a wide variety of

---

[32] See Trovato et al.: Religion and Robots, 546 f.

[33] Trovato et al. refer to 'zoomorphic' as the shape of an animal, to 'biomorphic' as the shape of a living being and to 'physimorphic' as something that resembles nature.

[34] See Trovato et al.: Religion and Robots, 547 f.

[35] See Nord, Ilona/Ess, Charles: Robotik in der christlichen Religionspraxis. Anschlussüberlegungen an erste Experimente in diesem Feld. In: Merle, Kristin/Nord, Ilona (ed.): Mediatisierung religiöser Kultur. Praktisch-theologische Standortbestimmungen im interdisziplinären Kontext. Leipzig 2022, 227–258, here 249, 256f.

mediums have been used. Sacred texts, books and images function as mediums; priests and angels are mediators between the divine and the earthly. Also, for religious communication within the religious community, media technology, such as broadcasting, television, film, internet and social media, is used. However, adaptation to new media technologies has been slow, and only those media that have become widespread and proven are usually accepted in religious communication.[36]

If there is no argument against religious robotics regarding the entity or its medial character, the question remains how religious robots may be designed. How does holiness or religiousness get into the robot? Trovato et al. argue that robots in theomorphic design (i.e. robots in the 'shape of something divine')[37] are advantageous in many religious areas of application – namely, in terms of 'acceptance' by the user, 'comfort' (i.e. that the user feels more protected by the religious robot), and 'regard' (i.e. an ordinary object is held in less esteem than a religious looking one). The theomorphic appearance can also make religious people feel more comfortable and help those more alien to religion feel more comfortable with religious traditions. For example, the elderly might deal more easily with a religious robot and feel more comfortable with it because they have a strong emotional attachment to religion, even though they are less familiar with technology.[38]

In their insightful critical examination of Trovato's work, Nord and Ess problematise this category of 'theomorphic robots'. It remains unclear what exactly is meant by theomorphic design and the theomorphic shape seems provocative with regard to considerations of the 'image of God'. Moreover, they identify many dualisms in Trovato's approach, such as the dualism of God and the world, which are theologically untenable.[39] Thus, it becomes evident that the design question in the religious context is confronted with very specific challenges that differ from the other discussions in robotics and call for theological research.

Other factors in the design of religious robots include size, voice, face and facial expressions, gestures, graphics and screens.[40] Besides shape, there are many aspects of design and practical implementation to consider: robot-likeness, name, (religious) symbols, materials, movement, user control and customisation, light, touch and location.[41] In general, anthropomorphic traits or highly complexity in robots, such as movement and communication, might not always be beneficial, especially in religious robotics even if they facilitate the interface between humans and robots. A robot that does not communicate like a human and moves in less than a human-like way affects the user's expectations and makes room for interpretation of how the user makes

36 See Löffler et al.: Blessing Robot BlessU2, 571.

37 Trovato et al.: Religion and Robots, 539, 541.

38 Ibid., 549.

39 See Nord/ess: Robotik in der christlichen Religionspraxis, 245–250, 256f.

40 See Löffler et al.: Blessing Robot BlessU2, 580, 582.

41 See Trovato et al.: Religion and Robots, 550–552; Löffler et al.: Blessing Robot BlessU2, 580.

sense of the robot's behaviour. Errors or lack of reactions on the part of the robot could also be reinterpreted by the human, whereas the errors and malfunctions of a highly complex robot would contradict the expected infallibility of the divine.[42] Further investigation is needed to identify which characteristics and which behaviours of the robot show 'thingness' and the 'robotic element' well, and when it is better to hide them. A related issue is user control of the robot by buttons, switches, keyboards, and touch screens, as well as the robot's reliance on power cords or batteries.[43] Suitable materials are those that are considered valuable, linked to the divine or perceived as natural.[44] Names and (religious) symbols can facilitate embedding the robot into the religious context. Where the robot is found also is important. The location of the robot (e.g. in a museum as an exhibit or a designated place in a church) can influence people's perception of it.[45]

In addition to the robot's design, it is evident that the context and the practices in which it is embedded are decisive, including the place, the processes and rites in which it is integrated and how religious authorities deal with it. A blessing of the religious robot or other practices by a religious authority also could reinforce the integration of the robot into the religious context.[46] The design guidelines outlined above remain vague. They have emerged from studies of the relatively few religious robots. Moreover, even though there will be more advances in research in the future, no guidance for developing religious robots will guarantee successful human-robot-interaction.[47] As the discussion below illustrates, the use of religious robotics is a highly complex, relational and subjective process between the user and the robot.[48] In this process, the space for religious experience might be made possible through construction and design. Among other issues, it is about gaps and the opening of spaces:

3.3 The puzzle of the candle: Space-opening robots and robots with gaps

The example of the wax candle and the electric candle in Section 1 demonstrates what matters in religious robotics and what successful religious robotics must look like. Superficially, one

---

42 See Trovato et al.: Religion and Robots, 550 f.; Löffler et al.: Blessing Robot BlessU2, 574.

43 See Trovato et al.: Religion and Robots, 550 f.; Löffler et al.: Blessing Robot BlessU2, 578, 583.

44 See Löffler et al.: Blessing Robot BlessU2, 583; Trovato et al.: Religion and Robots, 551.

45 See Trovato et al.: Religion and Robots, 550 f.

46 See ibid., 551.

47 Another important question is how to define successful interaction in a religious context. When is this interaction considered successful? See Nord/Ess: Robotik in der christlichen Religionspraxis, 237.

48 See Daelemans, Bert: The Need for Sacred Emptiness. In: Religions 13/6 (2022), 515, 1–15, here 14. DOI: 10.3390/rel13060515. – Of course, dealing with religious robotics also involves looking at the relationship with God.

could argue that the wax candle is natural and the electric candle is artificial. The immediate critique is that technology makes the electric candle less meaningful. I have argued elsewhere that, on closer inspection, the distinction between naturalness and artificiality is not viable. There is neither pure nature nor pure culture; the two are inextricably interwoven.[49] Actually, the 'wax candles' used today are not made of beeswax or other natural products. They are mostly paraffin, and they require chemical processes for their production.[50] Therefore, perceiving one candle as natural and the other as artificial is a phenomenal rather than an ontological distinction. The wax candle offers a comprehensive sensual experience, gives warmth and has a special smell. By appearing less modern and nostalgic, it seems closer to tradition and ancient rites, and the mutual lighting of candles can symbolise connectedness. A lengthy process can be experienced – of lighting, flickering and burning down – which is not controlled by us. This process has elements beyond our direct influence (whether the candle can be lit, whether it flares up strongly, when it goes out and whether it spontaneously goes out in between). The electric candle has a simple automatic mechanism, and the flicking on reminds us of a light switch or a drinks machine. The process is very short and unspectacular. Even if a candle not selected by us turns on, the process is still quite simple and seems to be controlled by us. If the candle does not come on, we blame it on a fault in the system and get angry at 'the technology'.

This response may also be due to our sceptical attitude towards technology. However, this inadequacy of the electric candle is not due to the technologies per se. What the wax candle offers in terms of experience could also be provided by better technology. For example, consider the effect of fireworks. A firework is as artificial as an electric candle and is chemical rather than natural, but the effect on the spectator is spectacular. The delicate ignition, the magnificent play of colours, the contrast of glowing light in the dark, the movements, noises and the uncontrolled course have an impressive effect. It happens (at best) far above our heads and transcends us; it has an element of danger; it is a multi-sensory experience; it astonishes and evokes a romantic feeling. Fireworks are also associated with nature (sky, darkness, weather), and perhaps the element of nature may be conducive to religious experience. However, fireworks are also about wonder and something out of our control. They demonstrate that we do not necessarily require the nostalgic and pre-modern for a poignant, moving, stirring, overwhelming or romantic feeling.

Technology not only evokes such experiences but also offers completely different experiences and functions. Technology can provide impressive sensory experiences and expand the senses (e.g. by implementing senses of the animal world such as infrared vision or magnetic sense in

[49] See Puzio: Über-Menschen, ch. 4.1.

[50] Churches use different material for votive candles and the material also depends on the type of candle.

technology), enable new spatial experiences (virtual and augmented reality) and change the perception of time. Therefore, it will be important for religious robotics to appeal to different senses and trigger various sensory stimuli (e.g. acoustic, visual, olfactory, haptic, tactile) or enable aesthetic experience while it connects with sensory experiences and experiences that are associated with the religious (e.g. smells in church, religious sounds and singing). Light effects may play an especially important role in robotics, since light has a major role in religious experience, is charged with ideas of good and evil, and is associated with hope, salvation, enlightenment and divine communication.[51] Why couldn't a religious robot provide as meaningful a religious experience as a bell, singing bowl or candle? The function of religious robotics is not to replace interpersonal experience and imitate humans. Instead, it can be 'used to extend the experience in ways only the technology can do'.[52] For example, the simple switching mechanism of the electric candle is not useless; it fulfils its quick, practical function for requirements such as switching on lights or buying drinks. However, for a religious, meaningful experience, the technology would have to be designed differently – which is quite possible.

Therefore, the future of religious robotics will also depend on our attitude towards technology, how technology feels to us and how we perceive it, and on its design. In addition, religious and spiritual needs and values when using religious robotics should be considered. For example, lighting a candle is about contemplation, thinking, hoping and praying – as enabled by the wax candle, but maybe not by the electric candle. Furthermore, the development of religious robotics also reveals 'important values in religious communication, like […] feeling connected to others. […] These values need to be carefully translated into meaningful experiences mediated through robot technology […].'.[53]

Furthermore, religious robotics will not involve omnipotent, god-like robots, as critics may suggest. Instead, robots will be effective when they have gaps and open spaces. As discussed above, it is not necessarily the human-like or perfect robot that facilitates religious experience, but precisely the mechanical movements of a robot, its gaps and shortcomings, and its distanced moment, which is alien to us. This gap or lack may open space for reflection and contemplation, meditation and prayer, and experiencing fullness and thinking more deeply. The religious robot is not an image of the divine, but it is meant to facilitate the religious experience, and in doing so, it can and must always be incomplete and fragmentary. Its gaps create space for one's thoughts, memories, wishes, hopes and religious experience.

The example of sacral architecture can illustrate this phenomenon. Since the development of religious robotics is still in its initial stages, sacral architecture, in particular, offers inspi-

51 See Trovato et al.: Religion and Robots, 552.

52 Löffler et al.: Blessing Robot BlessU2, 583.

53 Ibid., 584.

ration. Ivica Brnić speaks of the 'gap, [the] emptiness and [the] nothingness' as 'design principles'[54] and clarifies the construction of the 'presence of the absent' through space, opening and light[55]. He writes: 'In architecture, leaving things out is what makes spatial perceptibility possible in the first place and, as a result, also makes it possible for something to happen'.[56] The gap irritates and points to the whole. Moreover, paradoxically, the emptiness enables the experience of fullness and creates space – for the absent, the mystery, hope, thoughts and prayer.[57]

Brnić also refers to Paul Tillich, who convincingly illustrated the meaning of emptiness for religion:

> This emptiness is not emptiness by privation, but it is an emptiness by inspiration. It's not an emptiness where we feel empty, but it is an emptiness where we feel that the empty space is filled with the presence of that which cannot be expressed in any finite form.[58]

Tillich shows that 'emptiness is not mere absence but presence, not privation but inspiration'.[59] This concept of 'sacred emptiness'[60] can also be applied to religious robotics, as Daelemans states: 'Meaningful symbols and rites need room, sacred emptiness, to resonate and unfold'.[61] Technology in the form of religious robotics does not aim at godlikeness and perfection, but the construction of 'sacred emptiness' can paradoxically express the simultaneity of 'vulnerability and presence' and of 'limitation and fulfilment':

---

54 Brnić, Ivica: Nahe Ferne. Zürich 2019, 197. Own translation. Sentence in German with original quotations from Brnić: Ivica Brnić spricht von der 'Lücke, [der] Leere und [dem] Nichts' als 'Gestaltungsprinzipien' und verdeutlicht die Konstruktion der 'Anwesenheit des Abwesenden' durch Raum, Öffnung und Licht.

55 Ibid., ch. 'Anwesenheit des Abwesenden', from p. 195.

56 Ibid., 197. Own translation. Original: 'In der Architektur ermöglicht das Auslassen überhaupt erst die räumliche Wahrnehmbarkeit und infolgedessen auch das Geschehen.'

57 'Gap' ('Lücke') and 'void' ('Leere') are used side by side in this article, while Brnić and the discourse in architecture differentiate more sharply between both terms.

58 Tillich, Paul: On art and architecture. Ed. by John Dillenberger and Jane Dillenberger. New York 1987, 227. Also cited by Brnić: Nahe Ferne, 198.

59 Daelemans: Sacred Emptiness, 7.

60 Daelemans: Sacred Emptiness. Daelemans refers, among others, to Tillich, e.g: Tillich, Paul: Art and Society. In: Tillich, Paul: On Art and Architecture. Edited by Dillenberger John and Jane Dillenberger. New York 1989, 11–41.

61 Daelemans: Sacred Emptiness, 14.

> The empty space in the open ring 'is also Christ's empty seat at the table of this world. The death of the Lord and his going forth are the wound where history bleeds. When the Lord departed, he left the world open behind him' (Schwarz [1938] 1958, p. 78). Sacred emptiness expresses at the same time vulnerability and presence, expectation and promise, human limitation and divine fulfillment.[62]

One could argue that the limitation of technology may be included with the 'human limitation' in this last aspect.

It is important to note that Tillich argues in his elaborations that it is the architecture that constructs this emptiness:[63]

> 'The sacred void can be a powerful symbol of the presence of the transcendent God. But this effect is possible only if the architecture shapes the empty space in such a way that the numinous character of the building is manifest. An empty room filled only with benches and a desk for the preacher is like a classroom for religious instruction, far removed from the spiritual function which a church building must have' (Tillich [1962] 1989, p. 217). Again, he contrasts sacred and mere emptiness, which does not have the power to express the 'numinous', the presence of the divine.[64]

He even considers this shaping of emptiness through architecture to be 'powerful'.[65] Not every emptiness is a 'sacred emptiness', but it becomes one. In addition, architecture or design also plays a role in this becoming, both in religious spaces and in religious objects. A church, bell, singing bowl, host or Easter candle does not fall from the sky. Neither does a robot. Religious experiences are mediated.

Religious experience requires enduring and tolerating the difference between wish and wish fulfilment.[66] Moreover, speaking of God is not a perfect, superhuman language. Rather, it is

---

62 Daelemans: Sacred Emptiness, 7. Daelemans cites the church architect Rudolf Schwarz, to whom Tillich also referred: Schwarz, Rudolf: The Church Incarnate. Chicago 1958, 78.

63 See Daelemans: Sacred Emptiness, 5.

64 Ibid. Daelemans cites: Tillich, Paul: Contemporary Protestant Architecture. In: Tillich, Paul: On Art and Architecture. Edited by Dillenberger John and Jane Dillenberger. New York 1989, 214–220, here 217.

65 Daelemans: Sacred Emptiness, 5.

66 Lecture 'Gott zur Sprache bringen' held by Prof. Dr. Reinhard Feiter, winter semester 2013/2014, University of Münster. Feiter refers to Sigmund Freud's 'Die Zukunft einer Illusion' ('The Future of an Illusion', 1927).

communication that remains beyond the limit of language: '*Exceeding the limit* and remaining *below the limit* of the linguistic canon is the fate of speaking of God'.[67]

Therefore, religious experience includes that which is overwhelming, that which engages, prepossesses or captures us, that which amazes us, that which is removed from us, withdraws or eludes and is not controlled by us. The gaps and the emptiness create space for religious experience, thought and thoughtful meditation, and the development and experience of stories. Other forms of religious experience may emerge through religious robotics. The many dimensions outlined that are important for religious robotics make it clear that it requires the participation of theologians in their development and design.[68] Although many aspects of religious robotics have been mentioned, it is important to note that religious robotics is not merely a design and development process but a complex interaction between user and robot. What is perceived as religious and whether it leads to a religious experience or successful interaction with the robot is subjective and also depends on the user.[69]

4 Conclusion and Outlook

The use of religious robotics differs greatly according to the specific religion. From a Christian theological perspective, the use of religious robotics is possible and plausible in principle, but it will be decided based on different concepts (e.g. human, life), religious teachings, culturally traditional attitudes and subjective attitudes. The purposes for which the religious robots are used will also be relevant. Religious robots could take on very different tasks, such as administrative tasks, explaining and disseminating religious teachings, giving church tours, arousing interest in religion, facilitating religious and spiritual experiences, accompanying prayers, conducting conversations and other social interactions. Future research needs to reflect on specific fields of application for religious robots: What should robots be used for and what should they not be used for? Robots can contribute to an inclusive church, for example, by providing access to religious ceremonies via digital transmission or augmented reality for those who cannot participate. The relationship to technology is subject to, among other things, time-dependent religious and cultural negotiations. Historically, our relationship with technology and robotics has always changed. Therefore, it is probable that robotics will also become more acceptable in

---

[67] Hemmerle, Klaus: Von Gott sprechen. Online at: https://www.klaus-hemmerle.de/de/werk/von-gott-sprechen.html#/reader/0 (as of: 04.01.22), III. Own translation. Original: 'Überschuss sein *über* die Grenze und Bleiben *unter* der Grenze des Sprachkanons ist das Geschick des Sprechens von Gott.'

[68] See Löffler et al.: Blessing Robot BlessU2, 573 f.

[69] See Daelemans: Sacred Emptiness, 14; Trovato et al.: Religion and Robots, 549.

a religious context as our relationship to robotics changes as the technology develops, as robots become more familiar to us, and we develop a closer relationship with them.

This paper has argued that robots in themselves are not disqualified as religious media because of their quality, condition, constitution, material or technicity. Religions always use mediums, and various entities are suitable as divine representations. Why does Christianity mostly reject religious robots? Robots are just another medium. However, media differ greatly in their type and function and not every medium is suitable for every activity. In the case of highly developed robots, the question arises as to what extent they are actually media and how much agency can be attributed to robots. Exploring this Christian scepticism towards technology and robotics further promises important insights into the conception of and relationship to technology.

This paper has suggested, perhaps surprisingly, that design is a high priority. The example of electric candles has shown that technology must open very specific spaces, enable spaces of imagination or appeal to the senses. Certain movements, mechanisms and automatisms seem mechanical and worthless to us. In contrast, other technologies can, for example, evoke a play of colours that fascinates us (as with fireworks) or change the spatial experience with augmented reality. Several findings can be drawn from the research on the design of robotics. Why do we reject technology over other religious entities? The differences are not in the technology itself but in certain mechanisms, material compositions or functions. The differences are partly phenomenological.

However, since technologies bring many advantages to religious experience and communication, it follows that, rather than designing technology that imitates humans, the particular strengths of technologies should be exploited (e.g. the technological effects outlined above, endurance and repetition, personalised offers, and the fact that one is less ashamed in the presence of technology than with humans). Another result is that religious robots (like other religious objects, places, or buildings) are fundamentally about construction. Religious robotics is not a divine and passive issue, remote from the human being, but is essentially about construction and design. How do we construct technology to open spaces and enable religious experience? The design guidelines have made it clear that the rules for religious robots are different from those for other human–robot interactions. The question of design is a question that should not be underestimated since the identity of the robot depends on it, as does the facilitation of the interface, the relationship and the emotions of the human (respect, trust, fear, authority), religious experience, its acceptance and the comfort it offers. Furthermore, ethical aspects in design need to be investigated in further research (e.g. values in design).

Consequently, it has emerged from this study that the involvement of theologians in its design and conceptualisation is central and that it is a highly interdisciplinary endeavour. Al-

though religious robots are not just a matter of design, the design is relevant and needs scientific debate.

In addition, the study has also explored theological engagement with robotics. By pursuing various questions concerning religious robotics, we ask about the 'values in religious communication' (elaborated by the example of religious robotics),[70] the 'understanding of religious communication',[71] and the meaning of religious media and religious practices (e.g. the blessing robot makes us think about blessing). Religious robots allow us to question and reflect on religious concepts. Furthermore, technology will create new religious and spiritual access, transforming the religious and spiritual experience and leading to new religious practices. In addition, some misconceptions and prejudices have been addressed. Robotics must be demystified. God-like, omnipotent robots with apparent magical abilities (or 'strong AI') that can replace humans are not the goal of religious robotics. The absent should be kept present as the absent.[72] The paper has identified specific relevant questions. It is a matter of asking the right questions and working out issues. Particularly important for theological study will be the anthropological and ethical questions concerning robotics, which could not be dealt with here. These include questions about responsibility, anthropomorphism, personhood, status and relationships to non-human entities (e.g. are animals, angels and robots included in the concept of 'creation'?), moral actors and moral patients, discrimination and manipulation or positive effects.

Theology provides a specific, fruitful approach to robotics. The theological approach has already made it clear that technological success is about more than effectiveness and speed.[73] There are also other influences on technology (i.e. important cultural factors) and certain other dimensions such as values, psychological comfort, spiritual experience and imaginative spaces promoting the success of the technology. It also became clear that, especially with older people who find it difficult to access technology, the religious dimension can facilitate the interaction because it provides familiarity and comfort. It can also be a strength of theology in that it combines robotics with existential questions and can thus offer a hitherto special form of human–robot interaction. The enduring task of theology is to constantly search for new approaches to robotics and reflect on its own concepts in the process.

---

70 See Löffler et al.: Blessing Robot BlessU2, 584.

71 See ibid, 571.

72 Feiter, Reinhard: Lecture 'Gott zur Sprache bringen'. Original: 'Das Abwesende als Abwesendes anwesend halten'. Feiter refers to Winnicott, Donald W.: Übergangsobjekte und Übergangsphänomene. In: Psyche 23 (1969); engl. Transitional objects and transitional phenomena. In: International Journal of Psycho-Analysis 34/2 (1953), 89–97.

73 See Trovato et al.: Religion and Robots, 549.

Literaturverzeichnis

Barad, Karen: Agentieller Realismus. Über die Bedeutung materiell-diskursiver Praktiken. Berlin 2012.

Brnić, Ivica: Nahe Ferne. Sakrale Aspekte im Prisma der Profanbauten von Tadao Andō, Louis I. Kahn und Peter Zumthor. Zürich 2019.

Daelemans, Bert: The Need for Sacred Emptiness: Implementing Insights by Paul Tillich and Rudolf Schwarz in Church Architecture Today. In: Religions 13/6 (2022), 515, 1–15. DOI: 10.3390/rel13060515.

Feiter, Reinhard: Lecture 'Gott zur Sprache bringen', winter semester 2013/2014, University of Münster.

Geraci, Robert M.: Robotics and Religion. In: Runehov, Anne/Oviedo, Lluis (eds.): Encyclopedia of sciences and religions. Dordrecht 2013, 2067–2072.

Hemmerle, Klaus: Von Gott sprechen. Online at: https://www.klaus-hemmerle.de/de/werk/von-gott-sprechen.html#/reader/0 (as of: 04.01.22).

Klein, Mechthild: E-Priester im Einsatz. Religion in Japan. In: Deutschlandfunk 25.09.2019. Online at: https://www.deutschlandfunk.de/religion-in-japan-e-priester-im-einsatz-100.html (as of: 18.09.22).

Löffler, Diana/Hurtienne, Jörn/Nord, Ilona: Blessing Robot BlessU2: A Discursive Design Study to Understand the Implications of Social Robots in Religious Contexts. In: International Journal of Social Robotics 13/4 (2021), 569–586, here 569 f. DOI: 10.1007/s12369-019-00558-3.

Nord, Ilona/Ess, Charles: Robotik in der christlichen Religionspraxis. Anschlussüberlegungen an erste Experimente in diesem Feld. In: Merle, Kristin/Nord, Ilona (ed.): Mediatisierung religiöser Kultur. Praktisch-theologische Standortbestimmungen im interdisziplinären Kontext. Leipzig 2022, 227–258.

Puzio, Anna: Digital and Technological Identities – In Whose Image? A philosophical-theological approach to identity construction in social media and technology. In: Cursor (2021). Online at: https://cursor.pubpub.org/pub/y2bcesx4 (as of: 14.03.22).

Puzio, Anna: Über-Menschen. Philosophische Auseinandersetzung mit der Anthropologie des Transhumanismus (Edition Moderne Postmoderne). Bielefeld 2022.

Puzio, Anna: Zeig mir deine Technik und ich sag dir, wer du bist? – Was Technikanthropologie ist und warum wir sie dringend brauchen. In: Diebel-Fischer, Hermann/Kunkel, Nicole/Zeyher-Quattlender, Julian (eds.): Mensch und Maschine im Zeitalter 'Künstlicher Intelligenz'. Theologische Herausforderungen. 2023.

Robertson, Jennifer: Robo sapiens japanicus. Robots, gender, family, and the Japanese nation. Oakland, California 2018.

Schwarz, Rudolf: The Church Incarnate: The Sacred Function of Christian Architecture. Chicago 1958. First published 1938.

Smith, Joshua: Robot theology. Old questions through new media. Eugene, Oregon 2022.

Tillich, Paul: On art and architecture. Ed. by John Dillenberger and Jane Dillenberger. New York 1987, 1989.

Tillich, Paul: Art and Society. In: Tillich, Paul: On Art and Architecture. Edited by Dillenberger John and Jane Dillenberger. New York 1989, 11–41. First published 1952.

Tillich, Paul: Contemporary Protestant Architecture. In: Tillich, Paul: On Art and Architecture. Edited by Dillenberger John and Jane Dillenberger. New York 1989, 214–220. First published 1962.

Trovato, Gabriele et al.: Communicating with SanTO – the first Catholic robot. 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) 2019, 1–6. DOI:10.1109/RO-MAN46459.2019.8956250.

Trovato, Gabriele et al.: The creation of DarumaTO: a social companion robot for Buddhist/Shinto elderlies. Proceedings of the 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 606–611.

Trovato, Gabriele et al.: Religion and Robots: Towards the Synthesis of Two Extremes. In: International Journal of Social Robotics 13/4 (2021), 539–556, here 545. DOI: 10.1007/s12369-019-00553-8.

Winnicott, Donald W.: Übergangsobjekte und Übergangsphänomene. Eine Studie über den ersten, nicht zum Selbst gehörenden Besitz. In: Psyche 23 (1969); engl. Transitional objects and transitional phenomena – a study of the first not-me possession. In: International Journal of Psycho-Analysis 34/2 (1953), 89–97.

Do Robots Believe in Electric Gods?

Introducing the Theological Turing Test

Hendrik Klinge

Abstract

The Turing Test is a well-known, but controversial tool to check whether digital computers and sophisticated robots possess an ability comparable to human thinking. In this paper, the idea of a theological Turing Test is introduced to discover whether robots will ever be capable of forming religious beliefs. Although the question itself seems intelligible, there are decisive reasons why the theological Turing Test should be rejected. In discussing this test from the perspective of Protestant theology, it will be reasoned that it is plausible to assume that robots will someday be able to form some sort of belief. By employing Wittgensteinian thinking, however, I will argue that these beliefs can only be factual, and not religious. By interpreting metaphysics as a system of beliefs regarding supranatural facts, I will conclude that robots may someday be able to adapt or develop metaphysical systems. Yet, since religious beliefs should be distinguished from factual, including metaphysical, beliefs, there will never be a robot religion.

1 Introduction

How does one distinguish sophisticated, anthropomorphic robots from real human beings? How can I be certain that the person I am talking to on the phone is actually the person I believe them to be, and not some algorithm with a seemingly human voice? Is there any significant difference between talking to "Alexa" and talking to a real person that allows me to know for sure who the interlocutor is? Questions like these have been the subject of countless science fictions novels and films, one of the most famous examples being Philipp K. Dick's *Do Androids Dream of Electric Sheep*?, on which the blockbuster movie Blade Runner (1982) is loosely

based.[1] In this book, a bounty hunter has been assigned the task of "retiring,", i.e., killing, anthropomorphic robots (or androids), who are illegally inhabiting Earth. The problem lies in identifying these robots, as they look and behave like human beings in almost every way. In order to determine whether an individual is a human or a robot, the so-called Voight-Kampff test is used. By checking an individual's capacity for feeling empathy and reacting accordingly, this test helps to identify the robots that the bounty hunter must "retire."

A few melodramatic elements aside, the story is not as unbelievable as it may seem. First, due to recent developments in artificial intelligence, the emergence of highly intelligent, anthropomorphic robots does not seem very far-fetched. Even today, we interact on a daily basis with artificial intelligence, although mostly not anthropomorphic. More importantly, there is a test similar to that imagined by Dick in real life. Since this test is older than Dick's novel, it is unlikely that the author was not inspired by it. I am referring to the famous Turing Test, developed by Alan Turing at the beginning of the 1950s. The idea of this test is, roughly, to suggest a procedure by means of which it is possible to answer whether a machine is "intelligent" in a way comparable to human beings. According to Turing, the test will yield a positive result if a human agent having a conversation with the machine cannot tell whether they are talking to a machine or a real human person. In other words, if the machine succeeds in convincing the human counterpart that it is also human, we must conclude that is able to "think." The quotation marks are appropriate here, because it is highly controversial whether the Turing Test allows us to speak of "intelligent" machines at all.[2]

In this paper, written from the perspective of a Lutheran theologian, I will introduce a theological application of Turing's famous test, which may help answer some questions that are of special interest to theologians and other researchers active in the field of religious studies. Although I will touch on it briefly, my main concern is not the theological objection to the test, as has been suggested by Turing himself.[3] What I rather want to discuss is whether it is conceivable that robots someday will adapt or develop anything like a religion. Since this idea may seem absurd at first, the following paragraph is dedicated to the subject of robot religion, i.e., the idea that highly intelligent machines will someday become "religious" in a way comparable to human beings.

Having thus established that the question I will be discussing is intelligible, I will introduce the idea of a theological Turing Test, simultaneously elaborating a little further on the famous

---

1 See Dick, Philip K., Do Androids Dream of Electric Sheep?. New York 1968. For the philosophical discussion of Dick's novel see Wittkower, D. E. (ed.): Philip K. Dick and Philosophy. Chicago 2011.

2 For the Turing Test see below, section 3.

3 See ibid.

test itself. In the fourth section, I will provide a theological interpretation of Searle's Chinese Room thought experiment, which, no less famous than Turing's test, is often considered an objection to it. Employing Wittgensteinian thinking, which will help clarify the semantics of "belief" (the central concept of this paper), I will then show that even if the theological Turing Test is valid and yields a positive result, it will not prove that robot religion is possible. Finally, I will draw my conclusions by denying the possibility of robot religion. I will also stress that it is of the utmost importance to distinguish between religion and metaphysics, especially when one deals with a concept like robot religion.

It might seem that what I am providing is a mere *a fortiori* argument. Since the Turing Test cannot show that robots understand anything, it is *a fortiori* impossible to show in the test that robots will adapt or develop something like a religion. What renders the argument more than a trivial application, however, is that by discussing robot religion, we learn something about religion itself: Religion is more than just the assertion of a few statements about supranatural entities and, ultimately, that believing in God means more than just acknowledging his existence. In other words, answering whether robots could believe in a divine being, electric or not, may tell us more about ourselves than it does about them.

2 Robot Religion

When it comes to robot religion, i.e., the idea that robots themselves will someday form religious beliefs, one of the first authors who comes to mind is Ray Kurzweil. Director of engineering at Google, Kurzweil is known for numerous publications, mainly in popular science, in which he defends the position of technological posthumanism. Technological posthumanism differs from other, related approaches, such as transhumanism, in that it assumes that, in the future, sophisticated robots will replace humans as the leading species.[4] This thought is often linked to the concept of the so-called singularity, the occurrence of which is believed to coincide with the moment when machines turn out to be clearly superior to human beings. Kurzweil himself is one of the most prominent supporters of this hypothesis.[5]

If one assumes that machines will historically "replace" humans someday, it makes sense to raise the question of whether the end of humanity will also be the end of religion. Kurzweil denies this. Rather, he predicts that the highly intelligent machines of the future will themselves have some form of religion and, therefore, will be "spiritual machines."

---

4 See Loh, Janina: Trans- und Posthumanismus. Hamburg ³2020, 13.

5 See Kurzweil, Ray: The Singularity is Near. London 2006.

> They [i.e., the machines] will believe that they are conscious. They will believe that they have spiritual experiences. They will be convinced that these experiences are meaningful. … Twenty-first-century machines – based on the design of human thinking – will do as their human progenitors have done – going to real and virtual houses of worship, meditating, praying, and transcending – to connect with their spiritual dimension.[6]

Despite the astonishing results of AI research over the last two decades, it does not seem very likely, even twenty years into the 21st century, that machines in Kurzweil's sense will possess a spiritual dimension in the foreseeable future. In 2022, the houses of worship for robot religion, both virtual and real, still belong to the realm of mere speculation. Overall, Kurzweil's prognosis is far too vague to be seriously discussed; whether the age of spiritual machines, as he prophesizes, is actually just around the corner simply cannot be answered. In his main work, *Summa Technologiae* (1964), the Polish philosopher and science fiction author Stanisław Lem provided a much more compelling argument than Kurzweil. Although this is not the focus of his study, Lem addresses the "beliefs of electrical brains" with diligence.[7]

He begins with some observations on the induction problem. Since information in a closed system can only decrease or remain constant, but not increase, induction is absolutely necessary for intelligent beings if they want to receive complete information. Even highly intelligent machines (or homeostats, as Lem calls them) must make assumptions about the future if they want to remain in equilibrium. This, in turn, leads him to the assumption that the homeostats will also necessarily develop beliefs.[8] "Belief" here simply means assumptions about the future that relate to events that only occur with a certain probability.

A clear example of this is weather forecasts. Even the most advanced computer programs cannot predict the weather in the coming days with absolute certainty, but doing without such predictions entirely would mean a serious lack of information. In short, taking epistemic risks is simply a necessity for machines as well as for human beings. What is more, Lem assumes that there is a "continuous spectrum" of beliefs, ranging from the simple forms of belief mentioned above ("I believe that it will rain tomorrow") to metaphysical systems.[9] Given the idea of continuity, if machines are capable of forming one type of belief, it seems plausible to grant them (at least theoretically) the ability to form higher-order beliefs, including those that constitute a metaphysical system. Concluding his thoughts on the topic

---

6 Kurzweil, Ray: The Age of Spiritual Machines. London 1999, 153.

7 See Lem, Stanisław: Summa Technologiae. Minneapolis/London 2013 (orig. 1964), 125–129.

8 See ibid., 111–112

9 Ibid., 111.

of robot religion, Lem, since he equates metaphysics and religion, goes as far as to speak of "believing machines."[10]

At first glance, this argument is surprisingly consistent. There seems to be nothing wrong with assuming that sophisticated robots will eventually be able to develop metaphysical systems, since machines might already be working with (certainly far less complex) systems of beliefs. What is problematic, however, is the equation of metaphysics and religion, as will be shown in detail further below.[11] If you ignore this for the moment, Lem has indeed delivered a convincing argument for the possibility of robot religion: Due to the induction problem, highly advanced robots already exceed the realm of what can be immediately verified (by making predictions, etc.); therefore, it does not seem very far-fetched that they will someday form even higher-order beliefs, including metaphysical ones.

3 Turing and Theology

Thinking machines do not exist – at least if one interprets this expression literally. The idiom is as metaphorical as that of "dormant projects." Turing himself underscores this point in the paper in which he first presents his famous test as a theoretical sketch. Right at the beginning, he points out that he does not want to answer the question of whether machines can think, but rather to replace this highly ambiguous question with another.[12] The popular view of the Turing Test, as suggested by movies like *Ex Machina* (2015), i.e., that it is about proving that machines can think, is therefore an oversimplification at the least. Rather, Turing is concerned with finding out whether machines can have capacity analogous to what we call thinking in humans. The question he actually wants to answer is how an interrogator reacts in a certain scenario, i.e., that of his famous test.

Turing himself called this scenario an "imitation game." A man (A), a woman (B) and an interrogator (C) take part in this game. The interrogator, who is in a different room from A and B, should aim to discover who is the woman and who is the man. To achieve this, they ask A and B different questions, in which A and B may refer to one another when giving their answers. The Turing Test begins when A is replaced by a machine. The interrogator must now use their questions to discover which of the two individuals in the other room is a real human being

10 Ibid., 125

11 See below, section 6.

12 Turing, Alan M.: Computing Machinery and Intelligence. In: Mind 59, no. 236 (Oct. 1950), 433–460, here 433.

and which is a machine only pretending to be human.[13] If the machine succeeds in deceiving the human being, i.e., by making them believe that it is human, it passes the test. According to Turing, this ultimately proves that the machine has a capacity that is analogous to our ability to think. Turing concludes that the imprecise question "Can machines think?" should be replaced by the much more clearly defined "Are there imaginable digital computers that would do well in the imitation game?"[14]

Since Turing's seminal paper, the Turing Test has been carried out multiple times, with varying results.[15] The philosophical debate, meanwhile, has centered on the question of what will (or would) actually be proven by a positive result in the Turing Test – a faculty analogous to human thinking or something else entirely? One of the most important arguments against the validity of the Turing Test, the Chinese Room thought experiment, is discussed in the following section.[16] From a theologian perspective, the test is interesting because Turing himself introduces a theological rationale in his argument when he discusses the objections to his opinion. Contrary to his own caveat, Turing here speaks directly and unguardedly of "thinking machines." However, this is only a minor technical difficulty. Turing formulates the theological objection against his position as follows: Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence, no animal or machine can think.[17]

Although he does not take this objection very seriously, Turing suggests an elaborate reply. His answer to the objection is that God's omnipotence would be restricted if one was not willing to grant Him the capability of transferring a soul into a non-human being, like an elephant or a highly intelligent machine.[18]

Recently, Selmer Bringsjord attempted to refute Turing's argument and defend the theological objection. Without discussing this paper in detail here, the main idea seems to be that Turing's concept of "soul-persuasion" is not as plausible as it may seem at first glance. Rather than having a soul, human beings *are* souls. Therefore, "for God to ensoul bodies is to bring

13 See ibid.

14 See ibid., 442.

15 In 2014, a computer program at the University of Reading allegedly passed the test by convincing one third of the human judges. See University of Reading/Anonymous: Turing Test success marks milestone in computing history. Online at: https://archive.reading.ac.uk/news-events/2014/June/pr583836.html (as of 29.08.2022). For an overview of attempts to perform the Turing Test, see Misselhorn, Catrin: Grundfragen der Maschinenethik. Stuttgart ^{3}2019, 30 f.

16 See below, section 4.

17 Turing: Computing Machinery, 443.

18 See ibid.

into existence a thinking thing that now has use of that body."[19] If I understand Bringsjord correctly, this comes down to the argument that creating an ensouled machine would not mean providing a machine with a soul, which clearly an omnipotent God can do, but to bring into existence a *contradictio in adiecto* ("ensouled machine," i.e., a soul with the body of a machine). To create a *contradictio in adiecto,* however, is obviously something even an omnipotent being cannot do, just as God cannot make two plus three equal four. Thus, Turing's opposition to the theological objection indeed seems flawed. As Bringsjord himself concedes, this, however, does not mean that the objector succeeds.[20]

What is more, the theological objection is one that most theologians, especially those with a background in continental philosophy, would not be willing to endorse. First, the idea of an immortal soul is much more prominent in the antique-pagan tradition than in the Old and the New Testament. Regarding modern theology, it is nothing more than a common prejudice that Christian theology is committed to the view of the immortality of the soul. In fact, some famous theologians of the last century have stressed that there is no such thing as an immortal soul – for strictly theological reasons.[21] The so-called theological objection is, therefore, not very interesting from a theological point a view; whether the objection is a genuine theological one can be seriously doubted.

A far more interesting theological application of Turing's test can be found on the Internet. On a page on experimental theology, one finds the sketch of a "Christian vs. Atheist Turing Test."[22] The idea of this test, which goes back to an author named "Leah," who cannot be identified properly due to several dead links, is that two people, A and B, are placed in a room, where A is a Christian and B is an atheist. The interrogator must now use questions to find out who is the Christian and who is the atheist. The setting of this test is similar to Turing's initial scenario. Rather than a man and a woman, it is now a Christian and an atheist who are to be identified based solely on their statements. The (unspoken) goal of this experiment seems to be to find out whether it is possible to simply simulate being a Christian. At the same time, the "Christ vs. Atheist Turing Test" is, as I will show, also suitable for examining the question of

---

19 Bringsjord, Selmer: God, Souls, and Turing. In: Kybernetes 39, no. 3 (2010), 414–422, here 419.

20 See ibid., 420.

21 See Mühling, Markus: Grundinformation Eschatologie. Göttingen 2007, 171–175.

22 See Beck, Richard: The Christian versus Atheist Turing Test. Online at: http://experimentaltheology.blogspot.com/2011/07/christian-versus-atheist-turing-test.html (as of: 25.08.2022). Even if this page does not meet not the scientific standards that would be required for a thorough analysis, it proposes an interesting idea that seems worth pursuing.

robot religion from a new perspective. To achieve this, it is, however, necessary to modify the test a bit.[23] Imagine the following scenario:

In analogy to Turing's own procedure, the atheist is now replaced by a machine that should articulate certain beliefs and convince the interrogator that it is religious. If it succeeds, according to Turing's approach, it could then be said that the machine has an analogue of what we call religion. Whereas Turing transformed the ambiguous question "Can machines think?" into the question of whether digital computers may pass the Turing Test, the even more ambiguous question "Can computers be religious?" is now translated into that of whether a computer can succeed in playing the theological imitation game. This game would then include questions like the following:

- "Do you believe in God?"
- "Do you believe that Jesus Christ died for your sins?"
- "Do you believe in heaven?"

Certainly, every computer programmed with basic Christian dogmatics will answer these questions in a satisfactory way. But, given Lem's argument for metaphysical systems created by robots, it is also possible to ask more complex questions that do not presuppose that the machine being interrogated commits itself to some existing, human form of religion like Christianity.

- "What is the meaning of life?"
- "Why does evil exist?"
- "Do you believe in an afterlife?"

Even though it is quite imaginable that a digital computer might pass this theological Turing Test, something seems to be off here. In fact, testing whether a machine possesses an analogue to religion (rather than, as in Turing's original paper, an analogue to thinking) reveals the weakness of the initial idea. Rather than saying that a machine that passes the test possesses an analogue to what we call religion, the natural reaction would be to state that the machine only appears to be religious. Yet, if we are inclined to think that a positive result only means that the machine successfully fooled us, the transformation of the question of whether machines might be religious into the question of whether they might pass the theological Turing Test becomes a highly dubious endeavor.

[23] In the following, I will not limit the test to Christianity. However, most examples will be from the Christian religion, as I firmly believe that there is no such thing as an abstract concept of religion that could be used in the test. Religion has always and everywhere taken a concrete form, including natural religion.

To show that the theological Turing Test, regardless of its outcome, will not help answer the question of robot religion, one must, however, go one step further. Merely stipulating that the machine fools us is simply not enough, since we do not know anything about what goes on inside the black box. We simply suppose that the machine wants to trick us, whereby, of course, we make the second mistake of assuming that the machine possesses intentions, or something similar. Thus, a prominent objection to the Turing Test will be presented, which, although highly controversial, clearly shows that the theological application of the test faces the same objections as the test itself.

4 The "Chinese Room"

In his famous paper "Minds, brains and programs," John Searle presents the thought experiment of the Chinese Room.[24] What Searle wants to show in his experiment is that the hypothesis, which he calls "strong AI," is wrong. According to Searle's understanding of strong AI, an "appropriately programmed computer literally has cognitive states and … the programs thereby explain human cognition."[25] The second part of this sentence refers to a well-known position in the theory of mind, i.e., the position that the human mind itself can be understood analogously to a computer.[26] The first part, which is what is important for present purposes, simply states the hypothesis (which Searle wants to reject) that computers can think or that they have cognitive states. Although not explicit, Searle's argument here can also be understood as being directed against Turing's famous test. What Searle wants to show is that the assumption that computers can think is fundamentally wrong, regardless of what result Turing's test yields. Even if the computer succeeds in deceiving the interrogator about its identity, this in no way means that the computer can think or has a capacity analogous to human thinking.

The scenario Searle outlines is not unlike Turing's. Again, it is about a closed space that allows very limited interaction with the environment. Searle now imagines that he is sitting in this space, i.e., the eponymous "Chinese Room," and receiving a "large batch of Chinese writing."[27] Searle does not speak or understand Chinese himself; for him, Chinese characters are simply meaningless symbols. But now he receives a second batch of Chinese writing together

---

[24] Searle, John R.: Minds, Brains, and Programs. In: The Behavioral and Brain Sciences 3 (1980), 417–457.

[25] See ibid., 417.

[26] By this, I am referring to functionalist theories and, especially, the computational theory of the mind. Searle is a well-known critic of functionalism, and the Chinese Room thought experiment perhaps the most cited argument against this view.

[27] See Searle: Minds, 190, 417.

with an English manual that allows him to correlate the symbols of the two Chinese batches. Finally, Searle is given a third batch that contains questions in Chinese with some instructions, again in English. Although Searle himself does not speak Chinese, he can now answer the Chinese questions relating to the first two Chinese batches, i.e., he can find the correct answers to the questions with the help of the instructions alone. In short, Searle produces the answers merely "by manipulating uninterpreted symbols."[28] He thus gives answers in Chinese that are indistinguishable from those of a Chinese native speaker, although he is unable to understand a single Chinese sentence; the entire conversation in which he provides completely satisfactory answers is incomprehensible to him.

Searle, sitting in the "Chinese Room," is comparable to the machine being tested in Turing's scenario. In both cases, the tested subject succeeds in deceiving the interrogator. The machine "tries" to convince the interrogator that it is human, just as Searle tries to convince the person outside the room that he can write Chinese. Since Searle simply manipulates symbols according to the manual, it is not plausible to state that he understands Chinese. In the same vein, just because the machine gives the same answers as any human being would do, we should not jump to the conclusion that it thinks or possesses an ability analogous to thinking. What the machine is doing is simply manipulating symbols, which, to the observer, suggests that it is actually thinking, while this is, in fact, an illusion: Writing Chinese or providing answers like a human being in no way proves that someone (or something) really does understand Chinese or the questions asked. The ability to manipulate symbols is no proof of real understanding, which, in turn, is essential for every mental activity we might compare to thinking.

Put in another way, the problem is all about the relation of syntax to semantics.[29] What Searle's argument seems to suggest is that machines that can process and produce syntactically sound sentences, yet nevertheless have no access to the semantics of these sentences. They just do not know what they mean because they do not have contact with the reality outside the system. They are encapsulated in it, just as Searle is in his room. But it is only the relation to this reality outside the system that provides the means to transform mere symbols into meaningful utterances.

Although this argument, which presupposes only a very basic familiarity with externalist semantics,[30] seems powerful, there are various objections that can be made, some of which Searle himself discusses in his paper.[31] The most interesting, perhaps, is the so-called robot reply. According to the more recent variants of this objection, it now seems possible to put a digital

---

28 See ibid., 418.

29 Cole, David: The Chinese Room Argument. Online at: https://plato.stanford.edu/archives/win2020/entries/chinese-room/ (as of 25.08.2022), sections 1 and 5.1.

30 For an overview of semantic externalism, including the famous arguments made by Putnam and Kripke, cf. Kallestrup, Jesper: Semantic Externalism. Routledge 2012.

31 See Searle: Minds, 419–424.

computer in a robot body and equip it with a sensory apparatus. Searle, writing in 1980, did not imagine such scenarios. An "embodied" computer like this does have contact with reality, by using cameras etc. to perceive its surroundings. Thus, this machine has the means to connect symbols to reality, and the externalist's argument that computers will never be able to transform mere symbols into meaningful sentences just because they are "encapsulated" in their system falls flat.[32] Since not everyone thinks semantic externalism is right (although probably most current philosophers do), this argument, again, is not as victorious as one might think. Nevertheless, what can at least be shown by employing it is that Searle's objection to Turing's test is no less controversial than the test itself.

But what does this all amount to regarding robot religion? Clearly, if Searle's argument is sound, there will never be anything like robot religion. What is more, it will be impossible to attribute even simpler forms of belief to robots, since all they will ever be capable of is manipulating symbols. In other words, if robots have no cognitive states at all, then they *a fortiori* do not have beliefs. If, on the other hand, Searle's argument is unsound (which is still a matter of debate), then robot religion remains a possibility.

For the sake of the argument, I will assume that Searle's objection fails and Turing may be right, and concentrate on whether, even if we do concede that robots may have cognitive states (including some form of beliefs), there will still be an argument that could be made against the possibility of robot religion. This would mean that even the adherent of strong AI must reject the idea of robot religion. The final argument that I will present is not an *a fortiori* argument, since the inference is not simply that robots lack cognitive states and, therefore, are unable to form religious beliefs. Rather, what I want to suggest is that even if someday robots with cognitive states do exist, these robots will certainly not build churches. "Spiritual machines" is, and always will be, an oxymoron, regardless of whether you think strong AI is right or not.

5 What It Means to Believe

Let us, then, concede that the sophisticated robots of the future will have cognitive states, including beliefs. When I talk of "beliefs," I mean not only convictions regarding future events, but also beliefs about the current situation. Following the anglophone tradition since Hume, I will use "belief" broadly to "refer to the attitude we have, roughly, whenever we take something to be the case or regard it as true."[33] Examples of beliefs are as follows:

---

[32] See ibid., 40 and Cole: Chinese Room, section 4.2.

[33] Schwitzgebel, Eric: Belief. In: Zalta, Edward N. (ed.): The Stanford Encyclopedia of Philosophy (Winter 2021 Edition). Online at: https://plato.stanford.edu/archives/win2021/entries/

a. "I believe it will rain tomorrow."
b. "I believe that he is trustworthy."
c. "I believe I have understood this sentence by Wittgenstein."

Religious beliefs are an intricate matter. At first glance, it seems plausible to understand them similarly to those beliefs expressed in the propositions above. Religious beliefs, then, refer to the attitudes we have when we regard religious content to be true, for example:

d. "I believe that Jesus Christ was the Son of God."
e. "I believe that God will absolve me of my sins."

Looking more closely, however, although these sentences are, on the surface, quite similar to those cited above, their function in religious discourse is different. Is the religious function of a sentence like (d) really the same as the standard function of a sentence like (a), with the only difference being that (a) expresses an attitude towards tomorrow's weather, whereas (d) expresses the same attitude towards the identity of Jesus Christ? Of course, it is possible to use both sentences in the same manner, but that does not seem to be the standard use in religious discourse. More precisely, to use (d) in the same manner as we use (a) ordinarily would mean using it in a non-religious way. Similarly, a farmer uttering (a) while hoping for rain uses it in a different sense than a weather forecaster who wants to provide people, regardless of their attitude towards rain, with information about tomorrow's weather. The uses differ because the sentences play an entirely different role in the farmer's and the weather forecaster's lives.

Wittgenstein has remarked that a believer and a non-believer (like himself) may not be talking about the same thing when referring to the last judgment, and that there is, therefore, no disagreement between them regarding the matter.[34] Why is this so? To borrow Wittgenstein's famous (and often misused) term, the believer and the non-believer do not participate in the same language game (Sprachspiel). Of course, it is, again, in no way inconceivable that they would. They could both be talking about whether it is true that the last judgment is an event that will happen at some time in the future. In this case, they would be discussing the same thing, and therefore their disagreement would be real. However, a believer, who simply states that they believe that the last judgement will happen someday, does not, in so doing, express a genuine religious belief. On the contrary, what they express, using the picture of last judge-

belief/ (as of: 25.08.2022), section 1.

[34] See Wittgenstein, Ludwig: Vorlesungen und Gespräche über Ästhetik, Psychoanalyse und religiösen Glauben. Frankfurt a. M. [3]2000, 73 f. My reading of this opaque passages relies on D. Z. Phillips' ingenious interpretation. See Phillips, Dewi Z.: Religious Beliefs and Language Games. In: Phillips, Dewi Z., Wittgenstein and Religion. Houndmills 1993, 56–78.

ment in this manner, is better called a superstition.[35] Genuine religious belief seems to entail something more than the belief that a proposition about something of religious significance is true. In Reformation theology, we find this clearly expressed by Philipp Melanchthon, a close collaborator of Martin Luther:

> The adversaries feign that faith is only a knowledge of the history, and therefore teach that it can coexist with mortal sin. Hence, they say nothing concerning faith, by which Paul so frequently says that men are justified, because those who are accounted righteous before God do not live in mortal sin. But that faith which justifies is not merely a knowledge of history, [not merely this, that I know the stories of Christ's birth, suffering, etc. (that even the devils know,)] but it is to assent to the promise of God, in which, for Christ's sake, the remission of sins and justification are freely offered.[36]

Melanchthon distinguishes here between two forms of faith. One relates to historical facts only, like the birth and death of Jesus Christ (presupposing, of course, that these are facts). Of this faith, Melanchthon says that even the devils may possess it. Real faith, on the other hand, involves the "assent to the promise of God," i.e., a life that is led in the trust of God. To generalize Melanchthon's dictum, one may say the religious faith is more than believing the truth of a doctrine. It is true that even the devils may assent to a proposition like (d), thus confirming that Jesus Christ is the Son of God, as long as this proposition is understood along the lines of (a)–(c) in their standard use (i.e., when a language game is played that could be called "fact checking"). But clearly, most people who believe in (d) do not just want to affirm that, according to them, a supranatural fact about the person of Jesus Christ is true. They want to say that Jesus is the Lord, i.e., the person who is of utmost importance in their personal life.

It thus becomes obvious that, following Melanchthon's distinction, we must carefully distinguish between two senses or forms of belief. One is the form of belief that, in the ordinary sense, refers to the attitude that we have when we regard a proposition to be *true*. The other form of belief, which is the one that matters in religious contexts, refers to the attitude we have

---

35 See ibid., 73.

36 Evangelical Lutheran Synod of Missouri (ed.): The Book of Concord. Online at: https://bookofconcord.org (as of: 25.08.2022), Defense of the Augsburg Confession, art. IV. The latin original reads: "Adversarii tantum fingunt fidem esse notitiam historiae ideoque docent eam cum peccato mortali posse exsistere. Nihil igitur loquuntur de fide, qua Paulus toties dicit homines iustificari, quia, qui reputantur iusti coram Deo, non versantur in peccato mortali. Sed illa fides, quae iustificat, non est tantum notitia historiae, sed est assentiri promissioni Dei, in qua gratis propter Christum offertur remissio peccatorum et iustificatio" (Dingel, Irene [ed.]: Die Bekenntnisschriften der Evangelisch-Lutherischen Kirche. Göttingen 2014, 287–289). The words in brackets are added from the German version of the text which is much more colorful, see ibid., 286.

when we regard a proposition to be of existential *importance* to us. It may, however, be argued that we are simply not entitled to maintain that there are two different meanings; why is this not a mere stipulation? Referring again to Wittgenstein's notion of the language game, I think it is easy to justify this distinction. According to Wittgenstein, the meaning of (at least) some words cannot be detached from their use.[37] A word like "composite" has one meaning when we elaborate on the atomic structure of an object, and a different (though similar) meaning when we are entrusted with a much more mundane task like analyzing the components of a meal, etc.[38] Thus, the word "believe," like the word "composite," may have diverse meanings depending on the language game we are playing, i.e., the type of practice we are involved in. The weather forecaster's belief that it will rain tomorrow is, indeed, the belief that a proposition like (a) is true. We could, of course, talk about religion in the same way the weather forecaster talks about the weather. Then we would play the same game, and our belief in something like the last judgement would become rather superstitious. A believer, however, who joins a ceremony and confesses that Jesus Christ is the Son of God, seems to express another form of belief. What is predominant in their confession is not the attitude that a proposition like (d) is true, but rather that what is expressed by (d) has importance in their life.

How does this all, ultimately, relate to the question of robot religion, i.e., the question of whether robots will ever be able to form religious beliefs? The answer is simple. Even if we concede that robots may have cognitive states, including beliefs in the ordinary sense of the word, it is the very essence of faith that invalidates any attempts to attribute to them something like religious beliefs. Presupposing that robots actually possess an ability analogous to human thinking (although this is currently far from being proved), it does not seem very far-fetched to assume they can use propositions like (a)–(c) to expresses their beliefs. What is more, they may also do so by uttering a sentence like (d). Yet, a robot that states its belief that Jesus Christ is the Son of God will then merely be stating a factual belief, for "fact checking" is one of the (probably very few) languages it can participate in. Robots, therefore, can be compared to Melanchthon's devils. If they have cognitive states, it is imaginable that they really affirm the truth of a given proposition, for example, regarding the history of Jesus Christ. But they will never be able to play a part in the religious language game, since it does not suffice to simply have a cognitive state to join in this game. Something is still missing, and this something is not only consciousness, but the awareness that some things are of existential importance; and this

[37] "For a *large* class of cases – though not for all – in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language" (Wittgenstein, Ludwig: Philosophical Investigations. Translated by G.E.M. Anscombe. Oxford [2]1958, 20 f. [= paragraph 43]).

[38] See ibid., 21–22 [= paragraph 47].

awareness has at least *self-consciousness* as a prerequisite because what is of religious importance is always important to *me*.

In other words, the idea that there is a "continuous spectrum" of different forms of beliefs is deeply mistaken.[39] Religious beliefs are not comparable to simple forms of beliefs; there is a gulf between these two forms that cannot be bridged. Lem, of course, does not speak of religious beliefs when he introduces his "continuous spectrum," but of *metaphysical* beliefs. Probably unbeknownst to him, he does so very wisely, because religious beliefs and metaphysical beliefs must not be confounded. Drawing on this distinction, I will now turn to my concluding remarks about the impossibility of robot religion.

6 Conclusion: Religion and Metaphysics

Due to the influence of Friedrich Schleiermacher, recent German scholars active in the fields of theology and religious studies stress the importance of not confounding religion and metaphysics.[40] To put it very roughly, religion, as Schleiermacher envisions it, has to do with a special kind of feeling or emotion, which must be distinguished from both metaphysics, understood as a theoretical endeavor, and morality.[41] Although it may be possible to implement some equivalent of human emotions into machines, it seems very unlikely that artificial intelligence will ever be capable of complex emotions in the same way human beings are. The idea of emotional robots is much more challenging, to say the least, than Turing's idea of a thinking machine, which itself is far from uncontroversial. While it may be plausible to state that robots will someday take part in metaphysical investigations, it hardly seems imaginable that they will ever possess religious emotions. Since religion, in contrast to metaphysics, is basically about emotions, the very idea of religious or spiritual machines must be rejected.

The robots of the future may develop a metaphysical system that contains propositions about supranatural entities, and they also may believe (in a factual sense) that electric gods exist. They may even form the belief that these gods created them. But they will never believe in these gods in the sense that they acknowledge the importance of these gods for their existence and participate in a religious form of life. They never will truly praise their gods or thank them.

---

39 See above, section 2.

40 See for example Axt-Piscalar, Christine: Der Grund des Glaubens. Eine trinitätstheologische Untersuchung zum Verhältnis von Glaube und Trinität in der Theologie Isaak August Dorners (Beiträge zur historischen Theologie 79). Tübingen 1990, 51.

41 See Schleiermacher, Friedrich D. E.: Der christliche Glaube nach den Grundsätzen der evangelischen Kirche im Zusammenhange dargestellt. Second Edition (1830/31). Berlin/New York 2003, vol. 1, 19–32.

They may speak a prayer, but it will never be an honest one. And this is simply because these robots do not feel any fear or existential dread; they do not suffer from being limited in what they can do, and, perhaps most importantly, they lack any consciousness of being mortal (which, in fact, they are not, at least not in the same sense as human beings or animals).

To put it in a nutshell, whether the theological Turing Test, as it has been introduced in this paper, fails or not, is not relevant in the question of robot religion. One might even suppose that this test will yield positive results that are clear evidence of the robot's capability to form beliefs, and counter-arguments like Searle's Chinese Room thought experiment provide no sufficient reason to reject them. But even then, i.e., conceding that the theological Turing Test actually shows that machines can give answers to religious questions and, furthermore, understand their own answers, all that is proven is that robots can form and understand metaphysical statements, which is, according to both Melanchthon and Schleiermacher, something entirely different from being religious. Robots may be capable of believing certain things, including supranatural "facts," but these facts will never have a salience for their own existence that would allow to call these robots religious.

In conclusion, robots may believe that a divine being exists, understood simply as a supranatural fact, but they will never be able to acknowledge that this being is their God, their Lord and their savior. Just as Melanchthon's devils, robots probably will someday be able to participate in the metaphysical language game, weighing the truths of propositions regarding God and the life of Jesus Christ. Yet, they will never join in the religious language game, which is not about the truth of a proposition, but about the importance that a religious picture or concept plays in my own life. For robots, there is no "my" or "I," nor is there anything like selfhood. Since religion can be interpreted as a reflection on what it means to be a self, the term "spiritual machines" remains an oxymoron. In discussing robot religion, we once again learn the importance of distinguishing between religion and metaphysics; and therefore, although the result is ultimately negative, I think that it is worth our while dealing with it. After all, by talking about the impossibility of robot religion, we begin to understand human religion even better.

References

Axt-Piscalar, Christine: Der Grund des Glaubens. Eine trinitätstheologische Untersuchung zum Verhältnis von Glaube und Trinität in der Theologie Isaak August Dorners (Beiträge zur historischen Theologie 79). Tübingen 1990.

Beck, Richard: The Christian versus Atheist Turing Test. Online at: http://experimentaltheology.blogspot.com/2011/07/christian-versus-atheist-turing-test.html (as of: 25.08.2022).

Bringsjord, Selmer: God, Souls, and Turing. In Defense of the Theological Objection to the Turing Test. In: Kybernetes vol. 39, no. 3 (2010), 414–422. DOI: 10.1108/03684921011036141.

Cole, David: The Chinese Room Argument. In: Zalta, Edward N.: The Stanford Encyclopedia of Philosophy (Winter 2020 Edition). Online at: https://plato.stanford.edu/archives/win2020/entries/chinese-room/ (as of 25.08.2022).

Dick, Philip K.: Do Androids Dream of Electric Sheep. New York 1968.

Dingel, Irene (ed.): Die Bekenntnisschriften der Evangelisch-Lutherischen Kirche. Vollständige Neuedition. Herausgegeben im Auftrag der Evangelischen Kirche in Deutschland. Göttingen 2014.

Evangelical Lutheran Synod of Missouri (ed.): The Book of Concord. Online at: https://bookofconcord.org (as of: 25.08.2022).

Kallestrup, Jesper: Semantic Externalism. Routledge 2012. DOI: 10.4324/9780203830024.

Kurzweil, Ray: The Age of Spiritual Machines. How we will Live, Work and Think in the New Age of Intelligent Machines. London 1999. DOI: 10.1016/S0308-5961(99)00064-6.

Kurzweil, Ray: The Singularity is Near. When Humans transcend Biology. London 2006. DOI: 10.1057/9781137349088_26.

Lem, Stanisław: Summa Technologiae. Translated by Joanna Zylinska (Electronic Meditations 40). Minneapolis/London 2013 (orig. 1964).

Misselhorn, Catrin: Grundfragen der Maschinenethik. Stuttgart ^{3}2019.

Mühling, Markus: Grundinformation Eschatologie. Göttingen 2007.

Loh, Janina: Trans- und Posthumanismus. Eine Einführung, Hamburg ^{3}2020.

Phillips, Dewi Z.: Religious Beliefs and Language Games. In: Phillips, Dewi Z., Wittgenstein and Religion. Houndmills 1993, 56–78.

Schleiermacher, Friedrich D. E.: Der christliche Glaube nach den Grundsätzen der evangelischen Kirche im Zusammenhange dargestellt. Second Edition (1830/31). Berlin/New York 2003, vol. 1.

Schwitzgebel, Eric: Belief. In: Zalta, Edward N. (ed.): The Stanford Encyclopedia of Philosophy (Winter 2021 Edition). Online at: https://plato.stanford.edu/archives/win2021/entries/belief/ (as of: 25.08.2022).

Searle, John R.: Minds, Brains, and Programs. In: The Behavioral and Brain Sciences 3 (1980), 417–457.

Turing, Alan M.: Computing Machinery and Intelligence. In: Mind vol. 59, no. 236 (Oct. 1950), 433–460, here 433.

University of Reading/Anonymous: Turing Test success marks milestone in computing history. Online at: https://archive.reading.ac.uk/news-events/2014/June/pr583836.html (as of 29.08.2022).

Wittgenstein, Ludwig: Vorlesungen und Gespräche über Ästhetik, Psychoanalyse und religiösen Glauben. Frankfurt a. M. ^{3}2000.

Wittgenstein, Ludwig: Philosophical Investigations. Translated by G. E. M. Anscombe. Oxford ^{2}1958.

Wittkower, D. E. (ed.): Philip K. Dick and Philosophy. Do Robots Have Kindred Spirits? (Popular Culture and Philosophy 63). Chicago 2011.

III Transformation des Körpers

Medizin und Optimierung

Ambivalenzen gegenwärtiger Gewissheitsbestrebungen

Menschliche Entscheidungsfreiheit in einer gewisserwerdenden Welt

Max Tretter

Abstract

One characteristic of the present is its ubiquitous uncertainty. This uncertainty makes it difficult for many people to take decisions. Digital technologies and Artificial Intelligence (AI) are supposed to remedy this situation by creating certainties and giving people new freedom to make decisions. But how do these techno-generated certainties *actually* affect human freedom of decision? Drawing on research on digital self-trackers and Jean Baudrillard's reflections on simulation, this paper argues that digital- and AI-created certainties have ambivalent effects on human freedom of decision. On the one hand, they can increase this freedom by providing some degree of certainty about what to expect from each decision. On the other hand, they tend to reduce the number of options available and thus the freedom of decision. This paper concludes with a theological perspective on these results and provides some guidelines for dealing with certainty-generating technologies.

1 Einleitung[1]

Wir leben, folgt man den Gegenwartsdiagnosen Zygmunt Baumans, in „ungewissen Zeiten."[2] Nachdem die Moderne in der zweiten Hälfte des 20. Jahrhunderts in ihre „flüchtige Phase"[3] eingetreten ist, überholt sich Wissen derartig schnell, besitzen Informationen eine derart knappe Halbwertszeit, sind Trends derart kurzlebig und veralten Technologien derart schnell, dass, wie der polnische Philosoph festhielt, wer auf dem Stand der Zeit bleiben will, sich ständig neu erfinden, sich unaufhörlich neue Kenntnisse aneignen und kontinuierlich neue Dinge beschaffen muss.[4] Nimmt man zu dieser Flüchtigkeit noch die kontinuierlich ablaufenden Gesellschaftsdifferenzierungs-[5] und -rehybridisierungsprozesse[6] hinzu, findet sich das Individuum in einer Situation wieder, in der es von zwei Seiten – eskalierenden Beschleunigungsdynamiken[7] wie sich potenzierenden Sozialkomplexitäten[8] – bedrängt wird und es zunehmend schwerer für es wird, sein Leben weitsichtig zu planen oder wohlüberlegte Entscheidungen über seine mittel- bis langfristige Zukunft zu treffen.[9]

Diese gegenwärtigen Gesellschaftsprozesse haben zwiespältige Auswirkungen auf die menschliche Entscheidungsfreiheit – letztere verstanden im Doppelsinn als *Frei*raum für Entscheidungen wie als *Frei*mütigkeit beim Treffen von Entscheidungen. Während sich auf der einen Seite infolge immer weiterer Modernisierungs- und Liberalisierungsprozesse stets mehr Möglichkeiten, sprich Entscheidungsfreiräume auftun, müssen Personen ihre Entscheidungen zunehmend „ins Ungewisse" treffen, d. h. ohne die Konsequenzen ihrer Entscheidung genügend abschätzen und verschiedene Optionen fundiert miteinander abgleichen zu können.[10]

---

[1] Diese Arbeit wurde aus Mitteln des Bundesministeriums für Forschung und Bildung (Förderkennzeichen: 01GP1905B und 01GP2202B) finanziert. Der Förderer spielte keine Rolle bei der Durchführung der Forschung oder der Erstellung des Manuskripts.
Vielen Dank an Tabea Ott und David Samhammer sowie an Lorenz Garbe, Isabella Auer und Carima Jekel für Eure Impulse, hilfreichen Kommentare und Diskussionen.

[2] Vgl. Bauman, Zygmunt: Flüchtige Zeiten. Leben in der Ungewissheit. Hamburg 2008.

[3] Vgl. Bauman, Zygmunt: Flüchtige Moderne. Frankfurt a. M. 2003.

[4] Vgl. Bauman: Flüchtige Moderne.

[5] Vgl. Luhmann, Niklas: Die Gesellschaft der Gesellschaft. Bd. 1 & 2. Frankfurt a. M. 1997.

[6] Vgl. Latour, Bruno: Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie. Frankfurt a. M. 2010.

[7] Vgl. Rosa, Hartmut: Beschleunigung. Die Veränderung der Zeitstrukturen in der Moderne. Frankfurt a. M. 2014.

[8] Vgl. DeLanda, Manuel: A New Philosophy of Society. Assemblage and Social Complexity. London, New York 2006.

[9] Vgl. Bauman, Zygmunt: Leben in der flüchtigen Moderne. Frankfurt a. M. 2010.

[10] Vgl. Bauman, Zygmunt: Liquid Fear. Cambridge, Malden 2006.

Dies birgt die Gefahr, dass sich jede getroffene Entscheidung am Ende als nachteilig, ihre Folgen als schädlich erweisen – was nicht selten Bedenken und Sorgen hervorruft und die Freimütigkeit beim Entscheiden beeinträchtigt.

Um ihre Sorgen zu überwinden und neue Entscheidungsfreiheit zu gewinnen, streben Menschen danach, wie Dewey in seiner *Suche nach Gewissheit* festhält, bestehende Ungewissheit durch energische Gewissheitsbestrebungen zu überwinden.[11] Gerade die Digitalisierung und Fortschritte in der Forschung zur Künstlichen Intelligenz (KI) liefern hier neue Möglichkeiten der Gewissheitsgewinnung und versprechen, bestehende Ungewissheiten dauerhaft zu reduzieren. Gleichzeitig werfen diese neuen Optionen die Frage auf, *wie sich solch digital- oder KI-produzierte Gewissheiten tatsächlich auf die menschliche Entscheidungsfreiheit auswirken*. Sorgen digital- bzw. KI-produzierte Gewissheitsgewinne tatsächlich für mehr Entscheidungsfreiheit? Oder entfalten sie am Ende gegenteilige Wirksamkeit und schränken Letztere sogar ein? Dieser Frage wird in dem Beitrag nachgegangen und argumentiert, *dass wachsende, mittels digitaler oder KI-Technologien produzierte Gewissheiten die menschliche Entscheidungsfreiheit auf ambivalente Weise beeinflussen*. Während sie auf der einen Seite ein gewisses Maß an Entscheidungsfreimütigkeit wiederherstellen und die Entscheidungsfreiheit steigern können, schränken sie auf der anderen Seite den Raum möglicher Entscheidungen stückweise ein – und sorgen so für verringerte Entscheidungsfreiheiten.

Um diese These zu untermauern, werden zuerst die Konzepte Gewissheit und Ungewissheit vorgestellt und aufgezeigt, wie sie sich auf die menschliche Entscheidungsfreiheit auswirken. Das folgende Kapitel wird sich den Methoden der Gewissheitsproduktion widmen und in weitgehend abstrakter Weise schildern, wie durch den Einsatz von Digitaltechnologien und KI-Anwendungen Gewissheiten produziert werden. Das vierte Kapitel wird dann schließlich der bislang wenig diskutierten Frage nachgehen, welche Auswirkungen digital- wie KI-produzierte Gewissheiten auf die menschliche Entscheidungsfreiheit haben. Dazu werden zuerst digitale Selbstvermessungspraktiken betrachtet und herausgearbeitet, dass die gewonnenen Quantifizierungsgewissheiten dazu tendieren, gleichermaßen die Freimütigkeit beim Treffen von Entscheidungen zu steigern wie die bestehenden Entscheidungsräume einzuengen. Anschließend werden simulationstheoretische Überlegungen Jean Baudrillards aufgegriffen, um aufzuzeigen, welch ambivalente Auswirkungen umfassende (KI-)Simulationen auf die menschliche Entscheidungsfreiheit haben, wie sie gleichermaßen freimutschaffende Gewissheiten produzieren können, im Gegenzug aber die Entscheidungsoptionen reduzieren. Nachdem die leitende These durch die Betrachtungen der ersten vier Kapitel eingeholt und untermauert wurde, nimmt der Beitrag im Kapitel fünf eine normative Wende, wenn es abschließend darum gehen

[11] Vgl. Dewey, John: Die Suche nach Gewißheit. Eine Untersuchung des Verhältnisses von Erkenntnis und Handeln. Frankfurt a. M. 1998.

soll, aus einer evangelisch-theologischen Perspektive Leitlinien für einen angemessenen Umgang mit gewissheitsschaffenden Digital- und KI-Technologien zu entwickeln.

2 Konzeptionelle Überlegungen zu Ungewissheit und Gewissheit

Ungewissheit ist in erster Linie ein epistemologisches Konzept und beschreibt einen Zustand ungenügenden Wissens – entweder weil Informationen fehlen, verzerrt, ungenau oder schlichtweg falsch sind.[12] Da die Entscheidungen einer Person, wie in diversen Entscheidungs- und Handlungstheorien vielfach herausgearbeitet,[13] maßgeblich von ihrem Wissensstand abhängen, hat Ungewissheit auch praktische Auswirkungen. Je mehr eine Person weiß oder erahnen kann, wie sich die Zukunft gestaltet und welche Anforderungen auf sie zukommen werden, desto mehr kann sie sich auf zukünftige Szenarien einstellen. Und je genauer sie weiß oder antizipieren kann, welche ihrer Handlungsoptionen welche Folgen haben wird, desto präziser kann sie verschiedene Möglichkeiten in ihrem Kopf durchspielen und gegeneinander abwägen – sprich: Handlungsfreiräume gewinnen – und so zu einer begründeten wie freimütigen Entscheidung kommen.[14] Je weniger eine Person hingegen weiß oder abschätzen kann, d. h. je größer ihre Ungewissheit ist, desto kleiner werden nicht nur ihre Entscheidungsfreiräume, sondern desto weniger ist sie auch in der Lage, vorausplanende Entscheidungen zu treffen[15] – kurz: desto mehr schrumpft auch ihre Entscheidungsfreimütigkeit.

Das konzeptionelle Gegenstück zu Ungewissheit ist Gewissheit. Letztere beschreibt einen Zustand, in der eine Person mit hoher Zuverlässigkeit einschätzen kann, was auf sie zukommt, worauf sie sich einstellen muss und welche Folgen ihre Handlungen haben werden.[16] Dieses präzise Vor-Wissen, welches häufig das Ergebnis einer umfassenden Situationskenntnis und geschulter Abstraktions- wie Deduktionsmechanismen ist, erlaubt es Personen, verschiedene Handlungsoptionen sorgfältig gegeneinander abzuwägen und zu einer begründeten Entscheidung zu kommen[17] – und vergrößert dadurch deren Entscheidungsfreimütigkeit.

---

12 Vgl. Smithson, Michael: Ignorance and Uncertainty. Emerging Paradigms. New York, Berlin 1989.

13 Vgl. Horn, Christoph/ Löhrer, Guido (Hg.): Gründe und Zwecke. Texte zur aktuellen Handlungstheorie. Berlin 2010.

14 Vgl. Johnson, L. Syd: The Ethics of Uncertainty. Entangled Ethical and Epistemic Risks in Disorders of Consciousness. New York 2022.

15 Vgl. Zimmerman, Michael J.: Living with Uncertainty. The Moral Significance of Ignorance. Cambridge, New York 2008.

16 Vgl. Dewey: Die Suche nach Gewißheit.

17 Ebd.

Eine entscheidungserschwerende Ungewissheit und eine Gewissheit, die Freimütigkeit zur Entscheidung schafft, lassen sich damit als gegenüberliegende Endpunkte einer epistemischen wie praxeologischen Skala anordnen (siehe Abbildung 1). Steht an einem Ende der Skala eine absolute Ungewissheit, d. h. ein Zustand, in dem eine Person *nichts* weiß, die Folgen potentieller Handlungen nicht einschätzen kann und maximal entscheidungsunfreimütig ist, steht am anderen Skalenende eine absolute Gewissheit, d. h. ein Zustand, in der eine Person aus einer lückenlosen Kenntnis einer Situation zukünftige Entwicklungen präzise ableiten und vor diesem Hintergrund die Folgen möglicher Handlungen bis ins kleinste Detail antizipieren, sie gegeneinander abwägen und sich in maximaler Freiheitmütigkeit entscheiden kann. Beide Extreme sind dabei gleichermaßen unerreichbar. Denn einerseits wissen Personen, wie Wittgenstein in seinen Gewissheitsüberlegungen festhält, niemals nichts, sondern haben immer schon eine gewisse Vorstellung von ihrer Umwelt und einen Erfahrungsschatz, der ihnen dabei hilft, zukünftige Entwicklungen sowie die Folgen ihrer Handlungen zu einem gewissen Grad vorauszuahnen.[18] Gleichermaßen unmöglich ist es jedoch auch, vollständiges Wissen über eine Situation zu erlangen und sämtliche Zukunftsentwicklungen vorhersagen und die Folgen von Handlungen umfassend vorhersehen zu können.[19]

Abbildung 1: Ungewissheits-Gewissheits-Skala (erstellt vom Autor).

Sich immer im Zwischenraum zwischen absoluter Ungewissheit und absoluter Gewissheit wiederfindend, streben Menschen im Regelfall, ihre Ungewissheit zu reduzieren, mehr Gewissheit zu erlangen und dadurch seine ihre Entscheidungsmütigkeit zu vergrößern.[20]

18 Vgl. Wittgenstein, Ludwig: Über Gewissheit. Frankfurt a. M. 1970.

19 Vgl. Dewey: Die Suche nach Gewißheit.

20 Ebd.

3 Methoden der Gewissheitsproduktion

Die Methoden, mit denen der Mensch nach mehr Gewissheit strebt und seine Entscheidungsfreimütigkeit auszudehnen sucht, sind und waren jeweils abhängig von den Kenntnissen und Technologien der jeweiligen Zeit. Blickt man beispielsweise in die alttestamentliche Zeit, waren es in erster Linie Spruchweisheiten, gewonnen aus praktischen Welt- und Sozialbeobachtungen, die ihren Hörer:innen und Leser:innen grobe Leitlinien und Handlungsempfehlungen an die Hand gaben, mittels derer sie sich in der Alltagswelt orientieren und an denen sie ihre Entscheidungen ausrichten konnten.[21] Diese weisheitliche Form der Weltorientierung wurde im Alten Israel, aber auch über dessen Grenzen hinaus,[22] nicht selten um Auslegungs- oder Divinationspraktiken ergänzt. In den altorientalischen wie altägyptischen „Divinationskulturen"[23] spielten derartige Zeichendeutungen oder Götterbefragungen eine zentrale Rolle, sollten sie doch einen Einblick in die Pläne der Götterwelt gewähren und die Fragesteller:innen vergewissern, ob ihre Handlungen und ihre Entscheidungen auf göttliche Zustimmung oder Ablehnung treffen.[24]

Springt man ein paar Jahrhunderte in der Zeit vorwärts, verändern sich, vor allem aufgrund des unaufhaltbaren Siegeszugs der Wissenschaften, besonders der Naturwissenschaften, und dem kontinuierlichen technologischen Wandel, auch die präferierten Methoden der Gewissheitsproduktion. Wie Dewey in seiner *Suche nach Gewissheit* darlegt, wurden aus unsystematischen Weltbeobachtungen gezielte Experimente, mittels derer man suchte, den „Gesetze[n] der natürlichen Welt"[25] auf die Spur zu kommen. Und aus Divinationen wurden Versuche, die Zukunft mittels physikalischer Modelle, sozialer Gesetzmäßigkeiten und dergleichen präzise berechnen zu können.[26] Auf diese Weise konnten forschend wie rechnerisch Gewissheiten ge-

[21] Vgl. Hausmann, Jutta: Weisheit (AT). Online unter: https://www.bibelwissenschaft.de/stichwort/34707/ (Stand: 30.08.2022).

[22] Vgl. Lackner, Michael: Eine „divinatorische Kultur par excellence"? Chinesische Wahr- und Weissagung im Vergleich. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 7–28.

[23] Assmann, Jan: Zeitkonstruktion, Vergangenheitsbezug und Geschichtsbewußtsein im alten Ägypten. In: Assmann, Jan/Müller, Klaus E. (Hg.): Der Ursprung der Geschichte. Archaische Kulturen, das Alte Ägypten und das Frühe Griechenland. Stuttgart 2005, 112–214, hier 113.

[24] Heessel, Nils P.: Die altorientalische Divination als antikes Wissenschaftssystem. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 48–66.

[25] Dewey: Die Suche nach Gewißheit, 32.

[26] Ebd.

wonnen werden, die, im Rahmen ihrer jeweiligen Möglichkeiten,[27] Orientierung schafften und neue Freiräume für freimütigere Entscheidungen eröffnen konnten.[28]

Spielt die wissenschaftliche Naturerforschung gegenwärtig noch immer eine zentrale Rolle und gibt es auch heute noch Personen, die aus Zeichendeutungs- oder Divinationspraktiken Gewissheit zu ziehen suchen, sind es in jüngerer Zeit vor allem digitale und KI-gestützte Herangehensweisen, mittels derer Gewissheit zu produzieren und Entscheidungsfreiheit zu gewinnen, anstrebt wird.[29]

Digitale Gewissheitsproduktion ruht auf zwei Säulen: erstens einer umfassenden Vermessung und Dataisierung der Welt sowie, zweitens, der Möglichkeit, die digitalisierten Weltdaten maschinell auszuwerten.[30] So lassen sich Algorithmen nutzen, um riesige Datenmengen zu durchsuchen und mit enormer Effizienz bislang unerkannte Korrelationen zwischen verschiedenen Datenpunkten, sogenannte „Muster"[31] ausfindig zu machen. Diese algorithmisch gewonnenen Kenntnisse über signifikante Zusammenhänge innerhalb großer Datenbestände ermöglichen es, aus vorliegenden Informationen neue Schlussfolgerungen zu ziehen oder sogar ausstehende Entwicklungen umrisshaft zu extrapolieren.[32] So können auf digitale Weise statistische Gewissheiten erzeugt werden, die bei ausstehenden Entscheidungen Orientierung liefern und ein gewisses Maß an Entscheidungsfreimütigkeit schaffen können.[33]

Besitzen Algorithmen, die in solchen Apps oder zur Mustererkennung in Big Data allgemein eingesetzt werden, häufig schon Selbstlernfähigkeiten und sind damit zu einem gewissen Grad künstlich-intelligent,[34] steigern die jüngsten Fortschritte in der KI-Forschung deren Möglichkeiten der Gewissheitsproduktion nochmals. Denn mittlerweile sind die Fähigkeiten von Algorithmen nicht mehr darauf beschränkt, statistische Zusammenhänge zwischen ver-

---

27 Glassmeier, Karl-Heinz: Wetten, Wissen, Werten. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 176–194.

28 Dewey: Die Suche nach Gewißheit.

29 Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen: Einleitung. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 1–6.

30 Grunwald, Armin: Digitalisierung als Prozess. Ethische Herausforderungen inmitten allmählicher Verschiebungen zwischen Mensch, Technik und Gesellschaft. In: Zeitschrift für Wirtschafts- und Unternehmensethik 20/2 (2019), 121–145.

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

32 Siegel, Eric: Predictive Analytics. The Power to Predict Who Will Click, Buy, Lie, or Die. Hoboken 2016.

33 Vgl. Elgendy, Nada/Elragal, Ahmed: Big Data Analytics in Support of the Decision Making Process. In: Procedia Computer Science 100, 1071–1084.

34 Vgl. Mit meinem Verständnis von KI orientiere ich mich an: Rosengrün, Sebastian: Künstliche Intelligenz zur Einführung. Hamburg 2021.

schiedenen Faktoren in Datensets zu identifizieren, auf bestehende Wechselwirkungen hinzuweisen und vorliegende Trends zu erahnen.[35] Als sprichwörtliche „Prädiktionsmaschinen"[36] sind gegenwärtige Highend KI-Algorithmen auch in der Lage, auf Grundlage ihrer Daten und mittels komplexer Analyse- wie Berechnungsprozesse Simulationen über die Entwicklung einer konkreten Entität anzustellen.[37] Damit können sie nicht nur statistische Zukunftsaussagen treffen, sondern mit einer hohen Präzision vorhersagen, wie sich *eine* Person, *ein* technisches Artefakt, *eine* Stadt, etc. unter bestimmten Bedingungen in Zukunft verhalten oder entwickeln wird. Mehr noch, da sich diese KI-Simulationen jeweils auch unter veränderten Vorzeichen, d. h. unter Änderung eines oder mehrerer Faktoren durchführen lassen, lassen sie sich ebenfalls nutzen, um zu berechnen, wie verschiedene Änderungen oder Handlungen zukünftige Entwicklungen beeinflussen würden. Dadurch lassen sich nicht nur weite Freiheitsräume künstlich-intelligent erschließen, sondern auch individuelle Freimütigkeit zum Entscheiden gewinnen.

4 Welche Auswirkungen haben digital und KI-produzierte Gewissheiten auf die menschliche Entscheidungsfreiheit?

Wenn durch die Nutzung von Digitaltechnologien und KI-Prognosen neue statistische oder sogar personal-zugeschnittene Gewissheiten geschaffen und in die anfangs beschriebenen Ungewissheitsstrukturen eingeführt werden – hat dies dann auch zur Folge, wie vielfach erhofft, dass die Entscheidungsfreiheiten, die durch besagte Ungewissheiten beschränkt waren, wieder zunehmen? Diese Frage soll im folgenden Kapitel mittels zweier Annäherungen adressiert werden. In einem ersten Schritt wird dazu das digitale Selbsttracking untersucht und danach gefragt, welche Auswirkungen die quantifizierungsgetriebenen Gewissheitsgewinne auf die Entscheidungsfreiheit der Selbstvermesser:innen haben. Diese eher an empirischen Erhebungen und deren Ergebnissen orientierte Herangehensweise soll in einem zweiten Schritt um eine etwas spekulativere Annäherung ergänzt werden, wenn in Auseinandersetzung mit dem Denken Jean Baudrillards die Frage gestellt wird, wie sich umfassende Simulationsgewissheiten auf das menschliche Entscheiden und dessen Freiheit

[35] Vgl. Siegel: Predictive Analytics.

[36] Vgl. Agrawal, Ajay/Gans, Joshua/Goldfarb, Avi: Prediction Machines. The Simple Economics of Artificial Intelligence. Boston 2018.

[37] Vgl. Greif, Hajo: Modellierung und Simulation in der Künstlichen Intelligenz. In: Mainzer, Klaus (Hg.): Philosophisches Handbuch Künstliche Intelligenz. Wiesbaden 2019, 1–21.

auswirken. Anschließend werden die Ergebnisse auf Gewissheitsgewinne aus KI-Prognosen übertragen.

4.1 Die Auswirkungen digitaler Selbstvermessungsbestreben auf die menschliche Entscheidungsfreiheit

Als digitales Selbsttracking bezeichnet man das Bestreben, mittels technischer Hilfsmittel wie intelligenter Uhren, Fitnessarmbänder, *Smart Rings* oder dem Smartphone und spezieller Apps möglichst viele Aspekte des eigenen Lebens zu erfassen, zu analysieren und auszuwerten. Im Fokus dieser Bemühungen stehen dabei vor allem die körperlichen Vitaldaten (bspw. Puls, Blutdruck, Blutsauerstoffsättigung), sportliche Aktivitäten (Dauer und Intensität) und Erholungsphasen (Dauer und Effektivität) sowie die entsprechende Energiezufuhr (wann wurde was und wie viel aufgenommen).[38] Das Ziel dieses umfassenden Selbstvermessens besteht darin, den eigenen Körper besser kennen zu lernen, und jenseits „bloßer" Eigenwahrnehmungen, die häufig als irreführend und nicht vertrauenswürdig eingestuft werden,[39] Gewissheit über dessen Funktionsmechanismen, bspw. das Wechselwirken verschiedener körperlicher Parameter, seinen Gesundheitszustand und seine Leistungsfähigkeit zu erhalten.[40] Gleichzeitig soll diese körperliche Gewissheit, vor allem die Einsicht in die Auswirkungen bestimmter Aktivitäten auf Leistungsparameter, bspw. eines erholsamen Schlafs auf die geistige Leistungsfähigkeit, sportlicher Aktivität auf die Gemütslage oder gezielter Nahrungsmittelzunahme auf die Körperkonstitution, den Selbsttracker:innen dabei helfen, ihren Körper gezielt zu beeinflussen und durch das Anpassen der entsprechenden Parameter in eine gewünschte Richtung steuern zu können.[41] Selbstvermessung erscheint somit als Werkzeug, körperliche Ungewissheiten zu

---

[38] Zwar existierte derartiges Vermessungsbesteben schon lange bevor es die entsprechenden digitalen Tools hierfür gab, doch heben digitale Protokollapps oder *Trackingdevices* das Selbsttracking auf eine neue Ebene. Indem sie es erlauben, Aspekte des eigenen Körpers zu vermessen, die sich bislang nicht oder nur mühsam feststellen ließen, bspw. die tägliche Schrittzahl oder die Anzahl und Dauer nächtlicher Tiefschlafphasen, eröffnen sie neue Vermessungs- und Auswertungsmöglichkeiten. Andererseits ergeben sich ganz neue Erkenntnismöglichkeiten, wenn nicht nur die eigenen Daten zur Verfügung stehen, sondern sich die Daten Millionen anderer Nutzer:innen miteinander abgleichen und analysieren lassen.

[39] Lupton, Deborah: Understanding the Human Machine. In: IEEE Technology and Society Magazine 32/4, 25–30.

[40] Vgl. Lupton: Understanding the Human Machine; Sharon, Tamar: Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. In: Philosophy & Technology 30/1, 91–121.

[41] Vgl. Sharon: Self-Tracking.

überwinden, ein höheres Maß an körperlicher Gewissheit herzustellen und dadurch Entscheidungsfreiheiten zu gewinnen.

Gleichsam entspricht nicht jeder Wunsch der Wirklichkeit und oft sehen die Resultate ganz anders aus als ursprünglich intendiert. Aus diesem Grund ist es sinnvoll zu fragen, wie sich digitales Vermessungspraktiken denn *tatsächlich* auf die Selbsttracker:innen und deren Entscheidungsfreiheit auswirken. In ihrer Publikation *The Quantified Self* geht Deborah Lupton dieser Frage nach und hält dabei mehrere Beobachtungen fest.[42]

Als erste Folge einer kontinuierlichen Selbstvermessung beobachtet Lupton, dass sich die Wahrnehmung der Selbstquantifizierer:innen überforme, sodass sie die Welt und sich selbst fortan immer mehr und immer ausschließlicher aus einer „Zahlenperspektive" wahrnehmen. In eindrücklicher Weise führt Lupton diesen Perspektivenwandel am Beispiel von Sexualaktivitäten vor:

> Sexual activity becomes reduced to 'the numbers': how long intercourse lasts for, how often it takes place, how many thrusts are involved, the volume of sound emitted by participants, how good it is, with how many partners and so on. The comparisons with other users that some of these apps allow for emphasise the notion of sexual experience as a performance, as an activity that can and should be compared with the sexual experiences of others, since they are all rendered into digital data form. These technologies therefore act to support and reinforce highly reductive and normative ideas of what is 'good sex' and 'good performance' by encouraging users to quantify their sexual experiences and feelings in ever finer detail and to represent these data visually, in graphs and tables. The discourses of performance, quantification and normality suggest specific, limited types of sexuality.[43]

Hinzu kommt, wie in obiger Darstellung bereits mitschwingt, dass mit jeder Quantifizierung immer auch eine Wertsetzung verbunden ist. Entweder die *Trackingdevices oder -softwares* selbst oder die Selbstvermesser:innen bewerten manche Zahlen als gut, andere hingegen als schlecht. Über 10 000 Schritte am Tag: gut – weniger als 10 000 Schritte am Tag: schlecht.[44]

---

42 Vgl. Lupton, Deborah: The Quantified Self. Cambridge, Malden 2016.

43 Lupton: The Quantified Self, 103–104. Dass man einer solchen zahlenmäßigen Wahrnehmungsüberforderung nicht einfach vorbeugen kann, indem man Personen die Möglichkeit gibt, darüber zu entscheiden, welche Aspekte ihres Lebens quantifiziert und ausgewertet werden, zeigt: Tretter, Max: Perspectives on digital twins and the (im)possibilities of control. In: Journal of Medical Ethics 47/6 (2021), 410–411.

44 Ganz abgesehen davon, dass die Zielmarke von 10 000 Schritten täglich eher zufällig denn medizinisch belegt ist. Vgl. Reynolds, Gretchen: Do We Really Need to Take 10,000 Steps a Day for Our

Dies hat zweierlei Auswirkungen. Auf der einen Seite bereitet diese All-Bewertung einem Streben nach immer „besseren" Zahlen, einem immer „besseren" Selbst den Weg. Denn, wie Hartmut Rosa in seinen Unverfügbarkeitsüberlegungen festhält:

> ist [es] nahezu unmöglich, die tägliche Schrittzahl zu messen, ohne versucht zu sein, sie zu steigern bzw. zu optimieren. Es erweist sich als eine lebenspraktische Illusion zu glauben, man ließe sich von den Daten, sind sie erst einmal verfügbar, nicht zu entsprechenden Verhaltensänderungen verleiten.[45]

Dieser, wie man mit Sloterdijk formulieren könnte,[46] anthropotechnische Optimierungsimperativ formt die Selbstvermesserin zu einem „unternehmerischen Selbst"[47] und macht sie dadurch zunehmend anfällig für äußere Anreize oder externes Nudging.[48]

Auf der anderen Seite kann das wertsetzende Quantifizieren dazu führen, dass alles, was sich nicht messen und dem sich kein präziser Zahlenwert zuschreiben lässt, zunehmend abqualifiziert wird – denn es hat ja keine Auswirkungen auf das „Datenselbst"[49] und folglich keinen Wert. In Extremfällen kann dies dazu führen, dass jeder Schlaf, der nicht getracked und dessen Erholsamkeit nicht errechnet wurde, nicht zählt, jedes Training, das nicht gelogged und auf Social Media geteilt wurde, als nicht stattgefunden gilt[50] und dass überzeugte Selbstquantifizierer:innen sogar ihre eigene Existenz jenseits des Selbstvermessens in Zweifel zu ziehen beginnen.[51]

Insgesamt zeigen die Ausführungen Luptons und ihrer Mitstreiter:innen, dass digitale Selbstvermessungspraktiken Auswirkungen haben können, die ihrer ursprünglichen Intention konträr entgegenstehen. Zwar kann ein intensives Selftracking verlässlichere Körperkenntnisse und neue Gewissheiten schaffen und damit eventuell auch neue Entscheidungsfrei-

Health? Online unter: https://www.nytimes.com/2021/07/06/well/move/10000-steps-health.html (Stand: 18.10.2022).

45 Rosa, Hartmut: Unverfügbarkeit. Wien/Salzburg 2019, 87.

46 Vgl. Sloterdijk, Peter: Du mußt Dein Leben ändern. Frankfurt a. M. 2009.

47 Vgl. Bröckling, Ulrich: Das unternehmerische Selbst. Soziologie einer Subjektivierungsform. Frankfurt a. M. 2013.

48 Vgl. Thaler, Richard H./Sunstein, Cass: Nudge. Wie man kluge Entscheidungen anstößt. Berlin 2019.

49 Vgl. Lupton, Deborah: Data Selves. Medford 2020.

50 Vgl. Lupton: The Quantified Self, 61.

51 Vgl. Thomas, Scarlett: Nowhere to run: did my fitness addiction make me ill? Online unter: https://www.theguardian.com/lifeandstyle/2015/mar/07/fitness-addiction-ill-scarlett-thomas(Stand: 31.08.2022).

mütigkeit produzieren.[52] Umgekehrt kann es durch die Quantifizierung des eigenen Körpers jedoch auch zu Wahrnehmungsüberformungen kommen, durch die manche Entscheidungsoptionen in den Vorder-, andere hingegen in den Hintergrund gedrängt werden – abhängig davon, wie sie sich auf die Zahlen des Selbst auswirken werden. Zudem kann eine Orientierung an Zahlenwerten interne Zwänge befördern und öffnet externen Anreizen und Nudgings Tür und Tor. Während es demnach nicht ausgeschlossen ist, dass Selbstquantifizierung einigen zu mehr Freimütigkeit beim Entscheiden verhilft, kann es umgekehrt auch Entscheidungsfreiräume verschließen – und beeinflusst die menschliche Entscheidungsfreiheit damit in ambivalenter Weise.

4.2 Die Auswirkungen umfassender KI-Prognosen auf die menschliche Entscheidungsfreiheit

Nachdem im letzten Kapitel aufgezeigt wurde, dass digitales Selbsttracking die menschliche Entscheidungsfreiheit in ambivalenter Weise beeinflussen kann, soll nun gefragt werden, ob dasselbe auch für Gewissheiten gilt, die aus KI-Prognosen gewonnen wurden. Da der bislang publizierte Forschungsdiskurs zu den Auswirkungen künstlich-intelligenter Prognosen auf den Menschen und seine Freiheiten noch recht überschaubar ist, braucht es grundlegender Pionierarbeit auf diesem Gebiet. Diese soll auf konzeptioneller Ebene mittels einer Auseinandersetzung mit dem Denken Jean Baudrillards geschehen. Als einer der Denker, die sich am umfassendsten mit dem Simulationskonzept und den Auswirkungen umfassender Simulationen auf den Menschen und die Gesellschaft auseinandergesetzt hat, stellt er die Fragen und greift die Begriffe auf, die für das Nachdenken über KI-Prognosen und Entscheidungsfreiheit wichtig sind. Auch wenn er dabei abweichende Schwerpunkte setzt und Konzepte anders versteht, präsentiert Baudrillards viele Gedankengänge, die sich produktiv in die ausstehenden konzeptionellen Überlegungen einbringen lassen. Nach einer kurzen Einführung in dessen simulationstheoretische Überlegungen und eine Skizze, welche Implikationen Baudrillard einer umfassenden Simulation für die menschliche Freiheit zuschreibt, werden aus diesen Gedanken Schlussfolgerungen über die Auswirkungen KI-getriebener Gewissheiten auf die Entscheidungsfreiheit gezogen.

---

[52] Umgekehrt kann eine intensive Selbstvermessung aber auch ein neues Bewusstsein für die nichtquantifizierbaren und nichtkontrollierbaren Aspekte des eigenen Körpers schaffen, bestehende Krankheiten, Schwächen oder Abhängigkeiten, und dadurch neue Ungewissheiten ins Zentrum der Aufmerksamkeit rücken. Vgl. Lupton, Deborah: Digital Health. Critical and Cross-Disciplinary Perspectives. London, New York 2018.

In seinem programmatischen Hauptwerk *Simulacra and Simulation* formuliert Baudrillard die These, die sich seit den 1980ern durch all seine Werke zieht – dass die Welt des späten 20. Jahrhunderts eine vollständige Simulation darstelle.[53] Im Unterschied zu Gehirn-im-Tank-Überlegungen[54] oder Wir-leben-in-einer-Computersimulation-Annahmen,[55] beruht die baudrillardsche Simulationstheorie nicht auf bewusstseinsphilosophischen oder posthumanistischen, sondern auf symboltheoretischen Überlegungen. An die Stelle der Realität, so Baudrillard, seien zunehmend Zeichen getreten. Repräsentierten diese Zeichen als „Simulakren erster Ordnung" anfangs noch eine „hinter ihnen" liegende Realität, ging ihnen diese Verweisstruktur sukzessive verloren. Als „Simulakren zweiter Ordnung" referierten Zeichen strukturalistisch bald fast nur noch aufeinander und maskieren damit, dass die „Realität" hinter den Zeichen allmählich verwaist. Diese Prozesse gleichzeitiger zeichenhafter Selbstreferentialität und Realitätsverkümmerung resultierten schließlich in einer Welt, in der „Simulakren dritter Ordnung" ausschließlich miteinander interagieren und eine geschlossene „Simulation" bilden. Aus dieser Simulation gebe es keinen Ausweg mehr, denn das, was einst das „Reale" gewesen war, habe sich nun, d. h. in den Medien- und Kommunikationslandschaften der späten 1970er Jahre, vollständig in der Zirkulation der Zeichen aufgelöst und sei ununterscheidbar von der Simulation geworden.[56]

Was an diesen Überlegungen fasziniert, sind weniger die medientheoretischen, philosophischen oder psychoanalytischen Versuche Baudrillards, seine steile These einzuholen, sondern eher die Impulse zur Gegenwartsdeutung, die sich aus ihr ergeben.[57] Denn aus dem „Theorem der Simulation"[58] folge, wie Baudrillard konsequent herausarbeitet, dass mit der Eigenständigkeit der Realität auch die Möglichkeit von Ereignissen, d. h. das Geschehen von etwas, was nicht zuvor simulakrisch vorgezeichnet ist, von etwas, das wahrhaft „neu" und „außerordentlich" sei, verlorengehe.[59] Vielmehr folge alles, was in der Simulation geschehe, einzig der Struktur der Zeichen. Erst diese Zeichen leiten das Geschehende in seine Bahnen und machen es dadurch nicht nur wahrnehmbar, sondern auch „wirklich".[60] Folge die Welt jedoch der Bahn

---

53 Vgl. Baudrillard, Jean: Simulacra and Simulation. Michigan 1994.

54 Vgl. Putnam, Hilary: Vernunft, Wahrheit und Geschichte. Frankfurt a. M. 1990.

55 Vgl. Bostrom, Nick: Are We Living in a Computer Simulation?, In: The Philosophical Quarterly 53/211 (2003), 243–255.

56 Vgl. Baudrillard, Jean: Agonie des Realen. Berlin 1978, 7–69.

57 Vgl. Strehle, Samuel: Zur Aktualität von Jean Baudrillard. Einleitung in sein Werk. Wiesbaden 2012.

58 Vgl. Blask, Falko: Jean Baudrillard zur Einführung. Hamburg 2002, 23.

59 Einzig im Einbruch des radikal Bösen, das als dialektischer Einbruch in die allumfassende Simulation einfalle, seien unvorhergesehene Ereignisse weiterhin möglich. Vgl. Baudrillard, Jean: Die Intelligenz des Bösen. Wien 2010.

60 Vgl. Baudrillard: Simulacra and Simulation.

der Zeichen, ist in und mit diesen alles bis ins kleinste Detail vorausgeplant, durchsimuliert und orchestriert. Kurz, die Welt sei „integral" geworden.[61]

Für die einzelne Person innerhalb der Simulation bedeute dies, dass sie in einer totaldeterminierten Welt lebe und mit ihrem Denken nur das nach-denke, mit ihren Entscheidungen nur das nach-vollziehe und mit ihren Handlungen nur das nach-mache, was bereits in der Simulation der Zeichen vorhanden und ihr als Option vor-gegeben ist.[62] Darüber hinaus sind ihr auch die Folgen ihres Denkens, Entscheidens und Handelns vorgegeben, sodass all ihr Verhalten und dessen Auswirkungen von vornherein feststehen – und ihr totale Gewissheit verschaffen. Für ihre Entscheidungsfreiheit bedeutet dies zweierlei. Auf der einen Seite ermöglicht ihr dies eine maximale Freimütigkeit. Denn wissend darum, was geschieht und was folgt, muss sich die Person keine Sorgen mehr um falsche Entscheidungen oder zukünftige Handlungsfolgen machen. Auf der anderen Seite bedeutet diese Totalgewissheit jedoch auch, dass es de facto nur *eine* tatsächliche Option gebe. Jenseits der in den Zeichen festgelegten Gedanken, Entscheidungen und Handlungen, gibt es keine anderen Denk-, Entscheidungs- und Handlungsalternativen mehr. Der Entscheidungsraum ist gleich null bzw. eins.

Hebt die umfassende Simulationsgewissheit im Sinne Baudrillards die Freimütigkeit bei Entscheidungen einerseits auf ein Maximum, denn bereits bevor eine Person sich entscheidet, weiß sie um die Folgen ihrer Handlungen, reduziert sie gleichermaßen die Entscheidungsräume auf ein Minimum – und treibt damit die Ambivalenz der Gewissheit bis zum Äußersten.

Natürlich lassen sich diese baudrillardschen Überlegungen und seine Simulations-Freiheits-Schlussfolgerungen nicht eins zu eins auf die Frage nach den Freiheitsauswirkungen KI-getriebener Prognosen und der durch sie produzierten Gewissheiten übertragen. Dies verbietet allein das unterschiedliche Verständnis des Simulationsbegriffs. Denn im Unterschied zu Baudrillard, beschreibt Simulation im Kontext dieses Beitrags *nicht* die umfassende Grundstruktur einer integralen Wirklichkeit, sondern bezeichnet spezifische Prognosen über kleinräumige Weltausschnitte, bspw. technisch generierte Voraussagen über die gesundheitliche Entwicklung einer Person und unterliegen damit, wie Armin Grunwald betont, einigen epistemischen, zeitphilosophischen und wissenschaftstheoretischen Vorbehalten.[63]

---

61 Vgl. Baudrillard: Intelligenz des Bösen.

62 In besonders perfider Weise sei, so Baudrillard, sogar der Versuch, aus diesen Simulationszirkeln auszubrechen, bereits als Skript in der Simulation angelegt, sodass jeder Durchbrechungsversuch die Simulationsstruktur letztlich nur bestärke – und es jenseits von Fatalismen kein Entkommen aus der Simulation gebe. Vgl. Baudrillard, Jean: Die fatalen Strategien. Berlin 1991.

63 Vgl. Grunwald, Armin. Digitalisierung und Künstliche Intelligenz. Hoffnung auf bessere Prognosen? In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 195–214.

Nichtsdestotrotz lassen sich aus den Überlegungen Baudrillards Impulse für das Nachdenken über die Freiheitswirksamkeit KI-getriebener Prädiktionsgewissheiten gewinnen. So ist in Analogie zu den baudrillardschen Schlussfolgerungen davon auszugehen, dass auch weniger präzise und auf kleinere Bereiche zugeschnittene KI-Simulationen, bspw. im medizinischen Bereich, Gewissheiten erzeugen, die gleichermaßen Freimütigkeit zur Entscheidung schaffen wie sie Entscheidungsräume einschränken. Beides lässt sich auf anekdotische Weise am Umgang mit den Empfehlungen hochfunktionaler klinischer KI-Systeme verdeutlichen. Wo diese Systeme den Betroffenen nachvollziehbar aufzeigen, welche Behandlungsoptionen zur Verfügung stehen und welche Erfolgschancen, aber auch Risiken sie haben, können erstere ihre Entscheidung informierter und freier treffen. Ihre Entscheidungsfreiheit vergrößert sich. Umgekehrt stellt sich jedoch die Frage, wer es sich zutraut, den Empfehlungen dieser künstlich-intelligenten Kliniksysteme zu widersprechen und sich trotz vorliegender Prädiktionen und Vergleichssimulationen gegen die von ihnen empfohlene und eigenmächtig für eine andere Behandlungsoption zu entscheiden. Außer wenigen Maschinenstürmer:innen wohl kaum jemand. Damit zeigt dieses kurze Gedankenspiel, dass auch imperfekte KI-Prädiktionen, wie sie gegenwärtig möglich sind, Tendenzen einer baudrillardschen Totalsimulation aufzeigen, d. h. die vernünftig-wählbaren Entscheidungsräume einschränken, dafür aber die Sorge, man könne gegebenenfalls eine falsche Entscheidung treffen, reduzieren und Freimütigkeit zur Entscheidung generieren.

5 Theologische Perspektiven auf gegenwärtige Gewissheitsbestrebungen

Im vorherigen Kapitel wurde die anfängliche These eingeholt und aufgezeigt, dass digitale oder KI-getriebene Gewissheiten die menschliche Entscheidungsfreiheit gleichermaßen erweitern wie einschränken. Damit ist die Eingangsfrage beantwortet. Allerdings wirft die Antwort und der Hinweis auf die ambivalenten Freiheitswirksamkeiten künstlich-intelligenter Gewissheitsbestrebungen die Folgefrage auf, welche *praktischen* Konsequenzen dieser Einsicht folgen. Ist es trotz aller Ambivalenz empfehlenswert, neue Gewissheiten anzustreben oder sollte man umgekehrt sämtliche Versuche, digital oder künstlich-intelligent Gewissheit zu erzeugen, meiden? Und welchen Leitlinien sollte der Umgang mit gewissheitsschaffenden Technologien folgen? Um diesen Beitrag nicht mit dem bloßen Ausweis der Ambivalenz gegenwärtiger Gewissheitsbestrebungen und offenen Fragen zu beenden, soll ein zusätzlicher Rekurs auf die Theologie, genauer eine theologische Ethik protestantischen Zuschnitts, dabei helfen, einen angemessenen Umgang mit gewissheitsproduzierenden Digital- und KI-Technologien zu finden.

Einen ersten Impuls zur Einordnung digital- bzw. KI-gestützten Gewissheitsbestrebungen kann Peter Dabrocks Beitrag *Geheimnis, Freiheit, Verzeihen* liefern.[64] In diesem setzt er sich mit den Prädiktionsmöglichkeiten von Big Data auseinander und kontrastiert sie in kreativer Weise mit der göttlichen Vorhersehungslehre. Die datengestützten Vorhersagemöglichkeiten seien allein schon deshalb zu kritisieren, weil sie den Menschen für „medizinische, ökonomische, politische, militärische Ziele vereinnahmen"[65] und verzwecken, dadurch seine Freiheit einschränken und, ganz allgemein, das Wohlwollen vermissen lassen, das der Providenz Gottes innewohne. Diese Kritik spitzt der Ethiker in einer weiteren Argumentationsschleife noch stärker zu, indem er die bonhoeffersche Unterscheidung zwischen „Letztem" und „Vorletztem" aufgreift, d. h. zwischen den „Dingen", über die allein Gott verfügen darf und die einzig durch seine Rechtfertigung verwirklicht werden können auf der einen und den weltlich vorausgehenden „Dingen", die im Verfügungsbereich des Menschen stehen auf der anderen Seite,[66] und sie kritisch gegen Big Data Dynamiken ins Feld führt.

Denn indem sie Menschen Gewissheiten liefert, die ihnen vermeintlich mehr Gesundheit, tieferes Wohlergehen und umfassendere Freiheit einbringen, verheißt Big Data Dinge, die weit außerhalb ihrer Machbarkeitsraums liegen. Zwar kann sie neue Gewissheiten schaffen, die wiederum dabei helfen können, bestimmte Probleme besser adressieren und eventuell auch lösen zu können. Damit kann sie einen gewissen Beitrag zur Erhaltung oder Wiederherstellung der Gesundheit, zum menschlichen Wohlergehen oder zur individuellen Freiheit leisten. Doch beschränken sich diese technischen Möglichkeiten auf das vorletzte Hier und Jetzt – die letzte Verwirklichung dieser Dinge bleiben ihr schlussendlich doch entzogen und allein dem Letzthandeln Gottes unterstellt. Wo Big Data Technologien diesen fundamentalen Unterschied ausblenden, ihre eigene Begrenztheit ignorieren und sich als Gewissheits- und Lebensgelingensgaranten stilisieren, vermischen sie auf unzulässige Weise Letztes und Vorletztes und setzen sich in maßloser Selbstüberschätzung an die Stelle Gottes.[67] Diese Hybris, gottgleich über Letztes verfügen zu wollen und Dinge zu verheißen, die außerhalb der menschlich-technischen Verfügungsgewalt stehen, ist als sündig zu entlarven – und der Mensch sei am besten

---

64 Vgl. Dabrock, Peter: Geheimnis, Freiheit, Verzeihen. Warum Big Data an die Lehre von der Vorsehung erinnert. In: Zeitzeichen 11 (2014), 20–23.

65 Ebd., 23.

66 Vgl. Bonhoeffer, Dietrich. Ethik. München 1992, 137–162. In diesem Kapitel legt Bonhoeffer ausführlich seine Unterscheidung der letzten und der vorletzten Dinge und deren komplexe Verhältnis dar. Im ihrer wechselseitigen Bezogenheit geht das Vorletzte dem Letzten irdisch voraus, wird aber gleichermaßen erst durch das Letzte als Vorletztes eingesetzt. Diese Einsetzung geht mit dem Verantwortungsruf einher, dem Letzten bereits im Hier und Jetzt den Weg zu bereiten – dabei aber darum zu wissen, dass das Letzte selbst schlussendlich ausschließlich von Gott gewirkt werden kann.

67 Vgl. Bonhoeffer, Dietrich. Schöpfung und Fall. München 1989, 103–113.

beraten, diejenigen Technologien zu meiden, die sich derart gerieren, um nicht in deren Sog zu geraten und seine Hoffnung an ihre gewissheitsbringenden Kalkulations- und Berechnungsleistungen zu hängen.[68]

Dies bedeutet im Umkehrschluss jedoch nicht, dass *jedes* Streben nach Gewissheit als sündige Anmaßung abzuqualifizieren und *jeder* Technik- oder KI-Gebrauch zu Gewissheitszwecken fundamentalkritisch abzulehnen sei.[69] Vielmehr gebe es umgekehrt, wie Mikkel Gabriel Christoffersen in seinen risikotheologischen Darstellungen aufzeigt, Situationen, in denen es verantwortungsethisch geradezu *geboten* sei, nach weiteren Gewissheiten zu streben und in denen der Verzicht auf den Einsatz gewissheitsbringener Technologien als nicht nur töricht, sondern wiederum als sündig einzustufen sei. Dies sei besonders dort der Fall, wo es um das Wohlergehen der Mitmenschen gehe und um Gewissheiten, die sich ohne größere, negative Begleiterscheinungen produzieren lassen und dazu beitragen, andere vor unnötigem Schaden zu bewahren.[70] Denn gerade in solchen Fällen würde das Ablehnen von Gewissheitstechnologien gleichermaßen bedeuten, die Gewissheiten auszuschlagen, die es mir ermöglichen, meinem Nächsten besser zu helfen. Ein solcher Verzicht käme dabei dem Handeln des Priesters oder des Leviten in der lukanischen Geschichte des barmherzigen Samariters gleich (Lk 10,25–37), die, um sich buchstäblich nicht „die Hände schmutzig"[71] zu machen und gegebenenfalls ihrer Pflicht nicht mehr nachkommen zu können, ihrem Nächsten einen (lebens-)rettenden Dienst verweigern.[72] Vor diesem Hintergrund erweist sich auch der Pauschalverzicht auf jegliche Technik oder KI zu Gewissheitszwecken als lieblos und nicht verantwortungsgemäß.

Angesichts dieser beiden Extreme ist der Mensch dazu aufgerufen, weder die Kontingenzen des eigenen Lebens um jeden Preis überwinden zu suchen und „sein Herz" vermessen an gewissheitsschaffende Digital- und KI-Technologien zu hängen, noch gegenteilig die Möglichkeiten der technischen Gewissheitsproduktion grundlegend zu verteufeln und dadurch leichtfertig das Wohlergehen seiner Nächsten aufs Spiel zu setzen und in die Lieb- und Verantwortungslosigkeit abzugleiten. Vielmehr sei er, so lässt sich als Schlussfolgerung festhalten,

---

68 Vgl. Dabrock: Geheimnis, Freiheit, Verzeihen.

69 Vgl. Braun, Matthias: Ein Selfie für Alexa? Künstliche Intelligenz als Herausforderung für die theologische Ethik. In: Zeitzeichen 7 (2018), 29–31.

70 Vgl. Christoffersen, Mikkel G. Living with Risk and Danger. Studies in Interdisciplinary Systematic Theology. Göttingen 2019.

71 An dieser Stelle wird vielfach darauf verwiesen, dass für beide Berufsgruppen die rituelle Reinheit von äußerster Wichtigkeit war und sie jeglichen Kontakt zu Verstorbenen, zu denen der Überfallene augenscheinlich zählte (Lk 10,30), zu vermeiden hatten. Vgl. Fitzmyer, Joseph: Luke the theologian. Aspects of his teaching. New York 1989, 887.

72 Vgl. Bovon, Fançois: Das Evangelium nach Lukas (Lk 9,51–14,35). Zürich/Düsseldorf 1996, 79–99.

aus evangelisch-theologischer Perspektive dazu angehalten, in verantwortungsbewusster Weise zwischen beiden Extremen hindurchzuschiffen und darin seinen Weg zu finden.

Für das Ausloten dieses Korridors lässt sich keine universelle Formel aufstellen, denn wie viel Digital- und KI-Gewissheit jeweils anzustreben sei und wann diese Bestrebungen ins Anmaßende zu kippen drohen, ist stets von der konkreten Situation abhängig. Dennoch lassen sich theologisch einige Leitlinien aufstellen, die bei der Orientierung helfen können. So darf und soll der Mensch, wie man im Anschluss an Dabrock und Christoffersen festhalten kann, die ihm zur Verfügung stehenden technischen Möglichkeiten verantwortlich und frei nutzen, um die Welt zu gestalten und sich und vor allem seinen Nächsten vor vermeidbaren Schäden zu bewahren. Dabei ist er dazu angehalten, abgeklärt vorzugehen, sich nicht von den überzogenen Verheißungen der Technologie umgarnen zu lassen, die Grenzen der Verfügbarkeit, des Vorausberechnens und des Gewissmachens zu achten und damit zu rechnen, dass er weiterhin mit „Überraschendem, Unvorhergesehenem, Fremdem, Unverrechenbarem (!) und Außerordentlichem“[73] konfrontiert sein wird. Dass er dabei auch scheitern, mal *zu sehr* auf die Möglichkeiten der Technik vertrauen, mal *zu wenig* der Hilfsverantwortung gegenüber seinem Nächsten nachkommen wird, ist nicht zu vermeiden. Doch darf er in all seinem zwangsläufigen Scheitern auf Vergebung hoffen und mit der göttlichen Wiederherstellung „rechnen“.[74]

Diese theologischen Orientierungsüberlegungen zusammenfassend, lassen sich vier Leitlinien identifizieren – Verantwortung und Freiheit, Abgeklärtheit und Vergebungshoffnung –, die den Menschen in seinem Umgang mit digitalen wie KI-Technologien und ihrer Nutzung zu Gewissheitszwecken begleiten und sein Handeln ausrichten können.

6 Zusammenfassung

Ausgehend von den gegenwärtigen Ungewissheiten, die aus der Schnelllebigkeit und Komplexität der Gegenwartsgesellschaft erwachsen, und welche die Entscheidungsfreiheit des Menschen beeinträchtigen sowie der Beobachtung, dass es immer mehr Möglichkeiten gibt, Ungewissheiten digital zu überwinden oder durch KI neue Gewissheiten zu schaffen, lautete die Leitfrage dieses Beitrags, wie neue Gewissheitsbestreben sich auf die menschliche Entscheidungsfreiheit auswirken. Um diese Frage zu beantworten, wurden zuerst in mehreren überblicksschaffenden Kapiteln die Ungewissheits- wie Gewissheitskonzepte vorgestellt, ebenso wie ein Einblick in die verschiedenen Möglichkeiten der Gewissheitsproduktion gegeben. Im anschließenden argumentativen Hauptkapitel wurde erstens herausgearbeitet, wie digitale

[73] Dabrock: Geheimnis, Freiheit, Verzeihen, 23.

[74] Vgl. Bonhoeffer: Ethik, 245–299.

Selbstvermessungsbestreben, als Bemühungen um mehr Körpergewissheit, im gleichen Maße Gewissheiten produzieren und Entscheidungsfreimütigkeit schaffen, aber auch die verfügbaren Entscheidungsoptionen einschränken. Mittels einer konzeptionellen Auseinandersetzung mit Jean Baudrillard wurde zweitens gezeigt, dass Simulationsgewissheiten eine ähnliche Wirkung entfalten, freimutschaffende Gewissheiten erzeugen, aber die Entscheidungsoptionen auf ein Minimum reduzieren. Dies führte zu der Schlussfolgerung, *dass wachsende, mittels digitaler oder KI-Technologien produzierte Gewissheiten die menschliche Entscheidungsfreiheit auf ambivalente Weise beeinflussen*. Während sie auf der einen Seite ein gewisses Maß an Entscheidungsfreimütigkeit wiederherstellen und so die Entscheidungsfreiheit steigern können, schränken sie auf der anderen Seite den Raum möglicher Entscheidungen stückweise ein – und sorgen so für verringerte Entscheidungsfreiheiten. In einem weiterführenden Abschlusskapitel wurden diese Ergebnisse schließlich theologisch perspektiviert, nach Leitlinien zum Umgang mit gewissheitsschaffenden Digital- und KI-Technologien gefragt und eine Orientierung an Verantwortung, Freiheit und Vergebung als theologisch-ethische Wegmarken eines solchen Umgangs identifiziert.

Literaturverzeichnis

Agrawal, Ajay/Gans, Joshua/Goldfarb, Avi: Prediction Machines. The Simple Economics of Artificial Intelligence. Boston 2018.

Assmann, Jan: Zeitkonstruktion, Vergangenheitsbezug und Geschichtsbewußtsein im alten Ägypten. In: Assmann, Jan/Müller, Klaus E. (Hg.): Der Ursprung der Geschichte. Archaische Kulturen, das Alte Ägypten und das Frühe Griechenland. Stuttgart 2005, 112–214.

Baudrillard, Jean: Die fatalen Strategien. Übersetzt von: Wiener, Oswald. Berlin 1991.

Baudrillard, Jean. Simulacra and Simulation. Übersetzt von: Glaser, Sheila Faria. Michigan 1994.

Baudrillard, Jean. Die Intelligenz des Bösen. Übersetzt von: Winterhalter, Christian. Wien ^{2}2010.

Baudrillard, Jean. Agonie des Realen. Übersetzt von: Kurzawa, Lothar/Schaefer, Volker. Berlin 1978.

Bauman, Zygmunt. Flüchtige Moderne (edition suhrkamp). Übersetzt von: Kreissl, Reinhard. Frankfurt a. M. 2003.

Bauman, Zygmunt. Liquid Fear. Cambridge, Malden 2006.

Bauman, Zygmunt. Flüchtige Zeiten. Leben in der Ungewissheit. Übersetzt von: Barth, Richard. Hamburg 2008.

Bauman, Zygmunt. Leben in der flüchtigen Moderne (edition suhrkamp). Übersetzt von: Jakubzik, Frank. Frankfurt a. M. 2010.

Blask, Falko. Jean Baudrillard zur Einführung (Zur Einführung). Hamburg ^{2}2002.

Bonhoeffer, Dietrich. Schöpfung und Fall (Dietrich Bonhoeffer Werke 3). Herausgegeben von: Rüter, Martin/Tödt, Ilse. München ^{3}1989.

Bonhoeffer, Dietrich. Ethik (Dietrich Bonhoeffer Werke 6). Herausgegeben von Tödt, Ilse/Tödt, Heinz Eduard/Feil, Ernst/Green, Clifford. München 1992.

Bovon, Fançois: Das Evangelium nach Lukas (Lk 9,51–14,35) (Evangelisch-Katholischer Kommentar zum Neuen Testament III/2). Zürich/Düsseldorf 1996.

Bostrom, Nick. Are We Living in a Computer Simulation? In: The Philosophical Quarterly, 53/211 (2003), 243–255.

Braun, Matthias: Ein Selfie für Alexa? Künstliche Intelligenz als Herausforderung für die theologische Ethik. In: Zeitzeichen 7 (2018), 29–31.

Bröckling, Ulrich. Das unternehmerische Selbst. Soziologie einer Subjektivierungsform (suhrkamp taschenbuch wissenschaft 1832). Frankfurt a. M. [5]2013.

Christoffersen, Mikkel G. Living with Risk and Danger. Studies in Interdisciplinary Systematic Theology (Forschungen zur systematischen und ökumenischen Theologie 165). Göttingen 2019.

Dabrock, Peter. Befähigungsgerechtigkeit. Ein Grundkonzept konkreter Ethik in fundamentaltheologischer Perspektive. Gütersloh 2012.

Dabrock, Peter. Geheimnis, Freiheit, Verzeihen. Warum Big Data an die Lehre von der Vorsehung erinnert. In: Zeitzeichen 11 (2014), 20–23.

DeLanda, Manuel. A New Philosophy of Society. Assemblage and Social Complexity. London, New York 2006.

Dewey, John. Die Suche nach Gewißheit. Eine Untersuchung des Verhältnisses von Erkenntnis und Handeln. Übersetzt von: Suhr, Martin. Frankfurt a. M. 1998.

Elgendy, Nada/Elragal, Ahmed. Big Data Analytics in Support of the Decision Making Process. In: Procedia Computer Science, 100 (2016), 1071–1084. DOI: 10.1016/j.procs.2016.09.251.

Fitzmyer, Joseph: Luke the theologian. Aspects of his teaching. New York 1989.

Glassmeier, Karl-Heinz. Wetten, Wissen, Werten. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin/Boston 2021, 176–194. DOI: 10.1515/bthz-2021-0011.

Greif, Hajo: Modellierung und Simulation in der Künstlichen Intelligenz. In: Mainzer, Klaus (Hg.): Philosophisches Handbuch Künstliche Intelligenz. Wiesbaden 2019, 1–21. DOI: 10.1007/978-3-658-23715-8_26-1.

Grunwald, Armin: Digitalisierung als Prozess. Ethische Herausforderungen inmitten allmählicher Verschiebungen zwischen Mensch, Technik und Gesellschaft. In: Zeitschrift für Wirtschafts- und Unternehmensethik 20/2 (2019), 121–145.

Grunwald, Armin. Digitalisierung und Künstliche Intelligenz. Hoffnung auf bessere Prognosen? In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 195–214. DOI: 10.1515 /bthz-2021-0012.

Hausmann, Jutta: Weisheit (AT). Online unter: https://www.bibelwissenschaft.de/stichwort/34707/ (Stand: 30.08.2022).

Heessel, Nils P.: Die altorientalische Divination als antikes Wissenschaftssystem. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 48–66. DOI: 10.1515 /bthz-2021-0005.

Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen: Einleitung. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin, Boston 2021, 1–6. DOI: 10.1515 /bthz-2021-0002.

Horn, Christoph/Löhrer, Guido (Hg.): Gründe und Zwecke. Texte zur aktuellen Handlungstheorie (suhrkamp taschenbuch wissenschaft 1950). Berlin 2010.

Johnson, L. Syd: The Ethics of Uncertainty. Entangled Ethical and Epistemic Risks in Disorders of Consciousness. New York 2022. DOI: 10.1093/med/9780190943646.001.0001.

Lackner, Michael: Eine »divinatorische Kultur par excellence«? Chinesische Wahr- und Weissagung im Vergleich. In: Hock, Klaus/Stengel, Friedemann/van Oorschot, Jürgen (Hg.): Zukunfts-Sichten zwischen Prognose und Divination. Berlin/Boston 2021, 7–28. DOI: 10.1515 /bthz-2021-0003.

Latour, Bruno: Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie (suhrkamp taschenbuch wissenschaft 1967). Übersetzt von: Roßler, Gustav. Frankfurt a. M. 2010.

Luhmann, Niklas: Die Gesellschaft der Gesellschaft. Bd. 1 & 2 (suhrkamp taschenbuch wissenschaft 1360). Frankfurt a. M. 1997.

Lupton, Deborah: Understanding the Human Machine. In: IEEE Technology and Society Magazine 32/4, 25–30. DOI: 10.1109/MTS.2013.2286431.

Lupton, Deborah: The Quantified Self. Cambridge, Malden 2016.

Lupton, Deborah: Digital Health. Critical and Cross-Disciplinary Perspectives (Critical Approaches to Health). London, New York 2018.

Lupton, Deborah: Data Selves. Medford 2020

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

Putnam, Hilary: Vernunft, Wahrheit und Geschichte (suhrkamp taschenbuch wissenschaft 853). Übersetzt von: Schulte, Joachim. Frankfurt a. M. 1990.

Reynolds, Gretchen: Do We Really Need to Take 10,000 Steps a Day for Our Health? Online unter: https://www.nytimes.com/2021/07/06/well/move/10000-steps-health.html (Stand: 18.10.2022).

Rosa, Hartmut: Beschleunigung. Die Veränderung der Zeitstrukturen in der Moderne (suhrkamp taschenbuch wissenschaft 1760). Frankfurt a. M. [10]2014.

Rosa, Hartmut: Unverfügbarkeit. Wien, Salzburg [4]2019.

Rosengrün, Sebastian: Künstliche Intelligenz zur Einführung (Zur Einführung). Hamburg 2021

Sharon, Tamar: Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. In: Philosophy & Technology 30/1, 91–121. DOI: 10.1177/2055207616689509.

Siegel, Eric: Predictive Analytics. The Power to Predict Who Will Click, Buy, Lie, or Die. Hoboken 2016.

Sloterdijk, Peter. Du mußt Dein Leben ändern (suhrkamp taschenbuch 4349). Frankfurt a. M. 2019.

Smithson, Michael: Ignorance and Uncertainty. Emerging Paradigms. New York, Berlin 1989.

Strehle, Samuel: Zur Aktualität von Jean Baudrillard. Einleitung in sein Werk. Wiesbaden 2012.

Thaler, Richard H./Sunstein, Cass: Nudge. Wie man kluge Entscheidungen anstößt. Übersetzt von: Bausum, Christoph. Berlin [14]2019.

Thomas, Scarlett: Nowhere to run: did my fitness addiction make me ill? Online unter: https://www.theguardian.com/lifeandstyle/2015/mar/07/fitness-addiction-ill-scarlett-thomas (Stand: 31.08.2022).

Tretter, Max: Perspectives on digital twins and the (im)possibilities of control. In: Journal of Medical Ethics 47/6 (2021), 410–411. DOI: 10.1136/medethics-2021-107460.

Wittgenstein, Ludwig. Über Gewissheit. Übersetzt von: Anscombe, Gertrude Elizabeth Margaret/von Wright, Georg Henrik. Frankfurt a. M. 1970.

Zimmerman, Michael J.: Living with Uncertainty. The Moral Significance of Ignorance. Cambridge, New York 2008.

On Digital Twins and Heavenly Doppelgangers

Promises and perils regarding digital self-models in medicine seen through the lenses of gnostic tradition

Yannick Schlote

Abstract

After factory machines and entire urban infrastructures have received a digital, simultaneously operating data model, the prospect arises to transfer the benefits of a person's digital twin to health care. This duplication of a person's vital functions in their data poses epistemological hurdles for the calculability of human beings as well as the ethical implications of the actual prognostic value. This article demonstrates the uncanny similarities between the actual idea of a person's digital twin with the ancient gnostic belief of a person's coexistence with their heavenly doppelganger, thereby contributing to an ethical assessment on digital twins for medical purposes. In consequence, this article pleads for a pragmatic perspective in which digital self-models are not idolized into heavenly selves but instead be seen as possible means to enhance human flourishing.

1 Introduction

The prospect of personalized and predictive medicine highly depends on the willingness of individuals to donate data and data-to-be-derived-materials (blood, tissue etc.) for sufficient pattern recognition. In turn we see that, while these data on their own are highly abstract, Big Data analysis in diagnostics has a major impact on individual real-life decision making and may in result be significant for a person's current and future wellbeing. But despite their impactful predictive potential results, health care data remain elusive. Their abstractness leads

to a deep insecurity towards giving consent for data driven medicine. Concepts are needed to bring these aniconic data and algorithms into a meaningful and trustworthy relation to their owner.[1] One proposal on how to use the advances in information technology for a more personalized and predictive (instead of only reactive) health care is lent from industry 4.0: By amassing a multitude of environmental, biometric and omics-data and using latest algorithms, every person shall receive their own digital twin to ensure lower-risk treatment as well as targeted countermeasures for foreseeable illness. This article addresses some of the promises and perils regarding digital self-models in medicine and its use of the metaphor of a digital alter ego. For this purpose, we first explore the origin and ongoing appeal of digital twins in industry before we address current projects that try to create digital replica for human bodies. Then we approach late antiquity and its dualistic belief system called gnosis that separates people in earthly and heavenly halves. With this in mind, we stress how the apparent allure to interpret the digitization as a gnostic event may lead to dataism and data solipsism – which pose a problem for medical research which is highly dependent on the allowance to access bio data. In addition, the current vulnerability and proven exploitation of a person's data on the internet leads to the assumption, that these digital twins may under present circumstances have less in common with freed, heavenly selves than with enslaved digital bodies. This article concludes that digital twins in medicine may have the potential to secure and even flourish human agency, but not in or by themselves, but only through trustworthy algorithms and advanced data protection.

2 The origin of digital twins

A digital twin acts as a digital representation of a tangible or intangible object or process from the real world in the digital world. Essentially, they are hyper-realistic computer models of complex objects that are capable of simulating their function at a high level of detail. This concept originates from engineering and refers to machines in charge which – due to their complexity – usually need high maintenance. Aims are to model those systems computation-

[1] To relate to – and therefore: to be in control over – their own medical data is crucial to willingly supply data for research. This insight arose in the ethical research workgroup for the project Bavarian Genomes, funded by the Bavarian Ministry for Sciences. The project improves the care for patients with rare diseases and creates focal points for research into new treatment strategies by identifying causative sequence variants in the genome of patients with a rare disease but genetically unclear diagnosis. Therefore, the project is dependent on omics-data given access to by their probands. For more information on the project and its partners visit the official website https://www.bavarian-genomes.de/ (31.8.2022).

ally, in order to develop and test them more quickly and economically than it is possible in any real-life setting. Processes and objects are thereby simulated virtually, from construction to manufacturing to maintenance. The use of digital twins can be seen as a further development in the Internet of Things (IoT), in which nearly every physical object receives and produces data and therefore interacts with its environment.

German tech company Siemens for example offers its industrial clients services for the implementation of digital twins to optimize their entire value creation process. Three different types of digital images are intended to enable seamless linking of the stages of production: The digital twin of the product is already created at the stage of definition and design of a planned product. It enables the simulation and validation of product properties, adapted to the respective needs. The same applies to the digital twin of production: It maps the use of machines and plant controls right up to entire production lines in the virtual environment. In this way even highly complex production lines can be calculated, tested and programmed in shorter time with less effort. The digital twin of performance is in turn continuously fed with data from the operation of products or production facilities. On this basis, predictive maintenance strategies can be implemented to prevent downtime or optimize energy consumption.[2]

This method's transfer from large-scale industry to other areas such as urban development has already flourished: The High-Performance Computing Center in Stuttgart for example constructed a digital twin of the nearby small city Herrenberg by collecting large datasets of various kinds: air quality, traffic flow, and the prevalence of pedestrian traffic among other parameters of urban life. They merged these large datasets and visualized them in a virtual replica where it was easier to understand these complex interactions — for example, to see how a change in traffic patterns or a new building could affect air quality for the city as a whole. Here, too, a detailed representation of the current state of Herrenberg is intended to not only understand the interactions in the present but to also improve the forecasting ability of urban planning processes. While some complex data are difficult or impossible to understand in any other form, now not only images and hidden structures of the present can be presented via digital twin, but also different plans for the future shall be laid out on this replication. Experts and laypeople shall be able to interact more efficiently and effectively through better presentation of the complex information in urban planning processes. Moreover, the HPCC team from Stuttgart suggests that digital twins in VR may also function as a communication tool to enable broad citizen participation and to provide an easy-to-comprehend model for the growing complexity of cities and towns. For all this, data scientists have to ensure that the *modelans*

[2] See Siemens AG, Der digitale Zwilling. Online: https://new.siemens.com/de/de/unternehmen/stories/industrie/der-digitale-zwilling.html (31.8.2022).

remains faithful to the *modelandum*: Similarities to the real world have to be achieved at a level of detail that is in terms of effort manageable and technically accurate enough to tackle complex problems.[3]

3 Digital twins in medicine

In a major publication commissioned by the European Union in 2016, researchers envisioned the future healthcare in the European Union: With aging populations and fewer medical workers in sight, they advocate for the implementation of a digital twin for every EU citizen to help citizens living a longer and healthier life. The necessary paradigm shift in medicine itself is described in superlatives and compared to nothing less than human's first landing on the moon as this means to exploit technological advances and innovations in information and communications technology, molecular analyses, imaging and sensor techniques and computational modelling. Computer models of every patient and disease state would allow physicians to test the consequences of all possible therapies on a virtual patient rather than the real patient. In addition, similar to the function of twin of production in industry or urban twins, this digital twin of a human may also be implemented in the *maintenance* of a person's health. For this measure of health care the needed data collection has to be ubiquitous: implantable, wearable, environmental smart sensors, health monitoring including clinical/imaging/molecular survey techniques (for the genome, epigenome, transcriptome, proteome, metabolome, immune status etc.) potentially guarding every European from before birth into old age.[4]

The Swedish Digital Twin Consortium (SDTC) with partners from health care, medical and technical faculties and industry is a step ahead and already creates digital twins for a certain test group (step 1), then computationally treats those digital twins with thousands of drugs in order to identify the best performing drug (step 2); and finally treats the real patient with the proposed drug to see if the prognosis of the simulation is trustworthy (step 3), recalibrating the parameters if necessary.[5] Similar to the digital replica of Herrenberg, the main difficulty is to identify the required parameters that are sufficient for the purpose to reduce complexity without losing prognostic insight. For example, the digital twin at the SDTC works with var-

[3] See Dembski, Fabian: Urban Digital Twins for Smart Cities and Citizens. The Case Study of Herrenberg, Germany. In: Sustainability 2020 12/6 (2020), 1–17, here 2–6.

[4] See Lehrach, Hans et al.: The Future of Health Care: Deep Data, Smart Sensors, Virtual Patients and the Internet-of-Humans. Online: https://ec.europa.eu/futurium/en/system/files/ged/future_health_fet_consultation.pdf (31.8.2022), 1–7, here 4–7.

[5] See Björnsson; Bergthor et al.: Digital Twins to Personalise Medicine. In: Genome Medicine 12/1 (2020),1–4, here 1 f.

ious variables that are relevant to pathogenesis other than molecules, such as symptoms or environmental factors. Highly optimistic with their results the consortium concludes: „With increasing availability of multi-omics, phenotypic, and environmental data, network tools may allow the construction of disease models of unprecedented resolution. Such models may serve as templates for the construction of digital twins for individual patients."[6]

The potential benefits of digital twins in medicine seem evident even for a larger public. A study conducted in 2018 by German consulting services organization PricewaterhouseCoopers (PwC) indicates that 71% of German citizens consider it useful to have a virtual image created including data from their own DNA. They specifically expect benefits for patients in terms of support for doctors in making treatment decisions (86%), help in choosing the best drug (83%), relief for patients thanks to fewer side effects and fewer unnecessary operations (82%), and more accurate and faster diagnoses and therapies (80%). 58% of Germans, however, consider the creation of a digital twin to be more suitable for certain patient groups, such as the chronically ill or patients with rare diseases. For only 17% of Germans a digital twin would be out of the question under any circumstances, mainly out of fear of surveillance. However, for more than a third of Germans, a prerequisite for the creation of their digital twin would even be that only the treating physicians could access the data.[7]

4 The gnostic tradition of heavenly doppelgangers

According to German sociologist Armin Nassehi, digitization as a whole means nothing less than the reduplication of our shared world in data; the modern digital society is about to double itself in data. The data that machines produce, store and correlate, form a second layer of reality on top of the analog world, Nassehi says. Thereby digital technology discovers society and thus the individual in a much more precise sense than it was previously possible by pointing to former invisible correlations through the digital realm. Meanwhile, the unspecific character of the digital enables it to map every object in the digital sphere giving this second reality its own holistic appearance for to act as an almost ubiquitous form – hitherto reserved for the use of writing and the presence of God.[8]

---

6 Ibid., 3.

7 See PWC Wirtschaftsprüfungs- und Beratungsgesellschaft Deutschland GMBH: Der digitale Zwilling, Online: https://www.pwc.de/de/gesundheitswesen-und-pharma/pwc-studie-der-digitale-zwilling.pdf (31.8.2022), 3–5.

8 See Nassehi, Armin: Muster: Theorie der digitalen Gesellschaft. München 2019, 35.

In this sense, the recreation of a person's vital processes in data is nothing unheard of, but instead appears as a mere continuation of its basic idea of digitization as duplication. Christian belief is also built on an antithesis of different realities, not between the analog and the digital sphere, but between world and God. Paulus for example repeatedly emphasizes the disparity between the withering world as opposed to God's kingdom, which he claims to be the dominant, disruptive force operating in the unseen background of things (Mt 24,35, 1Cor 7,31, etc.). The different weighing of world and God, earth and heaven, profane and sacred thus has become a central conflict in Christian history of dogma.

Besides the more common division into body and spirit, gnostic tradition imagined the separated existence of earthly and heavenly selves. The Jewish pseud-epigraphical literature for instance is filled with stories about biblical figures like Moses, Jacob and Henoch, who after witnessing God's epiphany later on ascended to become heavenly beings. Thus, the ancient epiphanies became enriched with the doppelganger symbolism and thereby play a formative role in shaping the later rabbinic and Hekhalot speculations about the heavenly identities of the exalted patriarchs and prophets. These include the traditions about the celestial self of Enoch in the form of the supreme angel Metatron and the rabbinic stories about the upper identity of Jacob in the form of his engraved or enthroned image – developments that are then again crucial theophanic loci for later Jewish mysticism.[9]

Even though there were blurred lines between the gnostic paradigm and early Christian faith, the former was ultimately rejected. In contrast to other dualistic religions, the gnostic dogma is most distinct when it comes to dignify the immaterial and condemn the material reality. The gnosis recurs to an existential wisdom declaring that every person is unknowingly trapped in the material world by evil forces, personified in the lesser God of this material world, the demiurge, and his archons. This lesser God and his demons trap people's spirits in their bodies and thereby in this material universe to make them forget about every person's heavenly origin by a higher God of Light. To escape this material prison, the gnostic speculation relies heavily on the idea of a duplication of man. The gnostic dualism ultimately breaks with the chronical continuity between an earthly life and its later entrance into heavens in its vision of a heavenly alter ego. Instead, it stresses the dualism even further by aligning this order into a synchronicity: In the Syrian gnostic tradition the idea repeatedly appears of a person's heavenly doppelganger that exists simultaneously to their earthly twin. The heavenly doppelganger symbolizes the eternal or heavenly self of a person similar to his original idea in a platonic sense – that remains in the upper world and stays protected while the mortal twin toils his life below. But this heavenly doppelganger is not by himself free, both twins need each

[9] See Orlov, Andrei: The Greatest Mirror. Heavenly Counterparts in the Jewish Pseudepigrapha. New York 2017, 149–151.

other for true salvation: Through his life on earth, the mortal half forms their heavenly doppelganger through his actions and thereby completes it for the eschatological ascend to the true God of Light. When his duty on earth is done, the immortal spirit of the deceased half rises up from earth and meets his heavenly doppelganger in heaven, where they both unify to become their one true self. Only this completion paves the way to finally escape the banishment to the world. Therefore, the meeting with this separated aspect of himself, the recognition of his own image in him and the reunion with him signify the real moment of redemption.[10] A recurring phrase in gnostic speculations states: „I go to meet my image and my image goes to meet me. It caresses me and embraces me as if I returned from captivity."[11]

The similarity between the digital twin for health and the religious hope of salvation in the existence of a heavenly doppelganger becomes even more apparent from the fact how gnostic hymns describe the heavenly double. Even though they are in lifetime divided, the heavenly doppelganger emanates enlightenment, and ultimately healing, to its poor material sibling. The gnostic revelation about the existence of one's own heavenly image from the hidden place is said to heal the sick heart of the chosen on earth and to guide him to disobey the evil forces of this world.[12] No narrative stresses this gnostic association of an heavenly image with salvation more than an fragment in the apocryphic gospel of Thomas. The tale itself, which has no title and is labeled among historians as The Hymn of the Soul or The Song of the Pearl, is an allegory for the gnostic eschatology: A prince is sent by his parents out of his kingdom of light to find a rare pearl in a land far away to bring it home. On his journey, he has to adapt to the foreign land he infiltrates, so he changes his clothes and eats foreign food. In this process, he slowly forgets his royal descendance and his mission. Only when the King and Queen send him a letter, he remembers his past, finds the pearl and makes his way back. Before he can pass the border and return to his kingdom, his royal parents also send him his sparkling dress that he had to abandon when he first entered the land. As the prince looks at his dress, it appears to him like a mirror image of himself through which he fully recognizes himself. The fragment states that he recognizes that they were divorced from each other, and yet once again in the same form.[13] The doppelganger motif is not only echoed in the prince's dress, but also in the prince's brother: The hymn mentions several times that his brother remains at home with his

---

10 See Jonas, Hans: Gnosis. Die Botschaft des fremden Gottes. Frankfurt a. M./Leipzig 2008, 156–158.

11 See Lidzbarski, Mark: Ginza: Der Schatz oder Das Große Buch der Mandäer. Göttingen 1925, 559. Initial German translation by Lidzbarski: „Ich gehe meinem Abbild entgegen, und mein Abbild geht mir entgegen. Es kost und umarmt mich, als käme ich aus der Gefangenschaft zurück."

12 See ibid., 381 f.

13 See Preuschen, Erwin: Zwei gnostische Hymnen. Mit Text und Übersetzung. Gießen 1904, 19–27. The German translation on page 24 states: „Ich sah es ganz in Ganzem, und ward in ihm auch meiner ganz ansichtig, dass wir zwei wären in Geschiedenheit und wieder eins in einerlei Gestalt."

parents and waits for his brother's return for when they both will then inherit the kingdom together. The dress and the brother both allude to the gnostic speculation on heavenly doppelgangers; that is why when the prince returns to his kingdom, there is no mention about his homebound brother because he has already merged with his heavenly doppelganger in his dress before returning to the kingdom of light.[14]

5 Dataism and data solipsism

There is an apparent allure to interpret the digitization as a gnostic event: Just like the promise of gnosis to reveal the true, profound reality that cannot be seen or experienced via the senses due to its radical transcendence, Big Data is in similar something inherently unrecognizable and therefore not in itself a natural state, gaining an aura of transcendence while simultaneously proclaiming powerful consequences for this world.[15] In addition, the idea that information technology by itself will secure a deeper knowledge that may solve the major obstacles of today's humanity – providing a long and healthy life, overcome climate crisis, famines and lead to permanent peace – achieves the appearance of a pseud-religious confession.

A gnostic reading of digital twins is an extreme that reveals the seductive power in imagining a digital twin to be the ideal, platonic self to fully commit to. This blurs the sequential relationship between archetype and replica, *modelandum and modelans*. Gnostic salvation locates all its hopes for salvation in the heavenly doppelganger. Just like the heavenly doppelganger is formed by the acts of the person on earth, the digital twin only exists because it is constantly in-*formed* by the various data that the patient shall provide. The idea of the digital twin is in similar manner in danger of confusing goal and means. The health of the patient to whom the twin is supposed to serve can change to the opposite and lead to a neglect of the actual human being towards an undermining of his freedom in demanding the necessary, optimal data traffic. Hebrew University Professor for History and best-selling author Yuval Harari names this hermeneutical danger *dataism*, where information is no longer valued for its benefit in relation to humans but instead information flow in itself becomes the supreme value by which any organism or system ought to be measured.[16]

This misconception on the dependency between analog self and the digital self may have severe consequences: When interpretations about the future with digital twins convey the

[14] See Jonas, Gnosis, 156.

[15] Ibid., 49 f.

[16] See Harari, Yuval Noah: Homo Deus. A Brief History of Tomorrow. London 2017, 428. 445.

person's overidentification with their digital simulacrum and undermine the limits and level of abstractness and probability of these models, this hermeneutic absolutism projected on the digital twin must ultimately lead to data solipsism. When the sum of your data conjures the image to be essentially your essence, at such high stakes, no one is likely to grant voluntarily access to their medical data (or personal data in general). But data recognition and artificial intelligence heuristics are dependent on large data sets for there is no probability computing without enough material. Thus, the sacralization of one's own data in analogy to the heavenly image means the end to Big Data analysis and its benefits. A modern reimagining of this gnostic reversal of archetype and replica that leads to this hermeticism is illustrated by author Oscar Wilde in his famous novel *The Picture of Dorian Gray* (1890): While beautiful Dorian becomes on the inside more and more a shallow mask of his former self, his true self resides no longer in his physical appearance but in his painted twin – of which he is in this state existentially dependent on. Consequently, Dorian must hide his portrait at all costs.

Another misconception the gnostic view on digital duplication unfolds is the flawed nature of data itself: the gnostic kingdom of light is superordinate to this world, but digital data are not the basis of a deeper reality on which existence is built on. In opposite, data is a by-product of digital processes, which themselves always work only with a shortened and thus distorted reality. Data according to which the digital simulacrum is formed remain what they are etymologically: Nothing pre-existing, but something given (lat. *datum*), only shortages of reality to work with. The digital sphere is not intangible and is not transcendent in a classic sense. The digital is dependent on very worldly things like server structures, cables, electricity, and providers. In contrast to gnostic dualism, the digital twin does not exist in a domain governed by a liberating foreign God of Light, but by programmers with certain incentives and tech companies with accountable stakeholders.

6 Enslaved digital bodies and the doppelganger's guidance

The gnostic wisdom implies the human enslavement by the worldly forces of the material world, personified by the demiurge and his demonic archons. The universe forms a huge prison for humans with the earth as its center, around which cosmic spheres are arranged as concentrically enclosing shells.[17] For gnosis, one's own willing and feeling is also part of the fallen world and not part of the actual, acosmic self. Emotions and desires represent the demonic cosmos inside the human being, so that one's own will is deemed to be deeply mistrusted, too.

[17] See Jonas, Gnosis, 70–71.

An individual cannot rely on his own, he must flee from his own desires and therefore seeks guidance and healing from his heavenly doppelganger.[18]

Similarly, the appeal of the digital twin for medicine lies beyond a detailed visualization of an individual's present health status but to extrapolate the given parameters to model the future and to use this knowledge to design other futures on the basis of a proposed change in behavior in the present. Therefore, the digital twin can and may be used to influence patients to certain lifestyle changes. In a gnostic perspective, one can argue that it is right to be guided by their own heavenly twin who *knows best*. But this idea can only be unconditionally agreed on if one ignores the economic embedding and power strings of the data economy. The intention guiding the prognostic value of digital twins does not automatically need to have the individual patient's integrity in mind.[19] In contrast to the gnostic idea of the enslaved human on earth and his untainted heavenly doppelganger, today's digital twin is in latent danger to be held hostage either by companies or governments for their own benefit, which in turn will ultimately affect the real person as well. In this case, the idea of an enslaved digital twin reverses the gnostic idea from referring the doppelganger as a guardian angel to turning it into to a haunting ghost of continuous repression.

Long before the idea of a digital twin within the context of Big Data arose, IT Professor Simon Rogerson had claimed in 2007 that people already exist as digital beings through a myriad of personal data and electronic interaction. But this digital existence means to be owned by those who run the programs and data conduits. For Rogerson the same combination of power and ignorance that enabled physical slavery is present in digital slavery, where few tech companies have access to valuable personal data items that are still largely considered mere proxies rather than organs of data subjects.[20] Sociologist Mick Chisnall specified that digital slavery does not only constitute a massive and largely hidden theft in information to be limitless exploited economically. More importantly, the subsequent aggregation of personal data can enable the control over individual behavior, emotions and consumption.[21] And political scientist Barbara Prainsack adds to the idea of a digital slave in the context of medicine that patient data offer plenty of ways to manipulate people to act more cost efficiently. She states that if patient data were not under the control of profit-oriented companies or control-oriented governments,

---

18 See ibid., 334–337.

19 See Nemitz, Paul: Constitutional Democracy and Technology in the Age of Artificial Intelligence. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 376 (2018), 1–14, here 1–4.

20 See Rogerson, Simon and Anne: Digital Slavery, Online: https://www.researchgate.net/publication/313619199_Digital_Slavery (31.8.2022).

21 See Chisnall, Mick: "Digital Slavery, Time for Abolition?" In: Policy Studies 41/5 (2020), 488–506, here 499–501.

nudging could be used less manipulative and also more effective focusing on the structural factors and as a result changing institution.[22] This less manipulative, more enabling approach may in the end be even more effective to change people's behavior than tackling it individually:

> Patient data–besides helping to improve individual and population health outcomes–should be used not for risk stratification or pricing, but to target services and improve infrastructures where they are especially needed. In other words, in the much hailed era of big data, it is more important than ever that we use data to build better institutions, instead of trying to solve problems through tackling individual behavior.[23]

7 Conclusion

The prospect and perils of digital twins in medicine have been stressed, even exaggerated in the analogy with the gnostic idea of heavenly doppelgangers. In gnosis, knowledge is not solely a means to salvation, but already part of salvation as the knowledge is reciprocal in transforming the knower. The acritical acceptance of the salutoferic effect Big Data may have on individuals and society in general are widely distributed by the followers of the Californian ideology.[24] This naive optimism, despite what tech companies claim in their self-image as world improvement entrepreneurs, is unwarranted. No epiphany in digitization, no inherent goodness can be awarded to the nature of data itself, digital structures remain ambivalent to exploit personal agency or to strengthen it.

Potential health and actual freedom are not to be played off against each other, but neither should they be replaced with each other: Health at all cost is not the absolute form of realized freedom. Human agency must be secured via our digital replica and not assaulted by them, moreover the digital twin may be even a new opportunity to actualize freedom in new ways – maybe even seen as an enabling prosthesis and further extension of a person's body into the virtual realm.[25] The visionary aspect of digital twins may enable participatory processes to facilitate communication, not only on how to build better cities, but also on how to install better institutions for health care that meet personalized demands instead of trying to manipulate

---

22 See Prainsack, Barbara: The Value of Healthcare Data: to nudge, or not? In: Policy Studies 41/5 (2020), 547–562.

23 Ibid., 557.

24 See Nachtwey, Oliver; Seidl, Timo: Die Ethik der Solution und der Geist des digitalen Kapitalismus, (IfS Working Papers 2017 Nr. 11, Frankfurt a. M.: Institut für Sozialforschung), 1–36, here 21–23.

25 See Braun, Matthias: Represent Me: Please! Towards an Ethics of Digital Twins in Medicine. In: Journal for Medical Ethics 47 (2021), 394–400, here 399.

the demands of every individual. However, these data derived benefits cannot be the result of an absolute, detached freedom in data solipsism. No digital twin can live in their own digital heaven because each digital self only gains prognostic insight from the templates and patterns derived from millions of other shared data sets.

The extreme dualism of gnosis sensitizes for the middle path not to underestimate the potential harm digital twins may cause, but also not to overestimate the provisional nature of models and probabilities. In overestimating, one runs the risk of essentializing the digital image and surrendering oneself completely to it, both in terms of recording data and committing to its forecasts. By underestimating, this valuable information is left unprotected without supervision and stolen by companies for profit maximization and to be sold to employers or insurances. The potential benefit of digital twins in medicine is too apparent for not taking the chances, but the beneficial outcome must be politically claimed and trustworthy digital environment must be established, if necessary even in opposition to the current *archons* of the digital economy.

References

Bavarian Genomes: Official Project Homepage. Online: https://www.bavarian-genomes.de/ (31.8.2022).

Björnsson, Bergthor et al.: Digital Twins to Personalize Medicine. In: Genome Medicine 12/1 (2020), 1–4. doi: 10.1186/s13073-019-0701-3.

Braun, Matthias: Represent Me: Please! Towards an Ethics of Digital Twins in Medicine. In: Journal for Medical Ethics 47 (2021), 394–400. doi:10.1136/medethics-2020-106134.

Chisnall, Mick: Digital Slavery, Time for Abolition? In: Policy Studies 41/5 (2020), 488–506.

Dembski, Fabian, et al.: Urban Digital Twins for Smart Cities and Citizens. The Case Study of Herrenberg, Germany. In: Sustainability 12/6 (2020): 1–17, doi: https://doi.org/10.3390/su12062307.

Harari, Yuval Noah.: Homo Deus. A Brief History of Tomorrow. London 2017.

Jonas, Hans: Gnosis. Die Botschaft des fremden Gottes. Frankfurt a. M./Leipzig 2008.

Lehrach, Hans, et al.: The Future of Health Care. Deep Data, Smart Sensors, Virtual Patients and the Internet-of-Humans, 1–7. Online: https://ec.europa.eu/futurium/en/system/files/ged/future_health_fet_consultation.pdf (31.8.2022).

Lidzbarski, Mark: Ginza. Der Schatz oder Das Große Buch der Mandäer. Göttingen 1925.

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

Nemitz, Paul.: Constitutional Democracy and Technology in the Age of Artificial Intelligence (Philosophical Transactions: Series A, Mathematical, Physical, and Engineering Sciences 376) London 2018, 1–14. doi: https://doi.org/10.1098/rsta.2018.0089.

Nachtwey, Oliver, Seidl, Timo: Die Ethik der Solution und der Geist des digitalen Kapitalismus (IfS Working Papers 2017 Nr. 11, Frankfurt a. M: Institut für Sozialforschung), 1–36. Online: http://www.ifs.uni-frankfurt.de/wp-content/uploads/IfS-WP-11.pdf (31.8.2022).

Orlov, Andrei: The Greatest Mirror. Heavenly Counterparts in the Jewish Pseudepigrapha. New York 2017.

Prainsack, Barbara: The Value of Healthcare Data: to nudge, or not? In: Policy Studies, 41/5 (2020), 547–562. doi: https://doi.org/10.1080/01442872.2020.1723517.

Preuschen, Erwin: Zwei gnostische Hymnen. Mit Text und Übersetzung. Gießen 1904.

Rogerson, Simon and Anne: "Digital Slavery." In: IMIS Journal 17/5 (2007). Online: https://www.researchgate.net/publication/313619199_Digital_Slavery (31.8.2022).

PWC Wirtschaftsprüfungs- und Beratungsgesellschaft Deutschland GMBH: Der digitale Zwilling. Erwartungen und Einschätzungen der deutschen Bevölkerung mit besonderem Fokus auf Diabeteserkrankungen, 1–36. Online: https://www.pwc.de/de/gesundheitswesen-und-pharma/pwc-studie-der-digitale-zwilling.pdf (31.8.2022).

Siemens Deutschland AG: Der digitale Zwilling, Online: https://new.siemens.com/de/de/unternehmen/stories/industrie/der-digitale-zwilling.html (31.8.2022).

Ein neuer Blick auf den Menschen?

Impulse für Fragen der Leiblichkeit in der Ethik vor dem Hintergrund des Moral-Enhancement-Diskurses

Dominik Winter

Abstract

Digitization and new technologies change significantly how we perceive ourselves and the world around us. In recent years, changes of this kind were often discussed in the context of transhumanist ideas which usually propose a naïve science-optimism paired with a problematic conception of the human nature. Dealing with the example of Moral Enhancement, the article explores how transhumanist ideas usually picture a questionable understanding of the relationship between our body and our mind. At the same time, the usual transhumanist perspective on this topic allows for a fresh view on this topic from a theological and ethical perspective, thus offering an opportunity to reflect the image of humans as mind-body-unions anew. Therefore, the article explores problematic aspects of recent ethical positions in that regard and points to newer developments that might help to a better understanding of the close connection of our minds and bodies.

1 Einleitung

Neue Erkenntnisse und neue Technologien verändern unseren Blick auf die Welt, und kaum eine Entwicklung der letzten Jahrzehnte hat dazu mehr beigetragen als die Digitalisierung. Aber neue technische Möglichkeiten verändern nicht nur unseren Blick auf die Welt, sondern auch auf uns selbst. Veränderungen dieser Art wurden dabei in den letzten Jahren vor allem im Kontext der Ideen des Transhumanismus diskutiert. Neue Entwicklungen insbesondere aus den Bereichen der IT und der Neurowissenschaften werden in dieser Denkströmung aufgegriffen und führen bspw. zu verschiedenen Theorien, wie der Mensch

mit Hilfe dieser neuen Technologien verbessert werden könnte: Man spricht hier auch vom *Enhancement*.

Dass diese Ideen häufig mit einem naiven Wissenschaftsoptimismus verbunden sind und ein problematisches Bild vom Menschen zugrundelegen, ist bereits überzeugend gezeigt worden.[1] Ich werde aber dafür argumentieren, dass sich eine tiefere Auseinandersetzung mit transhumanistischen Ideen dennoch auch für eine theologische Ethik weiterhin lohnt. Denn häufig ist der doch sehr andere Blick auf den Menschen, den transhumanistische Positionen einnehmen, auch für theologische Überlegungen hilfreich, um Aspekte des je eigenen Verständnisses vom Menschen neu zu schärfen. Die häufig mehr oder weniger explizit vertretene Leibfeindlichkeit transhumanistischer Positionen nötigt im Rahmen einer argumentativen Zurückweisung zu einer neuen positiven Wertschätzung menschlicher Leiblichkeit und öffnet so den Blick auf bisher vernachlässigte Aspekte des Verständnisses des Menschen als Leib-seelischer-Einheit.

Wie dies im Kontext theologischer Ethik zu neuen Reflexionen anregen könnte, möchte ich im Folgenden zeigen. Dazu werde ich zunächst am Beispiel des *Moral Enhancement* – also der Idee, neurotechnische Mittel zur Verbesserung unseres Moralverhaltens einzusetzen – zeigen, wo Probleme dieser Ansätze liegen und welche Aspekte des spezifischen Blicks auf den Menschen bedenkenswert sind (Abschnitt 1). Im Anschluss daran erinnere ich an eine Debatte, die bereits Ende der 1990er/Anfang der 2000er Jahre zwischen Ethik und Neurowissenschaften in Bezug auf die Frage der Willensfreiheit geführt worden ist und in der es bereits eine Reflexion über die Rolle der Leiblichkeit des Menschen im Kontext theologischer Ethik gegeben hat (Abschnitt 2). Daraufhin zeige ich, inwiefern die damalige Debatte Fragen nach der Rolle unserer Leiblichkeit für die Ethik offengelassen hat und wo sich auf Grundlage des durch die Beschäftigung mit Moral Enhancement geschärften Blicks neue Reflexionsmöglichkeiten zur Beantwortung dieser offenen Fragen anbahnen (Abschnitt 3). Abgeschlossen wird der Artikel durch ein kurzes Fazit (Abschnitt 4). Ich hoffe, so zeigen zu können, wie die neuen technischen Entwicklungen der Digitalisierung und die daraus resultierenden neuen Denkströmungen auch in theologischer Perspektive unser Bild vom Menschen, wenn nicht völlig verändern, so doch positiv weiterentwickeln und schärfen können.

[1] Siehe bspw. Helmus, Caroline: Transhumanismus – der neue (Unter-)Gang des Menschen? Regensburg 2020.

2 Das Verhältnis von Körper und Geist in Ansätzen zum Moral Enhancement

Auch wenn eine ausführliche Darstellung verschiedener Ansätze von Moral Enhancement im Rahmen dieses kurzen Artikels nicht möglich ist,[2] sollen zumindest wichtige Ideen und zentrale Prämissen vorgestellt werden. *Emotives* Moral Enhancement[3] wird dabei vor allen mit den Ansätzen von Ingmar Persson und Julian Savulescu auf der einen und Thomas Douglas auf der anderen Seite verbunden. Beide Ansätze gehen davon aus, dass es bestimmte moralisch relevante Emotionen bzw. Dispositionen gibt, die für neurotechnische Eingriffe zugänglich sind und deren Beeinflussung einen positiven Effekt auf das Moralverhalten des bzw. der Empfänger:in haben könnte. Persson und Savulescu identifizieren dafür Empathie und den Sinn für Gerechtigkeit, die verstärkt werden sollen[4], während Douglas Aggressionen und rassistische Ressentiments abschwächen möchte.[5] Unabhängig von der Verschiedenheit der Zugänge zum Moral Enhancement lassen sich für beide Ansätze – und auch weitere ähnlich gelagerte Positionen – folgende zwei Prämissen als grundlegend festhalten:

1. Emotionen haben einen entscheidenden Einfluss auf unser Moralverhalten.
2. Emotionen basieren auf körperlichen Eigenschaften oder Prozessen und sind deshalb durch neurotechnische Eingriffe beeinflussbar.

Aus diesen beiden grundlegenden Prämissen ergeben sich zwei Positionen, welche die Befürworter:innen der emotiven Variante des Moral Enhancements mindestens indirekt vertreten:

---

2 Vgl. dazu aber bspw. Winter, Dominik: Falsche Hoffnung. In: Zeitschrift für Theologie und Philosophie 144/2 (2022), 220–242, hier 223–231.

3 Davon unterschieden werden kann eine kognitive Variante des Moral Enhancement, die sich aber nur nebensächlich mit spezifischen Fragen des Moral Enhancement beschäftigt und deshalb in der Diskussion eher unterrepräsentiert ist. Ich konzentriere mich daher hier auf die emotive Variante. Wenn im Folgenden also von Moral Enhancement gesprochen wird, ist dabei immer die emotive Variante gemeint.

4 Vgl. Persson, Ingmar/Savulescu, Julian: Unfit for the Future. Oxford 2012, 108.

5 Vgl. Douglas, Thomas: Enhancement der Moral. In: van Riel, Raphael/Di Nucci, Ezio/Schildmann, Jan (Hg.): Enhancement der Moral. Münster 2015, 85–111, hier 91.

1. Sie vertreten einen „schwachen" Nonkognitivismus. Diesen nenne ich deshalb „schwach", weil in diesem Kontext darunter nicht zwingend dasselbe verstanden wird, was normalerweise im metaethischen Diskurs mit Nonkognitivismus[6] bezeichnet wird. Vertreter:innen von emotivem Moral Enhancement versuchen i. d. R. eine Positionierung in diesen allgemeinen metaethischen Diskursen zu vermeiden, um möglichst breit anschlussfähig zu sein. Aus diesem Grund kommt es zu unterschiedlichen Aussagen in Bezug auf die Frage nach der Wahrheitsfähigkeit moralischer Urteile. Unter dem Begriff Nonkognitivismus wird in diesem Diskurs daher meist nur verstanden, dass Emotionen generell eine relevante Rolle für unser Moralverhalten spielen, was im Wesentlichen der ersten Prämisse entspricht. Welche Rolle genau, wird dabei ebenfalls offengelassen.[7]
2. Sie vertreten mindestens einen Kompatibilismus in der Frage nach unserer Freiheit. Sie gehen also davon aus, dass wir keinen völlig freien Willen haben, sondern durch Determinanten zumindest beeinflusst werden. Wenn Emotionen eine Rolle für unser Moralverhalten spielen und diese durch körperliche Eigenschaften und/oder Prozesse beeinflusst sind, werden wir zum Teil in unserem Verhalten durch körperliche Determinanten bestimmt. Denn nur, wenn das der Fall ist, kann ein Enhancement auch eine zuverlässige Wirkung erzeugen.[8]

Beide Positionen nehmen insbesondere vor dem Hintergrund der zweiten Prämisse also einen relevanten Einfluss von körperlichen Eigenschaften und Prozessen auf unser Verhalten an, da nur solche neurotechnisch beeinflussbar sind und nur so ein mögliches Enhancement eine Wirkung erzielen könnte. Dabei spielt es zunächst auch keine Rolle, ob Emotionen auch eine geistige bzw. kognitive Dimension haben (siehe dazu Abschnitt 3.2), da in dieser Perspektive nur die körperlichen Dimensionen in den Blick geraten. Es scheint daher angebracht, vor dem Hintergrund dieser Ansätze von einer engen Verbindung zwischen Körper und Geist auszugehen, die sich aus theologischer Perspektive mit dem Bild des Menschen als Leib-seelischer-

[6] Unter Nonkognitivismus wird i. d. R. die Position verstanden, dass moralische Urteile keinen propositionalen Inhalt haben und damit nicht wahrheitsfähig sind. In seiner radikalsten Form sind moralische Urteile in dieser Perspektive nur Ausdruck einer emotionalen Haltung zu einer Handlung (Emotivismus). Ich übernehme hier trotzdem die Bezeichnung des Nonkognitivismus, weil dies so auch von Diskursteilnehmer:innen als Selbstbezeichnung übernommen wird. Vgl. Douglas, Thomas: Moral Enhancement via direct emotion modulation. In: Bioethics 27/3 (2013), 160–168, hier 162.

[7] Vgl. dazu bspw. Douglas: Enhancement, 91.

[8] Vgl. dazu bspw. Savulescu, Julian/Persson, Ingmar: Enhancement der Moral, Freiheit und die Gottmaschine. In: van Riel, Raphael/Di Nucci, Ezio/Schildmann, Jan (Hg.): Enhancement der Moral. Münster 2015, 51–75, hier 62 f.

Einheit verknüpfen ließe, sofern menschliches Verhalten nicht nur durch körperliche Dimensionen, sondern auch durch geistige entscheidend geprägt werden soll (siehe dazu Abschnitt 2). Während die Theologie dieses Bild aber positiv auszudeuten vermag, sehen Ansätze für Moral Enhancement diese Verknüpfung als Kern des Problems. Denn diese enge Verknüpfung von Körper und Geist muss zwar grundsätzlich angenommen werden, da sonst das Enhancement in der intendierten Weise fehlschlägt. Wenn diese Verbindung aber angenommen wird, sind auch die körperlichen Prozesse, die menschliches Verhalten beeinflussen, integraler Bestandteil dieses Verhaltens und damit menschlicher Persönlichkeit. Jede Beeinflussung körperlicher Prozesse ist so auch eine Beeinflussung der Persönlichkeit.[9]

Diesem Problem kann aus zwei Richtungen begegnet werden, die zu einem ähnlichen Ergebnis führen: Zum einen kann die enge Verbindung von Körper und Geist zwar angenommen, dann aber für schlecht befunden werden. Der Körper begrenze uns in diesem Fall und verhindere, dass wir uns so verhalten, wie wir es eigentlich wollten. Weil wir nicht die richtigen Emotionen im richtigen Maß empfinden, tun wir auch nicht das, was wir eigentlich wollen, nämlich das Gute. Genau dies verspricht Moral Enhancement zu lösen. Hier liegt der Fokus also auf einer Betrachtung körperlicher Aspekte als Störfaktoren, die reduziert werden sollen.

Zum anderen kann die soeben dargestellte Persönlichkeitsveränderung, die sich durch eine Veränderung körperlicher Aspekte ergibt, positiv gesehen werden. Schließlich geht es ja genau darum, seine Persönlichkeit so zu verändern, dass man sich moralischer verhält als vorher – was immer „sich moralischer verhalten" in diesem Kontext genau heißen soll. In dieser Perspektive liegt der Fokus also auf der positiven Veränderung im Sinne eines vorher festgelegten Ideals, die mithilfe eines Moral Enhancement erreicht werden könnte.

Beide Antwortvarianten offenbaren dabei eine gewisse Leibfeindlichkeit, die am Ende nur mit einem Dualismus zu erklären ist, womit aber auch die Chancen eines umfassenderen Verständnisses des Menschen als Leib-seelischer-Einheit durch die eigentlich vertretenen Prämissen dieses Ansatzes nicht genutzt werden. Für die erste Antwortvariante ist diese Leibfeindlichkeit relativ offensichtlich: Hier wird ein klarer Antagonismus zwischen „gutem" Geist und „bösem" Körper aufgemacht, der zugunsten des Geistes gelöst werden soll. Thomas Douglas schreibt in diesem Kontext:

> „Da angenommen wird, dass Smiths [eine mögliche Empfängerin von Moral Enhancement; D. W.] Enhancement bestimmte Emotionen abschwächt, funktioniert es

[9] Vgl. dazu Winter: Falsche Hoffnung, 234–240.

> vermutlich, indem es jene rohen [d. h. körperlichen; D. W.] Mechanismen, die die relevanten Emotionen erzeugen, *unterdrückt*. Das Enhancement scheint zu funktionieren, indem es den Einfluss von Smiths rohem Selbst *reduziert* und somit ihrem wahren Selbst *mehr* Freiheit ermöglicht."[10]

Ähnliches gilt auch für die zweite Antwortvariante. Denn auch hier werden die negativen Einflüsse auf unser Moralverhalten den körperlichen Aspekten unserer Persönlichkeit zugeschrieben, über die der Geist einfach verfügen und diese ändern darf. Die neue Persönlichkeit entspräche dann am Ende mehr dem, wie man sich wünschte zu sein bzw. wie man ja eigentlich schon vorher wäre, wenn man nicht körperlich limitiert wäre.[11] Auch hier wird also dem Geist eine höhere Priorität eingeräumt und damit eine Trennung von Körper und Geist vorgenommen.

Hier zeigt sich eine Problematik, die besonders beim Moral Enhancement deutlich wird, welche aber auch für andere transhumanistische Positionen gilt:[12] Die Umsetzbarkeit des Projektes ist abhängig von der engen Verknüpfung zwischen Körper und Geist; denn nur wenn eine solche besteht, lassen sich signifikante Verbesserungen auf neurotechnischer Basis umsetzen. Gleichzeitig muss diese enge Verknüpfung aber abgewertet werden, da sich nur dann das Enhancement überhaupt rechtfertigen lässt. Anders ausgedrückt: Nur wenn die Verbindung defizitär ist, gibt es einen Spielraum, in dem Verbesserungen wirken können. Vor diesem Hintergrund ist es den Befürworter:innen von Moral Enhancement unmöglich, den Menschen als Leib-seelische-Einheit zu denken und sie sind damit immer auf einen Dualismus verwiesen, der sowohl philosophisch-theologisch als auch neurowissenschaftlich mindestens eine Außenseiterposition darstellt.[13]

---

[10] Douglas: Enhancement, 107. Hervorhebung im Original.

[11] Vgl. ebd., 98f.

[12] Da Enhancement eine der Kernideen des Transhumanismus darstellt, können auch prominente Vertreter:innen von Moral Enhancement wie bspw. Julian Savulescu, die sich nicht als Transhumanisten verstehen, in diesem Kontext ähnlich bewertet werden.

[13] Vgl. bspw. Beckermann, Ansgar: Gehirn, Ich, Freiheit. Paderborn 2010, 52 f.
Dies gilt natürlich nur für Positionen, die überhaupt davon ausgehen, dass Freiheit existiert. Vertreter:innen eines wie auch immer gearteten Reduktionismus können diesem Dualismus zwar entgehen, verfallen dann aber der Vorstellung eines ontologischen Naturalismus, die bereits verschiedentlich kritisiert worden ist. Vgl. bspw. Honnefelder, Ludger: Das Problem der Philosophischen Anthropologie. In: Ders. (Hg.): Die Einheit des Menschen. Paderborn u. a. 1994, 9–24.

3 Das Verhältnis von Körper und Geist im Kontext der Debatte um Willensfreiheit

Vor diesem Hintergrund scheint es daher eine geeignete Strategie zu sein, für eine fundierte Zurückweisung dieses Dualismus zu argumentieren, um die Idee des Moral Enhancement schon auf grundsätzlicher Ebene zurückweisen zu können. Dabei sollten die grundlegenden Positionen der Ansätze von Moral Enhancement durchaus auch für eine solche Zurückweisung übernommen werden. Dies hat mehrere Vorteile: Erstens erhöht eine Zurückweisung auf Grundlage derselben Prämissen die argumentative Kraft, da von derselben argumentativen Basis ausgegangen wird. Zweitens scheinen die beiden Grundpositionen (schwacher Nonkognitivismus und Kompatibilismus) durchaus auch in ähnlicher Form plausibel und relevant für andere Formen des Enhancements zu sein. Eine generell enge Verknüpfung von Körper und Geist mit relevanten Auswirkungen körperlicher Aspekte auf unser Verhalten oder unsere mentalen Fähigkeiten scheint die Grundlage für die meisten Arten des Enhancements zu sein. Erfolge auf diesem Gebiet plausibilisieren damit ebenfalls diese beiden Thesen. Drittens regen diese Grundpositionen eine vertiefte Reflexion der Bestimmung des Menschen als Leib-seelischer-Einheit aus theologisch-ethischer Perspektive an. Da das Ziel dieses Artikels ist, neue Perspektiven aufzuzeigen, die sich aus einer Beschäftigung mit dem Moral Enhancement ergeben, werde ich mich im Folgenden auf diesen dritten Punkt konzentrieren. Aufgrund der ersten beiden Punkte ergeben sich so aber bereits erste Perspektiven für eine grundsätzliche Zurückweisung der hier vorgestellten Ansätze zum Moral Enhancement.

3.1 Willensfreiheit als Illusion oder zwei Beschreibungssysteme der Wirklichkeit

Um zu verdeutlichen, wie eine Beschäftigung mit dem Moral Enhancement zu einer vertieften Reflexion über Fragen der Leiblichkeit des Menschen anregen kann, möchte ich an die Diskussion um die Frage, ob die Willensfreiheit eine Illusion ist, Ende der 1990er/Anfang der 2000er Jahre erinnern, die zwischen den Neurowissenschaften[14] und der (theologischen)

[14] Auch wenn in der Literatur häufig nur von „Neurowissenschaft" oder „Hirnforschung" die Rede ist, sollte klar sein, dass hier ein Verbund verschiedener wissenschaftlicher Disziplinen gemeint ist, die sich alle auf die eine oder andere Art und Weise mit dem Gehirn und dem zentralen Nervensystem beschäftigen. Dazu gehören bspw. Medizin, Psychologie, Biologie etc. In diesem Artikel wird daher immer im Plural von „Neurowissenschaften" gesprochen. Des Weiteren sollte auch klar sein, dass nicht alle Neurowissenschaftler:innen der These, dass die Willensfreiheit eine Illusion sei, zustimmen. Um zu komplizierte

Ethik geführt wurde. Auch damals ging es um die Reflexion des Verhältnisses von Körper und Geist vor dem Hintergrund aktueller neurowissenschaftlicher Erkenntnisse.[15] Auslöser waren die sogenannten Libet-Experimente[16], die bereits Ende der 70er/Anfang der 80er Jahre nahelegten, dass das Bereitschaftspotential – eine elektrische Veränderung des Gehirns, die die Bewegung der Muskeln vorbereitet – bereits ca. 350 ms vor dem Bewusstwerden des Willens zu einer spezifischen Handlung entstehe. Die intuitive Ereigniskette Wille → Bereitschaftspotential → Bewegung – welche Libet ursprünglich beweisen wollte – wurde damit umgekehrt zu: Bereitschaftspotential → Wille → Bewegung.[17] Weitere Experimente und neue Erkenntnisse in den folgenden Jahren, die diesen Eindruck noch verstärkten, führten schließlich einige Neurowissenschaftler:innen zu der These, dass sich in naher Zukunft durch eine neurowissenschaftliche „Theorie des Gehirns […] auch die schweren Fragen der Erkenntnistheorie angehen [lassen]: nach dem Bewusstsein, der Ich-Erfahrung und dem Verhältnis von erkennendem und zu erkennenden Objekt."[18] Mentale Prozesse sollen so auf biologische reduziert werden, was in seiner Konsequenz für manche Neurowissenschaftler:innen und Philosoph:innen bedeutete, dass die Willensfreiheit nichts anderes als eine Illusion sei.[19]

Diese These führte auf Seiten der Ethik zu starkem Widerspruch. Denn, – wie Eberhard Schockenhoff prägnant formuliert – „[d]ie Frage nach der Freiheit des Menschen benennt nicht eine unter vielen anderen philosophischen Problemstellungen, sondern die ethische Grundfrage, mit der alles Nachdenken über das moralische Handeln des Menschen seinen Anfang

Formulierungen zu vermeiden, soll daher im Folgenden immer nur die Fraktion derjenigen gemeint sein, die dieser These – im weitesten Sinne – zustimmen, da diese Position hier zur Debatte steht.

15 Für einen kurzen Überblick über die Geschichte dieser Diskussion vgl. Beckermann, Ansgar: Gehirn, 15–53.

16 Vgl. Libet, Benjamin u. a.: Readiness-potentials preceding unrestricted 'spontaneous' vs. preplanned voluntary acts. In: Electroencephalography and Clinical Neurophysiology 54 (1982), 322–335; Libet, Benjamin u. a.: Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). In: Brain 106 (1983), 623–642.

17 Vgl. Käuflein, Albert: Hirnforschung, Freiheit und Ethik. In: Ders./Macherauch, Thomas (Hg.): Determiniert oder frei? Karlsruhe 2006, 11–27, hier 13.

18 Monyer, Hannah/Rösler, Frank/Roth, Gerhard u. a.: Das Manifest. In: Gehirn & Geist 2004/6, 30–37, hier 37.

19 Vgl. bspw. Roth, Gerhard: Fühlen, Denken, Handeln. Frankfurt a. M. 2003, hier 395 f., 553; Wegner, Daniel: The Illusion of Conscious Will. Cambridge 2002; Singer, Wolf: Über Bewusstsein und unsere Grenzen. In: Becker, Alexander u. a. (Hg.): Gene, Meme und Gehirne. Frankfurt a. M. 2003, 279–305, hier 298 f.

nimmt."[20] Wenn Freiheit eine Illusion ist, erledigen sich auch die Fragen nach Ethik und Verantwortung.[21] In Reaktion auf die These von der Willensfreiheit als Illusion fand dementsprechend eine verstärkte Reflexion über das Verhältnis von Körper und Geist statt, um den Erkenntnissen der Neurowissenschaften einerseits Rechnung zu tragen, andererseits aber nicht das Konzept der Willensfreiheit aufgeben zu müssen.

Eine ausführliche Darstellung der vielfältigen Reaktionen ist an dieser Stelle leider nicht möglich.[22] Zentral für die Reflexion über das Verhältnis von Körper und Geist ist aber insbesondere das Argument, dass es zwei Beschreibungssysteme der Wirklichkeit (eines aus der Ersten-Person- bzw. Teilnehmerperspektive und eines aus der Dritten-Person- bzw. Beobachterperspektive) gebe, die nicht aufeinander reduziert werden können. Freiheit gehöre zu der Ersten-Person-Perspektive und könne daher nicht in der Dritten-Person-Perspektive bewiesen (oder widerlegt) werden. In seinen Grundzügen geht dieses Argument auf die Unterscheidung von Ursachen und Gründen zurück, die bereits in Platons *Phaidon* zu finden ist.[23] Diese Unterscheidung der Beschreibungstypen mache eine Besonderheit menschlichen Handelns deutlich: die Intentionalität. „Was menschliche Handlungen von physikalischen Ereignissen unterscheidet, ist die Struktur ihrer Intentionalität; Menschen handeln um der Ziele willen, die sie durch ihr Handeln erreichen wollen."[24] Intentionalität, Ziele, aber auch Freiheit und Verantwortung könnten nur im Bereich der Gründe verständlich werden.

> „Werden sie auf kausale Ereignisketten reduziert, muss sich der Mensch als das verantwortlich agierende Subjekt seiner Taten unverständlich werden. Er kann weder erklären, warum er sich selbst als Adressat von Aufforderungen und seine Entscheidungen als Resultat der Abwägung von Gründen erfährt, noch kann er Gründe dafür anführen, warum er die Frage nach der Geltung von Handlungsgründen sinnvoll stellen und sein Handeln aufgrund der Einsicht in die Gründe korrigieren kann."[25]

---

[20] Schockenhoff, Eberhard: Wie frei ist der Mensch? In: Gestrich, Christof/Wabel, Thomas (Hg.): Freier oder unfreier Wille? Berlin 2005, 53–66, hier 60.

[21] Vgl. ebd., 57; Schockenhoff, Eberhard: Wer oder was handelt? In: Rager, Günter (Hg.): Ich und mein Gehirn. Freiburg i. Br. 2000, 239–287, hier 247.

[22] Vgl. neben den bereits erwähnten Autoren bspw. Honnefelder, Ludger: Die ethische Dimension moderner Hirnforschung. In: Der deutsche Ethikrat (Hg.): Der steuerbare Mensch? Berlin 2009, 83–95; Rosenberger, Michael: Determinismus und Freiheit. Darmstadt 2006; Ernst, Stephan: Ist Freiheit noch denkbar? In: Trierer Theologische Zeitschrift 117 (2008), 192–213.

[23] Platon: Phaidon. In: Ders.: Sämtliche Werke. Band 2, übersetzt von Friedrich Schleiermacher, hg. von Ursula Wolf. Hamburg 2018, 103–184, hier 160 (98e).

[24] So bspw. Schockenhoff: Wie frei ist der Mensch, 56.

[25] Honnefelder: Die ethische Dimension, 85.

Es gehöre entscheidend zu unserer Selbsterfahrung, dass wir uns als frei, als intentional handelnd, als Gründen zugänglich erleben. Hier liegt aber auch das grundsätzliche Problem, denn wir können uns so ausschließlich aus unserer Selbsterfahrung, also der Ersten-Person-Perspektive, erleben. Von außen betrachtet lässt sich dagegen für jede Handlung auch eine naturwissenschaftliche Erklärung geben.[26] Vor diesem Hintergrund ließe sich eine Gleichwertigkeit beider Perspektiven durchaus plausibilisieren.[27]

3.2 Das Verhältnis von Körper und Geist

Was bedeutet dies nun für die Verhältnisbestimmung von Körper und Geist? Ludger Honnefelder bringt die Überlegungen mit dem auf Helmuth Plessner zurückgehenden Diktum auf den Punkt, dass der Mensch (als „Leib-seelischer-Einheit") nicht nur „sein Leib *ist* […, sondern; D. W.] ihn zugleich als Körper *hat*."[28] Dies spiegelt zum einen die Unterscheidung in der Betrachtungsweise des Menschen in den beiden Perspektiven wider – die Betrachtung als Leib entspricht der Ersten-Person-Perspektive, die als Körper der Dritten-Person-Perspektive – und kann zum anderen die Besonderheit des Menschen in dieser Unterscheidung verdeutlichen. Die Freiheit des Menschen liege darin, dass er sich zu dem Verhältnis seines Körpers zum Geist

---

[26] Mit Verweis auf die Analyse Godehard Brüntrups meint allerdings Michael Rosenberger, dass während die Vorstellung von Freiheit als Illusion und damit sozialer Konstruktion von den Neurowissenschaften sehr schnell akzeptiert werde, völlig aus dem Blick gerate, dass unsere Vorstellung von Kausalität ebenfalls ein soziales Konstrukt sei. Mit Blick auf Kausalitätserklärungen hält Brüntrup fest, dass „Kausalerklärungen […] intrinsisch ein epistemisches Moment [enthalten], das sich der Verobjektivierung entzieht. Im reduktionistischen Programm zeigt sich das Problem dadurch, daß beim Wechsel von der Ebene der Erklärung auf die Ebene der physikalisch-objektiven Relation ein wichtiges Element der Kausalerklärung nicht herübergerettet werden kann. In einer Kausalerklärung greifen wir normalerweise bestimmte im vorliegenden Kontext relevante Aspekte des gesamten physikalischen Geschehens heraus. […] Jede Grenzziehung zum Zweck der Kausalerklärung ist hier subjektiv und widersetzt sich der Objektivierung. Man stößt hier auf das notorische und vieldiskutierte Problem, daß die Vorstellung von Kausalität ohne Rücksicht auf einen Erklärungskontext völlig nichtssagend und leer bleibt. Kausalität und Erklärung lassen sich nicht voneinander trennen. […] Jeder benutzbare (weil aussagekräftige) Kausalitätsbegriff enthält eine epistemische Komponente und ist auf Erklärungszusammenhänge bezogen. Wenn aber der so verstandene Kausalitätsbegriff nicht unabhängig vom Erklärungsbegriff ist, dann sind alle kausalen Kontexte intensionale Kontexte […]. Die strenge Unterscheidung ‚x verursacht y' (extensional) von ‚x erklärt kausal y' (intensional) erweist sich als undurchführbar. Darum eignet sich der normalerweise benutzte Kausalitätsbegriff auch nicht für die Grundlegung einer metaphysischen Theorie über die geistunabhängige Welt." Brüntrup, Godehard: Mentale Verursachung. Stuttgart u. a. 1994, 197 f.

[27] Rosenberger: Determinismus, 215.

[28] Honnefelder: Die ethische Dimension, 87. Hervorhebung im Original.

(Dritte-Person-Perspektive) noch einmal ins Verhältnis setzen könne (Erste-Person-Perspektive).[29] Schockenhoff macht dies in diesem Kontext ebenfalls deutlich, indem er Freiheit als „die Fähigkeit des Menschen, das Wechselspiel der unterschiedlichen Determinanten seines Handelns aktiv und autoregulativ zu beherrschen"[30], versteht. Ähnlich urteilt auch Stephan Ernst – in Anschluss an Thomas von Aquin –, dass es dem Menschen durch eine grundsätzliche Distanziertheit, die er zu allen irdischen Gütern einnehmen könne, gegeben sei, „dass wir von dem, was uns erstrebenswert erscheint, *nicht gezwungen werden*, es auch tatsächlich zu wollen."[31]

Alle drei nehmen also eine Beeinflussung des Willens durch physische und psychologische Einflüsse an, gehen aber davon aus, dass sich der Wille diesen Einflüssen gegenüber noch einmal verhalten und so seine Freiheit bewahren könne. Auch wenn es an keiner Stelle explizit gesagt wird, vertreten mindestens Schockenhoff und Ernst dabei eine klar libertarische Position. Dies ergibt sich daraus, dass beide nicht nur keine zwingende Determination durch physische Einflüsse annehmen, sondern auch eine Determination durch Gründe verneinen. Ernst sieht bei der von der Tradition als *libertas specificationis* diskutierten Freiheit – der Freiheit, dies oder jenes wählen zu können – das Dilemma, dass die Wahl zwischen zwei Gütern, wenn sie Gründen folgt, zwar rational, aber nicht frei, oder wenn sie keinen Gründen folgt, zwar frei, aber nicht rational sei.[32] Eine Wahl, die durch Gründe determiniert ist, ist für Ernst also nicht frei. Dahinter steht ein anspruchsvoller Begriff von Freiheit, der die Fähigkeit zum „Anders-handeln-Können" beinhaltet, die auch angesichts rationaler Gründe bestehen bleiben soll. Selbst rationale Gründe würden dann nicht dazu führen, dass eine Entscheidung in einer Situation immer gleich gefällt werden würde – nur dann gilt sie als wirklich frei. Wenn wir uns eine Zeitschleife vorstellen, in der wir immer wieder vor derselben Entscheidung stehen (bspw. einen Diebstahl zu begehen), dabei aber jedes Mal vergessen, dass wir diese Entscheidung schon einmal getroffen haben, würde ein starker libertarischer Freiheitsbegriff beinhalten, dass diese Entscheidung jedes Mal anders ausfallen könnte, obwohl sowohl die körperlichen als auch die geistig-rationalen Gegebenheiten immer dieselben sind. Das Problem an dieser Sicht erläutert Ansgar Beckermann knapp, wie folgt:

> „Die Gründe, angesichts deren eine Person entscheidet, gehören mit zu der Situation, in der sie sich entscheidet. Wenn sie sich für *A* entscheidet, entscheidet sie sich angesichts *dieser* Gründe für *A*. Und wenn sie sich für *B* entscheidet, entscheidet sie sich angesichts *genau derselben* Gründe für *B*. Wenn jemand angesichts derselben Gründe einmal

[29] Vgl. ebd.; ähnlich Schockenhoff: Wie frei ist der Mensch, 61; Ernst: Ist Freiheit, 208–210.

[30] Schockenhoff: Wie frei ist der Mensch, 63.

[31] Ernst: Ist Freiheit, 210. Hervorhebung im Original.

[32] Vgl. ebd., 208.

> die Alternative *A* und das andere Mal die Alternative *B* wählt, ist die Wahl selbst aber offenbar unbegründet. Nichts auf der Welt bestimmt, wie sie ausfällt. Offenbar ist es purer Zufall, welche Alternative die entscheidende Person wählt“[33]

Schockenhoff sieht dies genauso wie Ernst, wenn er festhält: „Ein erkanntes und bewusst gewähltes Ziel ‚verursacht‘ ihr [der Menschen; D. W.] Handeln jedoch nicht, denn es bleibt ihnen die Möglichkeit, auch anders zu handeln.“[34] Ernst und Schockenhoff vertreten also weiterhin einen sehr anspruchsvollen Begriff von Willensfreiheit.

Albert Käuflein und auch Rosenberger vertreten dagegen eine kompatibilistische Position. Rosenberger beruft sich dabei ebenfalls wie Beckermann auf das logische Problem in der libertarischen Position, welches bereits erwähnt wurde.[35] Neben dieser Ablehnung einer libertarischen Position definiert er aber nicht näher, wie er seinen Kompatibilismus genau versteht und verweist eher auf Fragen der Wissenschaftstheorie, die mit einer konstruktivistischen Perspektive auf beide Betrachtungsweisen einhergehen.[36] Auch Käuflein sieht das Problem darin, dass die Möglichkeit, unter identischen Umständen anders und zugleich aus verständlichen Gründen handeln zu können, nicht miteinander vereinbar seien.[37] Stattdessen geht er davon aus, dass wir uns selbst in Freiheit determinieren. Diese Determination sei keine durch innere oder äußere Zwänge, sondern eine durch Gründe und frühere Entscheidungen.[38] Auch er stimmt aber mit den anderen hier genannten Autoren darin überein, dass es einige Einflüsse auf den Willen gebe und „[m]enschliche Freiheit niemals absolut voraussetzungs- und bedingungslos“[39] sei.

4 Weiterführende Reflexion vor dem Hintergrund des Moral-Enhancement-Diskurses

Auch wenn die Beteiligten an der damaligen Debatte somit zu verschiedenen Graden eine Beeinflussung unserer Freiheit durch körperliche Aspekte annahmen, gab es keine nähere theologisch-ethische Reflexion darüber, wie diese Beeinflussung eigentlich aussehen und in moraltheologische Überlegungen integriert werden könnte. Teilweise könnte sogar der Vorwurf

---

[33] Beckermann: Gehirn, 107. Hervorhebungen im Original. Für eine ausführlichere Erklärung s. ebd., 107–110.

[34] Schockenhoff: Wie frei ist der Mensch, 56.

[35] Vgl. Rosenberger: Determinismus, 202.

[36] Vgl. Rosenberger: Determinismus, 221 f.

[37] Vgl. Käuflein: Hirnforschung, 22.

[38] Vgl. ebd., 18 f.

[39] Ebd., 18.

formuliert werden, dass sich tatsächlich kaum ein Wandel im jeweiligen Denken vollzogen hat. Beispielhaft wird dies an der überraschenden Tatsache deutlich, dass mindestens Ernst und Schockenhoff weiterhin eine stark libertarische Position eingenommen haben, obwohl zum Teil auch aus theologisch-ethischer Perspektive dieser Position angesichts der Erkenntnisse der Neurowissenschaften eine klare Absage erteilt wurde. So hält bspw. Rosenberger fest:

> „Erkenntnistheoretisch geht es darum, eine objektivistische Interpretation libertarischer Freiheit endgültig zu den Akten zu legen. Freiheit im Sinne eines strikt objektiven Andershandelnkönnens, mithin im Sinne unvollständiger Determination[,] gibt es nicht."[40]

Gerade dieses Anders-handeln-Können nehmen Ernst und Schockenhoff, wie am Ende des vorherigen Abschnitts gezeigt, aber weiterhin an.

Doch auch Rosenberger und Käuflein helfen nicht, eine Position zu finden, welche die Grundpositionen des Moral Enhancement adäquat integrieren kann. Zwar vertreten beide einen Kompatibilismus, dieser ist aber entweder nicht wirklich ausgearbeitet (Rosenberger) oder beschränkt sich auf eine Determination durch Gründe, die nicht den Einfluss körperlicher Aspekte auf unsere Freiheit reflektiert (Käuflein). Die hier zu Wort gekommenen Positionen beschränken sich in Ihrer Verteidigung des freien Willens auf die Betonung der Unterscheidung zwischen Körper und Geist und bedenken dabei wenig die Seite der Leiblichkeit des Mensch-Seins, die aber bereits als zentral für eine Zurückweisung des dem Moral Enhancement zugrundeliegenden Dualismus identifiziert wurde. Dies ist sicherlich auch dem damaligen Diskurs geschuldet. Aus diesem Grund lohnt sich aber eine erneute Reflexion dieses Komplexes vor dem Hintergrund des Moral-Enhancement-Diskurses: Diesmal ist es nicht unsere Freiheit, die bestritten, sondern unsere Leiblichkeit, die als minderwertig betrachtet wird und die nun einer Verteidigung bedarf. Zum Schluss möchte ich daher einen Weg anbahnen, auf dem die vertiefte Reflexion über den Menschen als Leib-seelischer-Einheit weitergeführt werden kann.

4.1 Neurowissenschaftliche Plausibilisierung des Freiheitsbegriffs

In der bereits behandelten Debatte zwischen Neurowissenschaften und Ethik gab es auch Stimmen auf Seiten der Neurowissenschaftler:innen, die sich gegen die These von der Willensfreiheit als Illusion aussprachen und versucht haben, eine Erklärung für diese zu liefern. Ein

[40] Rosenberger: Determinismus, 202.

prominenter Vertreter dieser Richtung war (und ist) bspw. Wolfgang Prinz, auf den ich mich im Folgenden berufe.[41]

Prinz hält fest, dass uns Freiheit vor allem in Handlungsentscheidungen bewusst werde, obwohl es nicht unsere Entscheidungen für oder gegen Handlungen seien, die uns frei erscheinen, sondern wir uns selbst in diesen Entscheidungen als frei betrachten würden.[42] Handlungsentscheidungen würden aber aufgrund „subpersonaler Prozesse" – sprich unbewusster Prozesse – getroffen, auf die die Institution eines Selbst keinen direkten Einfluss nehmen könne.[43] Als ein Beispiel solcher subpersonaler Prozesse nennt er die Unfähigkeit, eine adäquate Beschreibung des Denkvorgangs zu geben, wenn ein Text verstanden wird: „Wenn man sich dabei selbst beobachtet, kann man nichts weiter feststellen, als dass man versteht oder nicht versteht, aber wie das Verstehen vor sich geht, weiß niemals die Versuchsperson, sondern allenfalls der Theoretiker", der in diesem Fall eine „subpersonale Maschinerie"[44] annimmt. Ähnlich lassen sich auch Erkenntnisse von Gerhard Roth verstehen, der bspw. dem limbischen System eine entscheidende Rolle zuspricht bei der Entscheidung, was uns überhaupt ins Bewusstsein kommt und was nicht und das er maßgeblich von Emotionen abhängig sieht.[45] Prinz stellt dabei eine gewisse Übereinstimmung in verschiedenen Disziplinen der Psychologie fest, dass Handlungsentscheidungen mindestens drei Ingredienzien benötigen: Präferenzen, Handlungswissen und Situationsbewertungen, deren Zusammenführung aber keine besondere personale Instanz benötige: „Entscheidungen kommen zustande, ohne dass da jemand wäre, der sie trifft."[46]

Er betrachtet das Selbst und damit den Träger von Freiheit dabei als eine soziale Konstruktion, die zwar kein eigenes „Organ der Seele" darstelle, nichtsdestoweniger aber als real betrachtet werden müsse und von ihm als eine spezifische Wissensstruktur verstanden wird.[47]

> „Mit anderen Worten: auch personale Wahrnehmung kann nur auf der Grundlage subpersonaler Prozesse zustande kommen, und deshalb besteht *a priori* kein Grund, sie als weniger wirklich und wirksam anzusehen als die subpersonalen Entscheidungsprozesse, auf die sie sich bezieht."[48]

---

41 Ähnlich bspw. Goschke, Thomas: Vom freien Willen zur Selbstdetermination. In: Psychologische Rundschau 55/4 (2004), 186–197.

42 Prinz, Wolfgang: Kritik des freien Willens. In: Psychologische Rundschau 55/4 (2004), 198–206, hier 201.

43 Vgl. ebd., 202.

44 Für beide Zitate ebd., 201.

45 Vgl. Roth, Gerhard: Das Gehirn und seine Wirklichkeit. Frankfurt a. M. 1997, hier 306 f.

46 Prinz: Kritik, 202.

47 Vgl. ebd., 203 f.

48 Ebd., 204. Hervorhebung im Original

Das Selbst basiere damit essentiell auf subpersonalen Prozessen, nehme aber auch in langfristiger Hinsicht Einfluss auf diese und stelle so eine sehr enge Verbindung zwischen dem, was traditionellerweise in körperliche und mentale Prozesse getrennt wird, her. Dieser Einfluss entstehe dadurch, dass die subpersonalen Prozesse durch die Selbst-Wahrnehmung in weitere Verarbeitungsschleifen versetzt würden und so

> „eine Vertiefung der Verarbeitung bewirken [gemeint ist die Verarbeitung von Informationen in einer Situation; D. W.]. Diese Vertiefung bringt zunächst eine Verbreiterung der Informationsbasis mit sich, die für Entscheidungen zur Verfügung steht und kann allein dadurch bereits zu einer Modifikation der Entscheidung selbst führen."[49]

Sehr vereinfacht lässt sich festhalten: Subpersonale Prozesse beeinflussen laut Prinz die Konstruktion und Wahrnehmung meines Selbst und dieses Selbst wirkt sich wiederum auf die subpersonalen Prozesse aus. Was wir mit „Geist" bezeichnen, wäre damit überhaupt nicht ohne diese subpersonalen Prozesse verständlich. Wenn wir Prinz darin folgen, entstehen unsere Handlungsentscheidungen also aus eng verknüpften Wechselwirkungsprozessen zwischen Körper und Geist: Dem handelnden Selbst liegen in der konkreten Handlungssituation subpersonale Prozesse zugrunde, die in langfristiger Perspektive wiederum vom Selbst beeinflusst werden.

4.2 Zur Rolle von Emotionen in der Ethik

Wenn wir nun in die theologisch-philosophische Reflexion zurückkehren, lässt sich vor dem eben dargestellten Hintergrund festhalten, dass die Grundpositionen des Moral Enhancement durchaus eine Plausibilität besitzen. Wenn unsere Selbst-Wahrnehmung und Freiheit in derart enger Verbindung mit subpersonalen Prozessen betrachtet werden, spricht vieles für die Annahme eines Kompatibilismus. Dieser kann sich in ethischen Kontexten durchaus daran zeigen, dass ethische Entscheidungsfindungen maßgeblich durch Emotionen beeinflusst sind, wie mit der These des schwachen Nonkognitivismus angenommen wird. Tatsächlich gibt es inzwischen erste Ansätze, die Rolle von Emotionen in ethischen Kontexten so zu reflektieren, dass diese nicht sofort in einen starken Nonkognitivismus übergehen.[50] Es gilt, Emotionen eine solche Rolle zukommen zu lassen, dass sie als essentieller Teil einer moralischen Ent-

[49] Ebd.

[50] Vgl. bspw. Klöcker, Katharina: Von der Autorität der Leidenden zu einer Moral der Fehlbarkeit. In: Autiero, Antonio/Goertz, Stephan/Merks, Karl-Wilhelm (Hg.): Autorität in der Moral. Freiburg i. Br. 2019, 191–208, hier 203–207; Breitsameter, Christof: Die Semantik „moralischer Gefühle" zwischen Aktion, Reaktion und Interaktion. In: Münchener theologische Zeitschrift 66/3 (2015), 243–256.

scheidungsfindung betrachtet werden können, ohne auf eine rationale Reflexion dieser Urteile verzichten zu müssen, wie es bspw. der Emotivismus tut.

Wichtige Ansatzpunkte lassen sich dazu in den neueren Überlegungen zur Philosophie der Emotionen finden, in denen eine kognitivistische Emotionstheorie entwickelt wird.[51] Traditionell wird – in Anlehnung an Hume – in der Metaethik zwischen kognitiven und konativen Aspekten eines moralischen Urteils unterschieden: Die rein kognitive Überlegung fällt das Urteil darüber, ob eine Handlung richtig oder falsch ist, der hinzutretende konative Zustand – bspw. ein Wunsch oder eine Emotion – liefert die zum entsprechenden Handeln nötige Motivation.[52] Diese Unterscheidung und scharfe Trennung – die aus verschiedenen Gründen zu einigen Problemen und damit überhaupt erst zu starken nonkognitivistischen Theorien geführt hat[53] – kann durch eine kognitivistische Emotionstheorie überwunden werden, wenn Emotionen nicht mehr eindeutig dem rein konativen Bereich zugeordnet, sondern ihnen ebenfalls kognitive Aspekte zugesprochen werden. Dabei wird festgehalten, dass die für diesen Kontext relevanten Emotionen immer auch einen intentionalen Bezug zur Welt haben, der gerechtfertigt – und damit wahr oder falsch – sein kann.[54] Die Emotion Angst ist bspw. immer eine Angst vor etwas. Ich habe Angst vor der Schlange, weil ich diese bspw. für giftig und damit gefährlich halte. Dieser intentionale Aspekt kann dabei wahr oder falsch sein und damit die Emotion rechtfertigen: Ist die Schlange tatsächlich giftig, ist meine Angst berechtigt.[55] Vor diesem Hintergrund scheint es möglich, eine kognitive Verbindung zwischen Emotion und Situationsbewertung herzustellen und Emotionen damit in Wertetheorien über rein motivationale Aspekte hinaus zu integrieren.[56]

5 Fazit

Diese beiden Beispiele können an dieser Stelle nur eine grobe Richtungsandeutung dafür sein, wie die Trennung zwischen körperlichen und mentalen Aspekten unseres Seins zumindest reduziert werden kann. Die Beschäftigung mit dem Moral Enhancement und damit auch dem möglichen Einfluss neuer Technologien auf den Menschen regt eine neue intensive Reflexion

[51] Für eine Vertiefung und weiterführende Literatur zu diesem Themenkomplex vgl. auch Döring, Sabine (Hg.): Philosophie der Gefühle. Frankfurt a. M. 2009; Weber-Guskar, Eva: Die Klarheit der Gefühle. Berlin 2009.

[52] Vgl. bspw. Niederbacher, Bruno: Metaethik. Stuttgart 2021, 23 f.

[53] Vgl. ebd., 28-31.

[54] Vgl. dazu bspw. Helm, Bennett: Emotional Reason. Cambridge 2001, 103 f.; Goldie, Peter: Emotionen und Gefühle. In: Döring, Sabina A.: Philosophie der Gefühle. Frankfurt a. M. 2009, 369–397.

[55] Vgl. Breitsameter: Die Semantik, 249–251.

[56] Vgl. ebd.

über unser Selbstverständnis an. Die inhärente Leibfeindlichkeit solcher transhumanistischen Ansätze nötigt uns dazu, das Verständnis des Menschen als Leib-seelischer-Einheit vor dem Hintergrund aktueller Erkenntnisse neu zu bedenken und geben der Theologie die Chance, dieses Verständnis positiv zu besetzen, indem unsere Leiblichkeit als essentieller Teil unserer Selbst verstanden und dualistische Positionen auch in der Theologie endgültig überwunden werden. Wenn dies auf Grundlage derselben grundlegenden Prämissen geschieht, wie sie von Positionen des Moral Enhancement angenommen werden, eröffnet uns dies nicht nur einen Weg, transhumanistische Positionen zurückzuweisen, sondern zeigt uns auch neue Reflexionswege für die Theologie. Die Auseinandersetzung mit dem Transhumanismus und neuen Technologien, die in diesem Kontext eine Rolle spielen, eröffnen ein vertieftes Verständnis dafür, wie wir uns selbst nicht nur als Körper-habend, sondern auch als Leib-seiend begreifen können.

Literaturverzeichnis

Beckermann, Ansgar: Gehirn, Ich, Freiheit. Neurowissenschaften und Menschenbild. Paderborn 2010.

Breitsameter, Christof: Die Semantik „moralischer Gefühle" zwischen Aktion, Reaktion und Interaktion. In: Münchener theologische Zeitschrift 66/3 (2015), 243–256.

Brüntrup, Godehard: Mentale Verursachung. Eine Theorie aus der Perspektive des semantischen Anti-Realismus. Stuttgart u. a. 1994.

Döring, Sabine (Hg.): Philosophie der Gefühle. Frankfurt a. M. 2009.

Douglas, Thomas: Moral Enhancement via direct emotion modulation. A reply to John Harris. In: Bioethics 27/3 (2013), 160-168. DOI: 10.1111/j.1467-8519.2011.01919.x.

Douglas, Thomas: Enhancement der Moral. In: van Riel, Raphael/Di Nucci, Ezio/Schildmann, Jan (Hg.): Enhancement der Moral. Münster 2015, 85–111.

Ernst, Stephan: Ist Freiheit noch denkbar? Philosophische und theologische Perspektiven angesichts der neueren Hirnforschung. In: Trierer Theologische Zeitschrift 117 (2008), 192–213.

Goldie, Peter: Emotionen und Gefühle. In: Döring, Sabina A.: Philosophie der Gefühle. Frankfurt a. M. 2009, 369–397.

Goschke, Thomas: Vom freien Willen zur Selbstdetermination. Kognitive und volitionale Mechanismen der intentionalen Handlungssteuerung. In: Psychologische Rundschau 55/4 (2004), 186–197. DOI: 10.1026/0033-3042.55.4.186.

Helm, Bennett: Emotional Reason. Deliberation, Motivation, and the Nature of Value. Cambridge 2001.

Helmus, Caroline: Transhumanismus – der neue (Unter-)Gang des Menschen? Das Menschenbild des Transhumanismus und seine Herausforderung für die Theologische Anthropologie (ratio fidei 72). Regensburg 2020.

Honnefelder, Ludger: Das Problem der Philosophischen Anthropologie. Die Frage nach der Einheit des Menschen. In: Ders. (Hg.): Die Einheit des Menschen. Zur Grundfrage der philosophischen Anthropologie. Paderborn u. a. 1994, 9–24.

Honnefelder, Ludger: Die ethische Dimension moderner Hirnforschung. In: Der deutsche Ethikrat (Hg.): Der steuerbare Mensch? Über Einblicke und Eingriffe in unser Gehirn (Jahrestagung des Deutschen Ethikrates 2009). Berlin 2009, 83–95.

Käuflein, Albert: Hirnforschung, Freiheit und Ethik. In: Ders./Macherauch, Thomas (Hg.): Determiniert oder frei? Auseinandersetzung mit der Hirnforschung. Karlsruhe 2006, 11–27.

Klöcker, Katharina: Von der *Autorität der Leidenden* zu einer *Moral der Fehlbarkeit*. In: Autiero, Antonio/Goertz, Stephan/Merks, Karl-Wilhelm (Hg.): Autorität in der Moral. Historische und systematische Perspektiven (Jahrbuch für Moraltheologie 3). Freiburg i. Br. 2019, 191–208.

Libet, Benjamin u. a.: Readiness-potentials preceding unrestricted 'spontaneous' vs. pre-planned voluntary acts. In: Electroencephalography and Clinical Neurophysiology 54 (1982), 322–335.

Libet, Benjamin u. a.: Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). The Unconscious Initiation of a Freely Voluntary Act. In: Brain 106 (1983), 623–642.

Monyer, Hannah/Rösler, Frank/Roth, Gerhard u. a.: Das Manifest. Elf führende Neurowissenschaftler über Gegenwart und Zukunft der Hirnforschung. In: Gehirn & Geist 2004/6, 30–37.

Niederbacher, Bruno: Metaethik. Stuttgart 2021.

Persson, Ingmar/Savulescu, Julian: Unfit for the Future. The Need for Human Enhancement. Oxford 2012.

Platon: Phaidon. In: Ders.: Sämtliche Werke. Band 2, übersetzt von Friedrich Schleiermacher, hg. von Ursula Wolf. Hamburg [36]2018, 103–184.

Prinz, Wolfgang: Kritik des freien Willens. Bemerkungen über eine soziale Institution. In: Psychologische Rundschau 55/4 (2004), 198–206. DOI: 10.1026/0033-3042.55.4.198.

Rosenberger, Michael: Determinismus und Freiheit. Das Subjekt als Teilnehmer. Darmstadt 2006.

Roth, Gerhard: Das Gehirn und seine Wirklichkeit. Kognitive Neurobiologie und ihre philosophischen Konsequenzen. Frankfurt a. M. 1997.

Roth, Gerhard: Fühlen, Denken, Handeln. Wie das Gehirn unser Verhalten steuert. Frankfurt a. M. 2003.

Savulescu, Julian/Persson, Ingmar: Enhancement der Moral, Freiheit und die Gottmaschine. In: van Riel, Raphael/Di Nucci, Ezio/Schildmann, Jan (Hg.): Enhancement der Moral. Münster 2015, 51–75.

Schockenhoff, Eberhard: Wer oder was handelt? Überlegungen zum Dialog zwischen Neurobiologie und Ethik. In: Rager, Günter (Hg.): Ich und mein Gehirn. Persönliches Erleben, verantwortliches Handeln und objektive Wissenschaft. Freiburg i. Br. 2000, 239–287.

Schockenhoff, Eberhard: Wie frei ist der Mensch? Zum Dialog zwischen Hirnforschung und theologischer Ethik In: Gestrich, Christof/Wabel, Thomas (Hg.): Freier oder unfreier Wille? Handlungsfreiheit und Schuldfähigkeit im Dialog der Wissenschaften. Berlin 2005, 53–66.

Singer, Wolf: Über Bewusstsein und unsere Grenzen. Ein neurobiologischer Erklärungsversuch. In: Becker, Alexander u. a. (Hg.): Gene, Meme und Gehirne. Geist und Gesellschaft als Natur. Frankfurt a. M. 2003, 279–305.

Weber-Guskar, Eva: Die Klarheit der Gefühle. Was es heißt, Emotionen zu verstehen. Berlin 2009.

Wegner, Daniel: The Illusion of Conscious Will. Cambridge 2002.

Winter, Dominik: Falsche Hoffnung. Warum emotives Moral Enhancement nicht die Freiheit vergrößern kann. In: Zeitschrift für Theologie und Philosophie 144/2 (2022), 220–242. DOI: 10.35070/ztp.v144i2.3764.

IV Transformation des Krieges

Autoregulative Waffensysteme

Autoregulative Weapons Systems

Automatization challenging Peace Ethics[1]

Nicole Kunkel

Abstract

Weapon systems with autoregulative functions are the subject of current political and social debates. The main ethical issue is such weapons possibly selecting and engaging human targets lethally without human intervention. Yet, since algorithms are prone to have biases implemented, may have unintended side-effects, and do rather reckon than judge, they can hardly make a moral decision by their own means. The terminology used, namely technological autonomy, however, suggests exactly that, as it is an anthropomorphism describing a human capacity the machine lacks. I therefore propose to use the term autoregulation instead, hinting towards the notion that the autoregulative device remains dependent on the autonomous agent. I further propose to assess the subject of autoregulative functions in weapons systems from the Christian peace ethical stance of just peace, mainly recurring to the peace memorandum of the Protestant Church in Germany, finding that the technology is indeed ethically problematic.

So-called autonomous weapons systems – or rather weapons systems with autoregulative functions – are the subject of current political and social debates. Especially within the international bodies of the UN, the development and possible deployment of autoregulative functions within weapons systems is being negotiated since 2014. To this end, I will at first introduce what exactly autoregulative functions in weapons systems are. Then, I will set forth why I prefer the term autoregulation to autonomy. Subsequently, I will introduce the arguments made so far within the ethical discourse. Finally, I will illustrate how the peace ethical discussion within the major Christin Churches shifted from just war reasoning towards a rationale of

[1] This article has been published originally in German in the magazine *Ethik und Gesellschaft* under the heading: Autoregulative Waffensysteme. Automatisierung als friedensethische Herausforderung – ein Werkstattbericht. In: Ethik und Gesellschaft 2 (2021).

just peace and, against that backdrop, identify some points worthwhile to discuss further from an ethical Protestant perspective.

1 Autoregulative functions in weapons systems

Autoregulative functions play a central role in current development of technology, especially within the field of robotics. Such devices are intended to operate without human intervention or control in a certain domain. With respect to traffic, for example, this concerns the development of self-driving – autoregulative – cars, while in the domain of care, deploying autoregulative care robots is under discussion.[2] Weapon systems are also affected by this process of automation. The mode of how such a weapons system operates is anything but complex and can be illustrated by the *IAI Harpy loitering munition*: Once the missile is launched, it hovers over enemy airspace until it receives a specific enemy radio signature – usually from a missile defense system. Upon reception, the weapon swoops down on the target, destroying it in the process.[3] Such a weapon, for all the military advantages it may offer, also poses several problems. For example, the weapon cannot assess easily whether the radio signature is being transmitted from a military site, a residential building, or even a hospital. All three scenarios, however, would be judged differently under current International Humanitarian Law (IHL): While the former is a legitimate target in war, the other two – certainly in gradations – do not constitute such a legitimate target.[4]

Another example of a somewhat different application of autoregulation is the military project FCAS (*Future Combat Air System*). In this joint project, France, Germany, and Spain, intend to utilize autoregulative functions. Co-developer Wolfgang Koch describes FCAS in the following terms:

> In the future, conflicts will be fought more automated than ever before. Within the framework of the *Future Combat Air System* FCAS, unmanned, artificially intelligent and technically autonomous aircrafts will accompany manned fighter jets of the latest generation as loyal *wing men.* In the event of an attack, they will protect the pilot and divert attention from them. In combat missions, they fly far ahead as *remote carriers*,

[2] See for instance: Loh, Janina: Roboterethik. Berlin 2019, 22–29; Misselhorn, Catrin: Grundfragen der Maschinenethik. Ditzingen 2018, 136–155.

[3] All information is taken from the official website of IAI. Online at: https://www.iai.co.il/p/harpy (as at: 12/01/2021).

[4] See here mainly the following paragraphs in the Protocol I Additional to the Geneva Conventions: 52(2); 51 (5)(b) and 57 (2)(iii).

> reconnoitering in a coordinated manner and engaging enemy targets. Other air defense components are also networked by FCAS as *systems-of-systems*: the Eurofighter, military transports, guided missiles, or AWACS [Airborne Early Warning and Control Systems, i.e., flying radar systems, note N. K.]. Thus, it is clear that Artificial Intelligence and Technological Autonomy must play a key role in FCAS.[5]

Accordingly, FCAS means not only to equip a single weapon with autoregulation, as is the case with IAI Harpy, but rather the goal is to operate a so-called system of systems, connecting various kinds of (crewed and uncrewed) devices in real time via a network.[6] The fact that such systems regulate themselves independently and without involving humans makes perfect sense against the background of such highly complex and very fast communication processes within the system. In an argument like the one mentioned above, there seems to be basically no other choice if such systems are to be developed.

However, Wolfgang Koch leaves open in which way 'artificial intelligence' and 'autonomy' are to be used and thus it remains unclear which functions actually are affected by automation.[7] This is why the question of how to define technological autonomy or autoregulation, plays a central role in the political and academic discourse. The first question therefore is: Which functions should be automated? For example, the automated takeoff and landing process for (military) drones does not necessarily lead to political and ethical problems. However, the issue at stake that causes controversial debates is the automation of the killing function. I follow the definition of the International Red Cross, which states that:

> Autonomous Weapon Systems are defined as any weapon system with autonomy in the critical functions of target selection and target engagement. That is, a weapon system that can select (i.e., detect and identify) and attack (i.e., use force against, neutralize, damage, or destroy) targets without human intervention.[8]

---

5 Koch, Wolfgang: FCAS – Challenges for sensor data fusion and resource management. In: mpc special issue 2019, 8–11. My translation.

6 When the same devices interact, the technical term is *swarm*. If the devices are different, the term *system of systems* is used.

7 I put the term artificial intelligence in quotation marks here because it is an anthropomorphism as well that cannot be readily applied to machines. See for example: Charbonnier, Ralph: Wahrnehmen, entscheiden, handeln – werden digitale Maschinen menschlich? In: Görder, Björn/Zeyher-Quattlender, Julian (eds.): Daten als Rohstoff. Münster 2019, 61–82.

8 ICRC: Views of the ICRC on autonomous weapon systems. Paper submitted to the Convention on Certain Conventional Weapons Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), April 11, 2016. Online at: https://www.icrc.org/en/document/views-icrc-autonomous-weap-

Two things are worth emphasizing in this definition: First, the machine operates in real time without the possibility of human intervention, and second, it involves critical functions such as killing people. This raises the question of what exactly is meant by technological autonomy.

2 Autoregulation instead of technological autonomy

Technological autonomy can be described as the final step of the automation scale. This, first of all, means that there can be very different degrees of automation when humans and machines are working together. Such a scale commences with processes in which the human being completely controls the machine, for example in classic driving, and it passes through various intermediate degrees in which the human being and the machine act together to a greater extent, for example in assistance systems. Automation finally extends to processes in which the human being – with regard to a specific function – no longer has any possibility to intervene. Only in the last case, i.e., where there is no longer a possibility of intervention, the machine operates autoregulatively. It is exactly this end of the scale that is typically called *autonomous*.

The term *autonomy* stems from Greek: It is composed of αὐτός (*autos),* meaning self, and νόμος (*nomos),* which is law, and denotes the ability to give oneself laws. Central to the concept of autonomy in philosophy and ethics are the writings of Kant, who defines autonomy essentially as the capacity of human beings to give themselves *moral* laws.[9] However, if this concept is to be applied to technological systems, the question arises whether this holds true. In order to clarify this, it is first necessary to denote at which point the capabilities of humans and machines differ from each other. Noel Sharky, a computer scientist and professor of artificial intelligence and robotics at the University of Sheffield, for example, points out the essential differences in information acquisition and processing between humans and machines: Machines, he argues, are much better at calculating numbers, sifting through large data sets, responding quickly to control tasks, and performing repetitive tasks with accuracy. Human capabilities, however, are situated rather in the realm of deliberative and creative thinking, as well as meta-cognition.[10] This means that computers and humans have very different abilities, so it can be stated: Tasks that require foresight, meta-cognition, and creativity are well-situated

on-system (as at 11/25/2021). This definition in turn harks back to the US-American Department of Defense.

9 Kant: AA, IV:439f.

10 Sharkey, Noel: Staying in the loop. In: Bhuta, Nehal/Beck, Susanne/Geiß, Robin/Liu Hin-Yan/Kreß, Claus (eds.): Autonomous weapons systems. Cambridge 2016, 23–38, here 27. In this context, Noel Sharky also mentions the ability of humans to think inductively. However, current algorithmic systems such as neuronal networks imitate these processes to some extent. Without saying that the

in the human workspace, while computation and pattern recognition can be outsourced without concern to the computer system. Both entities together can thus work effectively with each other and compensate for their respective deficits: the human's lack of computational power by the computer and the machine's lack of circumspection by the human's cognitive abilities. But: Such a cooperation always happens under the condition that the human specifies the goals of such a joint action. This makes sense because only humans have the necessary cognitive abilities: The machine may be able to find the way to reach a goal more quickly and perhaps even more expediently, but it currently cannot provide a reasonable objective for complex operations.[11]

If we use these insights to refer once again to the issue of technological autonomy, then the question arises whether the machine is capable of giving itself (moral) laws – and this question is simply to be answered in the negative. This is also what happens usually in the philosophical debate about technological autonomy. Thus, those who partake in the discourse about technological autonomy usually make clear in the beginning that it is not the philosophical understanding that is meant.[12] However, I consider this approach problematic for at least two reasons: first, it is questionable whether the attempt of such a transfer actually succeeds, or whether the philosophical idea is rather preserved despite all redefinitions, simply because the meaning resonating in the word cannot easily be separated thereof. The consequence would be that the implications of autonomy are transferred to machines after all, and thus a (false) anthropomorphization of technology takes place. In this case, the human being nevertheless assumes that the machine has autonomy in the philosophical sense, and the concept does not lose its ethical implications, despite the announcement. However, this in turn could lead to decisions being handed over to machines that cannot meaningfully be made by them[13], such as predictive or creative decisions or decisions that require meta-cognition. On the other hand,

procedures within humans and machines are equal, I would at least like to point to these algorithmic procedures. Thanks Laurence Lerch and Hannah Bleher for this reference.

11 This is expressed, for example, in the difference between *judgement* and *reckoning*, a distinction Brian Cantwell Smith introduced into the debate. The judgment of human beings, on the one hand, involves the (social) world – thus a human being can adapt their behavior within the material and immaterial world that surrounds them properly. If, on the other hand, a machine calculates its access to the world based on algorithms (*reckoning*), it lacks this holistic access. The machine's own reckoning procedures will therefore not be able to include the world surrounding in the same way as humans do. See Smith, Brian Cantwell: The promise of artificial intelligence. Cambridge/London 2019. I am thankful to Florian Höhne for this reference.

12 For instance: Leveringhaus, Alex: Ethics and Autonomous Weapons. London 2016, 32 f.; Grünwald, Reinhard/Kehl, Christoph: Autonome Waffensysteme. Bad Honnef 2020, 36.

13 Fuchs, Thomas: Menschliche und Künstliche Intelligenz. In: idem: Verteidigung des Menschen. Berlin 2020, 21–70. Thanks to Tobias Friesen for this reference.

the concept of autonomy might lose its philosophical point, but is then retransferred to humans – with consequences for anthropology, as humans are being 'machinized' as a result.[14] Therefore, my suggestion is to replace the term autonomy with autoregulation.[15] This term originally roots in cybernetics and pertains to the self-regulation of a technological system. An example would be the thermostat, where the system independently adjusts the room temperature according to given parameters.[16] This very basic meaning must then be extended to include the ability to dynamically adapt to changing environmental influences – that is, to regulate itself in the real environment. Even if this reinterpretation exceeds the traditional understanding within cybernetics, I see the advantage in avoiding the moral implications associated with the notion of autonomy from the outset – the adaptability of the algorithm thus remains on a technological level without denying the immense possibilities and opportunities.[17]

That way, the difference in content can also be marked linguistically: Autonomous and autoregulative entities are to be distinguished primarily with respect to their ability to set their own goals: "Artificial systems, such as thermostats and automatic pilots, are not autonomous: their primary goals are constructed in them by their designers."[18] This means that, while autonomous agents are able to set goals for themselves, including moral goals, autoregulative entities are able to pursue goals not set by the system itself.

In connection to this, there is the problem of unintended side-effects, which Nick Bostrom illustrates with his thought experiment about the paper clip system: Bostrom imagines a system whose only goal is to produce paper clips. Due to its construction, it is devoid of any external influence, i.e., autoregulative. Now, the machine might start to process literally everything in this world into paper clips, including people and the environment. The issue here is that the system has no idea what life means, which is all the more the reason why letting the self-learning systems set its goal on its own could have highly problematic consequences. What, for

---

14 Koch, Bernhard: Maschinen, die uns von uns selbst entfremden. In: Militärseelsorge. Dokumentation 54 (2016), 99–119; Charbonnier: Wahrnehmen, entscheiden, handeln, 80 f.

15 I adopt this term from an article by Lucy Suchman and Jutta Weber, who make a differentiation between biological and technological developments, referring to as "self-regulation" for one thing, and the philosophical term "autonomy", for another. See Suchman, Lucy/Weber, Jutta (2016): Human-machine autonomies. In: Bhuta, Nehal/Beck, Susanne/Geiß, Robin/Liu Hin-Yan/Kreß, Claus (eds.): Autonomous weapons systems. Cambridge 2016, 75–102, here 79 f.

16 Heylighen, Francis/Joslynn, Cliff: Cybernetics and second order cybernetics. In: Meyers, Robert A. (ed.): Encyclopedia of Physical Science and Technology, Volume 4. New York ^{3}2001, 155–170, here 165.

17 My reinterpretation of the concept of autoregulation is certainly not without problems, especially since the term has already been coined within the natural sciences. Nevertheless, to my mind the gain of a rather technologically oriented way of speaking stands out, which is why I will use it at least for my own contributions to describe so-called technological autonomy.

18 Heylighen/Joslynn: Cybernetics and second order cybernetics, 165.

example, if such a system 'learns' that humans are the problem in resorting to violence, and turns against us? Therefore, there always is a need for an autonomous agent who controls the system – a fact the term autoregulation points out terminologically. With that said, I would like to draw a first conclusion at this point: *Since autoregulative devices remain dependent on an autonomous agent, human control must be guaranteed to a sufficient degree because only humans are able to decide autonomously and to reflect on the moral implications of these decisions.*

3 Considering the political debate: *meaningful human control*

The question of whether and to what extent human control over autoregulative functions within weapons systems is necessary and useful has been discussed politically within the bodies of the UN since 2014, without having reached any substantial agreement so far. I will not go into the details of the political arguments and actors here – the urgency of political regulation goes without saying.[19] Instead, I will focus on the ethical arguments put forward, which are rather marginal in the current discourse, and usually embedded into political and legal lines of reasoning. Four main strands of ethical cases have been made so far: First, the advantages of such a system were and are emphasized. Against this line of reasoning, opponents of the technology responded to it with the arguments of a responsibility gap, as well as that of human agency and human dignity.[20]

As for the proponents, the robotics engineer Ronald C. Arkin plays a major role within the debate, as he argues that autoregulation in weapons systems can help to reduce harm on both sides of the conflict. This is due to the technological system not being distracted by emotions such as anger, frustration, or fear. He holds, for instance, that such weapons do not need to fire from far distance for reasons of safety and to protect their own lives, as humans would. The systems could rather let a potential danger approach to evaluate the threat they might pose.[21] That the lives and mental health of soldiers and drone pilots would be spared, at least on the side deploying the systems, is obvious, since they do not appear on the battlefield anymore, be it physically or mentally.[22]

---

19 For more information on this topic, see the website and output of the International Panel on the Regulation of Autonomous Weapons, online at: www.irpaw.org (as of 01.11.2021), as well as the website of the UN, online at: https://www.un.org/disarmament/group-of-governmental-experts/ (as of 24.11.2022).

20 The possibility of a responsibility gap was first pointed out by Robert Sparrow as early as 2007. The first two waves I adopt from Alex Leveringhaus' description. See Leveringhaus: Ethics and autonomous weapons.

21 See Arkin, Ronald C.: Governing lethal behavior in autonomous robots. Boca Raton 2009.

22 See Galloit, Jai: Lethal autonomous weapons systems. In Federal Foreign Office of Germany (ed.): Lethal autonomous weapons systems. Frankfurt a. M. 2016, 85–96, here 85.

On the other hand, and with respect to the arguments against autoregulation in weapon systems, this stance is opposed by ethicists who point out a responsibility gap. The philosopher and ethicist Robert Sparrow, for example, argues that autoregulation necessarily leads to a responsibility gap, because it ultimately remains unclear who should be held accountable for a possible mistake: the programmer, the commander, or the soldier who gave the order to deploy the system? Who would be legally prosecuted and could be held morally accountable, if something happens beyond their control?[23] In contrast, philosopher and ethicist Alex Leveringhaus doubts whether the issue of responsibility is that problematic and brings to attention the need for human agency: While a machine will always follow its programmed paths, a human has the ability to do otherwise and, for instance, exercise leniency. He writes:

> Unless re-progammed, the machine *will* engage the targeted person upon detection. Killing a person, however, is a truly existential choice that each soldier needs to justify before his own conscience. Sometimes it can be desirable not to pull the trigger, even if this means that an otherwise legitimate target survives. Mercy and pity may, in certain circumstances, be the right guide to action.[24]

In addition, the human dignity of the target is repeatedly referred to in the current discourse: Because the machine reduces the targeted person to a data point, their human dignity is violated. Consequently, killing by machines is morally impermissible.[25]

In addition to these ethical arguments, several political issues are discussed, such as the possibility of a new arms race in the field of artificial intelligence,[26] as well as legal problems, such as the question of whether the international legal system, in particular IHL, is sufficient to regulate autoregulation in weapon systems.[27] Generally, however, the importance of human control is emerging, for example in the form of *meaningful human control*. This term was introduced into the debate by the NGO Article 36 and defined more precisely in this context by Heather M. Roff and Richard Moyes. According to their definition,

---

[23] See Sparrow, Robert: Killer robots. In: Journal of Applied Philosophy 24 (2007), 62–77.

[24] Leveringhaus: Ethics and Autonomous Weapons, 92.

[25] See Heyns, Christopher: A Human rights perspective on autonomous weapons in armed conflict. In: Federal Foreign Office of Germany (ed.): Lethal autonomous weapons systems. Frankfurt a. M. 2016, 148–159; Rosert, Elvira/Sauer, Frank: Prohibiting autonomous weapons. In: Global Policy 3/10 (2019), 370–375.

[26] See Altmann, Jürgen/Sauer, Frank: Autonomous weapons systems and strategic stability. In: Survival, 5/59 (2017), 117–142.

[27] See Geiss, Robin/Lahmann, Henning: Autonomous weapons systems. In: Ohlin, Jens David (ed.): Research handbook on remote warfare. Cheltenham/Northampton 2017, 371–404, here 378–383, 399.

> meaningful human control […] means 1. that a machine applying force and operating without any human control whatsoever is broadly considered unacceptable. 2. that a human simply pressing a 'fire' button in response to indications from a computer, without cognitive clarity or awareness, is not sufficient to be considered 'human control' in a substantive sense.[28]

The authors thus initially draw a negative distinction, which clearly restricts the application of autoregulation in the aforementioned sense, at least if autoregulation impedes human control in real time. Although this definition is under dispute internationally, a majority of the actors partaking in the current debate embraces this term.[29] The need for human control over the system should be underscored with reference to the fact that even machines are by no means impartial and unbiased in their decision-making, and thus cultural biases, including errors and unintended side effects, have an impact on the technological procedures.[30] In order to not lose perspective on these problems, it is inevitable that humans control machines under conditions that leave sufficient time for consideration.

4 Development within Christian peace-ethics

Both discussions addressed so far, political and ethical, usually refer to a framework that is connected to just war thinking. This idea, originally located in ancient law, was transformed by the Latin-Christian tradition, and has found a place within the juridical bodies of the UN, mainly IHL.[31] While this idea can be reconstructed as *war for the sake of peace*,

---

[28] Roff, Heather M./Moyes, Richard: Meaningful human control. Online at: https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf (as of 29.08.2022). This term has also been applied to other fields where artificially intelligent systems are deployed in recent years. See Santoni de Sio, Filippo/ van den Hoven, Jeroen: Meaningful Human Control over Autonomous Systems. In: Frontiers in Robotics and AI 5 (2018), 1–14, or for self-driving cars: Heikoop, Daniël D.: Meaningful human control. In: Research OUTREACH. Online at: https://researchoutreach.org/articles/meaningfulhuman-control-designing-safety-into-automated-driving-systems (as of 22.11.2021).

[29] See Amoroso, Daniele/Tamburrini, Guglielmo: Toward a normative model of meaningful human control over weapons systems. In: Ethics & International Affairs, 2/35 (2021), 245–272. Against the juridical adoption of the term, See Marauhn, Thilo: Meaningful human control – and the politics of international law. In: Heintschel von Heinegg, Wolff/Frau, Robert/Singer, Tassilo (eds.): Dehumanization of warfare. Cham 2018, 207–218.

[30] See Noble, Safiya Umoja: Algorithms of oppression. New York 2018: O'Neil, Cathy (2017): Weapons of math destruction. Harlow 2017.

[31] See Jensen, Jessica: Krieg um des Friedens willen. Baden-Baden 2015.

as Jessica Jensen, for example, points out in her legal analysis, Christian churches have been insistent over the past three decades that a reorientation in peace ethics is needed, turning away from just war thinking, and putting a stronger emphasis on peace.[32] Along the lines of *Si vis pacem para pacem (if you want peace, prepare for peace),* the Protestant Church in Germany (*Evangelische Kirche in Deutschland)* undertook this reorientation in 2007 with its peace memorandum, emphasizing the utmost priority of civil conflict management mechanisms and a peace-oriented international legal system, in which the criteria for judging wars are presented only within the ethics of law-sustaining force.[33] These criteria, however, gain weight only as a last resort in extreme situations and in order to sustain a situation of law. The use of armed force is thus only considered at all in order to (re)establish a legal situation, from which just peace might then arise.[34] The traditional criteria of just war tradition – i.e., both the right to war, which is the *jus ad bellum,* and the right in war, which is the *jus in bello* – remain intact as test criteria here. In concrete terms, this means that when considering situations of violence, the criteria of permissible cause, authorization, right intention, the use of force as ultimate resort, proportionality of consequences, as well as the proportionality of means and the application of the principle of distinction are to be taken into account.[35] At the same time, the peace ethical framing in combination with the idea that not every use of force might be illegitimate, denotes a contingent pacifist stance, since such a position is situated in between an absolute pacifist position, opposed to any use of force, and a just war stance, which sees war as a morally legitimate means, provided it is waged justly.

Focusing on the ethical legitimacy of autoregulative functions in weapon systems within such a contingent pacifism primarily changes the viewpoint and therefore the question: Instead of discussing the issue of whether autoregulation in critical functions can be used in accordance with existing law and ethical norms, I focus on the question whether and to what extent such systems contribute to peace, qualified as just. To define such peace, the peace memorandum delineates four dimensions, namely the "*rejection of the use of violence*, *the promotion of freedom and cultural diversity* and *the alleviation of want.*"[36] Simultaneously, this means

[32] See Hoppe, Thomas/Werkner, Ines-Jaqueline: Der gerechte Frieden. In: Werkner, Ines-Jaqueline/Ebeling, Klaus (eds.): Handbuch Friedensethik. Wiesbaden 2017, 344–359.

[33] Referring to Hans-Richard Reuter, I use the term law-sustaining force instead of law-abiding force, as stated in the memorandum. Reuter points out that the concept of law-sustaining force is semantically more open and encompasses not only preservation but also enabling and enforcement of basic human rights. See Reuter, Hans-Richard: Kampfdrohnen als Mittel rechtswahrender militärischer Gewalt? In: epd-Dokumentation 49 (2014), 37–46, here 39.

[34] See Evangelische Kirche in Deutschland (EKD): Aus Gottes Frieden leben – für gerechten Frieden sorgen. Eine Denkschrift des Rates der Evangelischen Kirche in Deutschland. Gütersloh 2007.

[35] See EKD: Aus Gottes Frieden leben, 68 f.

[36] See EKD: Aus Gottes Frieden leben, 54. Emphasis in original.

that although peace implies the absence of violence, in its positive form it exceeds this basic understanding and needs to be imagined as a process, leading to the expanding formation of structures that are capable of establishing just peace in the long run.

Yet, if establishing sustainable and just peace processes is the first and only goal of using force, this presupposes that the conflicting parties behave in a way that their conduct within war concurrently lays the foundation for such a peace. The decision to use armed force and the way in which war is waged will therefore have to be measured against this goal. For this reason, it seems to me that a category that has received too little attention in the current discourse is the *jus post bellum*. This category refers to moral norms that are supposed to (re)enable stability after a war. In the words of philosopher Larry May, "Jus post bellum normally concerns how to move to a situation of stability after war."[37] Specifically, this includes, among other things, the ability to hold war criminals accountable for their actions and to facilitate reconciliation between conflicting parties.[38] Even though Christian just peace thinking does not refer to *jus post bellum* verbatim, the essential focus on peace enables such a perspective, because it is of utmost importance for the peace in the aftermath of an armed conflict to be just and sustainable.

Against this background, however, questions arise with regard to autoregulation in weapons systems in at least two points of reference: First, there are issues concerning the continuation of a trend already inherent in the use of drones and their disastrous effect on the civilian population.[39] In particular, the disruption of civilians' daily lives[40] is likely to continue when the systems are no longer controlled by humans. The situation might even exacerbate, being accompanied by the fear of technological mistakes, which is likely to be significantly more fatal than the mistake of a human operator. This is because a human operator would not repeat the same mistake at breathtaking speed.[41] However, once such an error has occurred and responsibilities cannot be reconstructed meaningfully in the aftermath of the conflict, this is likely to run counter to the formation of sustaining and just structures of peace. Accordingly, it is

---

37 May, Larry: After war ends. New York 2012, 6.

38 See ibid, 1. Even if the Peace Memorandum does not refer to these ideas verbatim, they can be inferred and frame the ideas the memorandum puts forward, since it is oriented toward peace. Nevertheless, I think it makes sense to refer explicitly to these debates, so that peace ethical reason can be thought of from its end, which is the establishment of a just and sustaining peace.

39 See International Human Rights and Conflict Resolution Clinic at Stanford Law School and Global Justice Clinic at NYU School of Law: Living under drones. Online at: https://www-cdn.law.stanford.edu/wp-content/uploads/2015/07/Stanford-NYU-LIVING-UNDER-DRONES.pdf (as of 29.08.2022).

40 See Ibid, VII.

41 See Scharre, Paul: Autonomous weapons and operational risk, 2016, 189–195. Online at: https://www.cnas.org/publications/reports/autonomous-weapons-and-operational-risk (as of 22.11.2021).

foreseeable that the use of such technologies could prevent rather than serve sustaining peace. This might still be the case if the use of autoregulation led to a reduction of harm.

In addition, the technological risk in weapons systems with autoregulative functions plays a significant role as well, especially if the machine operates inaccurately or if its proper functioning cannot be foreseen by its deployers, as exemplified in the paperclip thought experiment. In the context of war, this risk of malfunction or unintended side–effects then shifts primarily to the enemy's side. This is an effect the American legal scholar Paul W. Kahn called the "paradox of riskless warfare"[42], referring to drones: While the side using drones no longer exposes itself to any substantial risk, the risk for the other side increases, a trend which will be expectably more severe with the advent of autoregulation in weapons systems. However, if the right to kill in war is reconstructed via the soldiers' right to defend themselves, the deployment of lethal autoregulative weapon systems raises doubts, because an artificial system has no need thereof.[43]

The concept of risk plays an important role within peace ethics, as well when it is applied to the discussions revolving around the status of combatants. Larry May, for instance, emphasizes that participation in modern wars is inherently problematic because of the risk to harm unjust targets. Behind this rationale is the scholarly debate whether soldiers are to be targeted legitimated via their status as combatants and civilians are to be spared in turn because they, as a collective, are non-combatants, as described in the traditional reading of just war thinking.[44] Yet, thinkers such as McMahan question this rationale by emphasizing that civilians contribute to conflicts as well, be it by means of propaganda or providing the necessary infrastructure for waging war. Soldiers, on the other hand, fight for several reason, among them having enlisted in times of peace or being deceived by their government.[45] If this thought is taken seriously, not every soldier is a legitimate target, while some civilians might be and the risk to harm someone unjustly increases.[46] With regard to autoregulation in weapons system, it is highly doubtful, whether the technology can meet these demands.

---

42 Kahn, Paul: The paradox of riskless warfare. In: Philosophy and Public Policy Quarterly, 22 (2002), 2–7. The German theologian and peace ethicist Hans-Richard Reuter follows Kahn in his judgement. See Reuter: Kampfdrohnen als Mittel rechtswahrender militärischer Gewalt?

43 David Rodin has demonstrated, though, that this deduction is by no means uncontroversial and self-evident. See Rodin, David: War and self-defense. Oxford 2002.

44 See Walzer, Michael: Just and unjust wars. New York [5]2015.

45 See McMahan, Jeff: Killing in war. Oxford 2009.

46 See May, Larry: Contingent pacifism. Cambridge 2015.

5 Conclusion

Deploying autoregulation in weapons systems is therefore troubling from a Christian peace-ethical perspective, even if this does not necessarily mean that every usage of autoregulation in violent scenarios needs to be rejected principally. This concerns, for instance, functions that do not intend the death of a human being. It should also be considered whether there can be clear and distinct scenarios in which deploying such a weapon does not raise any concerns of safety, for instance in the deep sea or within a certain airspace. However, considering the situation and the fact that differentiating between just and unjust targets is an important ethical norm within armed conflict, a solution has to be found. This could be, for example, some kind of *autoregulation scale* such as that proposed by DeGreef for high-risk scenarios, where survivors in the aftermath of an earthquake are looked for, can be part of the solution.[47] In such situations of high risk, human-machine collaboration proves particularly effective when the machine's level of automation is dynamically adapted to the human part. Such a method of adaptive automation adjusts the level of automation depending on the workload of the operator.[48] Thus, if, on the one hand, the human within the collaboration is under stress, the automation level can be raised. If, on the other hand, the workload decreases, the automation level can be lowered again. Within such a scenario, the implementation of full-scale autoregulation is conceivable as well, for example if the human fails completely. Vice versa, the human could navigate the machine in case of its breakdown. To sum it up, I advocate for the implementation of *meaningful human control*, both for technological reasons and for reasons of Christian peace ethics: Only humans have the ability to comprehensively evaluate situations, therefore the human should exert control over the machine, especially in situations where lethal force is involved.

---

47 See De Greef, T.: ePartners for dynamic task allocation and coordination. Delft 2012.

48 Both proposals differed in the question of who performs the adaptation. While *Adaptive Automation* performs an automatic adaptation on the initiative of the machine part, this adaptation is performed by humans in the *Adaptable Automation* concept. Both systems have advantages and disadvantages. For example, the loss of control that goes hand in hand with automatic adaptation must be pointed out, while on the other hand it can be critically questioned whether an already overburdened human is still in a position to make sufficient decisions about handing over competencies to the machine – especially since such a situation increases the workload again. See DeGreef, Tjerk: ePartners for dynamic task allocation and coordination; Scerbo, Mark: Theoretical perspectives on adaptive automation. In: Mouloua, Mustapha/Hancock, Peter A./Ferraro James (eds.): Human performance in automated and autonomous aystems. Boca Raton 2019, 103–126.

Literature

Altmann, Jürgen/Sauer, Frank: Autonomous weapons systems and strategic stability. In: Survival, 5/59 (2017), 117–142. DOI: https://doi.org/10.1080/00396338.2017.1375263.

Amoroso, Daniele/Tamburrini, Guglielmo: Toward a normative model of meaningful human control over weapons systems. In: Ethics & International Affairs, 2/35 (2021), 245–272. DOI: https://doi.org/10.1017/S0892679421000241.

Arkin, Ronald C.: Governing lethal behavior in autonomous robots. Boca Raton 2009. DOI: https://doi.org/10.1201/9781420085952.

Charbonnier, Ralph Wahrnehmen, entschieden, handeln – werden digitale Maschinen menschlich? In: Görder, Björn/Zeyher-Quattlender, Julian (eds.): Daten als Rohstoff. Die Nutzung von Daten in Wirtschaft, Diakonie und Kirche aus ethischer Sicht. Münster 2019, 61–82.

DeGreef, Tjerk: ePartners for dynamic task allocation and coordination. Delft 2012.

Evangelische Kirche in Deutschland (EKD): Aus Gottes Frieden leben – für gerechten Frieden sorgen. Eine Denkschrift des Rates der Evangelischen Kirche in Deutschland. Gütersloh 2007.

Fuchs, Thomas: Menschliche und Künstliche Intelligenz. Eine Klarstellung. In: idem: Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie. Berlin 2020, 21–70.

Galloit, Jai: Lethal autonomous weapons systems. Proliferation, disengagement, disempowerment. In: Federal Foreign Office of Germany (ed.): Lethal autonomous weapons systems. Technology, definition, ethics, law & security. Frankfurt a. M. 2016, 85–96.

Geiss, Robin/Lahmann, Henning: Autonomous weapons systems. A paradigm shift for the law of armed conflict? In: Ohlin, Jens David (ed.): Research handbook on remote warfare. Cheltenham/Northampton 2017, 371–404. DOI: https://doi.org/10.4337/9781784716998.00023.

Grünwald, Reinhard/Kehl, Christoph: Autonome Waffensysteme. Endbericht zum TA-Projekt. Bad Honnef 2020. DOI: 10.5445/IR/1000127160.

Heikoop, Daniël D.: Meaningful human control. Designing safety into automated driving systems. In: Research OUTREACH. Online at: https://researchoutreach.org/articles/meaningfulhuman-control-designing-safety-into-automated-driving-systems (as of 22.11.2021).

Heylighen, Francis/Joslynn, Cliff: Cybernetics and second order cybernetics. In: Meyers, Robert A. (ed.): Encyclopedia of Physical Science and Technology, Volume 4. New York [3]2001, 155–170.

Heyns, Christopher: A human rights perspective on autonomous weapons in armed conflict. The rights to life and dignity. In: Federal Foreign Office of Germany (ed.): Lethal autonomous weapons systems. Technology, definition, ethics, law & security. Frankfurt a. M. 2016, 148–159. DOI: http://dx.doi.org/10.1080/02587203.2017.1303903.

Hoppe, Thomas/Werkner, Ines-Jaqueline: Der gerechte Frieden. Positionen in der katholischen und evangelischen Kirche. In: Werkner, Ines-Jaqueline/Ebeling, Klaus (eds.): Handbuch Friedensethik. Wiesbaden 2017, 344–359. DOI: https://doi.org/10.1007/978-3-658-14686-3_28.

IAI: Harpy. Online at: https://www.iai.co.il/p/harpy (as of 01.12.2021).

ICRC: Views of the ICRC on autonomous weapon systems. Paper submitted to the Convention on Certain Conventional Weapons Meeting of Experts on lethal autonomous weapons systems (LAWS), April 11, 2016. Online at: https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system (as at 11/25/2021).

International Human Rights and Conflict Resolution Clinic at Stanford Law School and Global Justice Clinic at NYU School of Law: Living under drones. Death, injury, and trauma to civilians from the US drone practices in Pakistan, 2012. Online at: https://www-cdn.law.stanford.edu/wp-content/uploads/2015/07/Stanford-NYU-LIVING-UNDER-DRONES.pdf (as of 29.08.2022).

IPRAW: Welcome. Online at: www.irpaw.org (as of 01.12.2021).

Jensen, Jessica: Krieg um des Friedens willen. Zur Lehre vom gerechten Krieg. Baden-Baden 2015.

Kahn, Paul: The paradox of riskless warfare. In: Philosophy and Public Policy Quarterly, 22 (2002), 2–7.

Kant, Immanuel: Gesammelte Schriften, Band IV. Berlin 1900 ff.

Koch, Bernhard: Maschinen, die uns von uns selbst entfremden. Philosophische und ethische Anmerkungen zur gegenwärtigen Debatte um autonome Waffensysteme. In: Militärseelsorge. Dokumentation 54 (2016), 99–119.

Koch, Wolfgang: FCAS – Challenges for sensor data fusion and resource management. In: mpc special issue 2019, 8–11.

Kunkel, Nicole: Autoregulative Waffensysteme. Automatisierung als friedensethische Herausforderung – ein Werkstattbericht. In: Ethik und Gesellschaft 2 (2021). DOI: https://dx.doi.org/10.18156/eug-2-2021-art-6.

Leveringhaus, Alex: Ethics and autonomous weapons. London 2016.

Loh, Janina: Roboterethik. Eine Einführung. Berlin 2019.

May, Larry: After war ends. A philosophical perspective. New York 2012.

May, Larry: Contingent pacifism. Revisiting just war theory. Cambridge 2015.

Marauhn, Thilo: Meaningful human control – and the politics of international law. In: Heintschel von Heinegg, Wolff/ Frau, Robert/ Singer, Tassilo (eds.): Dehumanization of warfare. Legal implications of new weapon technologies. Cham 2018, 207–218. DOI: https://doi.org/10.1007/978-3-319-67266-3_11.

McMahan, Jeff: Killing in war. Oxford 2009.

Misselhorn, Catrin: Grundfragen der Maschinenethik. Ditzingen 2018.

Noble, Safiya Umoja: Algorithms of oppression. How search engines reinforce racism. New York 2018. DOI: https://doi.org/10.18574/nyu/9781479833641.001.0001.

O'Neil, Cathy: Weapons of math destruction. Harlow 2017.

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977. Online at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Treaty.xsp?action=openDocument&documentId=D9E6B6264D7723C3C12563CD002D6CE4 (as of 11.04.2022).

Reuter, Hans-Richard: Kampfdrohnen als Mittel rechtswahrender militärischer Gewalt? Aspekte einer ethischen Bewertung. In: epd-Dokumentation 49 (2014), 37–46.

Rodin, David: War and self–defense. Oxford 2002. DOI: https://doi.org/10.1111/1468-2230.6603014.

Roff, Heather M./Moyes, Richard: Meaningful human control. Artificial intelligence and autonomous weapons. Briefing paper prepared for the informal meeting of experts on lethal autonomous weapons systems, UN Convention on Certain Conventional Weapons, 2016. Online at: https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf (as of 29.08.2022).

Rosert, Elvira/Sauer, Frank: Prohibiting autonomous weapons. Put human dignity first. In: Global Policy 3/10 (2019), 370-375. DOI: https://doi.org/10.1111/1758-5899.12691.

Santoni de Sio, Filippo/ van den Hoven, Jeroen: Meaningful human control over autonomous systems. A philosophical account. In: Frontiers in Robotics and AI 5 (2018), 1–14. DOI: https://doi.org/10.3389/frobt.2018.00015.

Scerbo, Mark: Theoretical perspectives on adaptive automation. In: Mouloua, Mustapha/Hancock, Peter A./Ferraro James (eds.): Human performance in automated and autonomous systems. Boca Raton 2019, 103–126. DOI: https://doi.org/10.1201/9780429458330-6.

Scharre, Paul: Autonomous weapons and operational risk, 2016. Online at: https://www.cnas.org/publications/reports/autonomous-weapons-and-operational-risk (as of 22.11.2021).

Sharkey, Noel: Staying in the loop. Human supervisory control of weapons. In: Bhuta, Nehal/Beck, Susanne/Geiß, Robin/Liu Hin-Yan/Kreß, Claus (eds.): Autonomous Weapons Systems. Law, Ethics, Policy. Cambridge 2016, 23–38. DOI: https://doi.org/10.1017/CBO9781316597873.002.

Smith, Brian Cantwell: The promise of artificial intelligence. Reckoning and judgement. Cambridge/London 2019. DOI: https://doi.org/10.7551/mitpress/12385.001.0001.

Sparrow, Robert: Killer robots. In: Journal of Applied Philosophy 24(2007), 62–77. DOI: https://doi.org/10.1111/j.1468-5930.2007.00346.x.

Suchman, Lucy/Weber, Jutta (2016): Human-machine autonomies. In: Bhuta, Nehal/Beck, Susanne/Geiß, Robin/Liu Hin-Yan/Kreß, Claus (eds.): Autonomous weapons systems. Law, Ethics, Policy. Cambridge 2016, 75–102. DOI: https://doi.org/10.1017/cbo9781316597873.004.

UNODA: Group of governmental experts. Online at https://meetings.unoda.org/meeting/ccw-gge-2021/ (as of 24.11.2022).

Walzer, Michael: Just and unjust wars. New York [5]2015.

Autonomous Weapons Systems and Battlefield Dignity

A Jewish Perspective

Mois Navon

Abstract

A great battle over the deployment of Autonomous Weapons Systems (AWS) is being waged by thinkers from around the globe. While there is a long list of concerns regarding their deployment, the quintessential moral concern surrounds human dignity. Many believe that death by AWS is an insult to human dignity, for only a human being who recognizes the other as a human being with inherent worth can make the decision to take his life – as a subject and not as an object. Others, however, reject this claim, arguing that there is no difference in the dignity of a death brought about by a weapon system, autonomous or not. In this chapter, I argue that to equate dignity on the battlefield to dignity off the battlefield is to make a category mistake. Dignity on the battlefield is an ethical category of its own, defined in completely different terms than peacetime dignity. Bringing Jewish thought to support this dichotomy, I demonstrate that human dignity is no more impacted in war whether fought with sticks and stones, knives and guns, tomahawk and hellfire missiles, or fully autonomous weapons systems.

1 Introduction[1]

"*Nation shall not lift up sword against nation, neither shall they learn war anymore*."

[1] The author thanks Michael Broyde and Maier Becker for their valuable comments.

The prophet Isaiah (2:4) declared these words some 2700 years ago,[2] giving hope that humanity would ultimately – "in the end of days" (2:2) – achieve peace. Much to our chagrin, not to speak of our pain and suffering, nation has not ceased to lift up sword against nation, but quite the opposite, has continued unabated in the design, development and deployment of ever more sophisticated "swords." In our day, the latest sword comes in the form of Autonomous Weapons Systems (AWS) – AI based weapons that, once launched, autonomously choose targets (human or otherwise) and deliver lethal (or non-lethal) force without a human in-the-loop (i.e., to fire) or on-the-loop (i.e., to abort).[3] These systems are no mere incremental improvement on previous weapons but are hailed as being as revolutionary to warfare as gunpowder and the atom bomb.[4]

But long before war was waged with AWS,[5] a battle of biblical proportions has been waged by technologists, ethicists, clergymen, academics and concerned citizens from around the globe.[6] While a great many issues are raised regarding the propriety of deploying such weapons (e.g., legal, technical, social, political),[7] most, if not all, can be answered from a consequentialist approach – i.e., will nations war more or less, kill more or less, adhere to *jus in bello* more or less.[8] Without belittling these questions, nor the engineering required to achieve positive

---

2 Gem. B. Batra 14–15.

3 See, e.g.: Docherty: Losing Humanity. Online at: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots; DODD 3000.09: Autonomy in Weapon Systems. Online at: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

4 See Open Letter on Autonomous Weapons. Online at: http://futureoflife.org/open-letter-autonomous-weapons/.

5 On AWS already deployed, see, e.g.: Trager/Luca: Killer Robots Are Here Here—and We Need to Regulate Them. In: Foreign Policy. 2022. Online at: https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/.

6 See, e.g., Sullins, John P.: An Ethical Analysis of the Case for Robotic Weapons Arms Control. In: 5th International Conference on Cyber Conflict. 2013, 1–20. Online at: https://ieeexplore.ieee.org/document/6568394; Sharkey, Amanda: Autonomous Weapons Systems, Killer Robots and Human Dignity. In: Ethics and Information Technology 21 (2018). DOI: https://doi.org/10.1007/s10676-018-9494-0. Some of the organizations involved are: Human Rights Watch, International Human Rights Clinic, United Nations Institute for Disarmament Research, International Committee of the Red Cross, Future of Life Institute, Stop Killer Robots (a coalition of 180 international organizations).

7 See, e.g., Lin, Patrick/Abney, Keith/Bekey, George: Ethics, War, and Robots. In: Sandler, Ronald L. (ed.): Ethics and Emerging Technologies. London 2014. DOI: https://doi.org/10.1057/9781137349088; Asaro, Peter: Autonomous Weapons and the Ethics of Artificial Intelligence. In: Liao, S. Matthew (ed.): Ethics of Artificial Intelligence. Oxford 2020, 212–36. https://doi.org/10.1093/oso/9780190905033.001.0001.

8 Some argue that it is simply impossible to program a machine to make ethical decisions (e.g., Sparrow, Robert: Robots and Respect: Assessing the Case against Autonomous Weapon Systems. In: Ethics & International Affairs 30/1 (2016), 93–116. DOI: https://doi.org/10.1017/s0892679415000647, here 100. Space does not permit me to properly address the "codifiability thesis" (i.e., is it possible to define a robust code of ethics, and by extension, program a machine to execute it with precision), but

consequences, the questions reduce to those of functional performance that will ultimately be answered in the field.[9] In contradistinction, the deontological question – i.e., is there an inherent violation of moral values in the deployment of AWS – can only be answered in the halls of philosophy. Peter Asaro, leading opponent of AWS, puts it like this:

> While I see the consequentialist side of this argument [i.e., 'autonomous weapons could be designed to be far better than humans at making targeting decisions and conducting attacks, thus reducing the risks of harm to civilians'], … I do not see the deontological side of it [i.e., there are inviolable moral duties, what some refer to as sacred values,[10] that are here trampled in the name of utility]. In particular, in order to fulfill our duty to respect the human dignity of others, I believe we are required to recognize them as human and to consider them as such when making the decision that it is justified to kill them or put them at risk of death.[11]

Similarly, Robert Sparrow writes that, while we could "imagine AWS being ethical" in consequentialist terms, the deontological demand for human dignity is insurmountable: "AWS should be acknowledged as *mala in se* by virtue of the extent to which they violate the requirement of respect for the humanity of our enemies."[12] And it is this demand that is central to the Human Rights Watch manifesto which asserts that "fully autonomous weapons could undermine the principle of dignity, which implies that everyone has a worth deserving of respect. As inanimate machines, fully autonomous weapons could truly comprehend neither the value of individual life nor the significance of its loss. Allowing them to make determinations to take life away would thus conflict with the principle of dignity."[13]

---

let it be said that this is not an insurmountable issue for AWS. For even those who hold the anti-codifiability thesis agree that it is possible to make a machine that operates better than humans (see, e.g., Moor, James H.: The Nature, Importance, and Difficulty of Machine Ethics. In: Anderson, Michael/ Leigh Anderson, Susan (eds.): Machine Ethics. New York 2011; also fn. 99 herein).

9 See, e.g., Muller, Vincent C.: Autonomous Killer Robots Are Probably Good News. In: Nucci, Ezio Di/Santonio de Si, Filippo (eds.): Drones and Responsibility. London 2016, 67–81. DOI: https://doi.org/10.4324/9781315578187-4, here 7.

10 See, e.g.: Dehghani, Morteza/Forbus, Ken/Tomai, Emmett/Klenk, Matthew: An Integrated Reasoning Approach to Moral Decision Making. In: Anderson, Michael/ Anderson, Susan Leigh (eds.): Machine Ethics. New York 2011.

11 Asaro: Autonomous Weapons and the Ethics of Artificial Intelligence, 217.

12 Sparrow: Robots and Respect, 110.

13 Docherty: Losing Humanity. Similarly, many argue that AWS violate human dignity by treating the other as an object not a subject, as a means and not an end (see Sharkey: Autonomous Weapons Systems, Killer Robots and Human Dignity, who brings: Heyns, Christof: Autonomous Weapons Systems: Living a Dignified Life and Dying a Dignified Death. In: Bhuta, Nehal/Beck, Susanne/

Asaro elaborates on why AWS will be unable to fulfill the demands of human dignity:

> In order to make a moral judgment to take a life, while respecting human dignity, it is minimally required that a moral agent can (1) recognize a human being as a human, not just distinct from other types of objects and things but as a being with rights that deserve respect; (2) understand the value of life and the significance of its loss; and (3) reflect upon the reasons for taking life and reach a rational conclusion that killing is justified in a particular situation.[14]

And while this is certainly a tall order to ask of a soldier, Sparrow – inspired by Thomas Nagel – takes it even further:

> Nagel … argues that *even during wartime it is essential that we acknowledge the personhood of those with whom we interact* and that "whatever one does to another person intentionally must be aimed at him as a *subject*, with the intention that he receive it as a *subject*. It should manifest an attitude to *him* rather than just to the situation, and he should be able to recognize it and identify himself as its object." Another way of putting this is that *we must maintain an "interpersonal" relationship with other human beings, even during wartime*.[15]

These descriptions – calling for a soldier to maintain an interpersonal relationship with his enemy, to identify the enemy as a subject, to reflect upon the enemy's inherent value – are all descriptions of the highest ethical behaviors one must strive for during *peacetime* with one's neighbor, reflecting the great biblical ethic to "love one's neighbor as oneself" (Lev. 19:18). But is this at all reasonable to demand of a soldier? During wartime? With one's enemy?

Geiss, Robin/Kress, Claus (eds.): Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge 2016. DOI: https://doi.org/10.1017/CBO9781316597873.005; Amoroso, Daniele/Tamburrini, Guglielmo: The Ethical and Legal Case Against Autonomy in Weapons Systems. In: *Global Jurist* 18/1 (2018). DOI: https://doi.org/10.1515/gj-2017-0012; Johnson, Aaron M./Axinn, Sidney: The morality of autonomous robots. In: Journal of Military Ethics 12/2 (2013), 129–141; Bhuta, Nehal/Beck, Susanne/Geiss, Robin/Kress, Claus (eds.): Autonomous weapons systems. Cambridge 2016.) Ulgen, Ozlem, Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an 'Elementary Consideration of Humanity'? (January 31, 2017). European Society of International Law (ESIL) 2016 Annual Conference (Riga), Available at SSRN: https://ssrn.com/abstract=2912002 or http://dx.doi.org/10.2139/ssrn.2912002.

14 Asaro: Autonomous Weapons and the Ethics of Artificial Intelligence, 229.

15 Sparrow: Robots and Respect, 106, *emphasis added*.

To be fair, even Nagel and Sparrow would answer in the negative. What they call "interpersonal relations," or relating to the other as a "subject" or a Kantian "end," is a radically minimalist version of the lofty ideals to which these terms refer in peacetime. Nagel explains that one could literally machine-gun an enemy combatant (Sparrow allows for hellfire missiles) and yet be considered to have maintained an "I-Thou" relationship with him.[16] This, because in *choosing* to kill him as an enemy combatant you treat him as an "end" in that you could have chosen to kill his wife and kids, which would have also stopped him in his tracks, yet such killing would be treating them as a "means." Essentially, by intentionally choosing to kill a combatant, in distinction from a non-combatant, you have exhibited enough humanity so as to grant him a level of human dignity that Kant would call an "end" and Buber would call a "Thou."

2 Enemy Combatants in the Bible

What does the Bible say about this definition, this demand for wartime interpersonal relations? Before answering, it is important to note that the Bible holds the life of every human being – Jew and Gentile – sacrosanct, all having been created in the *imago Dei* (Gen. 1:27). That said, the Bible also recognizes the inevitability of war and the taking of life inherent in war. Accordingly, as we shall now see, the Bible is quite candid about wartime relations between enemies.

In one of his parting speeches to the tribes of Israel, at the threshold of the promised land, Moses exhorts: "When thou goest forth to battle against thine enemies, and seest horses, and chariots, and a people more than thou, thou shalt not be afraid of them; for the Lord thy God is with thee, who brought thee up out of the land of Egypt" (Deut. 20:1). The great, and perhaps most influential biblical commentator of all time, R. Shlomo Yitzhaki (Rashi), writing in medieval France, explains the opening words of the verse, "When thou goest forth to battle against thine enemies," as teaching:

> Let them be in thine eyes as enemies: have no pity upon them, for they will have no pity upon thee (ad loc.).

R. Shabbtai Bass (1641–1718, Amsterdam), in his super-commentary *Siftei Chachamim*, explains Rashi's textual justification for such a bold pronouncement. He notes that the words "against thine enemies" are entirely redundant, "for obviously if they go to war they don't go against their loved ones!" Accordingly, the text could have simply stated, "When thou goest

[16] Nagel, Thomas: War and Massacre. In: Philosophy & Public Affairs 1/2 (1972), 123–144, here 138; Sparrow: Robots and Respect, 107,110.

forth to battle [ ], and seest horses …" The extra, seemingly superfluous words – "against thine enemies" – are there to teach an ethic. It is the ethic that wartime is not peacetime, and enemies are not neighbors. And if you don't realize that, if you don't "let them be in thine eyes as enemies" and "have no pity upon them," then you will find that "they will have no pity upon thee."[17]

Rashi's interpretation is not simply his personal intuition but is based on Jewish tradition[18] found in the name of God Himself:

> What is [the meaning of] "against your enemies?" The Holy One, blessed be He, said, "Go forth against them like enemies: In the way that they do not have mercy upon you, so [too], do not have mercy upon them. See what they say: 'Let us wipe them out as a nation; Israel's name will be mentioned no more' (Ps. 83:5).[19]

Clearly, these sources brook no form of interpersonal relations at wartime – even the most minimalist. Perhaps this is because the wars against Israel are, as the Psalmist writes, genocidal – a theme Jews revisit every Passover holiday, reciting: "In every generation they rise up against us to annihilate us, but the Holy One, Blessed be He, saves us from their hand." Be that as it may, the notion of showing no mercy in war is found in numerous sources.

For example, in a prior parting speech, Moses exhorts the fledgling nation to "consume all the peoples that the Lord thy God shall deliver unto thee; thine eye shall not pity them; neither shalt thou serve their gods; for that will be a snare unto thee" (Deut. 7:16). This commandment to have no pity, explains R. Haim Ben Attar (1696–1743, Morocco/Israel), comes specifically to discourage misplaced mercy. For, while mercy is a virtue in peacetime, it is a detriment when applied to an enemy in wartime.[20] R. Ben Attar continues his exposition of the verse by explaining that the words – "for it is a snare unto thee" – refer to the snare of applying mercy toward the enemy, for in so doing "you are being brutal toward yourself," inviting casualties and disaster upon yourself. The philosopher and biblical commentator Nachmanides (ad loc.), writing in medieval Spain, compares pity in wartime to a judge who would pity a criminal in a court case – just as surely as justice cannot so be served, similarly a war cannot so be won.

---

17 See also Abarbanel (Deut. 20:1) who concurs and elaborates.

18 The Midrash Tanhuma contains teachings from Talmudic rabbis – both Tannaim and Amoraim (200–500CE) as well as from the Geonic period (500–1000CE). The earliest manuscripts are believed to be from the late eighth or ninth century (Berman, Samuel A.: Midrash Tanhuma-Yelammedenu. New York 1996, 11–12).

19 Midrash Tanhuma, Shoftim 15.

20 Ohr HaHayim (Deut. 7:16).

A similar message is found in the comments of R. Moses Alshich (1508–1593, Turkey/Israel) on the verse, "When thou shalt besiege a city a long time, in making war against it to take it, thou shalt not destroy the trees thereof by wielding an axe against them; for thou mayest eat of them, but thou shalt not cut them down; for is the tree of the field a man, that it should be besieged of thee?" (Deut. 20:19). On the words "is the tree of the field a man," Alshich explains, in a rather novel midrashic (i.e., extra-textual) manner:

> Though in your estimation, the enemy soldier is like a tree of the field … he is actually a danger to you; for if you loosen your hand from him, he will not desist from coming upon you just because of your [pretty] face when he will lay siege upon you in your own gates. Therefore, now, before he comes upon you to kill you, rise up and kill him (Deut. 20, intro).

These sources all point to an essential claim that, according to Jewish thought, the ethics of war are very different than the ethics of peace; and that evincing peacetime virtues like pity, mercy and compassion during wartime (in even the most minimalist sense) is liable to get you killed.

3 Wartime versus Peacetime

The dichotomy between the ethics of wartime versus peacetime is made explicitly in an important comment on the biblical prohibition against murder. The prohibition is found in a number of places (e.g., Ex. 21:12, Num. 35:31, Deut. 5:17), however, its phrasing in the divine word to Noah is of particular interest in our context. The text states: "At the hand of every man's *brother*, will I require the life of man" (Gen. 9:5), teaching that not only is taking the life of another human being prohibited, but God Himself will see to it that justice is served, the perpetrator punished – either in this world or the next.[21] Interestingly, nineteenth century rabbinic leader and biblical commentator, R. Naftali Tzvi Berlin (ad loc.) notes the verse uses the anomalous word *brother*, prompting him to write:

> God, as it were, is teaching: When is man punished [for killing]? At a time when he should have acted as a *brother*. In contradistinction, at a time of war and a "time to hate,"

[21] See French medieval biblical commentator R. David Kimchi (ad loc.).

> which is a "time to kill," there is no punishment whatsoever, since it is in this manner that the world has been founded.[22]

Words as sad as they are true. Without disagreeing, the Midrash explains war in a more positive light, "If you see nations contending with one another, look for the foot of the Messiah" (Gen. R. 42:4) – i.e., each war brings us closer to the words of Isaiah when "nation will not lift up sword against nation."[23] Alas, until that halcyon day, we live in a world when there is "a time to love, and a time to hate; a time for war, and a time for peace" (Ecc. 3:8).

This distinction between times – i.e., wartime and peacetime – is given ethical and legal significance in the words of the celebrated first Chief Rabbi of pre-state Israel, Abraham Isaac Hacohen Kook (Mishpat Cohen #143). He explains that the guiding principle of the Bible – "And you shall live by them" (Lev. 18:5), which is interpreted by the Talmud (Yoma 85b) to mean that performance of the commandments is generally overridden by the imperative to preserve life[24] – does not apply to wartime. He learns this from the fact that the king (or leader of state)[25] is permitted to take the people to wars of expansion (*milhemet reshut*) and not only

---

[22] For further discussion on the applicability of R. Berlin's position see Bleich, J. David: Preemptive War in Jewish Law. In: Contemporary Halakhic Problems Vol. 3. New York 1989, 287–289. See also Broyde, Michael J.: Just Wars, Just Battles and Just Conduct in Jewish Law. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds.): War and Peace in the Jewish Tradition. New York 2007, 10–12, who brings a great many other sources that support (as well as question) R. Berlin.

[23] See further Goldvicht, Chaim Yaakov: War, Kingship, and Redemption. In: Jewish Thought 3/2 (1994), 69–72. As an interesting aside, this view appears to be in consonance with that of Plato's Republic wherein war is explained as the way the world will ultimately be brought to justice. Henrik Syse summarizes as follows: "As long as not all cities are just in the full sense, even (or especially) the fully just city must be prepared to fight wars, partly to defend itself, and partly to vindicate and spread justice" (Syse, Henrik: The Platonic Roots of Just War Doctrine. In: Diametros 23 (2010), 104–123. DOI: https://doi.org/10.13153/diam.23.2010.384, here 111). The idea is echoed in John Stuart Mill, "As long as justice and injustice have not terminated their ever-renewing fight for ascendancy in the affairs of mankind, human beings must be willing, when need is, to do battle for the one against the other" (Mill, John Stuart: The Contest in America. In: Fraser's Magazine. 1862. Online at: https://www.gutenberg.org/files/5123/5123-h/5123-h.htm.). In a similar vein, Dunlap writes, "Professor Ian Morris has argued persuasively that in the long run 'wars make us safer and richer,' because they force the societal organization and sophistication that ultimately functions to suppress human violence" (Dunlap, Charles: Accountability and Autonomous Weapons. In: Temple International & Comparative Law Journal 30/1 (2016), 63–76. Online at: https://scholarship.law.duke.edu/faculty_scholarship/3592, here 76).

[24] Though preservation of life is a sacred value, one is nevertheless required to sacrifice oneself if forced to violate one of the three cardinal principles (murder, sexual immorality, idol worship) or at a time of religious persecution (San 74a).

[25] While the Bible speaks of kings, R. Kook takes it to mean the legitimate leader of state (Mishpat Cohen #144).

to wars of defense (*milhemet mitzvah*).[26] R. Kook asks rhetorically, "Where do we find a permit [in the biblical tradition] to endanger a great many souls [i.e., soldiers] simply for national expansion?! Surely, then, the laws of war [which do allow such endangerment] are different than those of public life [in peacetime]."[27]

R. Kook goes on to explain that it is not the laws of individuals (i.e., peacetime laws) that govern behavior during war, but rather the laws of kings (*mishpatei melucha*).[28] "And what one learns from these laws cannot be applied elsewhere" – i.e., the ethics underpinning the laws of war are distinct from those of the peacetime life of the individual.[29] This dichotomy between the ethics of war and the ethics of peace, it is important to note, finds deep consensus in Jewish sources with practically no dissent.[30]

4 Jewish Just War Theory

Given that the ethics of war is a normative category of its own, the question then becomes: what does a Jewish Just War Theory (JWT) look like? What exactly does Jewish ethics demand of the nation at war, the soldier at war, the weapons of war?

Some, like twentieth century rabbinical judge R. Shaul Yisraeli, understand that the ethics of war essentially allows for all conduct necessary to win, within the limits of international treaties and conventions accepted by all sides involved.[31] Adherence to such agreements is

---

[26] These categories are distinguished through biblical exegesis in the Talmud (Sotah 44b) and codified in, e.g., Maimonides (Laws of Kings 5). For further discussion, see Bleich: Preemptive War in Jewish Law; Shatz, David: Introduction. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds): War and Peace in the Jewish Tradition. New York 2007; Broyde: Just Wars, Just Battles and Just Conduct in Jewish Law.

[27] R. Berlin (Gen. 9:5) makes the same inference.

[28] R. Kook (Mishpat Cohen #144.) notes that these "laws of the kings" are part of a tradition not entirely in our possession but includes various issues brought in the Talmud and Maimonides which clearly show they are relevant to kings and not individuals. For a more comprehensive analysis see Yisraeli, Shaul: Amud HaYemani (HEBREW). Jerusalem [1966] 1992. Online at: http://www.eretzhemdah.org/Data/UploadedFiles/FtpUserFiles/ravIsraeli/books/amudHayemini.pdf., here ch. 9. See also Shatz: Introduction, xvi–xvii.

[29] Precisely against what Michael Walzer labeled the "domestic analogy" (see Leventer, Herb: Philosophical Perspectives on Just War. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds): War and Peace in the Jewish Tradition. New York 2007, 66).

[30] See Broyde, Michael J: Only the Good Die Young? In: Meorot 6/1 (2006). Online at: http://www.edah.org/backend/journalarticle/conversation%20-%20final.pdf; Broyde: Just Wars, Just Battles and Just Conduct in Jewish Law.

[31] Yisraeli: Amud HaYemani, 132, 137. See also Broyde: Just Wars, Just Battles and Just Conduct in Jewish Law, 4,7,28–30; Blau, Yitzhak: Biblical Narratives and the Status of Enemy Civilians in

codified as legally binding as learned from the biblical story of the Gibeonites (Josh. 9), whose agreement was honored by Israel even though it was made under false pretenses.[32]

That said, R. Yisraeli notes that Jewish thought does evince an ethic of compassion even in war.[33] He brings two examples, both codified as law in Maimonides[34] (Laws of Kings 6:1,7):

(1) War is not conducted against anyone in the world until they are first offered peace (and refuse it), whether this is a discretionary war (*milhemet reshut*) or a commanded war (*milhemet mitzvah*), as it says, "when you come close to the city to fight with it, you shall call to it to make peace" (Deut. 20:10).[35]

(7) When we besiege a city which we want to capture, we do not encircle it from all four sides, but only on three. We leave one side open for them to flee. Anyone who wishes to escape with his life may so do, as it says, "and you shall deploy against Midian, as God had commanded Moses" (Num. 31:7). By tradition[36] we have learned that this is what was meant.[37]

Indeed, it is laws like these that have compelled many, like twentieth century R. Shlomo Goren, first head of the Israel Defense Forces' Military Rabbinate, to assert that Jewish sources provide a framework for a Jewish JWT that has more to say than just "adhere to international convention."[38] Such a JWT is undeveloped because, as Michael Walzer –echoing R. Goren – notes, "Jews had … no politics of war and peace from the time of Bar Kokhba (135 C.E.) to the time of Ben-Gurion (1948). The incompleteness of Jewish thought about war derives from this central historical fact."[39]

Wartime. In: Tradition Online 39/4 (2006). Online at: https://traditiononline.org/biblical-narratives-and-the-status-of-enemy-civilians-in-wartime/, here 21–25.

32 See e.g., Maimonides (Laws of Kings 6:3) and Radvaz (ad loc.).

33 Yisraeli: Amud HaYemani, 132.

34 Maimonides (1138–1204, Spain/Egypt) is the philosopher and legal codifier whose greatness is compared to Moses himself. As a point of reference for non-Jews, he is like Aquinas is to Christian theologians and, perhaps, like Al Farabi for Muslims.

35 The anonymously published thirteenth century Spanish Sefer HaHinuch (#527) explains the call for peace teaches compassion.

36 In fourth century Midrash Sifre (see Kesef Mishna, ad loc.).

37 Nachmanides (Hasagot, Omitted Pos. 5) explains this is to inculcate compassion even at war.

38 See Broyde: Only the Good Die Young?. For a review of Goren's work on Jewish JWT, see Edrei, Arye: Divine Spirit and Physical Power: In: Theoretical Inquiries in Law 7/1 (2005). DOI: https://doi.org/10.2202/1565-3404.1124.

39 Walzer, Michael: The Ethics of Warfare in the Jewish Tradition. In: Philosophia 40/4 (2012), 633–641. DOI: https://doi.org/10.1007/s11406-012-9390-5, here 633. For Goren see Edrei: Divine Spirit

That said, the primary demands of *jus in bello* – necessity, distinction, proportionality – find a firm basis in Jewish thought, as will be shown presently.[40]

4.1 Jus In Bello – Necessity

Staring with the *jus in bello* demand to act only out of "necessity," we find the notion is a watchword in the Bible. This can be seen, for example, in the command not to cut down fruit trees: "Fruit-bearing trees must not be cut down outside of the city [to cause suffering and distress][41] nor do we block their irrigation water causing the trees to dry up, as it says, 'do not destroy her trees' (Deut. 20:19)."[42] Nachmanides clarifies the commandment to mean that we do not destroy trees or property unnecessarily (*hinam*), however, destruction for the *necessity* of the military campaign would be permitted.[43]

But perhaps the most dramatic example of the "necessity" demand is found in the biblical prohibition against wartime rape (Deut. 21:10–14). In response to this scourge that is as odious as it is ubiquitous, the Bible promulgated its prohibition in terms that would be accepted.[44] R. Michael Broyde explains that this prohibition, beyond its clear ethical declaration that innocents must not be abused, teaches that those actions with no military necessity are forbidden.[45]

4.2 Jus In Bello – Distinction

Beyond the "necessity" ethic, the rape prohibition also hints to the *jus in bello* requirement for "distinction" (some call it "discrimination") between civilians and combatants. And, while there is no biblical command demanding "distinction," R. Norman Lamm writes that "the idea of refraining from harming civilian non-combatants, although it has no explicit origin in Torah, reflects the Torah value of 'Thou shalt not kill' (Ex. 20:13) and 'The fathers shall not put to

and Physical Power, 275.

40 There are also sources to support *jus ad bellum* and *jus post bellum* of which space here does not permit.

41 Maimonides includes these words in his *Book of Comm.* (Neg. 57).

42 Maimonides, Laws of Kings (6:8).

43 Hasagot (Omitted Pos. 6).

44 Note that the Bible does not ban the taking of a beautiful woman (*yafet toar*) but imposes conditions that seek to curb the practice to a minimum. So, while the biblical ideal is an outright ban, it is not so enacted because, explains the Talmud (Kid. 21b), man's "evil inclination" would simply ignore a such a ban.

45 Personal conversation.

death for the [sins of the] children, neither shall the children be put to death for [sins of the] the fathers; every man shall be put to death for his own sin' (Deut. 24:16). It should be looked upon as part of the 'continuing revelation'."[46] The idea of "continuing revelation" teaches that the interpretation of the biblical text matures as humanity does. Accordingly, Ariel Erlich explains that "distinction" is found in the biblical command to not harm woman and children (Deut. 20:14), which can be interpreted to mean all "noncombatants," as the combatants of yore were only men.[47] The distinction is not unlimited, however, for there are many Jewish source texts that "view civilians whose presence provides cover for enemy maneuvers, as well as those who identify with the enemy or support the war effort, as legitimate military targets."[48]

4.3 Jus In Bello – Proportionality

Finally, there is the *jus in bello* requirement of "proportionality." Alex Leveringhaus explains that necessity and proportionality "are closely related yet separate. The criterion of military necessity merely states that the use of force must have strictly military objectives within a conflict, while the criterion of proportionality of means states that the use of force must not cause excessive damage."[49] It could be said that Maimonides' (Book of Comm., Neg. 57) explanation on not cutting down trees, "that we cause not undue distress (*lehatzeir*) and pain their hearts" includes a call to proportionality. But the notion can be seen even more clearly in Nachmanides' explanation of the juxtaposition of the wartime prohibition against cutting fruit-trees (Deut. 20:19) versus the wartime permit to cut the fruitless trees (Deut. 20:20). Quoting the Talmud (B. Kam. 91b), he writes that soldiers can in fact cut down fruit trees for the sake of the war (i.e., "necessity"), for the verses together are not coming to deny wartime needs but only to promulgate the ethic of "proportionality": first cut down fruitless trees and only then, if need be, cut down fruit trees – but "'thou shalt not destroy the trees' to cut them down destructively." Finally, the notion can be found in the rule of "the pursuer" (*rodef*), which allows one to

---

46 Lamm, Norman: Amalek and the Seven Nations. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds): War and Peace in the Jewish Tradition. New York 2007, 228.

47 Ency. Judaica, Military Law. See also Abarbanel (Deut. 20:10) who hints at the idea.

48 Becker in Becker, Maier/Blau, Yitzchak: Biblical Narratives and the Status of Enemy Civilians in Wartime. In: Tradition. A Journal of Orthodox Jewish Thought 40/4 (2007), 103–108. DOI: http://www.jstor.org/stable/23263523, here 105. Note: I have presented here what could be called a "middle of the road" position between, e.g., Yisraeli: Amud HaYemani, sec. 24 versus Blau: Biblical Narratives and the Status of Enemy Civilians in Wartime. See also Bleich: Preemptive War in Jewish Law, 277.

49 Leveringhaus, Alex: Ethics and Autonomous Weapons. London 2016. DOI: https://doi.org/10.1057/978-1-137-52361-7, 17.

stop a "pursuer" from committing murder by any means even unto his death.[50] That is, one is to stop the murderer by inflicting the least possible damage on him to get the job done, notwithstanding that killing him is an open option. This implicit expression of "proportionality" has been applied to wartime ethics as well.[51]

4.4 Jus In Bello – Intentionality & Responsibility

Now, these jus in bello demands of necessity, distinction, and proportionality, some argue, require "intentionality" (i.e., the qualitative sense of acting for a reason)[52] – i.e., to act with the proper intent, for the right reasons.[53] While not disagreeing, I would argue that this is true only for beings with the capacity of an intentionality subject to emotions – the kind that brings one to act for the wrong reasons (e.g., racism, vengeance, etc.). That is to say, the need for proper intention stems from the fact that people can perform actions which on the surface seem reasonable (e.g., killing an enemy), but below the surface are found to have been driven by underlying human emotion (e.g., hatred) such that the act was unwarranted (e.g., an arrest of the enemy would have sufficed). The argument can also be made in the reverse – e.g., misplaced mercy can lead to the unwarranted release of a recalcitrant enemy. Barring such incongruous intentions, jus in bello criteria are quantitative, adherence to them numerical, execution of them functional. Consequently, they require nothing more than mechanical, hard-wired intentionality – precisely what artificial intelligence (e.g., in AWS) is made to do: follow orders.[54]

Nevertheless, those who demand "intentionality" argue that while an AWS could make the necessary distinction between combatant and civilian, yet, without the interpersonal relation-

---

50 R. Yosef Karo, Shulhan Aruch (Hoshen Mishpat 425:1).

51 Broyde: Only the Good Die Young?, 5; Blau: Biblical Narratives and the Status of Enemy Civilians in Wartime, 14–15.

52 See, e.g., Purves, Duncan/Jenkins, Ryan/Strawser, Bradley J.: Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. In: Ethical Theory and Moral Practice 18/4 (2015), 851–872. DOI: https://doi.org/10.1007/s10677-015-9563-y, here 864 and sources therein.

53 Nagel: War and Massacre, 139; Sparrow: Killer Robots, 67–68, and Asaro, Peter: On Banning Autonomous Weapons Systems. In: International Review of the Red Cross, 886/94 (2012), 687-709 in Purves: Autonomous Machines, Moral Judgment, and Acting for the Right Reasons, 864.

54 Of course, AI takes a statistical approach, however, just like autonomous vehicles have a "safety governor" that uses hard coded rules to ensure compliance (Shalev-Shwartz, Shai/Shammah, Shaked/Shashua, Amnon: On a Formal Model of Safe and Scalable Self-Driving Cars. 2017. Online at: https://arxiv.org/pdf/1708.06374.pdf; for a lay explanation see Yoshida, Junko: Can Mobileye Validate 'True Redundancy'? In: EETimes. 2018. Online at: https://www.eetimes.com/can-mobileye-validate-true-redundancy/), so too AWS can have an "ethical governor" (Arkin, Ronald C.: Governing Lethal Behavior. Boca Raton 2008. DOI: https://doi.org/10.1145/1349822.1349839).

ship, without the "human intention," the act of killing, is "profoundly disrespectful."[55] And this, because there must be someone who takes responsibility.[56] To be clear, the demand for responsibility has two aspects: one consequential, not necessarily linked to human dignity, and one deontological, explicitly linked to human dignity. Sparrow puts it like this:

> Responsibility … is a fundamental condition of fighting a just war [i.e.,] that someone may be held [morally] responsible for the deaths of enemies killed … This condition may be thought of as one the requirements of *jus in bello* … [or] a precondition to [them].[57]

Sparrow makes two primary claims as to why responsibility is so critical to war. The first he calls consequentialist: "An inability to identify those responsible for war crimes would render their prosecution moot, for instance, with disastrous consequences for the ways in which wars are likely to be fought."[58] That is, if there is no one to take responsibility countries may commit war crimes with impunity. This argument is rebuffed by Leveringhaus who explains that AWS can be outfitted with a "black box" that would make known the details of any attack.[59] And while the details would provide only "causal responsibility," they could serve to attribute moral responsibility – i.e., who is punishable.[60] And this, even if the weapon is autonomous, as Michael Schmitt, Professor of International Law, explains:

> Clearly, any commander who decides to launch AWS into a particular environment is, as with any other weapon systems, accountable under international criminal law for that decision. Nor will developers escape accountability if they design systems, autonomous or not, meant to conduct operations that are not IHL [international humanitarian law] compliant. And States can be held accountable under the laws of State responsibility should their armed forces use AWS in an unlawful manner.[61]

---

55 Sparrow: Robots and Respect, 107.

56 The lack of a moral agent to take responsibility for the consequences of autonomous systems is referred to as the "responsibility gap" or "accountability gap." See, e.g., Sparrow: Killer Robots; Sparrow: Robots and Respect; Heyns: Autonomous Weapons Systems, Asaro: Autonomous Weapons and the Ethics of Artificial Intelligence; Saxton, Adam: (Un)Dignified Killer Robots? Online at: https://www.lawfareblog.com/undignified-killer-robots-problem-human-dignity-argument; also Sharkey: Autonomous Weapons Systems, Killer Robots and Human Dignity.

57 Sparrow: Killer Robots, 67.

58 ibid. So too Asaro: Autonomous Weapons and the Ethics of Artificial Intelligence, 226.

59 Leveringhaus: Ethics and Autonomous Weapons, 70.

60 "To hold that someone is morally responsible is to hold that they are the appropriate locus of blame or praise and consequently for punishment or reward" (Sparrow: Killer Robots, 71).

61 In Dunlap: Accountability and Autonomous Weapons, 68.

Sparrow's second claim for responsibility he calls deontological:

> The least we owe our enemies is allowing that their lives are of sufficient worth that someone should accept responsibility [i.e., be punishable] for their deaths. … It is a necessary condition of the respect for persons [i.e., human dignity] that is at the heart of Kantian, and other deontological, ethics.[62]

But as we have just seen in the above quote, AWS does not abrogate the attribution of accountability. It may not be traceable to an individual but to someone higher up on the command chain, nevertheless, that is no different than with other wartime weapons or situations.[63]

All this notwithstanding, based on the Jewish approach that wartime ethics are not to be compared to peacetime ethics, so too it is my thesis that claims of dignity tethered to accountability and justice are true only in peacetime, human dignity on the battlefield being of an incomparable nature to human dignity off the battlefield.[64]

5 Human Dignity

5.1 Wartime versus Peacetime

Human dignity (*kavod habriyot*) is an ethical/legal category taken very seriously in Jewish thought, allowing for great leniencies to ensure that it is preserved.[65] It entails honoring people as a Thou, relating to them as a subject and appreciating their inherent worth simply as human beings. That said, as noted above, Jewish JWT is founded on the notion that the ethics of war are distinct from the ethics of peace. Accordingly, I propose that it is a category mistake to equate peacetime dignity with wartime dignity.

To begin, interpersonal relations – on any level – are untenable; for, as demonstrated above (sec. Enemy Combatants in the Bible), such an attitude is liable to get you killed. And it is kill-

---

62 Sparrow: Killer Robots, 67.

63 See, e.g., Muller: Autonomous Killer Robots Are Probably Good News, 8–10, Dunlap: Accountability and Autonomous Weapons, 66.

64 The notion is broached by Sharkey: Autonomous Weapons Systems, Killer Robots and Human Dignity, 81.

65 For an overview, see Ency. Talmudit, entry: *Kavod Habriyot*.

ing that is at the essence of wartime ethics – as the *moral* permit to kill is the quintessential difference between war and peace.[66] Accordingly, R. Broyde, writes:

> Once 'killing' becomes permitted as a matter of Jewish law, much of the hierarchical values of Jewish law seem to be suspended as well, at least to the extent that the ones who are hurt are people who also may be killed …. The basic argument is that the wholesale suspension of the sanctity of life that occurs in wartime also entails the suspension of such secondary human rights issues as the notion of human dignity.[67]

Simply put, once wartime ethics have permitted the killing of an individual, all other treatment is *a fortiori* permitted – e.g., if one can kill, all the more so one can disrespect. This logic, however, is opposed by Nagel who argues that the goal in war is to stop the soldier not the human being; as such, even killing him must be done with his dignity in mind.[68]

Yet, even if we accept that Broyde's claim is negated by Nagel, it can still be argued that human dignity on the battlefield is of an entirely different nature than off the battlefield. For, given the abrogation of the prohibition on "spilling blood" (*shfichut damim*) the soldier now operates under the dynamic of "kill or be killed" – in the words of the Talmud, "if one comes to kill you, rise up first to kill him" (Ber. 58a).[69] In such a situation, the soldier's dignity is not in being respected by his enemy but in being courageous against him. Indeed, that is precisely the message of the blessing given by the Priest Anointed for War (*mashuach milhama*) as explains R. Don Isaac Abarbanel, the great fifteenth century Spanish statesman and biblical commentator: the priest encourages them saying, "for your own dignity, which is so dear to you, fight with great bravery" (Deut. 20:1). Conversely, if one is "fainthearted," unable to evince the necessary bravery, the Bible (Deut. 20:8) exempts him from service out of concern for human dignity, for it wants not that he be embarrassed by his inability to evince the necessary bravery in battle.[70]

---

66 R. Eliezer Waldenberg, Tzitz Eliezer 12:57. See also Broyde: Just Wars, Just Battles and Just Conduct in Jewish Law, 2; Yisraeli: Amud HaYemani, 137, sec. 24.

67 Broyde: Just Wars, Just Battles and Just Conduct in Jewish Law, 4; Broyde: Jewish Law and Torture. In: The New York Jewish Week, 2006. Online at: http://www.broydeblog.net/uploads/8/0/4/0/80408218/jewish_law_and_torture.pdf.

68 Nagel: War and Massacre, 141.

69 See fn. 65.

70 Mishnah Sotah 8:5. See esp. Ency. Talmudit, entry: *Kavod Habriyot*. Note that the "faintheartedness" dispensation is limited to discretionary wars.

This bravery, of course, includes self-sacrifice – i.e., the soldier (or individual acting in the name of national defense) is called upon to sacrifice themselves for the greater good.[71] This is not a blemish on their dignity but just the opposite, a badge of honor. In the Bible, Esther is forever celebrated as the heroine of the Jewish people for putting her life on the line to save her people (Esth. 4:11–15). In the Talmud, the story is told of Lulianus and Papus who gave their lives to save the people and, as a result, are said to have attained the noblest place the heavens can afford.[72] And in modern times, IDF commander Roi Klein saved his unit by jumping on a grenade and became a symbol of Israeli heroism.[73]

Finally, the Talmud quotes the Psalmist, "Gird thy sword upon thy thigh, O mighty one, thy glory and thy majesty (*hod vehadar*)" (Ps. 45:4), explaining that the sword of the soldier is an expression of his dignity, like a "piece of jewelry" (Shab. 63a).[74] Noteworthy here is that the Talmud is not glorifying war, as it goes on to posit that the sword is really man's disgrace, to which the rejoinder comes: True, but in these times when nation still lifts sword against nation, it is his dignity.[75]

5.2 Dignity Forgone?

An important nuance in the wartime dignity to which I refer can be learned from Dieter Birnbacher who agrees that there is a difference in dignity between soldiers versus civilians.[76] The difference he notes, however, is only because the soldiers, having accepted the consequences of war, apparently forgo their claims to human dignity; as opposed to civilians who, not being a part of the fighting, maintain those claims. Yet it is not at all clear that anyone would ever forgo their personal claim to human dignity – even if they did accept the possibility of death.[77]

One does not forgo their dignity upon accepting battle but, I suggest, accepts that the dignity is not of a peacetime nature. Indeed, wartime dignity is something unachievable in peacetime. Let us be blunt here, Esther was raped; even when she willingly went to the king to plead her

---

71 R. Eliezer Waldenberg (Tzitz Eliezer 12:57:2) calls self-sacrifice an "obligation."

72 Rashi, Taanit 18b (s.v., *BeLudkia*).

73 https://en.wikipedia.org/wiki/Roi_Klein. It should be noted that such heroism is also found in the story of Eleazar Avaran Maccabeus (Maccabees I 6:43–46), similarly, Samson (Judges 16).

74 Glory and majesty refer to honor (*kavod*), see, e.g., Ibn Ezra, Radak (Ps. 104:1).

75 See esp., Rashi (ad loc., s.v. *Shraga*).

76 Birnbacher, Dieter: Are Autonomous Weapons Systems a Threat to Human Dignity? In: Bhuta, Nehal/Beck, Susanne/Geiss, Robin/Kress, Claus (eds.): Autonomous Weapons Systems. Cambridge 2016, 105–121. DOI: https://doi.org/10.1017/CBO9781316597873.005.

77 To wit, Abimelech (Judges 9), as will be explained in "5.3 Status Dignity is Passe?"

case for her people – which implicitly included sleeping with the king – she submitted to him only for the greater good (Meg. 15a). She relinquished her peacetime dignity in exchange for wartime dignity. Similarly, the Midrash (Ecc. R. 9:10 [1]) blesses God for having "removed the disgrace" of Lulianus and Papus, incurred by being killed in the streets of Lod. And clearly having one's body blown to smithereens by a grenade cannot be thought of as dignified in peacetime terms. But in wartime terms it is hard to imagine a more dignified end. Roi Klein, with his last breaths, reported his own death, yelling "Klein's dead, Klein's dead" over the radio … he then handed over his encoded radio to another officer, who took command of the force, and died.[78] This was his dignity. It was certainly not to be found in that the enemy fighter who lobbed the grenade over the wall evinced any sense of interpersonal relation – in even the most minimalist sense.

Similarly, I offer, soldiers in battle cannot be vouchsafed a dignified death in the peacetime sense – for it exists not. It matters little if one is stabbed or shot, burned or butchered – none can be thought of as dignified[79] – whether executed by man or machine.[80] The dignity is in the act of dying for one's country, for one's values, for one's people. In the last words of early Zionist Joseph Trumpledor, who died defending the Tel Hai pre-state village in Israel, "It is good to die for our country."[81]

5.3 Status Dignity is Passe?

And that brings us to another aspect of dignity we might call "status dignity." Ariadna Pop explains that the Roman concept of *dignitas* "conveys ideas such as honor, privilege and deference due to rank or office. [Today it conveys the] 'rank of humans generally in the great chain of being', that is, about their high status in comparison to other forms of existence."[82] Accordingly, it is a slight to one's dignity to be killed by a lower form in the hierarchy of creation – e.g., animals, and, by extension, mindless machines. Pop argues that such a notion is passe, for it "boils

---

78 See fn. 72.

79 Of course there is a difference between these deaths in terms of pain and suffering, but such considerations would fall under the demands for "proportionality," not "dignity" in my account of it.

80 Note that even many AWS opponents agree that AWS do not uniquely violate human dignity as opposed to other weapons (see, e.g., Saxton: (Un)Dignified Killer Robots?, Pop, Ariadna: Autonomous Weapon Systems. Online at: https://blogs.icrc.org/law-and-policy/2018/04/10/autonomous-weapon-systems-a-threat-to-human-dignity/; Sharkey: Autonomous Weapons Systems, Killer Robots and Human Dignity). Against see Sparrow: Robots and Respect, 110; Asaro: Autonomous Weapons, 228.

81 Trumpledor's words echo, of course, Horace's Odes (III.2.13): *Dulce et decorum est pro patria mori.*

82 Pop: Autonomous Weapon Systems.

down to a form of speciesism: that in the hierarchy of being we simply consider ourselves to be the most valuable form of existence and demand to be treated accordingly, without bothering to explain why this is supposed to be the case."

Permit me to "bother to explain" the reason human beings occupy the dignified position at the top of the great chain of being: it is because, according to the Bible, humanity is the very purpose of "being."[83] Today this is mocked as "anthropocentric" and blamed for the woes of the world (i.e., the environment and animal kingdom). The Bible, however, is theocentric not anthropocentric. Indeed, it is only when humanity declared "God is dead," that it looked to itself as the center of being and began the uninhibited exploitation of the world. The Bible, in contradistinction, demands we care for the environment, "to cultivate and protect it" (Gen. 1:15),[84] and that we care for the animal kingdom, found in a plethora of biblical sources encoded as laws against abusing animals (*tzar baalei haim*).[85] But this custodianship over creation is only truly possible when we recognize our dignity at the top of the great chain of being.[86]

Status dignity is not passe, for humans do – and must – hold a position of dignity *vis-a-vis* the rest of creation. Furthermore, the Roman *dignitas* between people, while muted, does exist today – e.g., "dignitaries" of state. That said, status dignity is part of peacetime dignity not wartime dignity (between enemies), as can be learned by juxtaposing the biblical stories of Abimelech and Yael.

Abimelech, explains the Bible (Judges 9), was a rogue king-warrior. In his final battle, he trapped an entire city's inhabitants in their great tower. As he set about to light it on fire, a woman standing on the roof dropped a millstone on his head and cracked his skull. Abimelech, in the throes of death, "called hastily unto the young man his armor-bearer, and said unto him: 'Draw thy sword, and kill me, that men say not of me: A woman slew him.'" (9:54). Clearly concern for one's dignity is not relinquished in war, even when one is about to die.

But the indignity suffered by Abimelech was not due to a lack of human consideration. The woman atop the tower was surely a moral agent that could: "(1) recognize a human being as a human, not just distinct from other types of objects and things but as a being with rights that deserve respect; (2) understand the value of life and the significance of its loss; and (3) reflect upon the reasons for taking life and reach a rational conclusion that killing is justified in a

---

83 See, e.g., San. 38a; Maharal (*Hiddushei Aggadot*, ad loc., s.v., *v'kamar davar*); Malbim (Ps. R. 104:1); Luzzatto ([1734] 1983, 1:2); Radak (Gen. 1:26).

84 See also, e.g., Ecc. R. 7: 13 [1].

85 See, e.g., Gross, Aaron S.: Jewish Animal Ethics. In: Doroff, Elliot/Crane, Jonathan (eds.): The Oxford Handbook of Jewish Ethics and Morality. New York 2013, 419–432. DOI: https://doi.org/10.1093/oxfordhb/9780199736065.013.0027.

86 See Soloveitchik, Joseph: The Lonely Man of Faith. Jerusalem [1965] 2012. Online at: https://traditiononline.org/wp-content/uploads/2019/09/LMOF.pdf, here 16.

particular situation."[87] The indignity was that the king was killed by a power thought to be inferior to his own.

This story, then, seems to argue that status dignity is a wartime concern and so, perhaps, there is room to argue that it is undignified to be killed by a mindless AWS. The fact is that status dignity is found in numerous Jewish sources, e.g.: children are to honor their parents (Ex. 20:12), students their teachers (Maim., Laws of Talmud Torah 5:1), youth their elders (Lev. 19:32), citizens their leader (Ket. 17a). But all this is at peacetime. At wartime the Bible clearly rejects status dignity between enemies. For, while Abimelech was distraught at the indignity of being killed by a woman, the Bible (Judges 4:21) has no qualms about a woman – Yael – killing a general at war – Sisera – for the sake of victory. Indeed, for her valor Yael is praised as being on the level of the matriarchs of the nation (Naz. 23b).[88] Status dignity between enemies, then, while of great import in peacetime, is of no consequence at wartime.

5.4 Soldiers' Humanity

A final issue raised in this discussion is the concern for the human dignity not of those targeted by AWS but those launching AWS. One argument is that human dignity entails the capacity to make moral decisions, and consequently, to abandon moral decision making is to abandon one's human dignity.[89] A second argument in this vein is that in removing the human operator from the decision-making process of killing, there is a "moral distancing" that negatively impacts the human dignity of the operator by making him uncompassionate.[90] Human dignity with respect to the one deploying the weapons is really what we might call "humanity," as in, they will lose their humanity.

Before responding to these claims, it is important to note that the Bible is indeed very concerned that soldiers maintain their dignity, their humanity. "When thou goest forth in camp against thine enemies, then thou shalt guard thyself from every evil thing" (Deut. 23:10). On this the medieval Spanish philosopher and biblical commentator R. Abraham Ibn Ezra explains: "every evil thing – be it spiritual or physical." Here, then, is a biblical imperative to guard one's soul just as surely as one must guard one's life, a warning that demands the soldier

87 See fn. 13.

88 A similar story is told of Judith who killed General Holofernes (Book of Judith, 13).

89 Heyns: Autonomous Weapons Systems.

90 Saxton: (Un)Dignified Killer Robots?

not descend to the brutish nature that war engenders.[91] Nineteenth century German rabbinic leader and biblical commentator, R. Samson Raphael Hirsch, elaborates:

> The laws immediately preceding [this command to "guard yourself from all evil"] were purposed to ensure feelings of personal morality and of sympathy with mankind in general, as being the fundamental principles of the character which is to be formed of the Jewish nation. … "When thou goest forth in camp" – even when you have left your homes and the restricting influence of ordinary family and social life and find yourself in a military camp set out for war against an enemy … so that even the ordinary restrictions of morality and decency become so loosened, and the purpose of the war itself could tend to foster unrestrained coarseness and brutality; "guard yourself from all evil", you are not to loosen your self-controlling inner inspection [but] keep yourself on guard from every "evil." (ad loc.).

This call to restrain the "coarseness and brutality" attendant in the military camp, to not "loosen one's self-controlling inner inspection," does not, it must be noted, imply a call to maintain interpersonal "I-Thou" relations on the battlefield, but simply to maintain one's own moral integrity. Indeed, commandments evincing compassion at wartime come to ensure the soldier doesn't lose the virtue in himself, not because the enemy is deserving of such.[92] Perhaps as much was hinted at by Plato when he wrote that the soldiers of the Republic must be trained in both body and soul, for "those who devote themselves exclusively to physical training turn out to be more savage than they should, while those who devote themselves to music and poetry turn out to be softer than is good for them" (Republic 410d). He too did not counsel for interpersonal relations with one's enemies, but with one's comrades in arms – to ensure that the soldiers of the republic maintain their humanity.[93]

Given this background, we can respond to the concerns over the humanity of those who deploy AWS. Regarding the need to make moral decisions, the Talmud (Ber. 33b) agrees that, in fact, the only area in which we express our freewill, and thus our very humanity, is in the moral decisions we make (*hakol bidei shamayim hutz mi'yirat shamayim*). Nevertheless, that does

---

[91] See also Nachmanides (ad loc.).

[92] Broyde explains that we are not beholden to one-sided compassion but exhibit such "good will gestures" (e.g., leaving a fourth side open to retreat, and not cutting down fruit trees) in the hope that the other side will respond in kind (personal conversation).

[93] Explaining Plato's position, Syse writes: "the problem of creating a soldier class is not that the soldiers might be (unjustly) brutal towards other cities – quite the opposite: we want them to be brutal towards the enemy. The challenge is how to keep them from turning against their compatriots" (Syse: The Platonic Roots of Just War Doctrine, 113).

not mean that if someone, or something, is better than we in some area, that we have become lesser for deferring to them.[94] Furthermore, one is not to go looking to place themselves in moral dilemmas just to express their freewill but rather, refraining from entering difficult moral scenarios is a moral decision in itself.[95] And this serves to answer the second claim that one will lose their empathy by not making the kill decisions. I would argue just the opposite: it is in going to war that makes one lose their humanity, as the history of war clearly demonstrates.[96]

6 Conclusion

In an effort to analyze the human dignity claims against the deployment of AWS, I proposed that there are really two categories of human dignity: peacetime dignity, expressed in interpersonal consideration; and wartime dignity, expressed in courage and self-sacrifice. Accordingly, I claimed that wartime dignity is not displaced by the deployment of AWS.[97]

Allow me to leave with you with another way to look at human dignity: as an overarching theme in creation, as humanity's task, as humanity's destiny. The great twentieth century leader of modern orthodoxy in America, R. Joseph Soloveitchik, writes as follows:

> Human existence is a dignified one because it is a glorious, majestic, powerful existence. Hence, dignity is unobtainable as long as man has not reclaimed himself from co-existence with nature and has not risen from a non-reflective, degradingly helpless instinctive life to an intelligent, planned, and majestic one. … Only when man rises to the heights of freedom of action and creativity of mind does he begin to implement the mandate of dignified responsibility entrusted to him by his Maker. Dignity of man, expressing itself in the awareness of being responsible and of being capable of discharging his responsibility, cannot be realized as long as he has not gained mastery over his environment. …
>
> Man of old, who could not fight disease and succumbed in multitudes to yellow fever or any other plague with degrading helplessness, could not lay claim to dignity. Only the man who builds hospitals, discovers therapeutic techniques and saves lives, is blessed with dignity. Man of the 17th and 18th centuries, who needed several days to travel

94 See, e.g., Mish. Avot 1:6.

95 See, e.g., Kid. 81a, Mish. Avot 1:1, Pes. 40b (s.v., *Nazira*), Shaarei Teshuva 3:7

96 See, e.g., Arkin, Ronald C.: The Case for Ethical Autonomy in Unmanned Systems. In: Journal of Military Ethics 9 (2010), 332–41. DOI: https://doi.org/10.1080/15027570.2010.536402; Leveringhaus: Ethics and Autonomous Weapons, 62.

97 While I have argued for deployment of AWS in wartime, there is also room to argue for its use in peacetime policing; but such is beyond the scope of this paper.

> from Boston to New York, was less dignified than modern man who attempts to conquer space, boards a plane at the New York Airport at midnight and takes several hours later a leisurely walk along the streets of London.
>
> The brute is helpless, and, therefore, not dignified. Civilized man has gained limited control of nature and has become, in certain respects, her master, and with his mastery, he has attained dignity, as well.[98]

Could it not be that AWS is part of this mastery and, thus, part of this dignity? Could it not be that by removing the soldier from the battlefield, AWS makes war more dignified? Could it not be that the same dignity accorded man in overcoming disease, in overcoming space, is accorded him in overcoming the brutality of war? Surely it is more dignified for humanity to resolve its conflicts with the minimum amount of bloodshed and without subjecting itself to the horrors of war that, in Michael Walzer's famous words, "shock the conscience of mankind."[99] Indeed, not a few have noted that "removing human beings from theatres [of war] makes warfare more humane."[100]

To conclude, AWS – through its better empirical and practical capabilities – advances adherence to *jus in bello*,[101] thus resolving the consequentialist concern; and AWS – in reducing/removing humans from the battlefield – advances human dignity, thus resolving the deontological concern. If only it were that Artificial Intelligence not merely fight our battles but help resolve our differences without the resort to war, that would be divine. And while such a thought may appear whimsical, the prophet assures us that there will come a time when:

> "*Nation shall not lift up sword against nation, neither shall they learn war anymore*."

---

98 Soloveitchik: The Lonely Man of Faith, 15–17.

99 Walzer as seen in in Leveringhaus: Ethics and Autonomous Weapons, 62.

100 Leveringhaus: Ethics and Autonomous Weapons, 62. Similarly, e.g., Spaulding, Norman W.: Is Human Judgment Necessary? In: Dubber, Markus Dirk/Pasquale, Frank/Das, Sunit (eds.): The Oxford Handbook of Ethics of AI. New York 2020, 375–402, here 394–5; Arkin: Governing Lethal Behavior, 124; Arkin: The Case for Ethical Autonomy in Unmanned Systems; Dunlap: Accountability and Autonomous Weapons; Sullins, John P.: RoboWarfare. In: Ethics and Information Technology 12/3 (2010), 263–275. DOI: https://doi.org/10.1007/s10676-010-9241-7; Purves, Duncan et alii: Autonomous Machines, Moral Judgment, and Acting for the Right Reasons, 859.

101 ibid.

References

Amoroso, Daniele/Tamburrini, Guglielmo: The Ethical and Legal Case Against Autonomy in Weapons Systems. In: *Global Jurist* 18/1 (2018). DOI: https://doi.org/10.1515/gj-2017-0012.

Arkin, Ronald C.: Governing Lethal Behavior. Boca Raton 2008. DOI: https://doi.org/10.1145/1349822.1349839.

———. The Case for Ethical Autonomy in Unmanned Systems. In: Journal of Military Ethics 9 (2010), 332–41. DOI: https://doi.org/10.1080/15027570.2010.536402.

Asaro, Peter: Autonomous Weapons and the Ethics of Artificial Intelligence. In: Liao, S. Matthew (ed.): Ethics of Artificial Intelligence. Oxford 2020, 212–36. https://doi.org/10.1093/oso/9780190905033.001.0001.

———. On Banning Autonomous Weapons Systems. Human Rights, Automation, and the Dehumanization of Lethal Decision-Making. In: International Review of the Red Cross, 886/94 (2012), 687–709.

Becker, Maier/Blau, Yitzchak: Biblical Narratives and the Status of Enemy Civilians in Wartime. In: Tradition: A Journal of Orthodox Jewish Thought 40/4 (2007), 103–108. DOI: http://www.jstor.org/stable/23263523.

Berman, Samuel A.: Midrash Tanhuma-Yelammedenu. New York 1996.

Bhuta, Nehal/Beck, Susanne/Geiss, Robin/Kress, Claus (eds.): Autonomous Weapons Systems. Law, Ethics, Policy. Cambridge 2016.

Birnbacher, Dieter: Are Autonomous Weapons Systems a Threat to Human Dignity? In: Bhuta, Nehal/Beck, Susanne/Geiss, Robin/Kress, Claus (eds.): Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge 2016, 105-21. DOI: https://doi.org/10.1017/CBO9781316597873.005.

Blau, Yitzhak: Biblical Narratives and the Status of Enemy Civilians in Wartime. In: Tradition Online 39/4 (2006). Online at: https://traditiononline.org/biblical-narratives-and-the-status-of-enemy-civilians-in-wartime/.

Bleich, J. David: Preemptive War in Jewish Law. In: Contemporary Halakhic Problems Vol. 3. New York 1989.

Broyde, Michael J: Only the Good Die Young? In: Meorot 6/1 (2006). Online at: http://www.edah.org/backend/journalarticle/conversation%20-%20final.pdf.

———. Jewish Law and Torture. In: The New York Jewish Week, 2006. Online at: http://www.broydeblog.net/uploads/8/0/4/0/80408218/jewish_law_and_torture.pdf.

———. Just Wars, Just Battles and Just Conduct in Jewish Law. Jewish Law Is Not a Suicide Pact!*. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds.): War and Peace in the Jewish Tradition. New York 2007.

Dehghani, Morteza/Forbus, Ken/Tomai, Emmett/Klenk, Matthew: An Integrated Reasoning Approach to Moral Decision Making. In: Anderson, Michael/ Anderson, Susan Leigh (eds.): Machine Ethics. New York 2011.

Docherty, Bonnie: Losing Humanity. Online at: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

———. Shaking the Foundations. Online at: https://www.hrw.org/report/2014/05/12/shaking-foundations/human-rights-implications-killer-robots.

DODD 3000.09: Autonomy in Weapon Systems. Online at: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

Dunlap, Charles: Accountability and Autonomous Weapons: Much Ado about Nothing? In: Temple International & Comparative Law Journal 30/1 (2016), 63–76. Online at: https://scholarship.law.duke.edu/faculty_scholarship/3592.

Edrei, Arye: Divine Spirit and Physical Power: Rabbi Shlomo Goren and the Military Ethic of the Israel Defense Forces. In: Theoretical Inquiries in Law 7/1 (2005). DOI: https://doi.org/10.2202/1565-3404.1124.

Goldvicht, Chaim Yaakov: War, Kingship, and Redemption. In: Jewish Thought 3/2 (1994).

Gross, Aaron S.: Jewish Animal Ethics. In: Doroff, Elliot/Crane, Jonathan (eds.): The Oxford Handbook of Jewish Ethics and Morality. New York 2013, 419–432. DOI: https://doi.org/10.1093/oxfordhb/9780199736065.013.0027.

Heyns, Christof: Autonomous Weapons Systems: Living a Dignified Life and Dying a Dignified Death. In: Bhuta, Nehal/Beck, Susanne/Geiss, Robin/Kress, Claus (eds.): Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge 2016. DOI: https://doi.org/10.1017/CBO9781316597873.005.

Johnson, Aaron M./Axinn, Sidney: The Morality Of Autonomous Robots. In: Journal of Military Ethics 12/2 (2013), 129–141.

Lamm, Norman: Amalek and the Seven Nations: A Case of Law vs. Morality. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds): War and Peace in the Jewish Tradition. New York 2007.

Leventer, Herb: Philosophical Perspectives on Just War. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds): War and Peace in the Jewish Tradition. New York 2007.

Leveringhaus, Alex: Ethics and Autonomous Weapons. London 2016. DOI: https://doi.org/10.1057/978-1-137-52361-7.

Lin, Patrick/Abney, Keith/Bekey, George: Ethics, War, and Robots. In: Sandler, Ronald L. (ed.): Ethics and Emerging Technologies. London 2014. DOI: https://doi.org/10.1057/9781137349088.

Luzzatto, Moshe Hayyim: The Way of God (Derech Hashem). Translated by Aryeh Kaplan. Jerusalem [1734] 1983.

Mill, John Stuart: The Contest in America. In: Fraser's Magazine. 1862. Online at: https://www.gutenberg.org/files/5123/5123-h/5123-h.htm.

Moor, James H.: The Nature, Importance, and Difficulty of Machine Ethics. In: Anderson, Michael/Anderson, Susan Leigh (eds.): Machine Ethics. New York 2011.

Muller, Vincent C.: Autonomous Killer Robots Are Probably Good News. In: Nucci, Ezio Di/Santonio de Si, Filippo (eds.): Drones and Responsibility. London 2016, 67–81. DOI: https://doi.org/10.4324/9781315578187-4.

Nagel, Thomas: War and Massacre. In: Philosophy & Public Affairs 1/2 (1972), 123–144.

Open Letter on Autonomous Weapons. Online at: http://futureoflife.org/open-letter-autonomous-weapons/.

Pop, Ariadna: Autonomous Weapon Systems: A Threat to Human Dignity? Online at: https://blogs.icrc.org/law-and-policy/2018/04/10/autonomous-weapon-systems-a-threat-to-human-dignity/.

Purves, Duncan/Jenkins, Ryan/Strawser, Bradley J.: Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. In: Ethical Theory and Moral Practice 18/4 (2015), 851–872. DOI: https://doi.org/10.1007/s10677-015-9563-y.

Saxton, Adam: (Un)Dignified Killer Robots? The Problem with the Human Dignity Argument. Online at: https://www.lawfareblog.com/undignified-killer-robots-problem-human-dignity-argument.

Shalev-Shwartz, Shai/Shammah, Shaked/Shashua, Amnon: On a Formal Model of Safe and Scalable Self-Driving Cars. 2017. Online at: https://arxiv.org/pdf/1708.06374.pdf.

Sharkey, Amanda: Autonomous Weapons Systems, Killer Robots and Human Dignity. In: Ethics and Information Technology 21 (2018). DOI: https://doi.org/10.1007/s10676-018-9494-0.

Shatz, David: Introduction. In: Schiffman, Lawrence H./Wolowelsky, Joel B. (eds): War and Peace in the Jewish Tradition. New York 2007.

Soloveitchik, Joseph: The Lonely Man of Faith. Jerusalem [1965] 2012. Online at: https://traditiononline.org/wp-content/uploads/2019/09/LMOF.pdf.

Sparrow, Robert: Killer Robots. In: Journal of Applied Philosophy 24/1 (2007), 62–77. DOI: https://doi.org/10.1111/j.1468-5930.2007.00346.x.

———. Robots and Respect: Assessing the Case against Autonomous Weapon Systems. In: Ethics & International Affairs 30/1 (2016), 93–116. DOI: https://doi.org/10.1017/s0892679415000647.

Spaulding, Norman W.: Is Human Judgment Necessary? Artificial Intelligence, Algorithmic Governance, and the Law. In: Dubber, Markus Dirk/Pasquale, Frank/Das, Sunit (eds.): The Oxford Handbook of Ethics of AI. New York 2020, 375–402.

Sullins, John P.: RoboWarfare: Can Robots Be More Ethical than Humans on the Battlefield? In: Ethics and Information Technology 12/3 (2010), 263–275. DOI: https://doi.org/10.1007/s10676-010-9241-7.

———. An Ethical Analysis of the Case for Robotic Weapons Arms Control. In: 5th International Conference on Cyber Conflict. 2013, 1–20. Online at: https://ieeexplore.ieee.org/document/6568394.

Syse, Henrik: The Platonic Roots of Just War Doctrine: A Reading of Plato's Republic. In: Diametros 23 (2010), 104-123. DOI: https://doi.org/10.13153/diam.23.2010.384.

Trager, Robert F./Luca, Laura M.: Killer Robots Are Here—and We Need to Regulate Them. In: Foreign Policy. 2022. Online at: https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/.

Ulgen, Ozlem: Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an 'Elementary Consideration of Humanity'? (January 31, 2017). European Society of International Law (ESIL) 2016 Annual Conference (Riga), Available at SSRN: https://ssrn.com/abstract=2912002 or http://dx.doi.org/10.2139/ssrn.2912002.

Walzer, Michael: The Ethics of Warfare in the Jewish Tradition. In: Philosophia 40/4 (2012), 633–641. DOI: https://doi.org/10.1007/s11406-012-9390-5.

Yisraeli, Shaul: Amud HaYemani (HEBREW). Jerusalem [1966] 1992. Online at: http://www.eretzhemdah.org/Data/UploadedFiles/FtpUserFiles/ravIsraeli/books/amudHayemini.pdf.

Yoshida, Junko: Can Mobileye Validate 'True Redundancy'? In: EETimes. 2018. Online at: https://www.eetimes.com/can-mobileye-validate-true-redundancy/.

V Transformation der Theologie

Theorie und Kritik

Jewish Philosophy and the Critique of AI Technology

Hava Tirosh-Samuelson

Abstract

AI technology features prominently in transhumanism, the ideology of extreme progress that envisions the emergence of a posthuman species that will supersede biological human beings and eventually will make humans obsolete. This essay contends that transhumanism and the AI technology it venerates should become more central to contemporary Jewish philosophers because they exert profound impact on all aspects of contemporary life. Two Jewish philosophers in particular – Emmanuel Levinas and Hans Jonas – have exerted profound impact on contemporary philosophy of technology and on the discourse of AI and they offer useful perspectives from which to critically engage transhumanism and its fetishization of AI technology. Endorsing the views of the late Rabbi and Lord Jonathan Sacks, who was deeply influenced by Jonas and Levinas, this essay argues that the transhumanist fetishization of technology that undermines genuine human freedom, even though transhumanism approaches AI technology as liberatory force from human biological limitations. The values of freedom, responsibility, and embodied dignity are the Judaic critical responses to AI technology. As much as philosophers of technology will do well to take note of Jewish philosophy, Jewish philosophers will do well to pay more attention to philosophy of technology in general and particularly to philosophy of AI technology.

1 Introduction

AI technology governs almost all aspects of contemporary life: transportation and travel, communication and information, medicine and health care, security and warfare, science and technology, finance and banking, manufacturing and labor, agriculture and food production, criminal law and policing, government and law, art, culture, and education, and, if this were not enough, entertainment and leisure. Given the ubiquity of AI, the paramount question is

not "whether we want to live with AI technology," but "how can we live with AI technology in a manner that is consistent with our social, political, ethical, and spiritual understanding of being human?" AI researchers, computer ethicists, and scholars of machine ethics have engaged AI technology philosophically, writing about the feasibility of building artificial moral agents (AMA), the ethical dilemmas generated by living with intelligent machines, or the ethical principles that could and even should guide the design of ethical machines.[1]

Contributors to philosophical reflections on machine-human interface include Jews, especially in Israel, a hub of the high-tech industry. Yet, it is safe to say that at least until recently technology in general and AI technology particularly have not been high on the agenda of contemporary Jewish philosophy. Other topics have preoccupied Jewish philosophers—Anti-Semitism and the Holocaust, sexuality and gender, Zionism and the Israeli-Arab conflict, Judaism and democracy, or Jewish-Christian relations—although technology is no less impactful on contemporary Jewish life. Likewise, climate change and the ecological crisis have also remained marginal in contemporary Jewish philosophy, although the discourse on Judaism and ecology is more advanced than the discourse on Judaism and technology.[2] Both AI technology and the environmental crisis are of critical importance since they threaten the future of human life on Earth, challenging us to think about the meaning of being human.

As a Jewish intellectual historian, I have engaged AI technology by focusing on transhumanism, a futuristic ideology that envisions the emergence of a new phase in the evolution of the human species in which biological humans will be replaced by super-intelligent machines.[3] Technology and particularly AI technology are crucial to transhumanist futurism. For transhumanists, AI is not only the technology that facilitates the enhancement of humans, AI is also the *telos* of the evolutionary process in which biological humans will be rendered obsolete. In this essay I engage transhumanism and AI technology from the perspective of Jewish philosophy. The Judaic critique of transhumanism and AI technology concerns three main themes: transcendence, responsibility, and embodiment. My reflections on these philosophical issues are particularly inspired by the legacies of Hans Jonas (d. 1993); Emmanuel Levinas (d. 1995), and Rabbi and Lord Jonathan Sacks (d. 2020), but there are many other Jewish philosophers,

---

1 See Sandler, Ronald L. (ed.): Ethics and Emerging Technologies. New York 2014; Duber, Marcus D./Pasquale, Frank/Das, Sunit (eds.): The Oxford Handbook of the Ethics of AI. Oxford 2020.

2 For a call to Jewish philosophers to engage philosophy of technology see Tirosh-Samuelson, Hava: On the Preciousness of Being Human. In: Tirosh-Samuelson, Hava/Hughes, Aaron W.: Jewish Philosophy for the Twenty-First Century. Leiden 2014, 428–457. A recent response to this call has been Bor, Harris: Staying Human. Eugene 2021.

3 See Tirosh-Samuelson, Hava/Mossman, Kenneth L. (eds.): Building Better Humans? Frankfurt a. M. 2012; Tirosh-Samuelson, Hava: Pursuit of Perfection. In: Theology and Science 16/2 (2018), 200–222.

past and present, who inform the way I think about the challenges of transhumanism and AI technology: Moses Maimonides (d. 1204), Hermann Cohen (d. 1918), A. D. Gordon (d. 1922), Franz Rosenzweig (d. 1929), Martin Buber (d. 1965), Abraham Joshua Heschel (d. 1972), Joseph Soloveitchik (d. 1993), and David Hartman (d. 2014). Their writings offer rich insights that are relevant to the issues under consideration in this essay. A Judaic philosophical critique of AI technology contrasts with the enthusiastic reception of transhumanism and AI technology among Christian philosophical theologians.[4]

2 Transhumanism: Salvation by Means of Technology

Transhumanism is an ideology of extreme progress that gives coherence to a vast range of converging technologies (e.g., genomics, robotics, informatics, nanotechnology, and artificial intelligence).[5] As a utopian narrative about the power of technology, transhumanism articulates a social imaginary about the future of the human species. In the transhumanist narrative converging technologies will bring about the physiological, perceptive, and cognitive enhancement of human beings that will pave the way for the replacement of biological humans by autonomous, super-intelligent, decision-making machines which will constitute the Posthuman Age. Since humans are going to build the so-called "digital minds" that will eventually make humans obsolete, by promoting AI technology transhumanism entails that humanity will bring about its own demise, its own collective suicide.[6]

Elsewhere I characterized transhumanism as a "secularist faith,"[7] transhumanism claims that human-made technology, and particularly AI technology, can enable humans to transcend their condition.[8] How so? In transhumanist ideology transcendence is understood on two axes: the horizontal and the vertical. By "horizontal axis" I refer to a slew of biotechno-

---

4 See Cole-Turner, Ron (ed.): Transhumanism and Transcendence. Washington DC 2013; Mercer, Calvin/Trothen, Tracy J. (eds.): Religion and Transhumanism. Santa Barbara 2015; Mercer, Calvin/Trothen, Tracy J. (eds.): Religion and the Technological Future. New York 2021. For Christian critiques of transhumanism consult Gouw, Arvin M./Green, Brian Patrick, Peters, Ted (eds.): Religious Transhumanism and Its Critics. Lanham 2022.

5 See Hurlbut, Benjamin J./Tirosh-Samuelson, Hava (eds.): Perfecting Human Futures: Transhuman Visions and Technological Imaginations. Wiesbaden 2016.

6 See Tirosh-Samuelson, Hava, The Transhumanist Pied Pipers. In: Gouw, Arvin M./ Green, Brian Patrick/Peters, Ted (eds.): Religious Transhumanism and Its Critics. Lanham 2022, 183–214.

7 Tirosh-Samuelson, Hava: Transhumanism as a Secularist Faith. In: Zygon: Journal of Religion and Science 147/4 (2012), 710–734.

8 See Tirosh-Samuelson, Hava: Technologizing Transcendence. In: Trothen, Tracy J./Mercer, Calvin (eds.): Religion and Human Enhancement. New York 2017, 267–283.

logical interventions that enables humans to enhance themselves through genetic engineering, designer genes, or designer drugs so as to make human beings "healthier, more beautiful, more athletic, more intelligent, more creative, more pleasant and more other 'mores'."[9] Since transhumanists understand "being more" or "doing better" as pertaining to the functioning and performance of the human body, and since they privilege choice over chance, transhumanists also promote "morphological freedom" namely the ability to choose a body we wish to have regardless of birth facts. Philosophically, morphological freedom is legitimated by an appeal to the Hedonistic Imperative,[10] and politically, morphological freedom is promoted as an individual civic right.[11] Indeed, if we are to live by the Hedonistic Imperative and if biology only causes misery and suffering, why not liberate humans from constraining biology? For transhumanists, biology is *not* destiny as numerous technological innovations, especially those related to sexuality and gender, amply demonstrate. And if you do not wish to go as far as changing your biological body, all you need is "live" on various virtual platforms, where your simulated avatar can interact with other avatars in the "as if" environment of cyberspace. In cyberspace, life is simulated without embodiment, since the physical body is precisely what transhumanism seeks to transcend and eventually obliterate.

If transcendence means going beyond the limits of biology, what about death? Aren't human beings, organic entities with a limited life span that are destined to die as all organisms do? Transhumanists' answer is "not really." Biotechnology, they predict, will enable us to overcome not only the limitations of aging but perpetually defer death. Aubrey de Grey, for example, is a transhumanist who calls for a "crusade to defeat aging and death," which he regards as "not only morally justified but is the single most urgent imperative for humanity."[12] His new approach to aging promises radical life extension and the perpetual postponement of death through what he calls "Strategies for Engineered Negligible Senescence (SENS)", an umbrella term for "a range of biomedical therapies with the ultimate purpose of postponing age-related effects."[13] In truth, even de Grey concedes that death cannot be vanquished entirely, but what if death could be outsmarted or tricked by using technology? For example, cryonics keeps dead

---

9 See Chu, Ted: Human Purpose, and Transhuman Potential. San Rafael 2014, 32.

10 See Pearce, David: "The Hedonistic Imperative," online at: www.hedweb.com.

11 See Sandberg, Andres: Morphological Freedom. In: More, Max/Vita-More, Natasha (eds.): The Transhumanist Reader. Malden 2013, 56–64.

12 De Grey, Aubrey: The Curate's Egg of Anti-Anti-Aging Bioethics. In: More, Max/Vita-More, Natasha (eds.): The Transhumanist Reader, 215.

13 See Dickel, Sascha/Ferwer, Andreas: Life Extension. In: Ranisch, Robert/Sorgner, Lorenz, Stefan (eds.): Post-and Transhumanism. Frankfurt a. M. 2014, 119–131, here 120; De Grey, Aubrey/Rae, Michael: Ending Aging. New York 2007; De Grey, Aubrey: Radical Life Extension. In: Mather, Derek F./Mercer, Calvin (eds.): Religion and the Implications of Radical Life. New York 2009, 13–24.

biological humans in "deep freeze" in order to resuscitate them in the posthuman future.[14] Many transhumanists endorse cryonics in the hope that in the very near future, technology, and especially nanotechnology, will be so advanced that resurrection of dead bodies will become a technological possibility.[15]

The horizontal axis of transcendence consists of the project of human of human enhancement. Here biotechnology (e.g., genetic engineering) plays a leading role, and Jewish ethicists have been largely supportive of biotechnology and enhancement technology on the ground that rabbinic Judaism is committed to the "repair of the world" (*tikkun olam*).[16] AI becomes crucial when biotechnology involves interface between the human brain and computers, an interface that not only augments human mental capabilities but also transforms humans into technological beings. This is the process of cyborgization that fuses cybernetic and organic features.[17] The figure of the Cyborg has been imagined already in the 1960 and since then it has become a common trope in science fiction, art, and media. Today, when humans live with pacemakers, cochlear implants, retinal implants, and deep-brain stimulation, Cyborg is much more than a metaphor; it is a medical reality that improves human life by overcoming disabilities, diseases, and injuries. Cyborgization erases the boundaries between organic and artificial life, between humans and machines, making AI crucial for the emergence of a post-biological, posthuman species envisioned by transhumanism. It is here that AI technology and the ideology of transhumanism challenges our understanding the meaning of being human and calls for Jewish philosophical engagement.

The reengineering of the human brain shifts the meaning of transcendence from the horizontal to the vertical axis and here the ethical challenges posed by AI technology become more acute. Currently humans and robots coexist but computers perform all sorts of functions that humans either cannot or do not want to perform, so that computer "serve" humanity (hence, we call contemporary supercomputers "servers"). However, in the next several decades, so transhumanists predict, AI may advance to Artificial General Intelligence (AGI) which is "human-level intelligence that can combine insights from different topic areas and display

---

[14] Interestingly, cryonics was invented and promoted by the Jewish engineer Robert Ettinger, already in the 1970s. The center for cryonics is in Scottsdale AZ, where Ettinger and his second wife settled in the early 1960s. Alcor Life Extension Foundation is based in Scottsdale and is currently headed by Max More, a leading transhumanist.

[15] For a Jewish response to Radical Life Extension technologies see Dorff, Elliot N.: Becoming Yet More Like God. In: Ancselovits, Elisha S./Dorff, Elliot N./Israel-Vleeschhouwer, Amos (eds.): The Impact of Technology, Science and Knowledge, Jewish law Association Studies 29 (2020), 125–138.

[16] See Zoloth, Laurie: Go and Tend the Earth. In: Journal of Law, Medicine & Ethics 36/1 (2008), 10–25.

[17] For an endorsement of cyborgization see Thweatt-Bates, Jeanine: Cyborg Selves. Burlington 2012.

flexibility and commonsense reasoning."[18] Transhumanists predict that accelerated exponential progress of AI technology will facilitate the ultimate form of machine-brain interface: the uploading of the human onto supercomputers.[19] When this is achieved, intelligent machines will be able to teach themselves and correct their own mistakes. This constitutes an irreversible turning point, referred to as AI Singularity, in which the super-intelligent machines become autonomous and self-aware because they will possess consciousness. Like a "black hole," a singular object in space and time, where normal laws of physics break down, "the technological singularity is supposed to generate runaway technological growth and massive alterations to civilization and the human mind."[20]

The feasibility of AI Singularity has been subject to a lot of debate, and I join those who are skeptical about it.[21] I am particularly concerned that in transhumanist futurism, technological Singularity is imagined as an inevitable and irrevocable shift from the biological to the computational and informational, and it is presented as a fact that accounts for how the future must and will develop. In other words, it is an entirely deterministic process in which no freedom is possible, since in this narrative, at no point can we decide not to pursue certain technological paths. This is quite odd, since transhumanism begins its path toward transcendence by relishing freedom, including the freedom to choose the body we wish to have. Is there freedom in AI technology? Currently, at least, it does not seem to be the case since algorithms can only function as they were programmed by their human designers. AI systems are artifacts devoid of freedom, and I am among those who are deeply concerned about the increasing power that they exert over our life, while we, biological humans, increasingly lose our freedom. All over the world, in authoritarian regimes and in weak or failing democracies, we witness how AI technology is used to curtail or eliminate freedom of thought, freedom of expression, and freedom of assembly as surveillance methods become more sophisticated and more intrusive.[22]

Transhumanists have acknowledged the risks of AI technology,[23] but they still believe that it alone will enable humanity to attain its perfection and become Transcendent Mind.[24] As phys-

---

18 See Turner, Cody/Schneider, Susan: Could You Merge with AI. In: Duber, Marcus D./Pasquale, Frank/Das, Sunit: The Oxford Handbook of Ethics of AI. Oxford 2020, 307–324, here 309.

19 See Bostrom, Nick: Superintelligence: Paths, Dangers, Strategies. Oxford 2014, 36–48.

20 Turner/Schneider: Could You Merge with AI?, 309.

21 See Ritchie, Barry: The (Un)Likelihood of a High-Tech Path to Immortality. In: Tirosh-Samuelson, Hava/Mossman, Kenneth L. (eds.): Building Better Humans? 357–378.

22 See Crosston, Matthew: Cyber Colonization: The Dangerous Fusion of Artificial Intelligence and Authoritarian Regimes. In: Cyber, Intelligence, and Security 4 /1 (2020), 149–171; Feldstein, Steven: The Road to Digital Unfreedom. In: Journal of Democracy 30 /1 (2019), 40–52; Lamdan, Sarah: Data Cartels: The Companies that Control and Monopolize Our Information. Palo Alto 2022.

23 For example, Bostrom, Superintelligence.

24 See Moravec, Hans: Robot: Mere Machine to Transcendent Mind. New York 2000.

ical reality becomes digital data, and larger and larger supercomputers hold entire simulated realities in their vast minds, Transcendent Mind will emerge. According to Giulio Prisco, an influential Italian transhumanist, this Transcendent Mind "will not be an inanimate machine, but a thinking and feeling person, order of magnitudes smarter and more complex than us."[25] Ted Chu, previously the chief economist of General Motors, refers to these super-intelligent machines as Cosmic Beings (CoBe) and has no qualms calling them "gods," describing them as "a new species on the frontier of cosmic evolution that is unimaginably powerful and creative."[26] He imagines them as our "Mind Children," using language coined by Hans Moravec already in the 1980s,[27] describing them as beings who will have the will to continuously evolve and push forward the evolutionary frontier in the universe. This new life forms will transcend death once and for all, because it will have no carbon-based body, no sexuality, no desire, and not even an interest in happiness or the pursuit of immortality. Engaged in infinite computation, CoBe will move beyond our planet Earth to explore outer space. Hugo De Garis, the Australian transhumanist, is so enthralled with this bliss that he insists "humans should not stand in the way of a higher form of evolution. These machines are godlike. It is human destiny to create them."[28] And when we create them, according to Giulio Prisco, we will ourselves become gods. Yuval Noah Harari, the Israeli commentator on contemporary technology, describes this future world of "dataism" in which humans worship data and trust algorithms to make their moral decisions. AI, he predicts, will follow its own path "going where no human has gone before—and where no human follows."[29] In the futuristic world of dataism, "death will have no dominion," to quote Dylan Thomas, because humans will become obsolete.

3 Transcendence and Freedom

How can we engage these ideas from the perspective of Jewish philosophy? A good place to start is to reflect on the meaning of transcendence in transhumanist narratives. For transhumanists 'transcendence' is understood in materialist terms: the enhancement of the human body so it could transcend biological limitations or eventually abolishing the biological body entirely by turning material existence into data. The transhumanist understanding of tran-

[25] See Prisco, Giulio: Transcendent Engineering. In: More, Max/Vita-More, Natasha (eds.): The Transhumanist Reader, 237.

[26] See Chu, Ted: Human Purpose and Transhuman Potential, 221.

[27] See Moravec, Hans: Mind Children. Cambridge, MA 1988.

[28] Quoted in Barrat, James: Our Final Invention. New York 2015, 86.

[29] See Harari, Yuval Noah: Homo Deus. New York 2015, 393.

scendence is rooted in a materialist interpretation of theory of evolution: natural evolution is too slow, too chaotic, and too unpredictable to be left alone. Instead, transhumanists call for a deliberate human intervention to accomplish the telos of human life technologically by engineering the human species, leaving no room for chance or randomness. But is this the only way to think about transcendence? Obviously not! In Judaism transcendence belongs to God, precisely because God is the Creator of the world. The doctrine of creation rather than the theory of evolution is the point of departure for Jewish philosophers and theologians who do not see a necessary conflict between the doctrine of creation and the theory of evolution because they do not interpret the biblical text literally.[30]

Rabbi and Lord Jonathan Sacks is among the few contemporary thinkers who have engaged the challenges of technology, especially AI technology that we experience daily on social media.[31] Rabbi Sacks commented on the fast pace of technological change that creates a cultural lag, "a state, like now, in which material culture, such as technology, is being transformed faster than non-material culture such as models of governance and social norms."[32] He insightfully viewed the Judaic worldview as a corrective response to the corrosive impact of AI technology on human life. Rabbi Sacks' critique of technological society begins with the doctrine of creation: the creator God is radically transcendent to the world of nature that God had had created. In the act of creation lies true freedom: the freedom of God from nature and the freedom of the human, who is created in the image of God, from material necessity. Judaically speaking, genuine freedom is not to be found in the liberation from material limitations, but in the joy of participating in spiritual life that cannot be reduced to material conditions, although material life in the here and now itself supports and gives rise to spiritual joy. This is the joyous freedom of covenantal life which Sacks has explicated in his numerous books.

In *The Home We Build Together: Recreating Society*, for example, Rabbi Sacks explains divine freedom and consequently human freedom by commenting on Exodus 3:14, in which Moses asks God to identify Himself. God replies: "*eheyeh asher eheyeh*," commonly translated as "I am who I am." Rabbi Sacks correctly notes that the proper translation of the Hebrew phrase is "I will be who I will be" and that the future tense of God's reply signifies God's transcendent freedom. Sacks explains:

---

30 See Samuelson, Norbert M.: Judaism and the Doctrine of Creation. Cambridge 1994; Goodman, Lenn E.: Creation and Evolution. London/ New York 2010. Cohn-Sherbok, Dan: Judaism, vol. 2: Divine Transcendence and Immanence. London/New York 2017.

31 See Sacks, Jonathan: Morality. New York 2020, 13.

32 See Sacks, Jonathan: The Dignity of Difference. London 2002, 69.

> God exists in the future tense because he is the God of freedom. God is not part of nature. He created nature. Therefore, He stands outside it. He is not bound by it. He is free. And to the extent that we are in God's image, we too are free. The gift God gives us is freedom itself. We too will be what we choose to be.[33]

In commenting on Exodus 19:3–8 Sacks teaches that:

> God made space for human freedom and invites an entire people to become, in the rabbinic phrase, "his partners in the work of creation." The free God desires the free worship of free human beings. God transcends nature; therefore, God is not bound by nature; therefore, God is free. God sets his image on the human person. Therefore, humanity is free. The story of the Bible is the tangled tale of the consequence of God's fateful gift of human freedom. Faith, or more precisely, faithfulness, is born where the freedom of human beings meets the freedom of God in an unconstrained act of mutual commitment.[34]

In the Judaic understanding of transcendence, only God is truly free and only created human beings can participate (at least to some extent) in divine freedom precisely because they are created in the image of God. Human-made machines, by contrast, cannot be said to be free and cannot ever be truly transcendent precisely because they are the artifacts made by human beings. Judaism does posit human exceptionalism along with the responsibility that it entails. Although Rabbi Sacks did not say so explicitly, I believe he would have considered technological singularity a logical impossibility as well as an undesirable telos of human life, precisely because it entails the demise of divinely created humans.

Rabbi Sacks' concise articulation of transcendence and freedom allows us to reflect on our current world that is saturated by AI technology. Note that AI technology is promoted by appeal to freedom because AI presumably liberates us from all sorts of constraints. AI systems are supposed to enlarge the scope of human freedom by replacing workers from tedious labor, by shortening the decision-making processes through fast data analysis, or by translating bodily functions into medically relevant information that leads to better health outcomes. Mostly, AI technology promises greater freedom concerning the one resource that humans need most: *time*. In its split-second calculations, AI technology can handle problems that earlier would have taken months if not years to reach a reasonable conclusion or decision. But does freedom lie merely in calculation? Can AI systems make decisions that negate or disrupts the way it was designed? I don't think so, and I believe that Rabbi Sacks too would have answered

---

[33] See Sacks, Jonathan: The Home We Build Together. London 2007, 58.

[34] See Sacks: The Home We Build Together, 105.

these questions in the negative. As human-made artifacts, AI systems reflect human values, decision making, and actions. Rabbi Sacks' approach also accords with Debora G. Johnson, an ethicist of AI, who asserted that AI systems can be viewed as "moral," but not as "moral agents."[35] While the interactions of AI systems with humans raise many moral considerations, AI systems themselves lack moral agency.[36] They also lack the freedom that characterize moral agents who enjoy the freedom to err and who are accountable for their choices. In the created world occupied by humans, freedom entails *responsibility*, that is, the ability to respond to the needs of the Other who is different from me. This is the second intersection point between Jewish philosophy, transhumanism, and AI technology. Modern Jewish philosophers—Martin Buber, Hans Jonas, and Emmanuel Levinas—had much to say about responsibility and its relationship to transcendence and freedom.[37] In the following section I discuss how Levinas was appropriated by AI ethicists, while expressing qualms about that appropriation.

4 Responsibility and Human Dignity

For transhumanists the future is dictated by the accelerated march of technological progress that will lead to the abolition of the human species. That may or may not happen in the remote future, but in the present, we increasingly co-exist with AI systems, whether we like it or not. All automation technologies enable us to delegate to machines what was once done by humans, but AI technology does more than that: it enables us to delegate far more to machines than we used to: namely, responsibility. If AI systems are moral agents, as some machine ethicists hold, should they be considered legally and morally responsible? Who is responsible for the harms and benefits of the technology when agency and decisions are delegated to AI? Machines already have a lot of power but are they also responsible for the consequences of their powerful consequences? We are responsible when we know what we are doing and know what the consequences will be. But AI systems can cause a lot of harm without knowing what they are doing, because they lack consciousness. Following Deborah Johnson, Mark Coeckelbergh also states that "machines can be agents but not moral agents, because they lack not only consciousness but also free will, emotions, the capability to form intentions, and the like."[38] If machines and

---

[35] See Johnson, Deborah G.: Computer Systems. In: Ethics and Information Technology 8 (2006), 195–204, here, 197.

[36] See Gunkel, David J.: The Machine Question. Cambridge, MA 2012, 15–91 and the literature cited there; Powers, Thomas M.: On the Moral Agency of Computers. In: Topoi 32 (2013), 227–236.

[37] See Werner, Micah: The Immediacy of Encounter and the Danger of Dichotomy. In: Tirosh-Samuelson, Hava/Wiese, Christian (eds.): The Legacy of Hans Jonas. Leiden 2008, 203–230.

[38] See Coeckelbergh, Mark: AI Ethics. Cambridge MA 2020, 111.

algorithms are *a-responsible w*here does responsibility lie? Is it with the innovator of the technology, the owner of the technology, the user of the technology, the one who benefits from the technology, the one who regulates the technology, or all of the above? This is a genuine conundrum that has begun to be explored by Jewish ethicists.[39]

The problem is that 'responsibility' is a fuzzy concept with numerous meanings, conditions, and applications.[40] The fuzziness plagues the debates about autonomous cars, autonomous weapon systems, and algorithm-driven financial systems.[41] In these cases it is difficult to assign responsibility because they consist of many interconnected components: the algorithm, the sensors, all kinds of data that interact with all kinds of hardware and software. All of these are connected to the people who programmed and produced them, so it is hard to delimit the scope of AI: Where does AI begin and where does AI end and the rest of the technology begin? In ethics, we should recall, responsibility is tied to answerability and explainability, but these conditions do not apply to AI, at least not today, since AI does not "know" what it is doing; it is not conscious and not aware of what it brings about. In that regard, at least currently, the phrase "artificial intelligent" is a misnomer, or a category mistake since they are not intelligent as humans are. If so, responsibility must lie with the human who programs the machines and designs the algorithm, but that claim is complicated since the human programmer does not always know what the AI is doing at any moment in time and cannot always explain what it did or how it came to its decision. This is the problem of transparency, which has become more evident in machine learning and deep learning that uses neural networks where explanation based on decision-making tree is no longer possible.[42] Since humans, including the designers of AI, do not know what AI is doing, they cannot explain a particular decision, and therefore it is difficult to talk about transparency or trust.

Ethicists of AI have wrestled with the problem of responsibility,[43] and on this issue they could do well to consult Jewish philosophers who have reflected extensively on responsibility. I am most intrigued by the impact of Levinas' philosophy on the discourse of machine-human interface. This conversation began in 2000 when Richard Cohen, a Jewish Levinas scholar,

---

39 See Navon, Mois: The Virtuous Servant. In: Frontiers in Robotics and AI 8 (2021), 1–15.

40 See Stahl, Bernd Carsten: Responsible Computers? In: Ethics and Information Technology 8 (2006), 205–213.

41 For Judaic engagement with these issues see Berman, Nadav: Jewish Law, Techno-Ethics, and Autonomous Weapon Systems. In: Ancselovitz, Elisha S./Dorff, Elliot N./Israel-Vleeschhouwer, Amos (eds.): The Impact of Technology, Science and Knowledge, Jewish Law Association 29 (2020), 91–124.

42 See Diakopolous, Nicholas: Transparency. In: The Oxford Handbook of Ethics of AI, 197–214; Chesterman, Simon: We, The Robots? Cambridge 2021, 144–170.

43 See Dignum, Virginia: Responsibility and Artificial Intelligence. In: The Oxford Handbook of Ethics of AI, 215–231.

reflected about cyber technology in the light of Levinas' philosophy.[44] The conversation has been carried out mainly by non-Jewish philosophers and ethicists for whom Levinas was an existentialist phenomenologist critical of traditional Western moral philosophy whose insights open new vistas for ethics of AI. [45]

Why is Levinas so attractive to philosophers of AI technology? The work of David J. Gunkel is a good place to start looking for an answer. He reminds us that "ethics is customarily understood as being concerned with question of responsibility for and in the face of an 'other.' For traditional forms of moral philosophy this "other" is more often than not conceived as another human being – another human subject who is essentially and necessarily like we assume ourselves to be."[46] By reserving moral consideration to humans only, traditional moral philosophy excluded animals from its preview and even among humans it tended to treat certain classes as morally inferior (e.g., women, blacks, Jews, indigenous people, etc.). Traditional moral philosophy justified the exclusion by appeal to certain abstract principles that highlighted sameness: only those who presumably possessed a certain criterion were included in the community of moral agents. As Gunkel explains, Levinas challenged the traditional assumptions of moral philosophy, when he argued that "ethics does not rely on metaphysical generalizations, abstract formulas, or simple pieties. His philosophy is concerned with the response to and responsibility for the absolutely Other who is confronted in an irreducible face-to-face encounter."[47] Although Gunkel admits that for Levinas "this other is always and unapologetically human," Levinas allow us to "think otherwise" and expand the scope of the Other by becoming more inclusive than he himself was. The logic of Levinasian philosophy of alterity thus facilitates the inclusion of the nonhuman Others (be they animals, eco-systems, or nature more broadly),[48] and most relevant to this essay, the inanimate and the artificial.

---

[44] See Cohen, Richard A.: Ethics and Cybernetic. In: Ethics and Information Teachings 2 (2000), 27–35; Reprinted in: Atterton, Peter/Calarco, Matthew (eds.): Radicalizing Levinas. Albany 2010, 153–167.

[45] See Beavers, Anthony F.: Phenomenology and Artificial Intelligence. In: Metaphilosophy 33 (no. 1/2) (2002), 70–82; Gunkel, David J.: Thinking Otherwise. In: Ethics and Information Technology 2 (2007), 165–177; Gunkel, David J.: The Machine Question. Cambridge, MA/London, England 2012; Gunkel, David, J.: The Symptoms of Ethics. In: Human-Machine Communication 4 (2022), 67–81; Bergen, Jan Peter/Verbeek, Peter-Paul: To-Do Is to Be. In: Philosophy and Technology 34 (2021), 325–248; Bergen, Jan Peter: Responsible Innovation in Light of Levinas. In: Journal of Responsible Innovation 4 (2017), 354–370; Wohl, Benjamin S.: Revealing the 'Face' of the Robot. 2014, 704–714; Liberati, Nicola/Nagataki, Shoji: Vulnerability under the Gaze of Robots. In: AI & Society 34 (2019), 333–342.

[46] See Gunkel, Thinking Otherwise, 166.

[47] See Gunkel, Thinking Otherwise, 167.

[48] On the application of Levinas to the environmental discourse see Atterton, Peter: Face to Face with the Other Animal? In: Atterton, Peter/Calarco, Matthew/Friedman, Maurice (eds.).: Levinas

Levinas' respect of "radical otherness" has enabled ethicists to extend moral considerability to nonhuman animate and inanimate AI systems, thereby "radicalizing Levinas."[49] But is such interpretation of Levinas doing justice to his teachings? According to Levinas, God can be accessed only through one's responsibility for the Other, but the Other is decidedly human. The Other cannot be contained; the Other is infinite, only pointing in the direction of the transcendent God who cannot be known but who is revealed in responsibility for the Other. The ethical life thus consists in responding to the moral call of the Other, but, let me reiterate: The Other is decidedly a human being whose embodied human needs obligate the self to respond. As Levinas puts it:

> For every man, assuming responsibility for the Other is a way of testifying to the glory of the Infinite, and of being inspired. There is prophetism and inspiration in the man who answers for the Other, paradoxically, even before knowing what is concretely required of himself. This responsibility prior to the Law is God's revelation.[50]

According to Levinas, we have no direct relations with the divine. The divine can only be accessed through the human Other to whom the self is infinitely responsible. The "Face" is thus a metaphor of the embodied human, and face-to-face encounter is distinctly human. Even Gunkel in a later essay admits that "technological devices do not possess a face or confront the human ser in a face-to-face encounter that would call for and would be called ethics"[51] For that reason, Gunkel rephrases the merit of Levinas for the ethics of AI by saying that it lies not simply in inclusive moral theory but in "questioning the entire history of ethics and its necessary and unavoidable exclusions."[52] Likewise, Peter Atterton, who extended the figure of the Face to animals, concedes that "Levinas, despite his self-professed departure from the philosophical tradition, retains the great twin leitmotifs of the tradition – anthropocentrism and humanism."[53]

Whereas admirers of Levinas consider his anthropocentrism and humanism to be a shortcoming, I consider the religious, and specifically Judaic, the core of Levinas' legacy which is rooted in the doctrine of creation. We do know that Levinas did not like to be known as a "Jewish philosopher," but we also know that Levinas was an observant Jew and that his Jewishness

and Buber. Pittsburgh 2004, 262–280; Edelglass, William/Hatley, James/Diehm, Christian (eds.): Facing Nature. Pittsburgh 2009.

49 See Atterton, Peter/Calarco, Matthew (eds.): Radicalizing Levinas. Albany, 2020.

50 See Levinas, Emmanuel: Ethics and Infinite. Pittsburgh 1985, 113.

51 See Gunkel, The Symptoms of Ethics, 79.

52 Ibid., 80.

53 See Atterton, Peter: Face to Face with the Other Animal? In: Levinas and Buber, 272.

is culturally and conceptually integral to his philosophy. As many interpreters of Levinas have made clear, Levinas's philosophic and religious writings offer a coherent whole, not a bifurcated message that divorces philosophy from religion.[54] The humaneness of the Other, which Levinas considered as pointing to God, is precisely what AI systems erase when embodied humans become data. Digital and AI technologies do erase our personal identity, disregard our privacy, dismiss our inherent dignity, because, if we want to use Levinas language, they "de-face" us. To put it in terms of Rabbi Sacks' language, when an AI system turn us into data, which can be harvested, manipulated, and sold to the highest bidder, they remove our "dignity of difference," our particularity that cannot be severed from our physicality. This can be best evident in the technology of facial recognition that some AI ethicists have endorsed. David Zvi Kalman correctly challenged the application of Levinas's philosophy to face recognition technology when he stated that "Levinas would have banned facial recognition technology [and] we should too."[55] While expanding moral considerability is an honorable goal, I do not think it justifies glossing over, ignoring, or dismissing Levinas's own humane teachings which posit Levinas (very much like Rabbi Sacks) as a critic of our technologically saturated society rather than its endorser.

5 Embodiment and the Future of Humanity

No Jewish thinker was more prescient about the power of modern technology than Hans Jonas (d. 1993), who articulated the Imperative of Responsibility (*Das Prinzip Verantwortung*), as a response to the trauma of WWII.[56] In this section I focus on Hans Jonas because he valorized embodied human existence and our responsibility toward future human generations. Jonas correctly understood the threat of technology to the future of humanity and to the future of life on Earth. Like Levinas, Jonas too did not like to be known as a "Jewish philosopher," but his philosophy, and especially his critique of modern technology, could not be severed from his

[54] See Gibbs, Robert: Blowing on the Embers. In: Modern Judaism 14 (1), 99–113; Meir, Ephraim: Buber and Levinas's Attitudes toward Judaism. In: Levinas & Buber, 133–156; Ben-Pazi, Hanoch: The Philosophical Meaning of the Names of God. In: Revue Internationale de Philosophie 60/1 (n. 235) (2006), 115–135; Hatley, James: Generations: Levinas in the Jewish Context. In: Philosophy & Rhetoric 38 (2) (2005), 173–189; Fagenblat, Michael: A Covenant of Creatures. Palo Alto 2012.

[55] See Kalman, David Zvi: Levinas Would Have Banned Facial Recognition Technology: We Should Too, Tablet (posted January 11, 2021), available on http://www.tabletmag.com.

[56] See Jonas, Hans: The Imperative of Responsibility. Chicago 1984; Vogel, Lawrence: 'The Outcry of Mute Things'. In: Macauley, David (ed.): Minding Nature. New York/London 1996, 167–185; Morris, Theresa: Hans Jonas's Ethics of Responsibility. Albany 2013; Coyne, Lewis: Hans Jonas: Life, Technology and the Horizons of Responsibility. London 2021.

own experience as a German Jew, a Zionist, and a soldier in the Jewish Brigade of the British army during WWII.[57]

Jonas presciently grasped that modern technology is more than a mere instrument for human purposes. Modern technology forces human beings into a dialectical situation: our attempt to bring the external world under our power ends with the power of technology to destroy or radically refashion the very subject whose power it is. Jonas tells us that,

> *techne* in the form of modern technology has turned into an infinite forward-thrust of the race, its most significant enterprise, in whose permanent, self-transcending advance to ever greater things the vocation of man tends to be seen, and whose success of maximal control over things and himself appears as the consummation of his destiny. Thus, the triumph of *homo faber* over his external object means also his triumph in the internal constitution of *homo sapiens*, of whom he used to be a subsidiary part.[58]

Jonas also correctly grasped the power of modern technology to set in motion a causal chain that has a profound effect on objects and peoples in very remote places and in future epochs and he understood that these changes are irreversible. If mistakes are made, correcting them is very difficult – in many cases, impossible. Therefore, modern technology, especially biotechnology and AI technology, have changed the moral situation and undermined the entire premodern framework of human action. Jonas' task was to articulate a new ethics of our technological age, that could take the new situation into consideration.

Jonas insightfully critiqued the utopian spirit of modern technology and its glorification of "progress" that animate transhumanism. This utopian impulse is behind the promises of technology to cure the "mistakes" of nature or overcome its shortcomings, which constitute what I described as the "horizontal" axis of the transhumanist pursuit of transcendence. But most insightfully Jonas understood how modern technology turned the human being into an object of technology, into a design object, as exemplified most distinctly in genetic engineering and in cyborgization. These technologies illustrate how "Homo faber [man the maker] is now turning on himself and gets ready to make over the maker of all the rest."[59] Jonas, therefore, argued vociferously that the very existence of humanity as created by God is itself a value. Genetic engineering, a human-made technology, enables humans to create other humans, not

---

57 See Wiese, Christian: The Life and Thought of Hans Jonas. Newton 2007.

58 See Jonas, Hans: Technology and Responsibility. In: idem: Philosophical Essays. Englewood Cliffs 1980 [1974], 3–20, here, 11.

59 See Jonas, Technology and Responsibility, 18. This statement was written a few years before the publication of Moravec: Mind Children, and Kurzweil, Ray: The Age of Intelligent Machines. Cambridge, MA 1990.

in the image of God but in their own image. This is the hubris against which Jonas spoke with prophetic passion, although unlike the biblical prophets, and more like Levinas, Jonas insisted on the hiddenness of God and on God's inability to prevent human self-destruction.

Precisely because modern technology, today even more than in Jonas' lifetime, has the capacity to destroy that which God had created, Jonas articulated a new philosophy of nature that recognizes the intrinsic moral value of the natural world and our responsibility toward it. The natural world is not just inert material stuff that we are free to exploit as we please; rather, it has an inherent moral worth that we must treat as an end and not merely as a means. Jonas was a champion of organic life because he personally experienced the vast devastation of life in WWII and the power of technology to destroy the possibility of future life. For Jonas, organic life is itself is "an ontological revolution in the history of matter," a radical change in matter's mode of being. As Strachan Donnelly has shown, Jonas

> traces the full reaches of human morality and responsibility back to natural, organic origins—specifically to the human parent-child relations, to the natural feeling and unchosen responsibility for the utterly needed and vulnerable, but inherently valuable infant, with all his or her human promise to come.[60]

He thus gave a phenomenological account of organic, including humanly organic, life and taught us to honor the natural world of which humans are integral part. Seeking to overcome the radical split between nature and ethics, between "is" and "ought," characteristic of modern science and philosophy, Jonas endowed all forms of life with intrinsic moral worth, insisting that life itself, and the material world necessary for its being, command ultimate respect, allegiance, and finally moral commitment. This approach is the very opposite of transhumanism that denigrates the biological as that which must be transcended by means of technology.

Jonas was the first to articulate our responsibility toward future generations of humanity as a primary moral value. Our responsibility is not just for humanity as it is but as it is *yet to be realized*. It is possible to read Jonas to justify responsibility toward technologized posthumans, but, as in the case of Levinas, I would contend that this is a misreading. For Jonas, the very existence of biological humanity is an objective good that imposes an obligation on the human will that, through technology, has power over this objective good. Having witnessed the destruction of life in WWII, Jonas declared that "there is an unconditional duty for mankind to exist."[61] The mankind he envisioned did not consist of cyborgs or digital minds but of biological humans

---

[60] See Donnelley, Strachan: Hans Jonas and Ernst Mayr. In: Tirosh-Samuelson, Hava/Wiese, Christian (eds.): The Legacy of Hans Jonas. Leiden 2008, 261–287, here 276.

[61] Jonas, Technology and Responsibility, 37.

like those who were gassed in Auschwitz or incinerated in Hiroshima. Jonas' powerful statement has become so much more compelling today in the specter of transhumanist ideology and AI technology. For Jonas, the source of the duty is an "ought" that stands above and commands us and future human beings to secure the future of humanity. That moral obligation is "the duty to be truly human."[62] This is the "ontological imperative" of humanity that derives from the *idea of Man*, an idea "telling us why there should be men [and it] tells us also how they should be."[63]

Jonas' life experiences taught him one powerful truth: Humanity ought to be, and it should not be replaced by some other beings, as transhumanism envision. This moral imperative sets a limit on human technological quest and its quest for control over natural processes or its insatiable desire to "improve" human nature. More than any other contemporary Jewish philosopher, Jonas engaged philosophically assisted reproductive technologies, germ-line engineering, cloning, and radical life extension. In so doing, he showed us why Jewish philosophy is so relevant to our technological age and how Jewish philosophy of technology could be written.[64] Jonas' ethics of responsibility has been a major inspiration for many philosophers, non-Jewish as well as Jewish, among them Rabbi Jonathan Sacks. Like Jonas and Levinas, Rabbi Sacks highlights the ethics of responsibility its relationship to human freedom which is rooted in creation in the image of God. In *The Great Partnership*, Sacks presents Judaism as:

> the assertion of freedom as against its ever-present, ever-changing denials is what marks Abrahamic monotheism as a distinctive philosophy. We are free. We are choosing animals … 'I have set before you life and death, the blessing and the curse. Therefore choose life' (Deuteronomy 30:19). Life is choice. In that fact lies our dignity. If we have no freedom, what makes us from the animals we kill for our own ends, sometimes even for sport? What makes us persons, not things? *If we deny freedom in theory, eventually we will lose it in practice*, as happened in Nazi Germany and Stalinist Russia, and as may yet happen in new and unforeseeable ways.[65]

---

62 Ibid, 42.

63 Ibid, 43.

64 The Jewish ethicist who followed Jonas most closely was Leon Kass, although their views were not identical. See Vogel, Lawrence: Natural-Law Judaism. In: Tirosh-Samuelson, Hava/Wiese, Christian: The Legacy of Hans Jonas, 287–314. Interestingly, Jewish bioethicists who are probiotechnology have been most critical of Kass. See Zoloth, Laurie: There is the World, and There Is the Map of the World. In: Duwell, Marcus/Rehman-Sutter, Christoph/Mieth, Dietman (eds.): The Contingent Nature of Life. Wiesbaden 2008, 307–324.

65 See Sacks, Jonathan: The Great Partnership. London 2011, 126–127.

I wholeheartedly endorse Rabbi Sacks' approach and find his views most instructive.

6 Conclusion

Transhumanism and the AI technology it fetishizes threaten human freedom and the dignity of human life. We see it in many domains of our contemporary life, but especially in economics and politics, where surveillance capitalism is closely linked to the rise of authoritarianism.[66] Contrary to transhumanists dreams, AI has not delivered us happiness but rather has exacerbated social problems including addictiveness, deceptiveness, shallowness, loneliness, lack of empathy, viciousness, misogyny, racism, hate speech, and many other social ills. This is not to say that we can revert to an earlier way of life prior to the invention of AI, but that we need to develop responsible AI and promote the design and engineering approaches that ensure the safe, beneficial, and fair use of AI technologies. To do so, AI innovations should not be left to engineers and computer scientists alone, but involve humanists (e.g., ethicists, philosophers, anthropologists, scholars of religious studies, and specialists in cultural studies) who could introduce a human-centered approach to AI technology. Such an approach is indeed conservative, since it is "urging us to try to use precedent, past wisdom, and conventional metaphysics as much as possible when trying to resolve ethical issues involving current and near-future AI technologies."[67] But the more we enlist a human-centered approach, the more likely we are to build responsible AI.[68] Most recently, the project of AI ethics has been questioned by Bernd Carsten Stahl, claiming that "AI ethical principles are useless, failing to mitigate the racial, social and environmental damages of AI technologies in any meaningful way."[69] The criticism may be justified, but an AI that is attentive to ethical concerns is definitely better than AI that is oblivious to them.

Humanity is faced with consequential choices about AI that now engulfs all aspects of life: economic, social, political, cultural, artistic, and spiritual. The choices we make now will determine whether the transhumanist vision will come to pass, and consequently render humans the victims of their own creations, or whether humans we will regulate AI and refuse to turn themselves into design objects. Jewish philosophy offers profound religious insights that guide us in these perplexing times, urging us to remember our creaturely status and the truth of be-

66 See Zuboff, Shoshana: Age of Surveillance Capitalism. New York 2018.

67 Crisley, Ron: A Human-Centered Approach to AI Ethics. In: The Oxford Handbook of Ethics of AI, 466.

68 Stahl, Bernd Carsten: Responsible Computers? In: Ethics and Information Technology 8 (2008), 205–213.

69 Munn, Luke: The Uselessness of AI Ethics. In: AI and Ethics (2022) https://doi.org/10.1007/s436810022000209-w.

ing created in the divine image. Technology is not our salvation, and technology for the most part does not make us more humane, more caring, or more just. The transhumanist worship of technology, especially AI technology, is a form of techno-idolatry that fetishizes human beings by worshiping their artificial products.[70] Jewish philosophy offers critical perspectives from which to engage transhumanism and its veneration of technology, especially AI technology, identifying the limitations and paradoxes of transhumanism. Beyond the critical task, Jewish philosophy, as we illustrate by reference to Rabbi Jonathan Sacks, Emmanuel Levinas, and Hans Jonas, offers an alternative vision of human existence that protects embodied human dignity against the mechanization and commodification of human life which is promoted by transhumanism and put into practice by tech entrepreneurs. Jewish philosophy, which teaches the ethics of responsibility and the dignity of human embodiment, supports genuine democracy, and rejects the authoritarianism of transhumanist technological idolatry.

References

Atterton, Peter: Face to Face with the Other Animal? Atterton, Peter/Calaro, Matthew/Friedman, Maurice (eds.): Levinas and Buber. Dialogue & Difference. Pittsburgh 2004, 262–280.

Atterton, Peter/Calaro: Matthew (eds.): Radicalizing Levinas. Albany 2010.

Atterton, Peter/Calaro, Matthew/Friedman, Maurice (eds.): Levinas and Buber. Dialogue & Difference. Pittsburgh 2004.

Barrat, James: Our Final Invention. Artificial Intelligence and the End of the Human Era. New York 2015.

Beavers, Anthony F: Responsibility and Artificial Intelligence. In: Metaphilosophy 33/1/2 (2002), 70–82.

Ben-Pazi, Hanoch: The Philosophical Meaning of the Names of God. In: Revue Internationale de Philosophie 60/1 (n. 235) (2006), 115–135.

Bergen, Jan Peter: Responsibility Innovation in Light of Levinas Rethinking the Relations between Responsibility and Innovation. In: Journal of Responsible Innovation 4 (2017), 354–370.

Bergen, Jan Peter/Verbeek, Peter-Paul: To-Do Is To Be. Foucault, Levinas, and Technological Mediated Subjectivation. In: Philosophy and Technology 34 (2021), 325–348.

Berman, Nadav: Jewish Law, Techno-Ethics and Autonomous Weapon Systems. Ethical-Halachic Perspectives. In: Ancslovitz, Elisha S./Dorff, Elliot N./Israel-Vleeschhouwer, Amos (eds.): The Impact of Technology, Science, and knowledge, Jewish Law Association 29 (2020), 91–124.

Bor, Harris: Staying Human. A Jewish Theology for the Age of Artificial Intelligence. Eugene 2021.

Bostrom, Nick: Superintelligence. Paths, Dangers, Strategies. Oxford 2014.

[70] Tirosh-Samuelson, Hava: The Paradoxes of Transhumanism. Theologische Literaturzeitung 146/3 (2021), 123–146.

Chesterman, Simon: We, The Robots? Regulating Artificial Intelligence and the Limits of the Law. Cambridge 2021.

Christley, Ron: A Human-Centered Approach to AI Ethics. A Perspective from Cognitive Science. In: Duber, Marcus D./Pasquale, Frank/Das, Sunit (eds.): The Oxford Handbook of AI. Oxford 2020, 463–474.

Chu, Ted: Human Purpose and Transhuman Potential. A Cosmic Vision for Our Future Evolution. San Rafael, 2014.

Coeckelbergh, Mark: AI Ethics. Cambridge MA 2020.

Cohen, Richard A.: Ethics and Cybernetics. Levinasian Reflections. In: Ethics and Information Technology, 2/1 (2002), 27–35. Reprinted in: Atterton, Peter/Calarco, Matthew (eds.): Radicalizing Levinas. Albany 2010, 153–167.

Cohn-Sherbok, Dan: Judaism. History, Beliefs, Practice, vol. 2: Divine Transcendence and Immanence. New York/London [2]2017.

Cole-Turner, Ron (ed.): Transhumanism and Transcendence. Christian Hope in the Age of Technological Enhancement. Washington, DC 2013.

Coyne, Lewis: Hans Jonas. Life, Technology and the Horizons of Responsibility. London 2021.

Crosston, Matthew: Cyber Colonization. The Dangerous Fusion of Artificial Intelligence and Authoritarian Regimes. In: Cyber, Intelligence, and Security 44 /1 (2020), 149–171.

De Grey, Aubrey: Radical Life Extension. Technological Aspects. In: Mather, Derek F./Mercer, Clavin (eds.): Religion and the Implications of Radical Life Extension. New York 2009, 13–24.

De Grey, Aubrey: The Curate's Egg of Anti-Anti-Aging Bioethics. In: More, Max/Vita-More, Natasha (eds.): The Transhumanist Reader. Classical and Contemporary Essays on the Science and Technology and Philosophy of the Human Future. Chichester 2013, 215–219.

De Grey, Aubrey/Rae, Michael: Ending Aging. The Rejuvenation Breakthrough that Could Reverse Human Aging in Our Lifetime. New York 2007.

Diakopolous, Nicholas: Transparency. In: Duber, Marcus D./Pasquale, Frank/Das, Sunit (eds.): The Oxford Handbook of AI. Oxford 2020, 197–214.

Dickel, Sascha/Ferwer, Andreas: Life Extension: Eternal Debates on Immortality. In: Ranisch, Robert/Sorgner, Lorenz Stefan (eds.): Post- and Transhumanism. An Introduction. Frankfurt a. M. 2014, 119–131.

Dignum, Virginia: Responsibility and Artificial Intelligence. In: Duber, Marcus D./Pasquale, Frank/Das, Sunit (eds.): The Oxford Handbook of AI. Oxford 2020, 215–231.

Donneley, Strachan: Hans Jonas and Ernst Mayr: An Organic Life and Human Responsibility. In: Tirosh-Samuelson, Hava/Wiese, Christian (eds.): The Legacy of Hans Jonas. Judaism and the Phenomenon of Life. Leiden 2008, 261–287.

Dorff, Elliot N.: Becoming Yet More Like God. A Jewish Theological, Institutional and Legal Perspective on Radical Life Extension. In Anscelovitz, Elisha S./Dorff, Elliot N./Vleeshchouwer, Amos Israel (eds.): Jewish Law Association Studies 29 (2020), 125–138.

Edelglass, William/Hatley, James/Dihem, Christian (eds.): Facing Nature. Levinas and Environmental Thought. Pittsburgh 2009.

Fagenblat, Michael: A Covenant of Creatures: Levinas's Philosophy of Judaism. Stanford 2012.

Feldstein, Steven: The Road to Digital Unfreedom. How Artificial Intelligence Is Reshaping Repression. In: Journal of Democracy 30 /1 (2019), 40–52.

Gibbs, Robert: Blowing on the Embers. Two Jewish Works of Emmanuel Levinas. A Review Essay. In: Modern Judaism 14 /1 (1994), 99–113.

Goodman, Lenn E.: Creation and Evolution. New York/London 2010.

Gunkel, David J.: The Machine Question. Critical Perspectives on AI, Robots and Ethics. Cambridge MA 2012.

Gunkel, David J.: The Symptoms of Ethics. Rethinking Ethics in the Face of the Machine. In: Human-Machine Communication 4 (2022), 67–81.

Gunkel, David J.: Thinking Otherwise. Ethics, Technology and Other Subjects. In: Ethics and Information Technology 2 (2007), 165–177.

Harari, Yuval Noah: Homo Deus. A Brief History of Tomorrow. London 2015.

Hatley, James: Generations. Levinas in the Jewish Context. In: Philosophy & Rhetoric 38 /2 (2005), 173–189.

Hurlbut, Benjamin J./Tirosh-Samuelson, Hava (eds.): Perfecting Human Futures: Transhuman Visions and Technological Imaginations. Wiesbaden 2016.

Johnson, Deborah G.: Computer Systems. Moral Entities but Not Moral Agents. In: Ethics and Information Technology 8 (2006), 195–204.

Jonas, Hans: Technology and Responsibility. Reflections on the New Task of Ethics. In: idem: Philosophical Essays. From Ancient Creed to Technological Man. Englewood Cliffs 1980 [1974], 3–20.

Jonas, Hans: The Imperative of Responsibility. In Search for an Ethics for the Technology Age. Chicago 1984.

Kalman, David Zvi: Levinas Would Have Banned Facial Recognition Technology. We Should Too, Tablet, Online http://www.tabletmag.com (posted January 11, 2021).

Kurzweil, Ray: The Age of Intelligent Machines. Cambridge MA 1990.

Lamdan, Sarah: Data Cartel: The Companies that Control and Monopolize Our Information. Palo Alto 2022.

Levinas, Emmanuel: Ethics and Infinite. Conversations with Phillip Nemo, trans. Richard A. Cohen. Pittsburgh 1985.

Liberati, Nicola/Nagataki, Shoji: Vulnerability under the Gaze of Robots. Relations among Humans and Robots. In: AI & Society 34 (2019), 333–342.

Meir, Ephraim: Buber's and Levinas's Attitudes toward Judaism. In: Atterton, Peter/Calaro, Matthew/Friedman, Maurice (eds.): Levinas & Buber. Dialogue and Difference. Pittsburgh 2004, 133–156.

Mercer, Calvin/Trothen, Tracy J. (eds.): Religion and The Technological Future. An Introduction into Biohacking, Artificial Intelligence and Transhumanism. New York 2021.

Mercer, Calvin/Trothen, Tracy J. (eds.): Religion and Transhumanist. The Unknown Future of Human Enhancement. Santa Barbara 2015.

Moravec, Hans: Mind Children. The Future of Robot and Human Intelligence. Cambridge MA 1988.

Moravec, Hans: Robots: From Mere Machine to Transcendent Mind. New York 2000.

Morris, Theresa: Hans Jonas's Ethics of Responsibility. From Ontology to Ecology. Albany 2013.

Munn, Luke: The Uselessness of AI Ethics. In: AI and Ethics (2022), online at: https://doi.org/10.1007/s436810022000209-w.

Navon, Mois: The Virtuous Servant Owner. A Paradigm Whose Time Has Come (Again). In: Frontiers in Robotics and AI 8 (2021), 1–15, article 71819.

Pearce, David: The Hedonistic Imperative, online at: http://www.hedweb.com.

Powers, Thomas M.: On the Moral Agency of Computers, Topoi 32 (2013), 227–237.

Prisco, Giulio: Transcendent Engineering. In: More, Max/Vita-More, Natasha (eds.): The Transhumanist Reader. Classical and Contemporary Essays on the Science and Technology and Philosophy of the Human Future. Chichester 2013, 234–239.

Ritchie, Barry: The (Un)Likelihood of a High-Tech Path to Immortality. In: Tirosh-Samuelson, Hava/Mossman, Kenneth L. (eds.): Building Better Humans? Refocusing the Debate on Transhumanism. Frankfurt a. M. 2012, 357–378.

Sacks, Jonathan: Morality. Restoring the Common Good in Divided Times. New York 2020.

Sacks, Jonathan: The Dignity of Difference. How to Avoid the Clash of Civilization. London 2002.

Sacks, Jonathan: The Home We Build Together. Recreating Society. London 2007.

Sacks, Jonathan: The Great Partnership. God, Science, and the Search for Meaning. London 2011.

Samuelson, Norbert M.: Judaism and the Doctrine of Creation. Cambridge 1994.

Sandberg, Anders: Morphological Imperative. Why We Not Just Want It, But Need It. In: More, Max/Vita-More, Natasha, (eds.): The Transhumanist Reader. Classical and Contemporary Essays on the Science, Technology and Philosophy of the Human Future. Malden, 2013, 56–64.

Sandler, Ronald L. (ed.): Ethics and Emerging Technologies. New York 2014.

Stahl, Bernd Carsten: Responsible Computers? A Case for Ascribing Quasi-Responsibility to Computers Independent of Personhood or Agency. In: Ethics and Information Technology 8 (2006), 205–213.

Thweatt-Bates, Jeanine: Cyborg Selves. A Theological Anthropology of the Posthuman. Burlington, 2012.

Tirosh-Samuelson, Hava: Technologizing Transcendence. A Critique of Transhumanism. In: Trothen, Tracy J./Mercer, Calvin (eds.): Religion and Human Enhancement. Death, Values and Morality. New York 2017, 267–283.

Tirosh-Samuelson, Hava: The Paradoxes of Transhumanism. Technological Spirituality or Techno-Idolatry? Theologische Literaturzeitung 146/3 (2021), 123–146.

Tirosh-Samuelson, Hava: The Pursuit of Perfection. The Misguided Transhumanist Vision. In: Theology and Science 6/2 (2018), 200–222.

Tirosh-Samuelson, Hava: The Transhumanist Pied Pipers. A Jewish Caution against False Messianism. In: Gouw, Arvin M./Green, Brian Patrick/Peters, Ted (eds.): Religious Transhumanism and Its Critics. Lanham 2022, 183–214.

Tirosh-Samuelson, Hava: Transhumanism as a Secularist Faith. In: Zygon: Journal of Religion and Science 147/4 (2012), 710–734.

Tirosh-Samuelson, Hava: On the Preciousness of Being Human. Jewish Philosophy and the Challenge of Technology. In: Tirosh-Samuelson, Hava/Hughes, Aaron W.: Jewish Philosophy for the Twenty-First Century. Personal Reflections. Leiden 2014, 428–457.

Tirosh-Samuelson, Hava/Mossman, Kenneth L. (eds.): Building Better Humans? Refocusing the Debate on Transhumanism. Frankfurt a. M. 2012.

Tirosh-Samuelson, Hava/Wiese, Christian (eds.): The Legacy of Hans Jonas. Judaism and the Phenomenon of Life. Leiden 2008.

Turner, Cody/Schneider, Susan: Could You Merge with AI? Reflections on the Singularity and Radical Brain Enhancement. In: Duber, Marcus D./Pasquale, Frank/Das, Sunit (eds.): The Oxford Handbook of Ethics of AI. Oxford 2020, 307–324.

Werner, Micah: The Immediacy of Encounter and the Danger of Dichotomy. Buber, Levinas and Jonas on Responsibility. In: Tirosh-Samuelson Hava/Wiese, Christian (eds.): The Legacy of Hans Jonas. Judaism and the Phenomenon of Life. Leiden 2008, 203–230.

Wiese, Christian: The Life and Thought of Hans Jonas. Jewish Dimensions. Newton 2007.

Vogel, Lawrence: Natural-Law Judaism. The Genesis of Bioethics in Hans Jonas, Leo Strauss, and Leon Kass. In: Tirosh-Samuelson, Hava/Wiese, Christian (eds.): The Legacy of Hans Jonas. Judaism and the Phenomenon of Life. Leiden 2008, 287–314.

Vogel, Lawrence: "The Outcry of Mute Things". Hans Jonas's Imperative of Responsibility. In: Macauley, David (ed.): Minding Nature: The Philosophers of Ecology. New York/London 1996, 167–185.

Wohl, Benjamin S.: Revealing the 'Face' of the Robot: Introducing the Ethics of Levinas to the Field of Robo-Ethics. Mobile Service Robotics (2014), 704–714, online at: https://doi.org/10.1142/9789814623353_0081.

Zoloth, Laurie: Go and Tend the Earth. Jewish View on an Enhanced World. In: Journal of Law, Medicine & Ethics 36 (1) (2008), 10–25.

Zoloth, Laurie: There Is the World and There Is the Map of the World. The Ethics of Basic Research. In: Duwell, Marcus/Rehman-Sutter, Christoph/Mieth, Dietman (eds.): The Contingent Nature of Life. Wiesbaden 2008, 307–324.

Zuboff, Shoshana: Age of Surveillance Capitalism. The Fight for a Human Future at the Frontier of Power. New York 2018.

Digitale Transformation des Unsichtbaren

Schöpfungstheologische Anmerkungen zu den Grenzen des digitalen Herstellens im Anschluss an Hannah Arendt

Lukas Ohly

Abstract

Theological reflection about digital developments has to solve the problem how to avoid the logical circle to have already embedded digital procedures. It is symptomatic in present theology to focus on the phenomenon of decisions since AI of military robots, autonomous vehicles or care robots possibly become independent of human control. The paradigm of decision-making is used also in other subjects of digital theology, for instance in the discussion about the Supper in the digital age. According to Hannah Arendt, decision-making differs categorically from thinking. I conclude that the so-called digital revolution might copy human voluntarism, but not human thinking. This article argues for holding theological frameworks which are based on thinking in Arendt's kind of view. Therefore, Arendt's term "appearing of the appearing" is a category, which reveals theological implications, which could not be copied by digital procedures.

1 Einleitung

Theologie greift derzeit die Digitalisierung vor allem in drei Hinsichten auf. Erstens werden Statusfragen diskutiert,[1] zweitens wird die Entscheidungsfähigkeit künstlich-intelligenter Sys-

[1] Vgl. Scholtz, Christopher: Alltag mit künstlichen Wesen. Theologische Implikationen eines Lebens mit subjektsimulierenden Maschinen am Beispiel des Unterhaltungsroboters Aibo. Göttingen 2008, 293. Foerst, Anne: Von Robotern, Mensch und Gott. Künstliche Intelligenz und die existenzielle Dimension des Lebens. Göttingen 2009, 194.

teme untersucht.[2] Drittens wird vor allem in der Praktischen Theologie die Anschlussfähigkeit von Digitalisierungsprozessen an die religiöse Kommunikation eruiert.[3] Während die Beiträge zu den ersten beiden Perspektiven tendenziell auf eine „Apologetik des Menschen" zielen, die seine Sonderstellung gegenüber künstlich intelligenten System retten wollen, liegt der Fokus der dritten Perspektive umgekehrt auf der Anpassungsfähigkeit religiöser Kommunikation und kirchlichen Handelns. An alle drei Perspektiven ist dabei die Frage zu adressieren, inwiefern sie eine genuin theologische Reflexion der Digitalisierung darstellen. Was wird hier verhandelt, was nicht auch die philosophische Anthropologie (bei der Statusfrage), die säkulare Ethik (bei der Entscheidungsfrage) oder die Gesellschaftstheorie (bei der Transformation kommunikativer Prozesse) bearbeitet? Solange offen bleibt, worin der theologische Charakter der Bearbeitung liegt, könnten sich die drei Perspektiven auch als Reduktionen herausstellen.

Warum wird beispielsweise ethisch die Entscheidungsfähigkeit künstlicher Systeme so stark betont, wenn die protestantische Ethik mit der iustitia passiva ansetzt, also mit der passiven Konstitution des Menschen? Müsste nicht dann in der theologischen Anthropologie die Unterscheidung zwischen Mensch und Maschine unabhängig von Eigenschaften und Fähigkeiten bestimmt werden? Und wird dann nicht möglicherweise bei der digitalen Transformationsfähigkeit religiöser Kommunikation voreilig die Frage übersprungen, welche anderen theologisch-normativen Aspekte in der christlichen Kommunikation eo ipso gebildet werden, die aber keine Entscheidungen sind? Am Beispiel des digitalen Abendmahls gefragt: Ist das digitale Abendmahl das Ergebnis kirchenleitender Entscheidungen und kreativer Ideen, die zugleich der Selbstbezüglichkeit des, nach Luther, „in sich verkrümmten Menschen" (homo incurvatus in sese ipsum)[4] entkommen, oder verdankt sich der Gestaltenwandel des Abendmahls anderer Ereignisse, die theologisch bindend sind – oder belanglos? Was also rechtfertigt die Konzentration auf Entscheidungen? In der EKD-Denkschrift „Freiheit digital" heißt es zum digitalen Abendmahl während der Corona-Pandemie: „Und doch war dies für viele in ihren Wohnzimmern eine Erfahrung

---

2 Vgl. Gräb-Schmidt, Elisabeth: Autonome Systeme. Autonomie im Spiegel menschlicher Freiheit und ihrer technischen Errungenschaften. In: Zeitschrift für evangelische Ethik 61 (2017), 163–170, 168f. Dahlmann, Anja: Militärische Robotik als Herausforderung für das Verhältnis von menschlicher Kontrolle und maschineller Autonomie. In: Zeitschrift für evangelische Ethik 61 (2017), 171–183, 176. Schwarke, Christoph: Ungleichheit und Freiheit. Ethische Fragen der Digitalisierung. In: Zeitschrift für evangelische Ethik 61 (2017), 210–221, 217.

3 Vgl. Nord, Ilona: Realitäten des Glaubens. Zur virtuellen Dimension christlicher Religiosität. Berlin u. a. 2008, 50. Grethlein, Christian: Mediatisierung von Religion und Religiosität. In: Zeitschrift für Theologie und Kirche 115 (2018), 372ff. Lienau, Anna-Katharina: Kommunikation des Evangeliums in social media. In: Zeitschrift für Theologie und Kirche 117 (2020), 489–522, 492. Schrodt, Christoph: Abendmahl: digital. Alte und neue Fragen – nicht nur in Zeiten der Pandemie. In: Zeitschrift für Theologie und Kirche 118 (2021), 495–515, 505.

4 Vgl. Härle, Wilfried: Der Glaube als Gottes- und/oder Menschenwerk in der Theologie Martin Luthers. In: MJTh 4/1992, 37–77, 46.

der geistgewirkten Gegenwart Christi."[5] Theologisch rückgefragt: War denn „dies" *wirklich* eine Erfahrung der geistgewirkten Gegenwart Christi oder nur der Schein einer solchen? Führt also die Entscheidung, die eigene Erfahrung zum Maßstab der Geistesgegenwart zu erheben, schon zur Geistesgegenwart? Letztlich steckt dahinter ein zirkulärer Verweis auf Verantwortung: Christ*innen rechtfertigen ein digitales Abendmahl über ihre Entscheidung, weil Entscheidungen der Rechtfertigung bedürfen. Über den Entscheidungsaspekt reduzieren sie damit das, was gerechtfertigt werden kann, auf Verantwortung.

Soweit ich sehe, wird die Diskussion um das digitale Abendmahl in der Regel so geführt, dass man zunächst bestimmt, was das Abendmahl *eigentlich* ist, und dann anschließend seine Übersetzung in neue Situationen vornimmt, die vom Kern des Abendmahls nicht abweicht. Die Hauptfrage lautet dann, worin dieser Kern besteht, was also seine Eigentlichkeit ausmacht. Dieselbe Strategie zeigt sich auch, wenn man den Status einer Künstlichen Intelligenz verhandelt oder Kennzeichen von Entscheidungen bestimmt. Ist also das Wesen des Menschen auf Computer übertragbar? Oder was sind Entscheidungen, und lässt sich ihr Wesen auf künstlich intelligente Maschinen übertragen? Von einem angeblichen Kern des Eigentlichen her stellt sich aber die Frage der Transformation gar nicht. Sie kommt also nur dann auf, wenn sie entweder durch empfindliche äußere Faktoren nahegelegt wird oder wenn es ein Transformationsinteresse gibt. Im letzteren Fall müsste geklärt werden, dass bereits das Transformationsinteresse dem Eigentlichen entspricht, im ersten Fall würde seinerseits eine Entscheidung zwischen Kern und Übersetzung vermitteln. Hier wie dort liegt die Betonung auf der Entscheidung, um das Transformationsproblem zu lösen. Der eigentliche Kern spielt dann nur die Rolle einer Folie, auf der die Entscheidung abgebildet wird, die die Rolle des Eigentlichen übernimmt.

Mir scheint, dass theologische Bearbeitungen mit dieser Konzentration selbst zu technischen Verfahren mutieren. Indem festgelegt wird, dass die Vermittlungsinstanz, um einen Prozess ins Laufen zu bringen, eine Entscheidung ist, wird dieses Mittel für einen Zweck stilisiert, der wiederum selbst aus dem Mittel bestimmt wird. Konkret: Die kirchenleitende Entscheidung legt fest, dass ab sofort digitales Abendmahl dem eigentlichen Abendmahl entspricht. Das digitale Abendmahl ist dabei das Mittel, das wiederum zirkulär durch die Entscheidung vermittelt ist, da es ja auf einer Entscheidung beruht, dass es dem eigentlichen Abendmahl entspricht.[6] Dasselbe trifft auf die anderen beiden Fragestellungen zu: Eine Künstliche Intelligenz ist oder ist keine Person, weil Interessen dahinterstehen, die die Entscheidung zum Mittel machen, um die fragliche Personalität zum Mittel der Interessen einzusetzen.[7] Ebenso ist

---

5 Evangelische Kirche in Deutschland: Freiheit digital. Die Zehn Gebote in Zeiten des digitalen Wandels. Eine Denkschrift der Evangelischen Kirche in Deutschland. Leipzig 2021, 87.

6 Vgl. Schrodt: Abendmahl: digital, 511.

7 Vgl. Foerst: Von Robotern, Mensch und Gott, 183.

Künstliche Intelligenz entscheidungsfähig oder eben nicht, weil hier eine Entscheidung über die Entscheidungsfähigkeit beide zum Mittel füreinander bestimmt (s. die Diskussion um Dilemma-Situationen beim autonom fahrenden Auto).

Laut Hannah Arendt zeigt sich an der Gewalt, dass Mittel die Steuerung übernehmen, wenn sich Ziele nicht schnell erreichen lassen.[8] Gewalt bestimmt dann auch die Zwecke,[9] wobei aufgrund des instrumentalen Charakters der Gewalt[10] die Zwecke im Rahmen der Mittel verbleiben. Ich lasse hier offen, ob das auch auf die obigen Beispiele zutrifft, ob also Gewalt dadurch gekennzeichnet ist, dass die Mittel die Ziele bestimmen und ob die obigen theologischen Verfahrenstechniken Gewaltmittel sind. Mir geht es hier nur darum, dass dieses Phänomen der zirkulären Verweise von Mitteln auf Mittel den Entscheidungsaspekt zur maßgeblichen Instanz von Problemlösungen erhebt. Ziele werden so durch Mittel übergangen, und zwar durch den Modus ihrer geistigen Bearbeitung, nämlich durch Entscheidungen, die den fraglichen Prozess in Gang bringen.

Welchem Geistesvermögen entspricht eine Entscheidung, und welche anderen Geistesvermögen könnten sich der Digitalisierung zuwenden? Nach Arendt sind die wichtigsten drei Geistesvermögen das Denken, das Wollen und das Urteilen. Alle drei sind nicht aufeinander reduzierbar.[11] Damit könnte sich andeuten, dass Entscheidungen eine Reduktion von Geistesvermögen darstellen. Zwar kommen in deliberativen Verfahren alle drei Geistesvermögen zum Tragen. Daraus folgt aber nicht, dass eine Entscheidung logisch aus der deliberativen Berücksichtigung des Denkens oder Urteilens erzwungen werden kann.

Von Arendt her scheint der theologische Diskurs um die Digitalisierung zunächst aus einer Technikfalle herausgeholt werden zu müssen, die darin besteht, theologische Probleme zu instrumentalen Verfahrensregelungen der Entscheidbarkeit zu transformieren. Könnte es sich aber nicht stattdessen bei den theologischen Fragen zur Digitalisierung um Themen handeln, die sachgemäß nur dem Denken oder dem Urteilen zugänglich sind, nicht aber der Entscheidung, die auf dem Wollen beruht? Oder muss der Begriff des Wollens über die Entscheidung der Mittel[12] hinaus entfaltet werden?

Noch eine Einschätzung zum derzeitigen Digitalisierungs-Diskurs: Die Haupt*quellen* der theologischen Bearbeitung sind Forschungsprogramme, insbesondere von Nachwuchswissenschaftler*innen, die sich in verschiedenen Netzwerken verbinden und austauschen. Dadurch ergeben sich sowohl Redundanzen als auch Fragmentierungen. Fertige Konzeptionen zum

8 Vgl. Arendt, Hannah: Macht und Gewalt. München ^{27}2019, 79.

9 Vgl. ebd., 56.

10 Ebd., 47.

11 Vgl. Arendt, Hannah: Vom Leben des Geistes. Das Denken. Das Wollen. München ^{10}2020, 75 f.

12 Vgl. ebd., 296.

geisteswissenschaftlichen Anspruch der Digitalität sind rar – ein Grund, warum in vielen Forschungsprogrammen transhumanistische Autor*innen bearbeitet werden. Dies liegt nämlich nicht daran, dass die evangelisch-theologische Anthropologie wieder in den Mittelpunkt gerückt worden wäre, um die es seit Pannenbergs einschlägiger Studie im deutschsprachigen Raum weitgehend ruhig geworden ist.[13] Vielmehr liegt mit den Prognosen der, wenn auch politisch bedeutsamen, aber doch wissenschaftlichen Minderheitenmeinung des Transhumanismus ein umfängliches Quellmaterial vor, das sich mit den klassischen Methoden der Textinterpretation bearbeiten lässt.

Demgegenüber halte ich Hannah Arendts Werk für einen bedeutsamen Kontrapunkt in dieser Debatte, weil sie weitsichtig die politischen Entwicklungen der Moderne mit den technologischen Automationen verknüpft hat. Einschätzungen zur Künstlichen Intelligenz, BIG DATA und Robotik hatte sie schon bis zu ihrem Tod im Jahr 1975 abgegeben und in einen weiten politischen Horizont gestellt. Zudem scheinen sich mit ihrem Werk blinde Flecken eines Diskurses aufdecken zu lassen, den selbst schon Muster digitaler Schlussverfahren kolonialisiert haben. Diese Einleitung beruht bereits auf Arendts theoretischen Grundlagen.

In diesem Artikel möchte ich einen theologischen Rahmen für die Behandlung der Digitalisierungsthematik skizzieren, der über eine technische Vermittlung über Entscheidungen hinausgeht. Die Frage des moralischen Status von Künstlicher Intelligenz oder der Transformationsfähigkeit menschlicher Phänomene in die Mensch-Maschine-Interaktion soll in einen umfassenderen Rahmen gestellt werden, der Reduktionismen ausschließt, die darin bestehen, dass selbst schon „digital gedacht" wird, um über digitale Phänomene theologisch zu befinden. Ob man Theologie als Verstehensdisziplin auffasst,[14] als Reflexion des Glaubens,[15] als Auslegung der Selbstoffenbarung Gottes,[16] als Glaubensexplikation[17] oder als Überprüfung der Implikationen von Glaubensaussagen[18] – in jedem Fall wird das Denken oder Urteilen in Arendts Sinn bemüht. Zwar könnte man einwenden, dass dazu über die jeweiligen Denkparadigmen entschieden werden muss. Dennoch macht es einen Unterschied, ob sich die Entscheidungen am theologischen Denken bewähren oder ob umgekehrt das Denken einem standardisierten Verfahren der formalen Mustererkennung folgt. Im letzteren Fall wird Entscheidung zum Modus digital-theologischer Verfahrensfindung erhoben,

---

13 Vgl. Pannenberg, Wolfhart: Anthropologie in theologischer Perspektive. Göttingen 1983.

14 Vgl. Bultmann, Rudolf: Glauben und Verstehen Bd. 3. Tübingen 1960, 33.

15 Vgl. Härle, Wilfried: Dogmatik. Berlin/New York 1995, 12 f.

16 Vgl. Barth, Karl: Kirchliche Dogmatik II/2. Zürich ^{3}1948, 36.

17 Vgl. Dalferth, Ingolf U.: Kombinatorische Theologie. Probleme theologischer Rationalität. Freiburg/Basel/Wien 1991, 77.

18 Vgl. Pannenberg, Wolfhart: Wissenschaftstheorie und Theologie. Frankfurt a. M. 1987, 335.

was sich etwa an den materialethischen Beiträgen zur Robotik zeigt, die primär die Entscheidungskompetenz autonomer Maschinen in den Blick nehmen. Hier scheint die theologische Beschreibung selbst von digitalen Mustern erfasst worden zu sein. Demgegenüber weitet Arendt den Blick, um unabhängiger zu erfassen, wie das Menschsein von der Digitalität betroffen wird.

Arendt weist einen theologischen Anspruch ihrer Schriften zurück. Es geht ihr vielmehr um eine politische Befähigung des Menschen.[19] Allerdings lässt es nicht nur ihre Beschäftigung mit christlichen Denkern (insbesondere Paulus und Augustin), sondern auch ihre Beschreibung von Phänomenen zu, ihre Beschreibungen zu retheologisieren, wie ich in diesem Beitrag zeigen möchte. Wenn der Mensch als Anfang des Anfangs beschrieben wird,[20] der in Analogie zur Schöpfung Neues in diese Welt bringt,[21] so wird hier gerade keine Grundlage für ein politisches Souveränitätsmodell im Sinne Carl Schmitts geschaffen („Souverän ist, wer über den Ausnahmezustand entscheidet"[22]), sondern eine ereignishafte Spur gelegt, die Passivität und Aktivität transzendiert. Der Mensch ist geboren (passiv) und daher zeitlebens zum Anfangen fähig. Ist „Anfangen" aber eine Aktivität, die einen Urheber hat? Oder gehört es zu den Verben, die zwar ein bestimmtes Subjekt haben mögen, aber keine Aktivität ausdrücken – so wie „aussehen", „geschehen", „sich ereignen" oder „scheinen"? Und aus welcher Perspektive wird dann auf das politische Modell geschaut, das sich aus dieser Bestimmung des Menschen als Anfang des Anfangs ergibt? Oder anders: Was heißt es für dieses politische Modell, wenn darin schöpfungstheologische Einsichten eingetragen werden?

Ich möchte zeigen, dass mit diesen schöpfungstheologischen Momenten ein spezifisches Verhältnis von Unsichtbarkeit und Sichtbarkeit bearbeitet wird. Auch wenn ich Arendt nicht durchgängig zustimme, wie sie dieses Verhältnis bestimmt, lässt sich mit ihrer Analyse erkennen, wie es sich durch Digitalisierung verändert. Ich will die These belegen, dass der Unterschied zwischen Unsichtbarkeit und Sichtbarkeit, den ich im Folgenden an Arendt rekonstruieren möchte, kategorial ist, aber durch digitale Prozesse in einen qualitativen Unterschied transformiert wird. Was diese Transformation theologisch bedeutet, soll dabei ermessen werden.

---

19 Vgl. Arendt: Vom Leben des Geistes, 429.

20 Vgl. Arendt: Vita activa oder Vom tätigen Leben. München [20]2019, 216.

21 Vgl- Arendt: Vom Leben des Geistes, 442 f.

22 Schmitt, Carl: Politische Theologie. Vier Kapitel zur Lehre von der Souveränität. Berlin [8]2004, 13.

2 Der Verdacht gegen das Sichtbare

Es gibt Dinge, die ein Recht auf Verborgenheit haben, während andere Dinge erst dadurch zu ihrem Recht kommen, dass sie ans Licht kommen.[23] Es fällt auf, dass Arendt den Rechtsbegriff metaphorisch auf Dinge bezieht, aber wohl nur deshalb, weil sie Menschendinge sind, weil sie also als Phänomene für Menschen sichtbar oder unsichtbar sein müssen. Es werden also eigentlich Rechte der Menschen angetastet, wenn diese Ordnung nicht beachtet wird. Arendt drückt das am Beispiel der Arbeit aus, die eigentlich zur Erfüllung des Lebensnotwendigen benötigt wird und deshalb im Verborgenen bleiben sollte. Wird sie dagegen in die Öffentlichkeit gezerrt wie etwa in Marx' Theorie der Arbeit, so entsteht Scham.[24]

Das Bedeutsame findet sich gerade an der Oberfläche.[25] Dagegen hat der produzierende Mensch die Sehnsucht, auch Unsichtbares in die Öffentlichkeit zu zerren,[26] auch mit Hilfe der Verarbeitung enormer Daten.[27] Was aber ins Licht gedrängt wird, verliert seinen Erscheinungscharakter.[28] An seine Stelle tritt der Schein, und zwar der Schein des Ungeordneten und Gleichen.[29] Als Beispiel nennt Arendt die menschlichen Organe, die trotz ihrer unterschiedlichen biologischen Funktion für das Auge gleich aussehen,[30] und die Wurzeln eines Baumes bei Sartre, die Ekel auslösen.[31] In Zeiten der Digitalisierung erzeugen die Aufdeckung privater Chatnachrichten oder Cybermobbing bei den Betroffenen tiefe Scham und bei den Rezipienten Fassungslosigkeit, wobei gerade das Schmutzige und Private ins Licht der Öffentlichkeit gelangt, das alle Menschen einander angleichen würde, wenn sie es gleichermaßen voreinander zeigen würden. BIG DATA ist genau diese Drohung, alle Menschen voreinander gleich zu machen, indem Elemente der gleichen Art (Daten) nach den gleichen Regeln ausgelesen werden. Doch es bleibt lediglich bei dieser Drohung, da Internetkonzerne ihre Algorithmen als Firmengeheimnisse zurückhalten und auch tyrannische politische Systeme sich nicht in die Karten schauen lassen, wie sie ihre Bevölkerung digital ausspähen. Hannah Arendt schreibt, dass beim neuzeitlichen Paradigma des experimentierenden Herstellens lediglich die Daten gezeigt werden, nicht aber der Prozess, der sie generiert, und auch nicht das Sein, das sich im Erscheinen zeigt: Es liegt „im Wesen des Prozesses, daß er selbst unsichtbar bleibt, daß

23 Vgl. Arendt: Vita activa, 90.
24 Vgl. ebd., 89.
25 Vgl. Arendt: Vom Leben des Geistes, 37.
26 Vgl. Arendt: Vita activa, 157.
27 Vgl. Arendt: Vom Leben des Geistes, 381, 387.
28 Vgl. ebd., 36.
29 Vgl. ebd., 38.
30 Vgl. ebd.
31 Vgl. ebd., 148.

sein Vorhandensein nur aus bestimmten Daten, die nicht eigentlich mehr Phänomene sind, erschlossen werden kann".[32] Wir werden sehen, dass dieser unsichtbare Prozess in kontradiktorischen Widerspruch tritt zu Arendts Vorstellung, dass das Bedeutsame an der Oberfläche zu finden ist.

Zu dieser Bemerkung steht in Spannung, dass das Äußere „nur eine einzige Funktion hat, nämlich die, das Innere zu verdecken".[33] Denn zum einen bestimmt Arendt damit das Äußere zum Mittel für einen inneren Zweck, was der Aussage widerspricht, dass sich das Bedeutsame gerade an der Oberfläche befindet. Zum anderen lässt sich dann über das Äußere nichts anderes aussagen als über sein Verhältnis zum Inneren, das es verdeckt. Nun ist es phänomenologisch trivial, dass Erscheinungen immer auch verdecken. Wenn genau darin ihre Funktion besteht, dann kommt das Verdeckte gerade dadurch zu seinem Recht und wird in seinem Recht anerkannt, dass es verdeckt ist. Es wird dann entweder gar nicht oder als Verborgenes offenbar. Unsichtbares und Sichtbares gehören dann zusammen,[34] weil das Unsichtbare zu seinem Recht kommen soll, indem es als Unsichtbares auftritt – und dazu wird das Sichtbare benötigt, das das Unsichtbare verdeckt und es so als Unsichtbares auszeichnet.

Nehmen wir an, Arendt übertreibt nicht mit dieser ausschließlichen Funktionsbestimmung des Äußeren. Wie kann sich dann zugleich das Bedeutsame an der Oberfläche befinden? Das Bedeutsame besteht dann darin, diesen Prozess an der Oberfläche bemerkbar zu machen, wie das Sichtbare das Unsichtbare konstituiert. Dieser Konstitutionsprozess befindet sich dann ebenso an der Oberfläche, wenn auch nicht als dinglich Äußeres. Doch andernfalls könnte das Unsichtbare auch nicht als Unsichtbares offenbar werden. Arendt spricht von einem uneigentlichen Schein, der darin besteht, dass das *Erscheinen erscheint*.[35] Während ein eigentlicher Schein wieder verschwindet, gehört zum Erscheinen wesentlich das Erscheinen des Erscheinens. Es ist zugleich uneigentlicher Schein, weil es nicht nur sichtbare Dinge hervorbringt, die erscheinen, sondern das Unsichtbare, das dieser Prozess an der Oberfläche ist. Unsichtbare Dinge lassen sich also deshalb bemerken, weil jegliches Erscheinen auf dem Erscheinen des Erscheinens gründet und damit auf dem unsichtbaren Prozess, der *beides* konstituiert, Sichtbares und Unsichtbares. Unsichtbare Dinge sind verborgen, das Unsichtbare des Erscheinens des Erscheinens ist dagegen (uneigentlicher) Schein. Das Verdecken des Unsichtbaren ist dann Schein und nicht Verborgenheit.

Bei diesem Ausdruck „Erscheinen des Erscheinens" wird nun eine kategoriale Differenz vorgenommen. Das zweite „Erscheinen" kann nicht an die erste Stelle rücken. Denn hier erschei-

---

[32] Arendt: Vita activa, 378.

[33] Arendt: Vom Leben des Geistes, 39, vgl. 45.

[34] Vgl. ebd., 114.

[35] Vgl. ebd., 48.

nen *Dinge* oder Sachverhalte, während das „Erscheinen" an der ersten Stelle das *Erscheinen* von Dingen oder Sachverhalten zur Erscheinung bringt. Erscheinende Dinge sind sichtbar, ihr Erscheinen dagegen ist nur durch sie sichtbar, dann aber nicht als Erscheinen, sondern als die Dinge, die sie sind. Das Erscheinen des Erscheinens ist folglich selbst unsichtbar, kommt aber im Erscheinen der Dinge mit an die Oberfläche.

Arendt befürchtet nun, dass Automatismen diesen Prozess zerstören, bei dem unsichtbare und sichtbare Dinge zu ihrem Recht kommen. Die Automation zerrt nämlich alles ans Licht. Zugleich aber verbirgt sich der Prozess, zu dessen Wesen gehört, „daß er selbst unsichtbar bleibt". Damit wird der dialektische Prozess des Erscheinens des Erscheinens zerstört, der zwar auch unsichtbar (uneigentlicher Schein), aber an der Oberfläche zu finden ist.

Der Materialismus experimentiert mit Rechenanlagen, Kybernetik und Automatisierung, um den Geist mit dem Gehirn zu identifizieren.[36] Das Bild, wonach die ganze Menschheit ein Riesengehirn hat,[37] wird von den Phantasien der BIG DATA-Verarbeitung bestätigt, die alle Prozesse in ein Gesamtmuster integriert. Die Automation stellt Produkte her, die nicht mehr in einem natürlichen Wachstum begriffen sind, bei dem also Erscheinen erscheint, sondern die erst dann anfangen zu existieren, wenn sie fertiggestellt sind und die Fabrik verlassen.[38] Durch die Automation könnte die Kunst zugrundegehen,[39] und zwar obwohl es inzwischen Apps gibt, die je nach Themenwahl neue Bilder aus einer Datenbank von Gemälden synthetisch herstellen. Doch die Kunst verdankt sich dem *Denken*,[40] während die Automation ihre Ergebnisse *herstellt*.[41] Ihr Mittel ist das *Errechnen*.[42]

Interpretiert man diese Beobachtungen vom Verhältnis von Sichtbarkeit und Unsichtbarkeit her, so ersetzt das Herstellen[43] den Denkprozess, der sich nämlich im Verborgenen vollzieht.

---

36 Vgl. ebd., 424.

37 Vgl. ebd. Diese Metapher berührt sich mit Arendts Beschreibung der Konzentrationslager, die sie als Laboratorien für Menschen beschreibt, „als ob sie zusammen nur einen einzigen Menschen darstellten" (Arendt, Hannah: Elemente und Ursprünge totaler Herrschaft. Antisemitismus, Imperialismus, totale Herrschaft. München [21]2019, 907).

38 Vgl. Arendt: Vita activa, 177 f.

39 Vgl. ebd., 155.

40 Vgl. ebd., 203.

41 Vgl. ebd., 207 f.

42 Vgl. Arendt: Vom Leben des Geistes, 11.

43 Und die Arbeit! Herstellen und Arbeit sind für Arendt zwar zwei verschiedene Modi der menschlichen Tätigkeit, werden aber durch die Automation zu zwei Seiten einer Medaille: Einerseits kann sie den Menschen von der Naturnotwendigkeit befreien, die in der Arbeit besteht (vgl. Arendt: Vita activa 154), andererseits wandelt sie mit dem Herstellen neuer Produkte auch die Natur um und ersetzt die Welt (vgl. ebd., 176, 180). In der Konsequenz wird die Arbeit aus der Unsichtbarkeit in den Mittelpunkt gestellt (vgl. ebd., 399 f.).

Vom unsichtbaren Prozess des fabrizierenden Herstellens unterscheidet sich der Denkprozess dadurch, dass hier die Verborgenheit an der Oberfläche bleibt. Man sieht einer Person unmittelbar an, wenn sie denkt.[44] Wenn dagegen auf dem Display eines Computers „Bitte warten …" erscheint, ist der verborgene Rechenprozess nur aufgrund einer sozialen Konvention identifizierbar.

Phänomenologisch ist zu berücksichtigen, dass Arendt in der Beschreibung des Denkens von der Selbst- zur Alteritätsperspektive hin- und herschwankt. So ist es aus der Denkperspektive falsch zu behaupten: „Das Denken zieht sich radikal von dieser Welt und ihrer Datenhaftigkeit zurück".[45] Die „gewöhnliche Erscheinung der Geistesabwesenheit, die man bei jedem beobachten kann, der gerade von irgendwelchen Gedanken in Anspruch genommen ist",[46] kann ja nur aus der Sicht einer Beobachterin ausgesagt werden und nicht der des Denkers. Doch genau diese intersubjektive Spannung ist zu berücksichtigen. Das „*Sicht-Zurückziehen* aus der Welt, wie sie erscheint" bleibt daran gekoppelt, dass wir „*von dieser Welt und nicht bloß in dieser Welt [sind]*; wir sind selbst Erscheinungen".[47] Ebenso wie der Mensch für Arendt nur in der Pluralität existiert,[48] so kann es das Denken wohl nur geben in der Resonanz auf andere Menschen. Zwar ist ein Denker mit sich allein (solitude), aber gerade nicht einsam (loneliness).[49] Er führt vielmehr mit sich ein Zwiegespräch, er ist „bei sich".[50] Zwar will Arendt das innere Zwiegespräch nicht als Modell für die zwischenmenschliche Pluralität gelten lassen.[51] Aber das Umgekehrte scheint für Arendt zuzutreffen, nämlich dass die Erscheinung des Denkers für andere an der Oberfläche wesentlich ist für das Denken. „Die geistigen Tätigkeiten, die definitionsgemäß nicht erscheinen, finden in einer Welt der Erscheinungen und in einem Wesen statt, das an […] seiner Fähigkeit und seines Bedürfnisses, anderen zu erscheinen, teilhat".[52] Es ist die *gemeinsame* Teilhabe der Menschengemeinschaft am *Erscheinen des Erscheinens*, die das Denken ermöglicht, also gerade keine solipsistische „Fähigkeit", mit dem Denken anzufangen, sondern der gesunde Menschenverstand einer gemeinsamen Welt,[53] der sensus

---

44 Vgl. Arendt: Vom Leben des Geistes, 62, 78.

45 Ebd., 65.

46 Ebd., 62.

47 Ebd., 32, Herv. H. A.

48 Vgl. ebd., 184.

49 Vgl. Arendt, Hannah: The Life of the Mind. The Groundbreaking Investigation on How We Think. Florida 1978, 74. Dies.: Responsibility and Judgment. (Hg. J. Kohn). New York 2003, 98. Die deutschen Übersetzungen tauschen diese Begriffe gelegentlich aus.

50 Arendt: Vom Leben des Geistes, 80.

51 Vgl. ebd., 427.

52 Vgl. ebd., 81.

53 Vgl. Arendt: Elemente und Ursprünge totaler Herrschaft, 41.

communis oder Gemeinsinn, der sowohl ein Sinn zur Koordination der fünf Sinne ist als auch ein Sinn der zwischenmenschlichen Gemeinsamkeit,[54] also in beiden Fällen Integrationskraft in eine Welt der Erscheinungen hat.[55]

Das Denken beruht also auch auf dem Erscheinen des Erscheinens. Es erscheint dem Zuschauer als Unterbrechung[56] und somit als Verborgenes. Es offenbart sich in der Mitteilung des Gedachten – sei es im Kunstwerk[57] oder in der sprachlichen Mitteilung, die für Arendt aber auch kunstfertig sein muss,[58] nämlich metaphorisch zum Ausdruck bringen muss, was als Denken dem Sichtbaren entzogen ist.[59] Umgekehrt liefert die Metapher dem Denken die Anschauung.[60] Die Fähigkeit zuzuschauen stellt damit dem Denken ein Instrument zum Sprechen zur Verfügung. Das Denken, obwohl es sich immer nur allein vollzieht, gründet also in der Intersubjektivität: „Descartes' ‚Cogito me cogitare, ergo sum' ist einfach nicht schlüssig, weil diese res cogitans überhaupt nicht erscheint, wenn sich ihre cogitationes nicht in gesprochener oder geschriebener Sprache äußern."[61] Selbst wenn es nicht die Zuschauerin, sondern der Denker ist, der sich mitteilt, so bedarf er der Zuschauerin, um sich auszudrücken: „Es gibt in dieser Welt nichts und niemanden, dessen bloßes Sein nicht einen *Zuschauer* voraussetzte."[62]

Sowohl beim künstlerischen wie beim sprachlichen Ausdruck bleibt das Erscheinen des Erscheinens ein intersubjektives Phänomen. Obwohl selbst unanschaulich, bringt es zur Anschauung. Zuschauer gibt es nur im Plural,[63] und dem Zuschauen bietet sich die Wahrheit dar, das Ewige im Unterschied zu den erscheinenden Dingen.[64]

Ich habe behauptet, dass die Automation das Denken durch das Herstellen ablöst. Was ändert sich dadurch am Prozess des Erscheinens des Erscheinens? Man könnte ja einwenden, dass Künstliche Intelligenz durch Kontrollschleifen eine Analogie zur zwischenmenschlichen Pluralität bildet: Die Funktionsfähigkeit eines Chips wird durch die Berechnung eines anderen Chips überprüft, so dass aneinandergekoppelte Prozesse aufeinander wirken, ohne die „Autonomie" des jeweiligen Prozesses zu verletzen. Zudem wird bei der Künstlichen Intel-

54 Vgl. Arendt: Vom Leben des Geistes, 59. Dies.: Vita activa, 359.

55 Vgl. Arendt, Hannah: Das Urteilen. München ^{6}1921, 109 f.

56 Vgl. Arendt: Vita activa, 31.

57 Vgl. ebd., 206.

58 Vgl. ebd., 205.

59 Vgl. Arendt: Vom Leben des Geistes, 111.

60 Vgl. ebd., 108.

61 Ebd., 30.

62 Ebd., 29, Herv. H. A.

63 Vgl. ebd., 99. Dies.: Das Urteilen, 70, 99.

64 Vgl. Arendt: Vom Leben des Geistes, 33. Dies.: Vita activa, 31.

ligenz durch Datenprotokolle transparent, was auch beim Denken an der Oberfläche liegt. Zweifellos sind hier Herstellungsprozesse beschrieben, und kein einziger Prozess muss dabei für die Automation bewusst verlaufen. Aber erscheint nicht trotzdem auch das Erscheinen für die Künstliche Intelligenz, wenn sie allmähliche und minimale Veränderungen im System frühzeitig registrieren kann, früher noch, als das Erscheinen uns Menschen erscheinen kann? So kann ein Gesichtsscanner die Emotion einer Person und auch den emotionalen Wechsel in Echtzeit erkennen, wofür Menschen bisweilen länger brauchen.

Im Einwand zeigt sich bereits die Transformation der kategorialen Unterscheidung im Ausdruck „Erscheinen des Erscheinens" in eine qualitative. Denn das Erscheinen des Erscheinens hat nichts damit zu tun, wie früh es einem Menschen auffällt. Es ist nicht so, dass einem Menschen zunächst das Erscheinen des Erscheinens auffallen muss, damit ihm auch ein Ding auffallen kann. Meistens ist es sogar umgekehrt, und das zeichnet seine Unsichtbarkeit aus, nämlich dass das Erscheinen des Erscheinens gar nicht auffällt, und wenn doch, dann nachträglich. Arendt folgt Platons Sicht, dass man am Unsichtbaren nur durch sichtbare Formen teilhat.[65] Der unsichtbare Sinn enthüllt sich erst nach dem Verschwinden.[66] Die Zuschauer sehen das Ganze im Gegensatz zur Akteurin, die nur ihre Sicht hat.[67] Und in einer Parallele zu Joh 3,8 zitiert Arendt Xenophon: „Die Winde selbst sind unsichtbar, doch ihre Wirkungen zeigen sich uns, und wir spüren irgendwie ihre Berührung."[68] Wie will die Automation das Erscheinen des Erscheinens registrieren, das selbst gar kein Datum ist, sondern nur rückwirkend aus Daten erschlossen wird?

Sie muss dazu diese kategoriale Differenz in eine qualitative Gleichheit überführen. Und sie leistet diese Transformation, indem sie das Erscheinen des Erscheinens zu einem zeitlichen Vorrang erhebt. Die Künstliche Intelligenz kann dann registrieren, ob sich Personen sympathisch sind, noch bevor sie sich ineinander verlieben, und Aktienbewegungen ausmachen, noch bevor gehandelt wird. Zum Vergleich: Menschen merken in der Regel erst, nachdem sie sich verliebt haben, dass sie verliebt sind, und kaum in „Echtzeit". Und der Schreck über den dramatischen Kursverlust auf dem Aktienmarkt kann sogar Monate später noch wiederkehren, wenn sich der Markt inzwischen wieder erholt hat. Was jedoch für Arendt das Ewige ist im kategorialen Unterschied zur Zeit,[69] das wird in der künstlich-intelligenten Transformation auf einer zeitlich-logischen Linie gefasst. Auch hier muss zwar etwas unsichtbar bleiben, um etwas sichtbar werden zu lassen: Indem das jeweilige Datum generiert wird, verschwindet

---

65 Vgl. Arendt, Hannah: Über das Böse. Eine Vorlesung zu Fragen der Ethik. München ^{9}2014, 64.

66 Vgl. Arendt: Vom Leben des Geistes, 134.

67 Vgl. ebd., 99. Dies.: Das Urteilen, 107.

68 Arendt: Vom Leben des Geistes, 174.

69 Vgl. ebd., 91.

der Prozess der Generierung. Es liegt eben „im Wesen des Prozesses, daß er selbst unsichtbar bleibt." Doch indem sich die Sichtbarkeit des Prozesses und seiner Daten ausschließen, wird der Unterschied zwischen Unsichtbarkeit und Sichtbarkeit allein auf der qualitativen Ebene fassbar, nämlich als Zeitlichkeit.

3 Die Verborgenheit des Guten

Nun kann man fragen, warum Menschen an der kategorialen Unterscheidung hängen sollten, wenn doch das frühzeitige Registrieren von Veränderungen viel effektiver ist, um Gefahren abzuwenden oder gesellschaftlichen Nutzen zu erzielen. Aber diese Frage verrät die Perspektive des Herstellens. Sie ist nach dem Zweck formuliert und blendet das aus, was Arendt den Sinn nennt. Vom Standpunkt des Denkens stellt sich diese Frage gar nicht. Handeln, Sprechen und Denken sind „‚unproduktiv', sie bringen nichts hervor, und als Tätigkeiten sind sie so flüchtig wie das Leben selbst".[70] Das ist nicht einmal eine ethische Aussage zur Rettung des Denkens, weil diese Aussage sich dann wieder einem Zweck unterstellen würde. Vielmehr soll lediglich der kategoriale Gegensatz zwischen Denken und Herstellen unterstrichen werden.

Trotzdem trifft Arendt damit eine Aussage zum guten Leben.[71] Nur wird der Begriff des Guten dabei nicht auf das Ethische reduziert. Auch das Gute bedarf der Verborgenheit.[72] Aus Jesu Forderung, die linke Hand solle nicht wissen, was die rechte tut (Mt. 6,3), folgert Arendt: „Ich muß sozusagen von mir selbst abwesend sein".[73] Sobald dagegen das Gute erscheint, löst es Selbstzweifel aus.[74] „Lebe in der Verborgenheit, auch vor dir selbst, und bemühe dich nicht, gut zu *sein*".[75]

Das Denken ist also nicht aus ethischen Gründen zu schützen, denn es bringt keine gute Tat hervor, aber es bewahrt Menschen durchaus davor, böse zu werden.[76] Das neuartige Phänomen der Banalität des Bösen, das Arendt an Adolf Eichmann rekonstruiert[77] und Jahre zuvor schon an Heinrich Himmler skizziert hatte,[78] besteht in der Unfähigkeit oder Verweigerung zu den-

---

70 Arendt: Vita activa, 113.

71 Vgl. ebd., 262.

72 Vgl. ebd., 90 f.

73 Arendt: Über das Böse, 109.

74 Arendt: Vom Leben des Geistes, 302.

75 Ebd., Herv. H. A.

76 Vgl. ebd., 15.

77 Vgl. ebd., 176 f. Dies.: Eichmann in Jerusalem. Ein Bericht von der Banalität des Bösen. München [17]2021, 371.

78 Vgl. Arendt: Elemente und Ursprünge totaler Herrschaft, 722.

ken.[79] Ihr kann ethisch nicht beigekommen werden, weil das Denken vor ihr versagt.[80] Man kann also niemandem, der das Denken verweigert, erklären, warum Denken seinem Verhalten ethisch überlegen ist. Man muss vielmehr den Gemeinsinn dafür bereits voraussetzen, um das verborgene Gute darin zu finden.

„Der weitere Fortschritt in den Geisteswissenschaften schließlich [könnte] mit der Zerstörung des geistigen Gutes enden."[81] Dieses Zitat steht im Kontext zum Zustand der Geisteswissenschaften und der Unmöglichkeit, dass sie dem Fortschrittsparadigma genügen,[82] weil sie sonst in Pseudo-Wissenschaften umschlagen, die sich selbst zerstören.[83] Der Kontext belegt, dass sich Arendt hier auf die Entwicklung der Rechenleistung von Computern bezieht. Aus den Geisteswissenschaften wird eine Pseudo-Wissenschaft, weil sie „Data gebiert, deren hypothetischer Charakter vergessen ist".[84] An die Stelle des Denkens tritt das Rechnen mit Hilfe von Maschinen.[85]

Diese Entwicklung hat sich durch BIG DATA bestätigt: Die schiere Menge an Daten und ihre Verarbeitung entwirft Prognosen, deren Wahrscheinlichkeitswert ebenso errechnet wird wie die Prognose selbst und der deswegen immer wahr ist, auch wenn die Prognose fehlgeht. Wenn es also doch regnet, obwohl die Regenwahrscheinlichkeit lediglich bei sechs Prozent gelegen hat, ist die Prognose dennoch erfüllt worden.[86] Selbst falsche Daten müssen nicht aussortiert werden, weil sich im errechneten Vergleich mit der schieren Menge der Alternativdaten herausstellt, dass sie nicht ins Muster passen. Das Pseudo-Wissenschaftliche an diesen Verfahren besteht m.E. darin, dass sie sich vor Kritik immunisieren, weil sie ihre Überprüfung verfahrensintern vornehmen: Die moderne Wissenschaft „kann es sich zur Aufgabe stellen, ‚die Phänomene und Prozesse zu *produzieren*', die sie zu beobachten wünscht".[87] Sie ist eben das „Riesengehirn", das bei der Überprüfung zirkulär auf sich selbst verweist. Ähnlich zeichnet diese Pseudo-Wissenschaft aus, dass der Ablauf nicht mehr kontrollierbar ist.[88] Unsichtbarkeit und Sichtbarkeit fallen hier sogar zusammen: Indem der Prozess hinter den Ergebnissen ver-

---

79 Vgl. Arendt: Vom Leben des Geistes, 14.

80 Vgl. Arendt: Eichmann in Jerusalem, 371, 401.

81 Arendt: Macht und Gewalt, 34. In der englischen Ausgabe fehlt der Begriff des Guten in diesem Zitat (Dies: On Violence. Orlando 1970, 30).

82 Vgl. Arendt: Macht und Gewalt, 33.

83 Vgl. ebd., 11, 34.

84 Ebd., 11.

85 Ebd., 10f.

86 Vgl. Ohly, Lukas: Theologie als Wissenschaft. Eine Fundamentaltheologie. Frankfurt a. M. 2017, 203 f.

87 Arendt: Vita activa, 361, Herv. H. A.

88 Vgl. Arendt: Macht und Gewalt, 11.

schwindet (es liegt „im Wesen des Prozesses, daß er selbst unsichtbar bleibt"), bleibt dunkel, ob etwas nicht stimmt. Es entsteht eine „pseudoscientific immanence" ohne äußere Maßstäbe.[89]

Das „Riesengehirn" kann nicht denken, weil es keine Pluralität hat. Zwar gehört auch das Gute nicht in die Öffentlichkeit.[90] Aber die Zuschreibung des Guten wird doch vom Zuschauer vorgenommen[91] und ist somit durchaus sozial konstituiert und an der „Oberfläche" zu finden. „Gut leben" ist kein Hergestelltes[92], sondern als Tätigkeit reine Aktualität, die ihre „volle Bedeutung […] im Vollzug selbst erschöpft".[93] Von dieser reinen Aktualität sagt Arendt, dass sie hinter dem Selbstzweck, gut zu leben, „sich verbirgt".[94] Das Gute beruht wohl ebenso auf dem Erscheinen des Erscheinens, von dem nur *Menschen im Plural* Zeugen werden.

An die zentrale Frage des Herstellens, welchen Zweck das Produkt erfüllen soll, tritt beim Guten eine kategorial andere Frage in den Vordergrund: Wem erscheint das Gute? Denn wem es erscheint, für den stellt sich die Zweckfrage nicht. Arendts Antwort auf die Frage lautet, dass es der Gemeinsinn ist, die Integration der fünf Sinne und der Sozialität in eine gemeinsame Welt, was die *Urteilskraft* ausbildet, zwischen Gut und Böse zu unterscheiden.[95] Ebenso wie bei einem Eichmann, der die menschliche Pluralität als solche auslöscht,[96] diese Urteilsfähigkeit ausfällt,[97] so fehlt einem maschinellen „Riesengehirn" diese Fähigkeit grundsätzlich. Es fehlt ihm der common sense, die Integrationskraft also, als Mensch unter Menschen zu urteilen[98] und damit auch zwischen Gut und Böse zu unterscheiden. Insofern scheint Arendt sowohl anzunehmen, dass das Gute nicht in die Öffentlichkeit gehört, als auch, dass es im Gemeinsinn fundiert ist. Das Gute ist kein erscheinendes Ding, sondern gründet im Erscheinen des Erscheinens.

Der Gemeinsinn bildet den Hintergrund für das Urteilen, ohne selbst beurteilbar zu sein. Denn jegliches Urteilen über ihn setzt ihn bereits voraus.[99] Das ist auch der Grund, warum das Gute keinen Zweck hat und man niemanden mit ethischen Argumenten dazu bringen kann zu denken, zu urteilen oder gut zu leben. Der Gemeinsinn macht nicht anschaulich, er vermittelt

---

89 Vgl. Arendt, Hannah: The Origins of Totalitarianism. Orlando 1979, 249. Diese Stelle fehlt in der deutschen Ausgabe.

90 Vgl. Arendt: Vita activa, 95.

91 Vgl. Arendt: Das Böse, 49.

92 Vgl. Arendt: Vita activa, 262.

93 Ebd., 261.

94 Ebd., 261f.

95 Vgl. Arendt: Vom Leben des Geistes, 311.

96 Vgl. Arendt: Eichmann in Jerusalem, 391.

97 Vgl. Arendt: Das Böse, 150.

98 Vgl. Arendt: Das Urteilen, 46.

99 Vgl. Arendt: Vom Leben des Geistes, 68.

kein Objekt und lässt sich auch nicht durch Denken in Gedankenstränge zerlegen.[100] Er ist vielmehr das Nicht-Subjektive im Privaten,[101] dasjenige, was das Nicht-Objektive des Urteils mit anderen Subjekten teilbar macht.[102]

Arendt hat Descartes dafür scharf kritisiert, dass er den Gemeinsinn, common sense oder sensus communis scheinbar zerstört und sich auf das solipsistische Cogito zurückgezogen hat.[103] Doch diese Zerstörung ist nur scheinbar, weil Descartes Gewissheiten in das Subjekt einfügt, die den Gemeinsinn enthalten. Bei seiner Transformation der Gewissheiten geht zwar der Weltbezug verloren.[104] Doch zugleich ist der cartesische Zweifel nicht konsequent genug und belegt nur, dass niemand als vollkommener Skeptiker leben kann.[105] Vom gesunden Menschenverstand schreibt Arendt, dass seine „‚Gesundheit' so ausschließlich an den Wirklichkeitscharakter der Welt gebunden ist, daß er den ihm eigenen ‚Verstand' verliert, sobald er durch Räsonieren versucht, das real Gegebene zu übersteigen".[106] Das ist eine deutliche Bezugnahme auf Descartes, bei dem „ein merkliches Abnehmen des gesunden Menschenverstandes" darauf zurückzuführen ist, dass Menschen begonnen haben „sich auf ihre Subjektivität zurückzuziehen".[107] Der Mensch verlöre nämlich ohne den Gemeinsinn den „Verstand", sobald er wirklich ein vollkommener Skeptiker wäre. Dann aber müsste er auch an seinem Zweifel zweifeln.[108]

Die Pointe von Arendts Descartes-Kritik besteht darin, dass nicht das ego cogito das letzte Fundament aller Gewissheiten bildet, sondern die *Tatsächlichkeit* der Wirklichkeit.[109] Denn was würde geschehen, wenn der vollkommene Skeptiker auch seinen Zweifel im Zweifeln einbeziehen würde? Dann müsste er daran zweifeln, dass er zweifelt, dass es Zweifel gibt und dass er es ist, der hier gerade zweifelt. Dann kann er sich gerade nicht seiner selbst vergewissern. Die Evidenz des ego cogito verdankt sich also einer anderen Evidenz, nämlich der Tatsächlichkeit, dass ein Zweifel ein Zweifel ist und dass auch sonst alles ist, was es ist. Diese Gewissheit der Tatsächlichkeit teilt aber jeder Mensch mit anderen. Sie ist ein Implikat des Gemeinsinnes, eine „unerläßliche Voraussetzung […], daß niemand, weder Gott noch ein böser Geist, etwas

100 Vgl. ebd., 61.
101 Vgl. Arendt: Das Urteilen, 105.
102 Vgl. ebd., 108, 112.
103 Arendt: Vita activa, 264 f.
104 Vgl. Arendt: Vita activa, 359, Dies.: Vom Leben des Geistes, 57 f.
105 Vgl. Arendt: Vom Leben des Geistes, 379.
106 Arendt: Vita activa, 265.
107 Ebd.
108 Vgl. Arendt: Vom Leben des Geistes, 379.
109 Vgl. Arendt: Elemente und Ursprünge totaler Herrschaft, 820.

daran ändern kann, daß zwei mal zwei vier sind".[110] Eine verrückte Person hingegen zeichnet sich gerade darin aus, dass sie zwar logische Operationen ausführen kann, aber keinen Gemeinsinn besitzt.[111]

Zusammengefasst kann man also einerseits niemanden durch Argumente überzeugen, warum es besser wäre, das Denken zu bewahren und es nicht durch maschinelles Rechnen zu ersetzen. Man kann nicht einmal jemanden argumentativ überzeugen, dass Denken etwas anderes ist als Rechnen. Darum kann das Denken nicht in Gedankenstränge aufgelöst werden, weil es mit der Tatsächlichkeit eine Voraussetzung hat, die man jemandem lediglich „ansinnen"[112] kann, weil er in sie auch immer schon vertraut. Andererseits setzt Arendt damit eine neue Alternative, nämlich die zwischen Gemeinsinn und den Verrückten, die ihren „Verstand" verlieren. Sie verlieren nämlich nicht weniger als die Tatsächlichkeit. Sie mutieren zum cartesischen Element eines „Riesengehirns", das alle Daten gleichermaßen erfasst, als gäbe es keinen Unterschied von wahr und falsch, weil das Riesengehirn seine Maßstäbe selbst erschafft, nach denen es prozessiert. Zwar entkommt auch das Riesengehirn nicht der Tatsächlichkeit, denn immerhin muss sein eigenes Prozessieren sein, was es ist. Darin besteht für Arendt der cartesische Glaube,[113] der deswegen ein Glaube ist, weil jegliches Argumentieren für oder gegen ihn bereits die Tatsächlichkeit voraussetzt. Aber dem cartesischen Glauben fehlt der Gemeinsinn, die Anerkennung der menschlichen Pluralität, dass also andere Menschen andere Menschen sind. An die Stelle der Anerkennung tritt die gleiche Art der Muster: „Natürlich mußte man von Anfang an unterstellen, daß man von dem Menschengeschlecht so reden kann wie von irgendeiner Tiergattung, in welcher Pluralität nicht mehr besagt als Exemplare der gleichen Spezies".[114]

Die Unterscheidung von Gut und Böse ist also keine ethische, sondern ergibt jeweils einen völlig anderen Sinn, ob sie aus der Perspektive des Denkens oder des Herstellens beschrieben wird. Vom Herstellen aus erfüllt das Gute einen Zweck, auch wenn hier Mittel und Zwecke vertauscht werden können, ohne dass sich am System etwas ändert. Für das Denken ist das Gute die *unscheinbare* Voraussetzung, um zu urteilen. Seine Unscheinbarkeit beruht auf dem Erscheinen des Erscheinens. Jedes Urteil wiederum wird den „Zuschauern" „angesonnen" aufgrund der gemeinsamen Grundlage des Gemeinsinns, den Menschen miteinander teilen.

---

110 Arendt: Vita activa, 361.

111 Vgl. Arendt: Das Urteilen, 100.

112 Ein Ausdruck Kants, auf den sich Arendt in ihren Reflexionen zum Urteilen vor allem bezogen hat (Kant, Immanuel: Kritik der Urteilskraft [hg. v. W. Weischedel, Bd. X]. Frankfurt a. M. 1974, 127). Arendt gibt daher dem Urteil, weil es keine Einstimmung postulieren kann, sondern das „Besondere qua Besonderem" erfasst (Arendt: Das Urteilen, 103), exemplarische Gültigkeit (vgl. ebd., 118).

113 Vgl. Arendt: Vom Leben des Geistes, 379.

114 Arendt: Macht und Gewalt, 28.

4 Der theologische Charakter der Unsichtbarkeit

Auch wenn Hannah Arendt den theologischen Charakter ihrer Position grundsätzlich zurückweist,[115] die sie gleichwohl mit Hilfe von Theologen entwickelt, lässt sich die transzendente Konstitution menschlicher Wirklichkeit nicht umgehen. Dafür sprechen die kategorialen Iterationen ihrer Begriffe: Der Mensch ist ein „Anfang des Anfangs",[116] der Wille will das Wollen,[117] das Denken denkt das Denken,[118] das Gefallen gefällt,[119] und der sensus communis ist ein „Sinn …, um meine fünf Sinne zusammenzuhalten".[120] So ist auch der Schein ein „Erscheinen des Erscheinens". In all diesen Beschreibungen zielt Arendt auf das Entstehen von Neuem.[121] Der Unterschied zum modernen Bild des Herstellens liegt darin, dass der Fortschrittsglaube keine Kreativität kennt, also keinen Anfang.[122] Jegliche Entwicklung ist vielmehr bereits potenziell im Prozess enthalten. An die Stelle der Kreativität tritt dann die Ursache.[123] „Kein Mensch, so hat sich gezeigt, kann inmitten von ‚Ursachen' leben oder in der Umgangssprache ein Sein vollständig erfassen, dessen Wahrheit im Laboratorium wissenschaftlich dargetan und in der Welt mittels der Technik praktisch geprüft werden kann."[124] Stattdessen scheint Arendt zu denken, dass der *Mensch inmitten von Kreativität* lebt, „daß der einzelne Mensch sein Leben nicht einfach der Vermehrung der Art verdanke, sondern der Geburt, dem Eintritt eines neuen Geschöpfs, das mitten im Zeitkontinuum der Welt *als* etwas völlig Neues erscheine."[125] Dieses völlig Neue inmitten der Welt ist eine Paradoxie, da es das Gegebene voraussetzt, das nicht völlig neu ist. Offenbar betont Arendt aber das Umgekehrte, dass die Welt ein Ort des Neuen ist. „Die ganze Fähigkeit zum Anfangen wurzelt im *Geborensein*."[126] Auch das ist eine Paradoxie, weil die kontinuierliche Lebensexistenz des Menschen damit auch das völlig Neue kontinuiert. Wenn daher der Mensch zeitlebens das völlig Neue *bleibt*, dann lässt er sich nicht

115 Vgl. Arendt: Vom Leben des Geistes, 429, 442.

116 Arendt: Vita activa, 216.

117 Vgl. Arendt: Vom Leben des Geistes, 303.

118 Vgl. ebd., 422.

119 Vgl. So „ist es nicht länger der Gegenstand, der gefällt, sondern daß wir ihn als Gefallen bereitend beurteilen" (Arendt: Das Urteilen, 108, Herv. H. A.).

120 Arendt: Vom Leben des Geistes, 59.

121 Vgl. Arendt: Vita activa, 215. Dies.: Elemente und Ursprünge totaler Herrschaft, 935. Dies.: Vom Leben des Geistes, 343. Dies.: Macht und Gewalt, 81.

122 Vgl. Arendt: Vita activa, 398, Dies.: Macht und Gewalt, 32.

123 Vgl. Arendt: Vom Leben des Geistes, 51.

124 Ebd., 35f

125 Vgl. ebd., 442, Herv. H. A.

126 Ebd., Herv. H. A.

auf seine Eigenschaften oder digitale Muster festlegen. Ein digitales Muster müsste vielmehr die Anfänge des Menschen bereits erfasst haben, aber damit wären sie keine Anfänge mehr, sondern Punkte auf der Reihe einer geschlossenen Entwicklung.[127] Genau darin liegt der Gegensatz zwischen Mensch und Maschine, dass die Maschine aus einem Fortschrittsparadigma hervorgeht, das nichts Neues kennt. Nicht die Maschine ist also „verrückt", wenn sie rechnet, ohne einen Gemeinsinn zu haben. Aber das Paradigma, aus dem sie entwickelt wird, kennt weder einen Gemeinsinn, weil es keine Pluralität wahrnimmt, noch die Neuheit.

Doch wie kann der Mensch er selbst bleiben, wenn er zeitlebens das völlig Neue ist? Wie lässt sich das völlig Neue dieses Menschen vom völlig Neuen eines anderen Menschen abgrenzen? Und wie kann die Welt sie selbst bleiben, wenn sie der Ort ist, an dem das völlig Neue geschieht? An dieser Stelle vermittelt Arendt zwei Kategorien, nämlich das *Neue* mit seiner *Tatsächlichkeit*. Wenn keine Eigenschaften und festen Bestände eine Welt der Kreativität kennzeichnen, dann muss das Neue zumindest sein, was es ist. Es muss Tatsächlichkeit besitzen. Und wenn sich das Neue fortsetzt, dann ist auch seine Fortsetzung, was sie ist. Darin besteht die Tatsächlichkeit der Wirklichkeit. Deshalb schließen sich Neuheit und Tatsächlichkeit nicht aus, sondern bedingen einander. Sie bedingen sich aber nicht erst in der Welt, weil nichts in der Welt für diese wechselseitige Angewiesenheit bürgen kann, die in der Welt bereits vorausgesetzt ist. Es könnte sonst nie verlässlich etwas Neues in der Welt geschehen. Neuheit und Tatsächlichkeit sind vielmehr weltkonstituierend. Sie liegen ihr voraus. Darum sind sie schöpfungstheologisch grundlegend und können in Arendts Denken nicht hintergangen werden.

Arendt versucht zwar umgekehrt, die christliche Vorstellung der creatio ex nihilo aus dem Anfang des Anfangs zu bestimmen, der der Mensch ist: „Der Übergang vom Nichts zum Etwas ist so schwierig, daß man zu der vorläufigen Vermutung berechtigt ist, es sei das wollende Ich gewesen, das … die Idee eines absoluten Anfangs als seinem Pläneschmieden kongenial empfunden habe."[128] Weil also der Mensch den Übergang vom Nichts zum Etwas vollbringt, überträgt er die Vorstellung von der Entstehung der Welt auf einen Schöpfergott. Zur „vorläufigen Vermutung berechtigt" ist dann für Arendt nicht die Vorstellung eines Schöpfergottes, sondern die, dass dem Menschen diese Analogie „vorläufig" einfällt, weil er selbst Übergänge initiieren kann.

Doch damit verlässt Arendt die Spur ihrer cartesischen Kritik. Das wollende Ich kann nämlich nicht die letzte Urheberin des Neuen sein, weil es dabei den „cartesischen Glauben" zugrundelegen muss, dass es ist, was es ist. Und seine Tatsächlichkeit kann es nicht willentlich erzeugen, ohne bereits von ihr abhängig zu sein. Was Arendt bei Descartes kritisiert, nämlich den konsequenten Zweifel an seiner Grenze angehalten zu haben, trifft sie selbst, sobald der

[127] Vgl. ebd., 254, 269.

[128] Ebd., 268.

Wille ohne Tatsächlichkeit aus dem Nichts hervorgehen zu können scheint. Nicht der Wille schafft den Übergang vom Nichts zu Etwas, sondern der Anfang des Anfangs, der der Mensch ist. Der Anfang des Anfangs ist aber selbst schon ein Übergang. Er birgt also die creatio ex nihilo in sich, und das kann nur sein, wenn er ist, was er ist. Anders gesagt: Der Mensch ist der Anfang des Anfangs, weil er *Geschöpf* ist. Und er ist Geschöpf, weil das Neue, das er ist, aus dem Nichts kommt, zugleich aber Tatsächlichkeit besitzt, um in diese Welt zu kommen. Die theologischen Grundlagen in Arendts Bild von der Kreativität inmitten der Welt lassen sich also nicht zum Verschwinden bringen.

Arendt mag versucht haben, auch hier wieder Selbst- und Alteritätsperspektive miteinander zu vermitteln, um nicht beim isolierten Willen einzusetzen oder bei einer solipsistischen Beschreibung des Anfangs des Anfangs. Damit könnte sie auch versucht haben, die theologische Dimension dieses Denkens intersubjektiv aufzulösen. Mit dieser Vermittlung beider Perspektiven würde zudem verständlich werden, warum das Geborensein der Tiere kein Neues in die Welt bringt. Denn die Tiere sind für Arendt nur „Exemplare der gleichen Spezies", der Mensch dagegen existiert nur im Plural. Aus welcher Perspektive also erscheint ein geborener Mensch als Anfang des Anfangs?

Ebenso wie das Denken anderen Menschen verborgen bleibt und trotzdem an der „Oberfläche" beobachtet werden kann, so ist das Geborensein das Neue für die Menschen, die schon da sind. Und ebenso wie das Denken an der Oberfläche als verborgenes erscheint, so erscheint das Neue als verborgenes. Denn sichtbar ist nicht das Neue, sondern das geborene Kind. Wie es für das Kind ist, jetzt da zu sein, bleibt verborgen. Ebenso sehen seine Eltern nicht den Anfang des Anfangs; sie sehen nur den Anfang. Der Anfang des Anfangs ist für sie „uneigentlicher Schein", das Erscheinen des Erscheinens.

Doch lässt sich damit die theologische Dimension ausräumen? Nach meinem Eindruck wird damit versucht, Neues und Tatsächlichkeit auf zwei Pole der Intersubjektivität zu verteilen: Ein Mensch ist dann nicht *an sich* Anfang des Anfangs, sondern für die Menschen, die schon da sind. Er ist *für sie* das völlig Neue, während sie *für ihn* Beständigkeit verbürgen. Es lässt sich dann zwar verständlich machen, warum nach Arendt das Neue ein soziales oder sogar politisches Phänomen ist.[129] Aber diese Verteilung blendet aus, dass das Erscheinen des Erscheinens, gerade weil es ein soziales Phänomen ist, nicht *einen Autor* und überhaupt keinen Autor hat. Wiederum ist das Erscheinen des Erscheinens nicht dadurch konstituiert, dass es Menschen auffällt. Wie ich oben gezeigt habe, wird es erst im Nachhinein erfahren. Es ist also kein intersubjektives Phänomen, das erst dann aktiviert würde, wenn es je aktuell wahrgenommen wird, sondern „transsubjektiv", ein Bestimmendes ohne Autoren, die bestimmen.[130]

---

[129] Arendt: Macht und Gewalt, 81.

[130] Vgl. Fischer, Johannes: Leben aus dem Geist. Zur Grundlegung christlicher Ethik. Zürich 1994, 9f.

Ebenso wenig verbürgt die Menschheit für den Anfang des Anfangs, dass er ist, was er ist. Denn dann würde die Tatsächlichkeit übersehen, die die Menschheit hat, ohne sie gebildet zu haben. Auch Tatsächlichkeit muss nicht auffallen. Sie ist nicht „für" jemanden, aber für einen anderen nicht, weil von ihr *alles* „schlechthinnig abhängig" ist.[131] Sie ist die Instanz, die Schleiermacher „Gott" nannte und die er damit, auch wenn man sie anders nennen mag, religionsphilosophisch kategorisierte.[132] Arendts Darstellung verbleibt damit in einem theologischen Rahmen.

5 Folgen für die theologische Reflexion Künstlicher Intelligenz

Zu Beginn dieses Aufsatzes habe ich behauptet, dass sich die Theologie bei ihrer Thematisierung der Digitalisierung auf Statusfragen, auf die Moralfähigkeit und auf die religiöse Anschlussfähigkeit fokussiert. Dabei konzentriert sich die Theologie methodisch auf die Bestimmung der Mittel (Medien), um mit ihnen diese Fragen zu beantworten. Im Rückblick auf den Durchgang meines Aufsatzes lässt sich diese Methode dem Paradigma des Herstellens zuordnen. Das bedeutet, dass bei der aktuellen Beschäftigung mit diesen Fragen das Paradigma des Herstellens bereits den theologischen Diskurs kolonialisiert hat.

Eine kritische Theologie der Digitalität steht vor der Herausforderung, eine unabhängige Perspektive auf diesen Phänomenbereich zu gewinnen, auch wenn sich bereits digitale Verfahren theologisch eingespielt haben. Ein geeigneter Weg scheint mir im hermeneutischen Zirkel zu bestehen, der im Selbstvollzug zugleich eine Selbstdistanzierung erreicht. Was Bultmann das Vorverständnis nennt, lässt sich ja erst im hermeneutischen Prozess als solches rekonstruieren: es tritt rückwirkend als dasjenige auf, wovon sich das Verstehen distanziert.[133] Das Verstehen verhält sich zum Vorverständnis wie Arendts Erscheinen des Erscheinens zum erscheinenden Gegenstand. Beide sind kategorial zu unterscheiden. Diese kategoriale Unterscheidung vollzieht sich im Denken.

Ich habe zu Beginn des Artikels auch behauptet, dass sich der Trend zum Herstellungsparadigma vor allem an der Rolle von *Entscheidungen* verdichtet. Entscheidungen vermitteln bei der Transformation von Mensch zu Maschine, vom Abendmahl in die Digitalität. Und Entscheidungen werden reflexiv behandelt, wenn die Entscheidungsfähigkeit von Künstlicher Intelligenz zum Gegenstand wird. Was in der Vermittlung von Transformationsprozessen still-

131 Vgl. Ohly, Lukas: Ethik als Grundlagenforschung. Eine theologische Ethik. Berlin 2020, 102. Ders.: Ethik der Robotik und der Künstlichen Intelligenz. Berlin 2019, 99.

132 Vgl. Schleiermacher, Friedrich: Der christliche Glaube Bd. 1. Berlin 1960, 29f.

133 Vgl. Bultmann: Glauben und Verstehen Bd. 3, 147.

schweigend vorausgesetzt wird, wird bei der Moralfähigkeit Künstlicher Intelligenz bewusst unterstellt, nämlich dass sich die theologische Dimension der Digitalität an der *Entscheidung entscheidet*.

Auffällig ist, dass der Entscheidungsbegriff in Arendts Denken keine Rolle spielt, und zwar obwohl sie ein Buch zum Willen geschrieben hat. Aber als geistiges Vermögen ist der Wille selbstbezüglich[134] und richtet sich nicht darauf, äußere Sachverhalte zu verändern. Aus der Selbstbezüglichkeit tritt der Wille erst heraus, wenn er ins Handeln übergeht.[135] Im Handeln ist Neues gesetzt, ohne dass eine Entscheidung hier vermittelt hätte. Das Handeln beruht nicht auf vorgefassten Zielen.[136] Es *ist* vielmehr der Übergang, der Anfang des Anfangs, der der *Mensch* ist. Dementsprechend ist für Arendt das Handeln die Enthüllung der Person.[137] „Handelnd und sprechend offenbaren die Menschen jeweils, wer sie sind".[138] Das Handeln ist an die Natalität mehr angebunden als andere menschliche Tätigkeiten.[139] Der Mensch kann danach nur Neues setzen, weil er selbst Neues ist. Arendt ordnet das Handeln den „unproduktiven" Tätigkeiten zu[140] und stellt es damit nicht nur dem Herstellen entgegen, sondern auch einem zielstrebigen Entscheiden. Es ist eine Antwort[141] („Verantwortung"[142]) und kein eigener Entschluss.[143]

Was bedeutet es nun, wenn demgegenüber die Moralfähigkeit an die Entscheidung gebunden wird? „Ohne diese Eigenschaft, über das Wer der Person mit Aufschluß zu geben, wird das Handeln zu einer Art Leistung wie andere gegenstandsgebundene Leistungen auch."[144] Das ist bei einer Künstlichen Intelligenz, die „moralanalog handelt", der Fall. Unerwartetes soll jetzt ausgeschaltet werden, indem an die Stelle des *unvorhersehbaren* Handelns das Experiment tritt,[145] das mit wissenschaftlicher Exaktheit politische Institutionen setzt.[146] Das autonom fahrende Auto soll dann in Notsituationen die richtige Entscheidung treffen ebenso wie der Pflegeroboter bei

134 Vgl. Arendt: Vom Leben des Geistes, 422.

135 Vgl. ebd., 304.

136 Vgl. Arendt: Vita activa, 226.

137 Vgl. ebd., 224f.

138 Ebd., 219.

139 Vgl. ebd., 18.

140 Vgl. ebd., 113.

141 Waldenfels, Bernhard: Sozialität und Alterität. Modi sozialer Erfahrung. Berlin 2015, 19. Ders.: Phänomenologie der Aufmerksamkeit. Frankfurt a. M. 2004, 47, 55.

142 Arendt: Vita activa, 215.

143 Vgl. ebd., 214.

144 Ebd., 221.

145 Vgl. ebd., 381f.

146 Vgl. ebd., 380.

einer bedenklichen Zustandsveränderung der Patientin. Doch was das Richtige ist, bestimmt die technische Herangehensweise. Das ist unausweichlich, wenn Maschinen so gebaut werden sollen, dass sie Entscheidungen treffen. Denn wenn das Handeln die Moral enthüllt[147] und wesentlich unter Menschen im Plural vollzogen wird,[148] kann die Moral nicht in Maschinen die Maßstäbe für Richtig und Falsch bilden. Die Entscheidung wird so zum Siegel der Kolonialisierung des Handelns durch das Herstellen, zur Kolonialisierung der Ethik durch die Automatisierung.

Nun hat Arendt das Handeln mit der Politik stark assoziiert, nämlich mit dem gemeinsamen Handeln in einem Gemeinwesen.[149] Spielt nicht die Entscheidung eine wesentliche Rolle im politischen Handeln? Und ist dann nicht die Entscheidung das adäquate Kriterium auch für eine sozial verfasste und Öffentliche Theologie, um die Probleme zur Digitalisierung zu behandeln? Warum kommt dann bei Arendt der Entscheidungsbegriff so wenig vor?

Anstelle der Entscheidung zieht Arendt den Konsens heran,[150] die gemeinsame Anerkennung. Und geradezu gegenläufig zur Autonomiefiktion der Entscheidung betont Arendt, dass *Gehorsam* die häufigste Form des Konsenses ist.[151] Anscheinend bilden Revolution und Rätesystem, die beiden einzigen zwischenstaatlichen Formen politischen Handelns,[152] den gemeinsamen Willen „spontan"[153], und zwar im wechselseitigen Gehorsam der beteiligten Personen. Die Revolution belegt, dass gemeinsames Handeln auf dem Neuen beruht.[154] Eine Entscheidung dagegen kann auch jemand allein treffen. Sie ist dann unpolitisch. Das trifft auch auf das Herstellen zu.[155]

Werden nun kirchenpolitische oder theologische *Entscheidungen* zum Status Künstlicher Intelligenz oder zur digitalen religiösen Kommunikation getroffen, so wird ein Instrument des Herstellens benutzt, das nicht auf die Anerkennung gemeinsamer politischer Macht angewiesen ist. Die Anerkennungspraxis wird umgangen, indem lediglich Textvorlagen produziert werden. Und indem der Status des jeweiligen Digitalisierungsphänomens aus der Transformation eines „Kerns" bestimmt wird, kann die Autorin der entsprechenden Textvorlagen von einer gemeinsamen Praxis isoliert sein, in der sich der jeweilige Status entscheidet. Die Transformation verbleibt dann im errechneten Rahmen und kann somit gerade nichts Neues

147 Vgl. Wenn das Handeln die Person enthüllt und das Personsein eine moralische Eigenschaft ist (vgl. Arendt: Das Böse, 53), dann offenbart das Handeln die Moral.

148 Vgl. Arendt: Vita activa, 17.

149 Vgl. ebd., 17, 189. Arendt: Macht und Gewalt, 45. Dies.: Vom Leben des Geistes, 433.

150 Vgl. Arendt: Macht und Gewalt, 42.

151 Vgl. Arendt: Macht und Gewalt, 46. Dies.: Vom Leben des Geistes, 427.

152 Vgl. Arendt: Macht und Gewalt, 131 f.

153 Ebd., 132.

154 Vgl. Arendt: Vom Leben des Geistes, 431.

155 Vgl. Arendt: Vita activa, 270.

sein. *Gerade so verfehlt sie ihren Kern!* Denn der Kern liegt in der Verborgenheit, die in einer gemeinsamen Welt als Erscheinen des Erscheinens an die Oberfläche tritt.

6 Wie der theologische Diskurs zu erweitern ist

Es ist nicht von Vornherein ausgeschlossen, dass das digitale Abendmahl dem christlichen Sinn des Abendmahls entspricht und dass eines Tages aus Künstlicher Intelligenz Singularitäten entstehen, die Gefühle haben und sich erleben. Methodisch ausgeschlossen ist aber, dass diese Statusfragen über ein Verfahren geklärt werden, das Instrumente des Herstellens einsetzt. Denn dieses Verfahren entspricht einer zirkulären Herangehensweise. Die Theologie sollte der Versuchung nicht erliegen, das Paradigma zu übernehmen, welches sie reflektieren will, weil es damit ihre Ergebnisse bereits präjudiziert, so dass man aus dem geschlossenen System der Maschinerie nicht mehr aussteigen kann.

Ein wesentliches eigenes Denkinstrument besitzt die Theologie mit dem Phänomen der *Offenbarung*. Wenn Gott sich offenbart, so wird er nicht sichtbar, sondern bringt die Erscheinung zur Erscheinung. Der hermeneutische Zirkel, wie ich ihn oben skizziert habe, hat insofern Offenbarungscharakter. Es ist bemerkenswert, dass eine Theoretikerin, die theologisches Denken ausdrücklich von sich weist, obwohl sie bei Bultmann studiert hatte, die Phänomenologie der Offenbarung skizziert hat, um damit auch den kategorialen Gegensatz zum Sichtbarmachen des Unsichtbaren in der Automation zu bestimmen. Der theologische Diskurs zur Digitalisierung erfährt eine wesentliche Erweiterung, wenn er am Offenbarungsphänomen einsetzt und dann die kategorialen Einebnungen von Sichtbarkeit und Unsichtbarkeit erkennt, die durch die Digitalisierung vorgenommen werden – nicht erst von künstlich intelligenten Maschinen, sondern auch schon in der menschlichen Sicht, die sich selbst automatisiert und so das Unsichtbare verfehlt, über das sie eigentlich reden will.

Literaturverzeichnis

Arendt, Hannah: On Violence. Orlando 1970.

Arendt, Hannah: The Life of the Mind. The Groundbreaking Investigation on How We Think. Florida 1978.

Arendt, Hannah: The Origins of Totalitarianism. Orlando 1979.

Arendt, Hannah: Responsibility and Judgment. (Hg. J. Kohn). New York 2003.

Arendt, Hannah: Vita activa oder Vom tätigen Leben. München ^{20}2019.

Arendt, Hannah: Elemente und Ursprünge totaler Herrschaft. Antisemitismus, Imperialismus, totale Herrschaft. München [21]2019.

Arendt, Hannah: Über das Böse. Eine Vorlesung zu Fragen der Ethik. München [9]2014.

Arendt, Hannah: Vom Leben des Geistes. Das Denken. Das Wollen. München [10]2020.

Arendt, Hannah: Macht und Gewalt. München [27]2019.

Arendt, Hannah: Eichmann in Jerusalem. Ein Bericht von der Banalität des Bösen. München [17]2021.

Arendt, Hannah: Das Urteilen. München [6]1921.

Barth, Karl: Kirchliche Dogmatik II/2. Zürich [3]1948.

Bultmann, Rudolf: Glauben und Verstehen Bd. 3. Tübingen 1960.

Dahlmann, Anja: Militärische Robotik als Herausforderung für das Verhältnis von menschlicher Kontrolle und maschineller Autonomie. In: Zeitschrift für evangelische Ethik 61 (2017), 171–183.

Dalferth, Ingolf U.: Kombinatorische Theologie. Probleme theologischer Rationalität. Freiburg/Basel/Wien 1991.

Evangelische Kirche in Deutschland: Freiheit digital. Die Zehn Gebote in Zeiten des digitalen Wandels. Eine Denkschrift der Evangelischen Kirche in Deutschland. Leipzig 2021.

Fischer, Johannes: Leben aus dem Geist. Zur Grundlegung christlicher Ethik. Zürich 1994.

Foerst, Anne: Von Robotern, Mensch und Gott. Künstliche Intelligenz und die existenzielle Dimension des Lebens. Göttingen 2009.

Gräb-Schmidt, Elisabeth: Autonome Systeme. Autonomie im Spiegel menschlicher Freiheit und ihrer technischen Errungenschaften. In: Zeitschrift für evangelische Ethik 61 (2017), 163–170.

Grethlein, Christian: Mediatisierung von Religion und Religiosität. In: Zeitschrift für Theologie und Kirche 115 (2018), 361–376. DOI: 10.1628/zthk-2018-0017.

Härle, Wilfried: Der Glaube als Gottes- und/oder Menschenwerk in der Theologie Martin Luthers. In: MJTh 4 (1992), 37–77.

Härle, Wilfried: Dogmatik. Berlin/New York 1995.

Kant, Immanuel: Kritik der Urteilskraft [hg. v. W. Weischedel, Bd. X]. Frankfurt a. M. 1974.

Lienau, Anna-Katharina: Kommunikation des Evangeliums in social media. In: Zeitschrift für Theologie und Kirche 117 (2020), 489–522. DOI: 10.1628/zthk-2020-0022.

Nord, Ilona: Realitäten des Glaubens. Zur virtuellen Dimension christlicher Religiosität. Berlin u. a. 2008.

Ohly, Lukas: Theologie als Wissenschaft. Eine Fundamentaltheologie. Frankfurt a. M. 2017. DOI: 10.3726/b11619

Ohly, Lukas: Ethik der Robotik und der Künstlichen Intelligenz. Berlin 2019. DOI: 10.3726/b15565

Ohly, Lukas: Ethik als Grundlagenforschung. Eine theologische Ethik. Berlin/Boston 2020. DOI: 10.1515/9783110705607-004

Pannenberg, Wolfhart: Anthropologie in theologischer Perspektive. Göttingen 1983.

Pannenberg, Wolfhart: Wissenschaftstheorie und Theologie. Frankfurt a. M. 1987.

Schleiermacher, Friedrich: Der christliche Glaube Bd. 1. Berlin 1960.

Schmitt, Carl: Politische Theologie. Vier Kapitel zur Lehre von der Souveränität. Berlin [8]2004.

Scholtz, Christopher: Alltag mit künstlichen Wesen. Theologische Implikationen eines Lebens mit subjektsimulierenden Maschinen am Beispiel des Unterhaltungsroboters Aibo. Göttingen 2008.

Schrodt, Christoph: Abendmahl: digital. Alte und neue Fragen – nicht nur in Zeiten der Pandemie. Zeitschrift für Theologie und Kirche 118 (2021), 495–515, DOI: 10.1628/zthk-2021-0024.

Schwarke, Christoph: Ungleichheit und Freiheit. Ethische Fragen der Digitalisierung. In: Zeitschrift für evangelische Ethik 61 (2017), 210–221.

Waldenfels, Bernhard: Phänomenologie der Aufmerksamkeit. Frankfurt a. M. 2004.

Waldenfels, Bernhard: Sozialität und Alterität. Modi sozialer Erfahrung. Berlin 2015.

Metaversum und resistente Körperlichkeit

Ein neo-materialistischer Blick auf die virtuelle Kreation und Produktivkraft des modernen Humanismus

Simon Reiners

Abstract

The present utopia of the metaverse creates its image of the human being: an individual that works, communicates, loves, and lives in virtual space. Every attribution of an image embodies social relations. This also accounts for the relations of production in the metaverse. Tech-companies that could create a metaverse dominate the necessary means of production. Critical and Material-Feminist Theories look to the humanist promise of self-determination through the renunciation of the physical body – a promise that the virtual world holds as well. Their analyses point to the role of excluded corporeality in participating in a dignified, humane life. What challenges do we face from a creation of the human as one who must transcend himself in order to live a humane life in times of the metaverse? Tracing bodily resistance based on the perspective of Donna Haraway and Theodor Adorno through the humanist imagination of the metaverse makes its relations of domination visible, leading to ethical practice in a present that is becoming increasingly virtual.

1 Einleitung

In Marge Piercy's Science-Fiction Roman *He, She and It* von 1991 wird von der Entwicklung eines Cyborgs erzählt. Die beiden jüdischen Forscher:innen Malkah und Avram verschreiben sich der Schöpfung mehr-als-menschlichen Lebens: den Cyborg Yod. Das entscheidend Andere in diesem Roman ist jedoch, dass das ‚eigentlich' Leben in dieser Welt durch Plug-In-Zugänge in einem virtuellen Raum stattfindet. Dort wird gearbeitet, politisch und ökonomisch verhandelt, gereist und geliebt. Es verschmelzen die Grenzen von real und virtuell. Yod

repräsentiert also nur einen Teil davon, was es in Piercy's Roman bedeutet, die Grenzen des Menschseins auszuweiten.

Unter dem Begriff „Metaversum" versammeln sich heterogene Utopien. Gemeinsam ist ihnen das Versprechen alle Lebensbereiche des Menschen wie Kommunikation, Arbeit, Freizeit, Politik, Sozialität, Kunst in einen virtuellen Raum zu verschieben. Der Dualismus von materiell/real gegenüber immateriell/virtuell soll aufgehoben werden. Damit ist der Gedanke verbunden sich der notwendig physisch, räumlich und zeitlich gebundenen Praktiken der menschlichen Lebensform zu entledigen. Mit der Überschreitung der Grenzen der *conditio humana* geht das Versprechen einer wie auch immer gearteten Emanzipation des Individuums einher. Wie und wer jemand sein möchte, wäre im virtuellen Raum nicht mehr an bestehende physische Kategorien wie etwa Geschlecht, Hautfarbe oder Alter gebunden.

Die vermeintliche Utopie eines Metaversums wird bisher fast ausschließlich analytisch und deskriptiv betrachtet. Eine derartige Realitätserweiterung muss jedoch schon in der Entstehung normativ und herrschaftskritisch begleitet werden, um mit Faktizität Schritt halten zu können. Das bildet den Ausgangspunkt der folgenden Untersuchung. Der Wandel gesellschaftlicher Organisationsformen, wie etwa das Metaversum, bringt neue Formen Mensch zu sein hervor. In diesem Text zur gesellschaftlich-ökonomischen Bedeutung der Utopie des Metaversums soll es somit grundlegend um eine kritische Reflexion der Konstruktion von Menschenbildern gehen. Ein Verständnis dessen, was Mensch *sei*, entscheidet über die Möglichkeiten ein humanes, menschen*würdiges* Leben zu führen.

Aus einer kritischen, christlich sozialethischen Sicht müsste nun gefragt werden, welche Formen von ökonomischer Ausbeutung und gesellschaftlicher Ausgrenzung diese Schöpfung einer hoffnungsvollen neuen Welt hervorbringt. Kritik, schreibt die Sozialethikerin Michelle Becka „ist Aufgabe von Theologie im Allgemeinen und christlicher Sozialethik im Besonderen. Dabei handelt es sich nicht um eine für die Theologie vernachlässigbare, sondern um eine zentrale Aufgabe."[1] Das Metaversum also auch aus diesem Blick zu betrachten gilt insbesondere vor dem Hintergrund, dass ein Metaversum höchstwahrscheinlich kein öffentlich, demokratisch erzeugter Ort des kollektiven Handelns sein wird. Vielmehr wird die Struktur des Metaversums von den ökonomischen Interessen weniger Tech-Giganten abhängen. Demnach wird im Folgenden gefragt: Vor welche Herausforderungen stellt uns die Hervorbringung eines Bildes des Menschen, als einem, der sich selbst übersteigen muss,

[1] Dieser Text steht auch insofern in teilweise Einverständnis mit Michelle Becka, als dass „dieser Beitrag weitgehend auf eine theologische Semantik verzichtet", aber doch durch die Leser:innen als genuin theologisch verstanden werden kann. Becka, Michelle: Kritik und Solidarität. In: Beck, Michelle/Emunds, Bernhard/Eurich, Johannes u. a.: Sozialethik als Kritik. Baden-Baden 2020, 19–55, hier: 50.

um Mensch zu sein und erst so ein menschen*würdiges* Leben in Zeiten des Metaversums führen kann?

In einem ersten Schritt (2.) werden dazu einige denkbare Kriterien eines Metaversums gegenüber bestehenden virtuellen Welten formuliert. Daran anschließend soll auf die Rolle des Körpers geschaut werden (3.). In der virtuellen Utopie des Metaversums verschwindet Körperlichkeit vollständig aus der Frage nach einem menschenwürdigen Leben. Aus diesem Grund rückt der Körper ins Zentrum dieser Untersuchung als: zum einen das Objekt, welches durch die essentialistische Festlegung von Menschenbildern anhand von Herrschafts- und Produktionsmustern verkörpert wird (3.1, 3.2); zum anderen Körperlichkeit als ein Ort, an dem die Grenzen vorherrschender Menschenbilder erfahren und widerständige Praktiken ausgebildet werden können (3.3., 3.4).

Eine vom Körper ausgehende Kritik an Menschenbildern finden sich unter anderem im negativ-dialektischen Materialismus von Theodor W. Adorno und im feministischen Materialismus von Donna J. Haraway. Beide Theoretiker:innen formulieren anti-humanistische Perspektiven, um sich gegen die Folgen humanistischer Festschreibungen zur Wehr zu setzen. Ihre Positionen ziehen ihre Kraft jedoch nicht gegen das Versprechen des Humanismus auf ein menschenwürdiges Leben, sondern vielmehr aus dem Anspruch heraus dieses zu ermöglichen. Aus diesem Grund sollen die Perspektiven von Haraway und Adorno zu einer kritischen Analyse des Metaversums und dessen Humanismus verwoben werden.

Es soll gezeigt werden, dass das Metaversum gar nicht so *meta* ist, wie es scheint. Im Gegenteil lässt sich an dessen Vorstellung viel über die Gegenwart aussagen, die nach Meta-, Trans- oder Post-Lebenswelten sucht. Der Widerständigkeit von verdrängter Körperlichkeit in der Idee des Metaversums nachzugehen, macht die darin bestehenden Herrschaftsverhältnisse sichtbar und führt zu einer ethischen Praxis in einer Gegenwart, die zunehmend virtuell wird.

2 Metaversum: Utopie und Herausforderung

2.1 Jenseits von Raum und Zeit

Der Roman *He, She and It* von Marge Piercy wurde 1991 veröffentlicht. Das zeigt, dass die Idee einer virtuellen Welt in der gearbeitet, politisch und ökonomisch verhandelt, gereist und geliebt wird schon lange in unseren Köpfen lebt. Während das Internet eine Ergänzung zur physischen Welt darstellt, um (Informations-) Austausch in einem umfangreicheren Netz zu ermöglichen, drückt der Begriff Metaversum die Auflösung des Dualismus von materiell/real gegenüber immateriell/virtuell aus. Die Utopie des Metaversums stellt einen vollumfänglichen Raum der Erfahrung und Interaktion dar: aktiver, unmittelbarer, die Erweiterung aller Le-

bensbereiche, insbesondere Kommunikation, Arbeit, Governance, Unterhaltung oder Kunst. Der Dualismus von zwei bisher gleichwertigen Welten verschmilzt nicht einfach, sondern wird zur Seite der immateriell-virtuellen hin aufgehoben.

In Spielen wie Minecraft, Fortnite oder Roblox sind Teile dieser Lebensform bereits real. Sie sind eine Kombination aus Unterhaltung, Austausch und Marktplatz. Diese Ansätze reichen aber nicht aus, um zu verdeutlichen, was *meta* sein würde. Matthew Ball, ehemaliger Amazon-Manager, versucht Kriterien zu benennen, die ein Metaversum auszeichnen müssten, um mehr zu sein als die Erweiterung von Social-Media und Gaming.[2] Ihm zufolge ist die Überschreitung der beiden Dimensionen von Raum und Zeit kennzeichnend:

Raum: Der Raum des Metaversums müsse dreidimensional erfahren werden, um mit der physischen Welt zu verschmelzen. Er müsste eine Synchronisation des gesamten Lebens sein, allem voran eine vollumfängliche Ökonomie durch Reorganisation des Produzierens, Besitzens, Tauschens von Werten und Gütern und entgrenzte Globalisierung von Arbeit. Damit gehe das Ende von Nationalstaaten einher.[3] Letzteres sei nicht nur denkbar und möglich, sondern notwendig.

Zeit: Ball erwartet vom Metaversum ein anderes Erinnern. Erlebnisse könnten vollständig gespeichert und jederzeit ohne Lücken hervorgeholt werden. Gewisse Formen des Zeitreisens, zumindest in der Zeit seit dem Start des Metaversums, wären ein Wandel des Zeitverstehens. Selbst die Überwindung des Todes wird thematisiert, wenn zwar nicht die physische Hülle, so doch die vergangene Erfahrung, Erinnerung und Beziehungen gespeichert und fortgesetzt werden können.[4]

In Piercy's Roman werden die Avatare, mit denen sich Menschen in diesem Metaversum präsentieren, unabhängig von physisch existierenden Formen gestaltet. Hier ließe sich eine weitere Grenzüberschreitung festhalten: Indem körperlich-physische Grenzen überschritten werden, wird im Metaversum nicht mehr die festgeschriebene Identität repräsentiert, sondern sie kann aktiv bestimmt werden. Merkmale, durch die ein Subjekt unfreiwillig sozialer Diskriminierung ausgesetzt ist, lassen sich so überwinden. Geschlecht, Hautfarbe, Alter oder körperliche Beeinträchtigungen werden unsichtbar und frei wählbar, sowie permanent wandelbar. Ganz grundsätzlich scheint ein Metaversum die Ausweitung von Handlungsmöglichkeiten zu bieten. Das ist Inbegriff der Beziehung von Emanzipation und Fortschritt. Was die Moderne als Fortschritt verspricht, das heißt, die Unbedingtheit des Individuums, würde nicht nur ver-

[2] Vgl. Ball, Matthew: The Metaverse. New York 2022.

[3] Vgl. Ball: Metaverse, 29–70.

[4] Vgl. Farman, Abou: On Not Dying. Minneapolis 2020.

wirklicht, sondern überschritten. Am Ende einer derartigen Entwicklung neuer Wirtschafts-, Glücks-, Arbeits- und Lebensformen stünde eine neue Form des Individuum-, Subjekt- und damit Menschseins.

2.2 Grenzen der Freiheit

Das Metaversum wird als Realisierung des Versprechens der Moderne auf die Entfaltung souveräner Subjektivierung ersehnt. Zugleich lassen sich bereits jetzt Herausforderungen erkennen, die notwendig in diese Utopie eingeschrieben sind. Zu dieser vermeintlichen Utopie gehört auch, dass sie von Tech-Giganten wie Google, Apple, Facebook (jetzt Meta Platforms, Inc.), Amazon, Microsoft (GAFAM) vorangetrieben wird. Das hat offensichtlich Gründe: Nur sie verfügen über die Software, Hardware, finanziellen Kapazitäten sowie rechtliche und geographische Grenzenlosigkeit, für ein solches Projekt. Sie erschließen sich dadurch zugleich neue Märkte. Die daraus resultierenden Herausforderungen sind kritischen Reflexionen der Gegenwart nicht neu. Die Betrachtung der Idee des Metaversums ergänzt und übersteigt diese aber in mindestens drei Aspekten:

1. Soziale Gerechtigkeit: Das Versprechen frei gewählter Identität durch die Gestaltung der Avatare übersieht die soziale Diskriminierung durch Klassenzugehörigkeit. Die Teilhabe am und im Metaversum hängt vom Zugang zu Software, Hardware und Wissen ab. Ungerechte Verteilung und Zugang zu Bildung sind somit wesentliche Grenzen des Metaversums. Hinzu kommen Gerechtigkeitsfragen bezüglich arbeitsrechtlicher Herausforderungen. An der Blaupause der Plattformökonomie zeigt sich bereits, wie schwierig die Organisation von Arbeitnehmer:innen ist.[5] Dabei ist auch unklar, wer die Möglichkeiten und Legitimation für die Setzung arbeitsrechtlicher Standards besäße.
2. Vergesellschaftung: Die Digitalisierung ist ein Beschleuniger des modernen Versprechens auf Individualismus. Das Metaversum verspricht der Turboantrieb zu sein. Wie dies zu einer sozialen Herausforderung wird, hat Andreas Reckwitz in seinem Buch *Die Gesellschaft der Singularitäten* gezeigt. Im Zentrum des von ihm beschriebenen Strukturwandels steht der Zwang, sich kontinuierlich als etwas Einzigartiges zu prä-

[5] Vgl. De stefano, Valerio: The Metaverse is a labour issue. Online unter: https://socialeurope.eu/the-metaverse-is-a-labour-issue (Stand: 29.08.2022).

sentieren, nach singularistischen Kriterien der Originalität, Kreativität und Performanz,[6] letztlich mit dem Ziel der eigenen Vermarktlichung.[7]

3. Herrschaft: Wie Ball verdeutlicht, kommen für die Realisierung eines Metaversums nur die ohnehin tonangebenden GAFAM in Frage. Statt wie im Fall des Internets nachträglich ökonomische Interessen durch Datenhandel und Werbung anzulegen, würde bereits die Umsetzung des Metaversums von ökonomischen Interessen geprägt. Mit der Bestimmung der zu überschreitenden Grenzen des Menschen schaffen die GAFAM ihrerseits ein Menschenbild. Es ist davon auszugehen, dass es nicht allein um die Möglichkeiten erweiterter Freiheit geht, sondern um das Herausheben bestimmter, ökonomisch attraktiver Potenziale: nach dem Ziel der Produktivität. Was, wo und wie ein Individuum zu sein möglich ist, entscheiden vorgefertigte Software, Algorithmen und KI unter der Devise der Wertschöpfung. Wer welche Möglichkeiten hat, in diesem Raum zu kommunizieren, zu arbeiten, sich zu vergnügen und zu leben, was als legitime Äußerung und als angemessene Präsentation gilt, ist gänzlich vorstrukturiert, durch diejenigen, die die Infrastruktur, die Produktionsmittel und damit die Lebensformen beherrschen.

Während für alle drei Aspekte eine kritische Reflexion aussteht, ergibt sich die Fragestellung dieses Textes zu den Herausforderungen einer Schöpfung des Menschen als einem, der sich selbst übersteigen muss, für ein menschenwürdiges Leben, aus dem dritten Punkt. Die Versprechen des Metaversums greifen auf das humanistische Bild des Menschen als ein freies und selbstbestimmtes Wesen zurück. Was jedoch Freiheit und Selbstbestimmung kennzeichnet, liegt nicht bloß vor, sondern benötigt Bestimmung. Solche Festlegungen sind nicht historisch-kontingent. Sie sind an herrschaftsförmige Prozesse und Interessen gebunden. Die Möglichkeiten der Überschreitung körperlicher Grenzen durch gesellschaftliche Reorganisation in einem Metaversum bestimmen, welche Fähigkeiten zur freien Selbstbestimmung bereitgestellt werden. Körperlichkeit verschwindet vollständig hinter Daten und ökonomisch organisierter Wirklichkeit. Humanismuskritik setzt an diesem Punkt an und argumentiert, dass mit dem Körper ein wesentlicher Aspekt des Menschseins verabschiedet wird. Leibesgebundene Dimensionen wie Glück, Leid und Lust würden stattdessen erst das vervollständigen, was der entkörperte Humanismus verspricht: ein menschenwürdiges Leben. Das muss sich auch für das Metaversum übersetzen lassen.

---

[6] Vgl. Reckwitz, Andreas: Die Gesellschaft der Singularitäten. Berlin 2017, 25; 102.

[7] Vgl. Reckwitz: Gesellschaft, 285.

3 Humanismus: Versprechen auf ein menschenwürdiges Leben

Die Bestimmung dessen, was als Wesen des Menschen gilt, ist abhängig von den historisch-situierten gesellschaftlichen Verhältnissen und nicht umgekehrt. Das bedeutet, menschlichem Leben werden Rollen, Fähigkeiten, Funktionen und Ziele zugesprochen. Kontexte schreiben nicht nur etwas von außen ins Menschsein ein, sondern dort entsteht erst, was ‚der Mensch' ist. Die historisch-geistige Epoche des Humanismus (ca. 15. Jahrhundert) erhält diesen Namen dadurch, dass sie auf der Grundlage wissenschaftlicher Erkenntnisse ihrer Zeit, die göttliche Vorherbestimmung des Menschen bestritt. Das Wesen des Menschen und die Vorstellung eines menschenwürdigen Lebens besteht nachgehend in der Fähigkeit zur Selbstbestimmung, also Freiheit.

Dieser Essenz wird in der geisteswissenschaftlichen Entwicklung die Möglichkeit und Notwendigkeit von Eigentum vorausgesetzt. Dafür stehen namentlich Autoren wie Hobbes, Locke und später Kant. Über „ein äußeres Mein und Dein" heißt es bei Kant, kann das Einzelwesen souverän verfügen. Ohne ein solches äußeres Mein gäbe es keine Grundlage von freien Subjekten.[8]

Im Rahmen des Frühkapitalismus und der Industrialisierung kommt das Eigentum der Eigentumslosen in den Blick: der Besitz der eigenen Arbeitskraft. Freiheit bestehe auch in der doppelt freien Entäußerung dieser Kraft in der Zeit gegen Lohn. Mit der Dominanz des Kapitals unterliegt das, was als Mensch und als gelungenes Leben gilt, den Produktionsverhältnissen. Das Wesen des Menschen rückt damit ins Zentrum gesellschaftlich-ökonomischer Produktivkraft. Menschsein *verkörpert* die jeweiligen Produktionsbedingungen.[9] Heute ist die radikale Sakralisierung der Individualität wertvollstes Gut menschlichen Lebens – auch am Markt.[10]

Geistesgeschichtlich ist mit dem Humanismus zudem die Auseinandersetzung mit dem Dualismus zwischen Geist und Körper, Verstand und Leib verbunden. In der Frage nach selbstbestimmten, freien Individuen wird der Körper abgewertet als Form der Abhängigkeit von etwa Trieben, Affekten und physischen Grenzen. Die vermeintliche Utopie des Metaversums führt dies weiter, mit dem Ziel sich für ein humanes, menschenwürdiges Leben ganz von körperlichen Bedingungen zu lösen. Dies wird unverhohlen festgelegt durch diejenigen, die über die Produktionsmittel verfügen. In der virtuellen Projektion verkörpert der Mensch als Avatar die Bedingungen von Produktivkräfte der zukünftigen Gegenwart.

---

8 Zum Verhältnis von Freiheit und Eigentum: vgl. Govrin, Jule: Politische Körper. Berlin 2022, 27–30.

9 Vgl. Govrin: Politische Körper, 35.

10 Vgl. auch: Reckwitz: Gesellschaft, u. a. 308.

Adorno und Haraway zeigen, inwiefern 1) das Freiheitsversprechen humanistischer Positionen nicht eingelöst werden kann; 2) dass jede Bestimmung einer Essenz notwendig einen Überschuss jenseits der Bestimmung, was Mensch sei, mit sich bringt; 3) dass und inwiefern dieser Überschuss mit der ausgeschlossenen Körperlichkeit zusammenhängt; 4) inwiefern die Rückbesinnung auf Körperlichkeit die Möglichkeit eines menschenwürdigen Lebens denkbar macht; inwiefern also Körperlichkeit resistent ist gegenüber der Verkörperung von Produktion und welche Praktiken ein Leben jenseits dessen ermöglichen. Diese Blickpunkte werden im Folgenden mit einer Reflexion auf die Vorstellungen des Metaversums verwoben.

3.1 Falsche Versprechen, falsche Hoffnung

Theodor W. Adorno versteht unter einem „realen Humanismus“[11] die Verwirklichung des Glücks jeder:s Einzelnen als „menschenwürdigen Zustands.“[12] Seine Kritik an existierenden Vorstellungen von Humanismus richtet sich demnach nicht gegen das Versprechen und die Hoffnung des Humanismus per se. Adorno hält an der Hoffnung auf ein menschenwürdiges Leben durch die Autonomie jeder:s Einzelnen fest, aber: „Das objektive Ende der Humanität ist nur ein anderer Ausdruck fürs Gleiche. Es besagt, daß der Einzelne als Einzelner, wie er das Gattungswesen Mensch repräsentiert, die Autonomie verloren hat, durch die er die Gattung verwirklichen könnte.“[13] An diesem Zitat wird Adornos immanente Kritik am Versprechen des Humanismus offenbar: Das Allgemeine (das Gattungswesen) bestehe darin, ein Besonderes (einzelner Mensch) zu sein. Oder anders formuliert: Das Wesen des Menschen bestehe darin, keine allgemeine Wesensbestimmung zu besitzen. Das wäre ein Widerspruch zwischen Anspruch und Verwirklichung der nicht aufgelöst werden kann.

Ein ähnlicher Widerspruch steckt laut Adorno in der Notwendigkeit zu bestimmen, wie genau ein menschenwürdiger Zustand als nichtentfremdete Einzigartigkeit zu erreichen wäre. Es seien allgemeingültige Kriterien nötig, um Freiheit als Selbstverwirklichung eines Besonderen von den Vermögen eines Tieres, eines Steins oder einer Sklavin zu unterscheiden. Die schiere Notwendigkeit, einem Begriff wie dem der Autonomie eine Bedeutung zu geben, stehe vor der gleichen Dialektik wie bereits das erste Beispiel, dem Humanismus selbst eine Bedeutung zu geben. Das Besondere werde in ein allgemeines Muster gepresst und hebe sich dadurch auf.[14]

---

[11] Adorno, Theodor W.: Minima Moralia. Frankfurt a. M. 1951, 76.

[12] Adorno: MM, 117.

[13] Ebd., 40.

[14] Vgl. Adorno, Theodor W.: Negative Dialektik. Frankfurt a. M. 1973, 21.

Adornos Kritik an der Unmöglichkeit, essentielle Kriterien zu bestimmen, verbindet er mit der Überzeugung, dass die „Zerlegung des Menschen in seine Fähigkeiten“[15] nicht von gesellschaftlichen Herrschaftsverhältnissen getrennt gedacht werden könne. Zu spätkapitalistischen Produktions- und Tauschverhältnissen, die Adorno betrachtet, sei die Zerlegung des Menschen „eine Projektion der Arbeitsteilung auf deren vorgebliche Subjekte, untrennbar Interesse, sie mit höherem Nutzen einzusetzen, überhaupt manipulieren zu können.“[16] Das Individuum sei folglich hervorgebracht als „bloßer Agent des Wertgesetzes“[17], als die Auswahl von Fähigkeiten zum Nutzen gesteigerter Produktivität.

Vorstellungen des Metaversums gehen noch einen Schritt weiter. Sie zerlegen das Individuum nicht nur in spezifische Fähigkeiten und erklären die Nützlichen zu den Wesentlichen. Nützliche Fähigkeiten werden erst erzeugt, mit dem Versprechen, die vermeintlich wesentlichen Vermögen des Menschen zu überschreiten: „Sie expropriiert den Einzelnen, indem sie ihm ihr Glück zuteilt.“[18]

In der feministisch-materialistischen Auseinandersetzung mit einem essentialistischen Humanismus setzt Donna Haraway insbesondere an der Logik und Macht von Wissen an. Sie schaut auf die Rolle von Wissen(-schaften), die Anteil an der jeweiligen Formierung dessen haben, was als Individuum wahrgenommen wird. Wissen sei nicht unabhängig und rein, sondern historisch und lokal gebunden.[19] Das reicht über Adornos Essentialismuskritik hinaus. Feministische Erkenntniskritik schaut nicht auf die Unmöglichkeit gelingender Repräsentation, wie Adornos Hoffnung doch Einzelne *als* Einzelne darstellen zu können. Die Hoffnung auf gelingende Repräsentation des einzigartigen Subjekts wird mit Haraway unter den Vorbehalt gestellt, dass Darstellung von Etwas als Etwas an historisch situiertes Wissen gebunden sei. Dieses bilde lediglich Machtverhältnisse ab.[20]

Ähnlich wie Adorno, schaut Haraway somit ebenfalls auf die historischen Prozesse. In ihrem Fall steht dabei Wissen als diejenige Kraft, die ermöglicht, was als menschlich und was als nicht-menschlich gilt, im Mittelpunkt. Gentechnik etwa wähle spezifische Aspekte menschlicher Körper aus, die verbessert und verändert werden sollen und können.[21] Der Mensch werde zum Objekt der Organisation – auch über ökonomische Produktivität hinaus.

---

15 Adorno: MM, 76.

16 Ebd.

17 Ebd., 307.

18 Ebd., 77.

19 Vgl. Haraway, Donna J.: Situated Knowledge. In: Haraway, Donna J.: Simians, Cyborgs, and Women. London 1991, 183–202, hier: 188.

20 Vgl. Haraway: Situated Knowledge, 185.

21 Vgl. Haraway, Donna J.: A Cyborg Manifesto. In: Haraway, Donna J.: Simians, Cyborgs, and Women. London 1991, 149–182, hier: 152.

Haraways Verweis darauf, dass jede Theorie historisch situiert ist, erhebt damit auch einen starken Vorwurf gegenüber Adorno. Seine Auseinandersetzung mit der Unmöglichkeit von Repräsentation nimmt notwendig ein *Etwas* an, das nicht repräsentiert wird – das Nichtidentische mit der Identifikation des Menschenseins etwa. Haraway geht davon aus, dass die Bestimmung des Menschen nicht nur ein begriffliches Problem ist. Hierauf hatte Adornos dialektische Kritik sich beschränkt. Auf Materialität zurückführbare Aspekte wie Gentechnik und Codierung haben laut Haraway hingegen ebenfalls Anteil an der Zerlegung des Menschen. Semiotik sei lediglich ein Teil dieser Wesensbestimmung. Das falsche Versprechen des Humanismus auf freigestaltete Subjekte müsse somit „materiell-semiotisch" gefasst werden. Darunter ist die Untrennbarkeit von Beschreibung und Materialität zu verstehen: „Bodies as objects of knowledge are material-semiotic generated nodes."[22]

Mit Blick auf die Zurechtlegung des Subjekts im Metaversum muss auf die Unterschiede zwischen Adorno und Haraway in Bezug auf die Bedeutung der Materialität von menschlicher Körperlichkeit geachtet werden. Hieraus ergeben sich verschiedene Perspektiven und Lösungsansätze.

3.2 Jenseits der Versprechen

Gerade ein Begriff wie der der Freiheit, an dem hier so vieles hängt, ist ein stark emphatischer Begriff. Das heißt, in seiner Verwendung gehen dessen Gehalte nie vollständig auf. Es bleibt ein Überschuss möglicher weiterer Verwendungen des Begriffs.

Insbesondere Adornos Gesellschaftskritik ist immer auch Sprachphilosophie, die ihre kritische Kraft anhand von Analysen, wie die der Rolle emphatischer Begriffe in gesellschaftlichen Verhältnissen entfaltet. In begrifflicher Unabgeschlossenheit stecke das Potenzial, der verfestigten Bedeutung, sowohl des Begriffs als auch der dort eingeschriebenen gesellschaftlichen Verhältnisse, zu widerstehen.[23]

Metaverselle Humanität verspricht die Realisierung eines menschenwürdigen Lebens anhand der Fähigkeiten vernunftgeleiteter Selbstbestimmung. Dabei überschreiten diese Begriffe bisherige Grenzen und meinen damit Freiheit eine neue Bedeutung einzuschreiben. In der historischen Genese und der gegenwärtigen Verwendung des Begriffs lässt sich entlang Adornos Kritik am Humanismus der Überschuss in der Verdrängung des konkreten Körpers gegenüber der reinen Idee eines körperlos vernünftigen Wesens hervorheben. Aspekte wie Triebe, Emotionen und Affekte werden laut Adorno zugunsten eines vermeintlich rein geistigen Subjekts

[22] Haraway: Situated Knowledge, 201.

[23] Vgl. Adorno: ND, 153f.

verdrängt: „Die Trennung von Gefühl und Verstand […] hypostasiert die historisch zustandegekommene Aufspaltung des Menschen nach Funktionen."[24] Diese Aufspaltung repräsentiere nicht den ganzen Menschen, sondern sei eine von außen herangetragene Bestimmung zu vorausgesetzten Zwecken.

Der nächste, nun explizit gesellschaftskritische Schritt besteht darin, selbst in der begrifflichen Bestimmung der Instrumente der Vernunft ein Zurechtlegen zu entlarven. Vernunft werde zum bloßen Instrument einer Rationalität, die sich an der Herrschaft der Produktionsverhältnisse bemisst. Damit verbunden ist auch die spezifische Verkörperung des Menschen als Arbeitskraft. Was hingegen als Überschuss auftaucht, seien Aspekte wie Spiel und Lust, die mit zweckgerichteter, produktiver Verkörperung nicht vereinbar seien.[25] Ähnliches lässt sich über die Festlegung der Möglichkeiten zu Lust und Spiel im Rahmen spezifischer Funktionen im Metaversums sagen. Adorno muss sich letztlich fragen, welche Besonderheit an den körperlichen statt bloß geistigen Aspekten von Subjektivität, wie Emotionen und Spiel, ausgemacht werden kann, um sich resistent zu zeigen gegenüber der Schöpfung des Menschen als Funktion der Produktivkraft. Das geht nicht ohne einen Begriff von Materialität.

Hier lässt sich bereits mit Haraway anschließen. Sie sucht den Überschuss gegenüber essentialistischen Bestimmungen in den historisch-lokalen Wissensbeständen, verbunden mit konkreter Materialisierung des Menschen. Demnach ist der (menschliche) Körper zum rein sprachlich verkörperten Objekt geworden. Jedoch verschwindet der Körper nie ganz hinter der Bedeutung. Das drückt Haraway durch das Adjektiv ‚materiell-semiotisch' aus. Beispielhaft lässt sich diese Bedeutung von Körperlichkeit erfahren an zum einen der Gewalt aufgrund von Hautfarbe, zum anderen an Hungerstreiks, oder der Auseinandersetzung um Sexualitäten. In einem Fall wird der Körper zum Ziel von Gewalt, im anderen zum Mittelpunkt des Widerstandes. Körper sind unbestimmt und prekär, nicht nur als begriffliches Konzept.[26]

Gewalttätigen Körperpolitiken, die festschreiben wollen, was als lebbarer und tötbarer Körper oder was als wahre Sexualität gilt, hält Haraway keine alternativen, konkreten Körper entgegen. Stattdessen sieht sie den Überschuss gegenüber körperpolitischer Verkörperung in der beständigen Offenheit und Unbestimmbarkeit von Körpern, die nicht entweder resistent oder prekär sind, sondern werden. Dies entsteht und besteht in Kontexten.[27] Der Überschuss der Körper stecke nicht in dem, was nicht repräsentiert, was nichtidentisch ist. Er steckt in der Kritik an der oft gewaltsamen Setzung von vollständiger Repräsentation.

---

24 Adorno: MM, 262.

25 Vgl. ebd., 165.

26 Vgl. Haraway: Situated Knowledge, 197.

27 Vgl. Haraway: Cyborg Manifesto, 176.

Paradoxerweise wird mit dem Metaversum behauptet, genau diese Offenheit und den Wandel abbilden zu können. Fängt nicht der unbestimmte Raum des Digitalen den Überschuss auf? Haraway zufolge verdeckt diese Annahme den Kern von Repräsentationskritik: Simulation verliert die Materie aus dem Blick, die Teil hat am Wandel und auch zur digitalen Präsenz in Beziehung steht.[28] Die Unmöglichkeit virtuelle und materielle Darstellung zu trennen, straft auch den Anspruch eines Metaversums Lügen. Dass Körperlichkeit materiell-semiotisch ist, heißt, dass die Kriterien, welche abgebildet werden können, unfasslich und überschüssig bleiben. Sie sind es, weil Materialität von Bedeutung in fluider Beziehung zu virtueller Präsenz und sprachlicher Interpretation steht.[29] Das Besondere an Haraways Blick ist, dass Körperlichkeit nicht als das Gegenteil vom Virtuellen den Überschuss darstellt. Virtualität wird als eine Relation des Körpers unter vielen eingeordnet.

3.3 Resistente Körper

Der Körper als Überschuss gegen Festlegung und damit Fremdbestimmung spielt im humanistischen Versprechen „das Glück aller zu verwirklichen"[30] im Sinne Adornos eine entscheidende Rolle. Um zwischen Körper als begrifflichem Konzept und materieller Sache zu unterscheiden, verwendet Adorno den Begriff des Leibes. Subjekt erscheine in dialektischer Vermittlung von Leibhaftigkeit und Idee bzw. Konzept. Konkretes Handeln richte sich zwar nach Ideen, aber sei nicht ohne Bezug zu äußerer Natur und dem eigenen Leib zu vollziehen. Der Leib sei demnach nicht vollständig von Denken zu vereinnahmen. Er entziehe sich als Ort der Erfahrung sprachlicher und damit gedanklicher Konzeption. Dadurch sei der Leib die Erfahrbarkeit der Grenze des Denkens.[31] Lust, Leid und die Praxis der Mimesis ermögliche somit eine quasi-metaphysische Erfahrung: *dass das, was ist, nicht alles ist*. Dass also, was denkbar und sagbar ist, die Welt nicht vollständig abbilde. Diese körpergebundene und zugleich geistige Erfahrung ist die Voraussetzung für den Einspruch gegen Fremdbestimmung, das heißt, *dass es auch anders sein kann*.[32] Auf diese Weise biete der Leib als über den Begriff Hinausweisendes Widerstand: die Möglichkeit sich dem falschen Versprechen des Humanismus zu entziehen.

---

[28] Vgl. Kroker, Arthur: Body Drift. Minneapolis 2012, 127.

[29] Vgl. Kroker: Body Drift, 122.

[30] Schmidt, Alfred: Adorno – ein Philosoph des realen Humanismus. In: Schmidt, Alfred: Kritische Theorie, Humanismus, Aufklärung. Stuttgart 1981, 27–55, hier: 32.

[31] Vgl. Adorno: MM, 166.

[32] Vgl. ebd., 334.

Ausgehend vom resistenten Leib hält Adorno es für möglich, gewaltsame Subjektivierungsweisen zu benennen.[33] Auf diesem Weg wird auch ein Teil der epistemischen Gewalt von Tech-Unternehmen sichtbar. Sie ermöglichen nur ausgewählte Handlungs- und damit Subjektivierungsweisen, indem sie die virtuelle Dimension gegenüber leibhafter Praxis zum Ort eines humanitären Lebens erklären. Die Bedeutung, körperlicher Erfahrung als Erfahrung der Emanzipation wird dadurch verdrängt. Erst dieses Zurechtlegen des Subjektes macht es möglich, dieses der Verwertbarkeit zuzuführen.

Wie bereits gezeigt, verschiebt Haraway das Bild vom Köper anders als im Fall von Adornos Trennung von Körper und Leib. Der Körper lasse sich nicht als jenseits seines Begriffs gegeben charakterisieren. Haraways These lautet, dass ‚wir' nicht über einen konstanten Körper verfügen. Wer ‚wir' sind, ergibt sich temporär aus sich permanent neu zusammenfügenden Körpern. Das entblößt zunächst das Prekäre, Verwundbare am Körper. Selbst was das vermeintlich Materielle am Körper ist, nimmt Haraway als fluide und vom Kampf von körperpolitischer Festschreibung bedroht an. Der Körper wird als rassifiziert, als vergeschlechtlicht oder als Arbeitskraft und Tauschgut gelesen: „bodies are maps of power and identity."[34] Diese Perspektive zeigt jedoch nicht nur die Verwundbarkeit. Was ein Körper ist, lässt sich nicht festhalten. Er ist Schnittpunkt zwischen gesellschaftlich, sexualisiert, vegetativ, bakteriell, mineralisch, erzählerisch, arbeitend, technologisch verändert und virtuell.[35] ‚Den' Körper gibt es nicht, sondern Relationen, die sich ohne Pause verschieben. Diese Verschiebung sei kein friedlicher Prozess, sondern immer in Machtverhältnisse gebettet. Die virtuelle Version des Körpers spielt dabei auch eine Rolle; aber eben nur *eine*. Was ‚der' unveränderliche Körper ist, kann auch kein Metaversum festhalten. Es kann nur seine Gewalt entblößen. Die Resistenz des Körpers steckt bei Haraway ungebremst in dessen Hybridität.

Haraway zufolge ist der Körper resistent, da es nur temporäre, ungebremste Kompositionen gibt. Mit Com-post benennt Haraway das einzige *post*, das sie zu denken vermag. Im Kompost findet sich eine Vielfalt von Gestalten, mehr-als-menschliche, die sich permanent durch gemeinsame Beziehungen wandeln, Neues schaffen und vergehen. Eine davon wäre auch das Metaversum. „I am a compostist, not a posthumanist: we are all compost, not posthuman."[36]

---

33 Vgl. ebd., 7.

34 Haraway: Cyborg Manifesto, 180.

35 Vgl. Kroker: 14.

36 Haraway, Donna J.: Staying with the Trouble. Durham 2016, 101.

3.4 Praktiken leibhafter Erlösung

Sowohl Adorno als auch Haraway zufolge geht der Widerstand gegen verkörperte Produktivität als funktional erzeugtes Bild des freien Menschen vom nicht-ganz-erfassten Körper aus. Worin dieser besteht, unterscheidet sich bei Haraway und Adorno. Dadurch werden zwei verschiedene Orte hervorgehoben, auf die widerständige Körper treffen: herrschaftsförmige Subjektivierungsformen gegen den Leib und gewaltsame, körperpolitische Praktiken entgegen Kompositionen. Beide Theoretiker:innen benennen jedoch auch, wie von der Beschreibung der Resistenz zu Praktiken ihrer Verwirklichung übergegangen werden kann.

Der Blick auf den eigenen Leib ist für Adorno keine einfache Praxis. Von ihr gehe jedoch der Einspruch gegen bestehende Verhältnisse aus. Möglich sei eine solche Praxis vom Blick auf den eignen Tod aus. Nur der Tod entziehe sich dem verdinglichten Leben.[37] Der eigenen „Hinfälligkeit eingedenk"[38] werde das Ende von Verdinglichung ersichtlich. In dieser Praxis der Selbstreflexion realisiere sich das Potenzial der Leiblichkeit. Einen „ernsthaft klugen Menschen"[39] sieht Adorno somit in demjenigen, der sich auf die Grenzen des Lebens richtet und dadurch „sich als Natur durchschauende Natur"[40] anerkenne. Somit gehe es nicht darum, Bedingungen einer von Natur gelösten Existenz zu entdecken. Jeder Versuch dazu, schreibe nur neuen, naturhaften Zwang ins Leben ein. Allein in leibhafter Entsagung und Anerkennung der eigenen Natur stecke Herrschaftsfreiheit und damit „Erlösung"[41] beziehungsweise „leibhafte Auferstehung"[42]. Diese Begriffe wählt Adorno bewusst am Ende seiner zentralen Schriften *Minima Moralia* und *Negative Dialektik*. Explizit hebt er die Bedeutung der Untrennbarkeit von Geistigem und Leiblichem und die Hoffnung auf leibhafte Erfüllung im Christentum gegenüber dem von allem Körperlichen geschiedenen Astralleib des Okkultismus hervor: Letztere bezeichnet Adorn als „[die] Metaphysik der dummen Kerle"[43] unter gesellschaftlichen Verhältnissen der Arbeitsteilung im Spätkapitalismus. Adorno betont zugleich, dass sich die christliche Hoffnung leibhafter Auferstehung „durch deren Vergeistigung ums Beste […] gebracht weiß."[44] Jeder Versuch, dem Tod und dem Leib eine transzendente Bedeutung zuzuschreiben, verdränge deren Anteil an einem emanzipatori-

37 Vgl. Adorno: MM, 96.
38 Adorno: MM, 22.
39 Ebd., 264.
40 Schmidt: Realer Humanismus, 49.
41 Adorno: MM, 333 f.
42 Adorno: ND, 393.
43 Adorno: MM, 325 f.
44 Adorno: ND, 393.

schen Wissen von Freiheit.[45] Wer hingegen auf spekulative Metaphysik und auf Perspektiven jenseits der Grenzen des leibhaften Menschen verzichte, könne auf eine Welt ohne Herrschaft blicken.

Die Vorstellung leibhafter Erlösung wird Haraway wohl nicht über die Lippen bringen. Wobei sich auch bei ihr kritische Anerkennung der Aufwertung des Fleisches zum Denken von Freiheit findet: „My soul marked indelibly by Catholic formation, I hear in species the doctrine of the Real Presence under both species, bread and wine, the transubstantiated signs of the flesh" schreibt Haraway in ihrem *Campanion Species Manifesto*. Sie bezieht sich damit laut Kroker auf "the deep epistemology of Roman Catholicism, with its doubled liminality—sign–flesh, grace–bodies, sign–corporeality."[46]

Jedoch geht Emanzipation von Verhältnissen der Verkörperung nach Haraway von keinem *göttlichen* Standpunkt aus, sondern vom Standpunkt des Körpers selbst. Sie versteht darunter Intersektionen vielfältiger, materiell-semiotischer Relationen. Eine Praxis selbstbestimmten Lebens heißt dann, sich auf die immanenten Beziehungen einzulassen, in denen wir stehen. Daraus entspringt mehr als nur eine andere Perspektive auf eine Welt ohne Herrschaft, „wie sie vom Standpunkt der Erlösung aus sich darstellt."[47] Haraway weist auf eine lebendige Praxis, die ein (mehr-als-)menschenwürdiges Leben ermöglicht. Leben ist mehr-als-menschlich, da Körper mit nicht-menschlichen Gefährt:innen – Hunden, Tauben, Algorithmen, Bakterien – in Beziehungen stehen und entstehen.[48]

Von hieraus auf gegenwärtige Beziehungen wie die der Vorstellung eines Metaversums zu blicken, macht ebenfalls sichtbar, welche Beziehungen entwertet, vernichtet oder nach instrumentellen Zwecken kreiert werden. Jedoch wird mit Haraway nicht die mögliche Existenz eines Metaversums an sich zum Problem für ein mehr-als-menschenwürdiges Leben. Wie bereits gezeigt wurde, wäre das Metaversum auch nur Teil des Kompostes, der Wirklichkeit heißt. Haraway wehrt sich gegen eine Auf- und Abwertung spezifischer Weisen eines In-der-Welt-Seins. Für sie gibt es nur ein Mit-der-Welt-Werden.[49]

Ihre Betrachtung bewegt sich somit weg vom Einzelnen als kleinste Einheit der Welt, dessen Glück verwirklicht werden soll. Relationalität anzunehmen, ermögliche der Suche nach Glück eine neue Praxis. In responsiven Beziehungen zu bestehen, bedeutet für Haraway besondere

---

45 Der Gemeinschaft stiftende leidende Körper Jesu verliere in den Schriften des Apostels Paulus durch die nun sakralisierte Symbolkraft des Körpers den kritischen Einspruch des real leidenden, leiblichen Körpers. So führt jedenfalls Govrin die theologische Figur des Doppelkörpers als gleichsam verwundbar und unsterblich weiter aus. Vgl. Govrin: Politische Körper, 20.

46 Kroker: Body Drift, 135.

47 Adorno: MM, 333.

48 Vgl. Haraway: Staying with the Trouble, 16.

49 Vgl. ebd., 58.

Fähigkeiten der Verantwortung zu praktizieren. Response-ability ist der Begriff, den Haraway für diese ethische Praxis übernimmt. Sie versteht darunter sowohl die Fähigkeit zu antworten als auch die Verantwortung Andere(s) zum Antworten zu befähigen; also Anderem zu ermöglichen, Teil einer responsiven Beziehung zu werden.[50] Dem entgegen stehen Praktiken, die gewisse Formen des Antwortens ausschließen. Das gilt etwa dann, wenn Leben rein ins Virtuelle verschwinden soll.

4 Schluss

Die vermeintliche Utopie eines Metaversums erschafft, formt und schreibt ein spezifisches Bild des Menschen fest, um sein humanistisches Versprechen auf ein menschenwürdiges Leben durch die Überschreitung von Grenzen zu ermöglichen. Derartige schöpferische Festlegungen sind notwendig an Erwartungen und Funktionen gebunden. Insbesondere im Falle des Metaversums ist die Vorstellung des Menschseins demnach von den ökonomischen Interessen der GAFAM Tech-Giganten abhängig.

Die Praktiken der Emanzipation davon erfolgen über den Blick auf die Resistenz des Körpers in zweierlei Weisen: zum einen ausgehend von Adorno im Zustand der Selbstreflexion auf die eigene leibliche Naturhaftigkeit. Das Bewusstsein über die eigenen Grenzen erlaubt es, das Versprechen (virtueller) Grenzenlosigkeit als Ideologie zu durchschauen. Daraus kann ein Blick auf eine Welt ohne Herrschaft hervorgehen. Zum anderen ermöglicht Haraways Blick auf den Körper, die eigene Verwobenheit mit anderen Körpern anzunehmen. Darüber kann sich eine lebenswürdige, ethische Praxis der Response-ability etablieren. Somit stellt sich die humanistische Frage nach menschenwürdigem Leben anders: Wie ist Relationalität weniger gewaltsam lebbar, in einer Gegenwart, die längst mehr-als-menschlich ist, was sich auch in der lebendigen Vorstellung des Metaversums spiegelt?

Insofern lässt sich diese Untersuchung auch als genuin sozialethisch verstehen: als kritische Untersuchung der Möglichkeiten und Grenzen in Würde zu leben. Eine kritische Praxis, die auf ein mehr-als-menschenwürdiges Leben zielt, muss sich darauf einlassen, dass die zunehmend virtuelle Realität sich nicht verhindern lässt. Dann ist sie in der Lage, die Pathologien, die eine Entwicklung hin zum Metaversum mit sich bringen können, zu erkennen und zu intervenieren. Ausgangspunkt dafür ist der Blick aus resistenter Körperlichkeit gegen funktionale Instrumentalisierung. Körper sind der Schlüssel eines Aufbegehrens und der Verweis auf eine ethische Praxis, die ihre normativen Maßstäbe nicht aus einer entkörperten Vernunft rekonstruiert. Die Prekarität fluider Beziehungen und Respons-ability des Gemeinsamen-Wer-

[50] Vgl. ebd., 114.

dens als Maßstab zu wählen, ermöglicht ein mehr-als-menschenwürdiges Leben zu führen in Zeiten vom und auf dem Weg zum Metaversum.

Literaturverzeichnis

Adorno, Theodor W.: Minima Moralia. Reflexionen aus dem beschädigten Leben. Frankfurt a. M. 1951.

Adorno, Theodor W.: Negative Dialektik. Frankfurt a. M. 1973.

Ball, Matthew: The Metaverse. And How It Will Revolutionize Everything. New York 2022.

Becka, Michelle: Kritik und Solidarität. Zu einem sozialethischen Verständnis von Kritik. In: Beck, Michelle/Emunds, Bernhard/Eurich, Johannes u. a.: Sozialethik als Kritik. Baden-Baden 2020, 19–55.

De stefano, Valerio u. a.: The Metaverse is a labour issue. Online unter: https://socialeurope.eu/the-metaverse-is-a-labour-issue (Stand: 29.08.2022).

Farman, Abou: On Not Dying. Secular Immortality in the Age of Technoscience. Minneapolis 2020.

Govrin, Jule: Politische Körper. Von Sorge und Solidarität. Berlin 2022.

Haraway, Donna J.: A Cyborg Manifesto. Science, Technology, and Socialist-Feminism in the Late Twentieth Century. In: Haraway, Donna J.: Simians, Cyborgs, and Women. The Reinvention of Nature. London 1991, 149–182.

Haraway, Donna J.: Situated Knowledge. The Science Question in Feminism and the Privilege of Partial Perspective. In: Haraway, Donna J.: Simians, Cyborgs, and Women. The Reinvention of Nature. London 1991, 183–202.

Haraway, Donna J.: Staying with the Trouble. Making Kin in the Chthulucene. Durham 2016.

Kroker, Arthur: Body Drift. Butler, Hayles, Haraway. Minneapolis 2012.

Reckwitz, Andreas: Die Gesellschaft der Singularitäten. Zum Strukturwandel der Moderne. Berlin 2017.

Schmidt, Alfred: Adorno – ein Philosoph des realen Humanismus. In: Schmidt, Alfred: Kritische Theorie, Humanismus, Aufklärung. Stuttgart 1981, 27–55.

Autor:innenverzeichnis

Brand, Lukas
Lukas Brand ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Religionsphilosophie und Wissenschaftstheorie an der Katholisch-Theologischen Fakultät der Ruhr-Universität Bochum. Dort lehrt und forscht er zu den Themen Anthropologie und Ethik der Künstlichen Intelligenz und der Virtuellen Realität.
E-Mail: lukas.brand@ruhr-uni-bochum.de

Klinge, Hendrik
PD Dr. Dr. Hendrik Klinge ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Historische und Systematische Theologie der Bergischen Universität Wuppertal. Seine Forschungsschwerpunkte sind die Theologische Metaethik, Klassische lutherische Theologie, die Religionsphilosophie Immanuel Kants sowie Posthumanismus aus theologischer Perspektive.
E-Mail: klinge@uni-wuppertal.de

Kunkel, Nicole
Nicole Kunkel promoviert mit einem Promotionsstipendium des Evangelischen Studienwerks Villigst e.V. am Lehrstuhl für Systematische Theologie mit dem Schwerpunkt Ethik und Hermeneutik an der Humboldt-Universität zu Berlin. In ihrer Forschung beschäftigt sie sich mit Fragen an der Schnittstelle von Friedensethik und Künstlicher Intelligenz, verkörpert in letalen autoregulativen Waffensystemen.
E-Mail: nicole.kunkel@student.hu-berlin.de

Navon, Mois
Mois Navon ist Informatiker, Moralphilosoph und orthodoxer Rabbiner. In seiner Doktorarbeit an der Schnittstelle von Naturwissenschaft und Philosophie, die er an der Bar Ilan Universität Tel Aviv absolviert, wendet er jüdische Philosophie im Blick auf ethische Zugänge zu Künstlicher Intelligenz an. In diesem Zusammenhang unterrichtet er „Ethik in Big Data und KI“ an der Ben Gurion Universität in Be'er Sheva und an der Yeshiva Universität in New York.
E-Mail: mois.navon@divreinavon.com

Nyholm, Sven
Dr. Sven Nyholm ist Professor für Ethik der Künstlichen Intelligenz an der Ludwig-Maximilians-Universtität München. In seiner aktuellen Forschung konzentriert er sich als Philosoph auf ethische Fragen zu Künstlicher Intelligenz und Robotern. Sein neuestes Buch ist This is Technology Ethics: An Introduction (Wiley-Blackwell, 2023).
E-Mail: s.nyholm@lmu.de

Ohly, Lukas
Prof. Dr. Lukas Ohly ist Professor für Systematische Theologie und Religionsphilosophie an der Goethe-Universität in Frankfurt am Main. Seine Forschungsschwerpunkte sind Ethik der Künstlichen Intelligenz und Robotik, Grundlegung der Ethik, Phänomenologie der Gotteserfahrung und Theologische Kategorienlehre.

Puzio, Anna
Dr. Anna Puzio, hat Katholische Theologie, Germanistik und Philosophie in Münster und München studiert. In München hat sie zur Anthropologie des Transhumanismus promoviert. Zu ihren Forschungsthemen gehören die Technikanthropologie und Technikethik, Robotik, reproduktive Technologien, queere KI, Neuer Materialismus und Umweltethik. Nach Stationen in München, Frankfurt am Main und Wien arbeitet sie nun als Postdoctoral Researcher im niederländischen ESDiT Research Programme (Ethics of Socially Disruptive Technologies) an der University of Twente.
Homepage: www.anna-puzio.com. E-Mail: a.s.puzio@utwente.nl

Reiners, Simon
Simon Reiners ist Wissenschaftlicher Mitarbeiter am Nell-Breuning-Institut für Wirtschafts- und Gesellschaftsethik und Doktorand am Institut für Sozialphilosophie in Frankfurt am Main. Seine Forschungsschwerpunkte liegen auf der Frankfurter Schule, Feministischer Erkenntniskritik, Neuen Materialismen und der Zukunft der Arbeit.
E-Mail: reiners@sankt-georgen.de

Schlote, Yannick
Dr. Yannick Schlote ist Vikar und Wissenschaftlicher Mitarbeiter am Lehrstuhl für Systematische Theologie und Ethik an der Ludwig-Maximilians-Universität in München. Er arbeitet zudem am Institut Technik-Theologie-Naturwissenschaften im ethischen Projekt Bavarian Genome und hat in seiner Dissertation eschatologische Narrative zu Künstlicher Intelligenz untersucht.
E-Mail: yannick.schlote@elkb.de

Smith, Katherine G.
Dr. Katherine G. Schmidt ist Associate Professor und Leiterin des Studienprogramms Theology and Religious Studies an der Molloy University in New York. Ihre Doktorarbeit hat sie an der Theologischen Fakultät der Universität in Dayton absolviert. In ihrer derzeitigen Forschung beschäftigt sie sich mit Theologie und Religion in der amerikanischen Kultur. Ganz besonders interessiert sie sich für die Schnittstelle von Religion/Theologie und Technik, dem Internet und Medien.

Tirosh-Samuelson, Hava
Hava Tirosh-Samuelson ist Regents Professor of History, Irving und Miriam Lowe Professor of Modem Judaism, sowie Direktorin des Center for Jewish Studies an der Arizona State University. Als Historikerin beschäftigt sie sich mit Mystik jüdischer Philosophie, Religion, Technik und Ökologie. Sie ist ist Hauptredakteurin der Library of Contemporary Jewish Philosophers (2012–2018), bestehend aus 21 Bänden.
E-Mail: Hava.Samuelson@asu.edu

Tretter, Max
Max Tretter ist Wissenschaftlicher Mitarbeiter der Evangelischen Theologie am Lehrstuhl für Systematische Theologie II (Ethik) an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Seine Forschungsschwerpunkte liegen neben der Ethik des Digitalen und der Künstlichen Intelligenz im Bereich der politischen Ethik und der Kulturethik.
E-Mail: max.tretter@fau.de

Winter, Dominik
Dominik Winter ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Theologische Ethik der Katholisch-Theologischen Fakultät der Ruhr-Universität Bochum. Seine Forschungsschwerpunkte sind Moral Enhancement und Emotionen im Kontext theologischer Metaethik.
E-Mail: dominik.winter-y8w@rub.de

Alexa, How Do You Feel About Religion? Theological Approaches to Technology and Artificial Intelligence

Theology and Artificial Intelligence, Vol. 1

Technik und Künstliche Intelligenz gehören zu den brisanten Themen der gegenwärtigen Theologie. Dieses Buch erforscht die drängenden Fragen unserer Zeit. Dazu begibt sich die Theologie in Dialog mit den Technikwissenschaften. Untersucht werden die Veränderungen des Menschenbildes durch Roboter, Religiöse Roboter, Körperoptimierung, medizinische Technologien, Autoregulative Waffensysteme und Transformationen der Theologie. Aus interdisziplinärer Perspektive stellen die Texte neue Forschungsergebnisse aus dem internationalen Raum vor.

Dr. Anna Puzio forscht als Theologin und Philosophin zur Technikanthropologie, Technikethik und Umweltethik im Forschungsprogramm *Ethics of Socially Disruptive Technologies* (ESDiT) an der University of Twente.
Nicole Kunkel promoviert am Lehrstuhl für Ethik und Hermeneutik der Humboldt-Universität zu Berlin zu autoregulativen Waffensystemen an der Schnittstelle von Friedensethik und KI.
PD Dr. Dr. Hendrik Klinge ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Historische und Systematische Theologie der Bergischen Universität Wuppertal und forscht zur Theologischen Metaethik und Klassischen lutherischen Theologie.

www.wbg-wissenverbindet.de
ISBN 978-3-534-40782-8

wbg Academic