
Brand | Kutz | Winter [Hrsg.]

Instrumente Gottes

Soziale Roboter und KI-Tools
in religiösen Kontexten

ALBER THEOLOGIE



Alber Theologie

Band 4.1

Lukas Brand | Martin Kutz
Dominik Winter [Hrsg.]

Instrumente Gottes

Soziale Roboter und KI-Tools
in religiösen Kontexten

ALBER THEOLOGIE



Die Open-Access-Veröffentlichung dieses Titels wurde durch die Dachinitiative »Hochschule digital Niedersachsen« des Landes Niedersachsen ermöglicht.

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

1. Auflage 2025

© Die Autor:innen

Publiziert von

Verlag Karl Alber – ein Verlag in der
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.verlag-alber.de

Gesamtherstellung:

Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-495-99382-8

ISBN (ePDF): 978-3-495-99383-5

DOI: <https://doi.org/10.5771/9783495993835>



Onlineversion
Nomos eLibrary



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Inhaltsverzeichnis

Lukas Brand, Martin Kutz, Dominik Winter

Instrumente Gottes

Soziale Roboter und KI-Tools in religiösen Kontexten 9

Praktisch-Theologische Perspektiven 31

Marcel Scholz

BINA48 and Religious Education in the Context of Artificial Intelligence 33

Steffi Fabricius

Robocatio: Digitale Zeugen des Glaubens?
Roboter und KI im modernen Religionsunterricht 53

Marius de Byl

„Roboter, unterrichte für mich!“
Religionspädagogische Erkundungen in den Kontexten
Künstlicher Intelligenz und Sozialer Robotik. 75

Simon N. Balle

Limitations for Pastoral Robots 97

Ralf Peter Reimann

Praying with Avatars
A Response to Germany's First AI Worship Service and its
Theological and Anthropological Implications. 109

Ilona Nord und Thomas Schlag

Robotik in der christlichen Gemeindegarbeit

Praktisch-theologische Ausgangspunkte, Reflexionshorizonte
und Perspektiven 123

Systematisch-Theologische Perspektiven 155

Theresa M. Thunig

The Final Frontier

Investigating the Impact of Science Fiction on our
(Technological) Future 157

Ramy Abdin

What are Robots?

An Islamic Approach from a Metaphysical Understanding of the
Human Being 173

Jonas Simmerlein

Doomed to fail?!

Three Inherent Problems Haunting Religious Robotics 197

Max Tretter

“Where do all the calculators go?”

Methodological Considerations About how to Think About
Robots and Their Afterlife 215

Noreen Herzfeld

The Banality of Artificial Evil 241

Hille Haker

**Der Einsatz sozialtherapeutischer Roboter in der
Betreuung und Pflege älterer Menschen**

Theologisch-Ethische Anfragen 257

Außertheologische Perspektiven 279

Manishk Gawande et al.
From SanTO to CeleSTE
A Theomorphic Device for Religious Practice 281

Hildelith Leyser and Kate Cooper
Motion, Mirror Neurons and the Human–Robot
Relationship 297

Autorinnen und Autoren 315

Lukas Brand, Martin Kutz, Dominik Winter

Instrumente Gottes

Soziale Roboter und KI-Tools in religiösen Kontexten

1. Theologie nach der spirituellen Maschine

Steht die Theologie angesichts der qualitativen Veränderung der gegenwärtigen digitalen Technik erneut vor einer paradigmatischen Neuausrichtung zentraler Fragestellungen?¹ Spätestens seit der Aufklärung setzt sich die Theologie mit ihrem Verhältnis zu den anderen Wissenschaften auseinander. Diese Auseinandersetzung war umso drängender geworden, nachdem sich die Naturwissenschaften als eigenständige Disziplinen von ihrer Verbindung zur Theologie und Offenbarung als primärer Erkenntnisquelle bei der Erforschung der Natur gelöst hatten. In der Auseinandersetzung mit den Naturwissenschaften schienen vor dem Hintergrund dieser Veränderung Glaubensüberzeugungen in Frage zu stehen, die zum Beispiel die Entstehung des Menschen und die Ordnung des Kosmos berühren. Heute beschäftigt sich besonders die systematische Theologie auch mit der Frage der epistemologischen Reichweite, etwa der Biologie oder der Physik. Im Zuge derselben Entwicklung hat sich spätestens im 20. Jahrhundert auch die theologische Ethik, die bis dahin vorwiegend der theologischen Betrachtung des Menschen und seines Handelns in der Welt galt, zunehmend der Auseinandersetzung mit den Humanwissenschaften zugewandt. Die Reflexion des vorwiegend materialistischen und rationalistischen Menschen- und Gesellschaftsbildes der Biologie, Medizin, aber auch der Politik und Wirtschaft ist heute selbstverständlicher Teil der theologischen Ethik. Seit einigen Jahren beschäftigen sich die Sozialethik und die Fächer der praktischen Theologie aber auch vermehrt mit den Erfor-

1 Vgl. *Platow*: Theolo:KI.

dernissen und Lebensrealitäten der Menschen im Zeitalter der Globalisierung und Digitalisierung. Abgesehen von einigen offensichtlichen Beispielen wie der Atomtechnik, hielt man bis in die zweite Hälfte des 20. Jahrhunderts hinein technische Artefakte und ihre Auswirkungen für relativ beherrschbar. Nicht zuletzt mit dem Bericht des *Club of Rome* (1972) wurde der Menschheit jedoch erstmals bewusst, dass die Gestaltung der ökonomischen, politischen, natürlichen und sozialen Systeme einer globalen Anstrengung bedarf.² Die wechselseitige Abhängigkeit dieser Systeme ist maßgeblich von der Reichweite der Technik geprägt,³ deren Platz heute weitgehend von einer vernetzten Digitaltechnik eingenommen wird. Diese wiederum wird, dem Trend der letzten Jahre folgend, auch in den kommenden Jahren schrittweise einer auf künstlicher Intelligenz (KI) basierenden Digitaltechnik weichen.⁴

Die heute zunehmend relevant gewordenen digitaltechnischen und KI-gestützten Artefakte wirken sich aber nicht nur global auf das Klima oder internationale Wirtschaftsbeziehungen und Finanzströme aus,⁵ sondern auch unmittelbar auf alle menschlichen Lebensbereiche und Praktiken, einschließlich der Religion, sowie auf das menschliche Selbst- und Weltbild im Allgemeinen. Wenn der Theologie die Weisung Gottes an die Menschen, Walter:innen der Schöpfung zu sein (Gen 1,26), eine innewohnende Maxime ist, wird sie sich in Zukunft auch den technischen Praktiken und Artefakten sowie deren Zwecksetzungen zuwenden müssen, mit denen Menschen gegenwärtig sich und ihre Umwelt gestalten und verwalten. Nur so kann die Theologie auch in Zukunft noch überzeugend darauf Einfluss nehmen, dass diese Artefakte dem göttlichen Auftrag angemessen sind.

Während sich die Digitalisierung offenbar im Laufe des vergangenen Jahrzehnts auch in der deutschen Theologie als Gegenstand und Methode fest verankert hat,⁶ ist der gerade in der Konsolidierungs-

2 Vgl. *Meadows/Meadows/Randers* u.a.: *Limits of Growth*, 9.

3 Vgl. ebd., 129–155.

4 Vgl. *Bughin/Hazan/Ramaswamy* u.a.: *The Next Digital Frontier?*; *Holmström*: *From AI to digital transformation*.

5 Vgl. *Bender/Gebbru/McMillan-Major* u.a.: *Can Language Models be Too Big?*

6 Vgl. etwa *Beck/Nord/Valentin*: *Theologie und Digitalität*; *Pirker/Pišonić*: *Virtuelle Realität und Transzendenz*, *Kistler/Puzio/Riedl* u.a.: *Digitale Transformationen*.

phase sich befindende Bereich der Robotik und künstlichen Intelligenz nicht einfach nur ein logischer nächster Schritt der bisherigen Entwicklung, sondern eine transformative Technik, die entschieden mehr leistet, als z.B. die Möglichkeiten der zwischenmenschlichen Kommunikation zu skalieren. Die Theologie wird sich daher im 21. Jahrhundert auf allen Ebenen vermehrt ihrem Verhältnis zu den Welt- und Menschenbildern auch der Ingenieurwissenschaften, das heißt den Wissenschaften der Kybernetik (der Regelungsmechanismen),⁷ des Technikdesigns und der Technikproduktion widmen und ihr Interesse den technischen Artefakten zuwenden müssen, die auf den Grundlagen der besagten Wissenschaften aufbauen und menschliche Praxis in ihrer Gänze unmittelbar mitgestalten.

2. Heilige Maschinen gestern und heute

Bereits im 16. Jahrhundert gab es mit dem „Mechanischen Mönch“⁸ einen auf die religiöse Praxis ausgerichteten Automaten. Die sich auf Rädern bewegende Figur eines Heiligen fügte dem Instrumentarium der Devotionalien buchstäblich eine neue, „automatische“ Dimension hinzu. Heute berührt aus dem Bereich der Ingenieurwissenschaften insbesondere das Feld der *Social Robotics*, also die Entwicklung von Robotersystemen, die soziale Interaktionen simulieren sollen, unweigerlich auch den Bereich der Religion. Dabei werden digitale Systeme nicht einfach nur um religiöse Inhalte ergänzt oder spirituelle Praktiken und Angebote „digitalisiert“. Vielmehr werden vermehrt computerbasierte Robotersysteme entwickelt, die wie der mechanische Mönch die religiösen Ausdrucksmöglichkeiten erweitern sollen. Im Sinne der Akteur-Netzwerk-Theorie⁹ wird damit das Netzwerk spiritueller Assoziationen um technische Akteure erweitert, die sich explizit an Nutzer:innen in einem religiösen Kontext¹⁰ oder an ihr spirituelles Bedürfnis, etwa nach Unsterblichkeit, Trans-

7 Vgl. Wiener: *Cybernetics and Society*.

8 Vgl. King: *Clockwork Prayer*.

9 Vgl. Latour: *Akteur-Netzwerk-Theorie*.

10 Vgl. Trovato u.a.: *The Creation of SanTO*; Trovato u.a.: *Communicating with SanTO*.

zendenz oder schlicht der Wiedervereinigung mit Verstorbenen richten.¹¹

Der Schwerpunkt der Entwicklung von Robotik-Systemen für die religiöse Praxis lag zunächst in ostasiatischen Ländern, in denen Roboter mit weniger Vorurteilen behaftet zu sein scheinen: Beispiele, die hier Bekanntheit erlangt haben, sind etwa Xian'er, der im chinesischen Longquan-Tempel Fragen der Besucher:innen zum Buddhistischen Glauben beantwortet,¹² Mindar, der als Verkörperung eines Bodhisattva gilt und im japanischen Kodaiji-Tempel den Gästen buddhistische Lehren und Sutren vorträgt,¹³ oder der handelsübliche, für Mensch-Maschine-Interaktionen entwickelte und personalisierbare, humanoide Roboter Pepper, der auf einer japanischen Fachmesse für die Durchführung von buddhistischen Beerdigungsriten ausgestattet und angeboten wurde.¹⁴ Im christlichen Kulturbereich waren der Segensroboter BlessU-2¹⁵ – ein umfunktionaler Geldautomat mit Armen und Gesicht, der im Rahmen der Weltausstellung zum Reformationsjahr 2017 Nutzer:innen auf Anfrage einen Segen zusprach – und Gebetsassistenten wie SanTO – eine einem Heiligen nachempfundene, computergesteuerte Figur, die über ein Sprachprotokoll Nutzer:innen Fragen zum Glauben und spirituelle Anliegen beantworten kann¹⁶ – vergleichsweise neue Phänomene.

Die Verbreitung von großen, öffentlich zugänglichen Sprachmodellen forciert seit 2022 auch die Entwicklung von spirituellen Begleitern für den Alltag und Anwendungen für dezidiert religiöse Kontexte im Internet bzw. in Form von im wesentlichen digitalen Anwendungen. Dabei sind nicht nur an Religion interessierte Computer- und Robotikwissenschaftler:innen an der Entwicklung beteiligt: Immer häufiger werden diese Systeme auch von technisch ausgebildeten oder versierten Theolog:innen bzw. Angestellten im kirchlichen Bereich oder kirchennahen Organisationen unterhalten, so zum Beispiel der virtuelle Luther-Avatar¹⁷ oder der digitale Apo-

11 Vgl. Brand/Zweig: Conundrum; Watts/Wilks: Spiritual Conversation; Riesewick/Block: Digitale Seele; Figueroa-Torres: Affection as a service.

12 Vgl. Ke: Robot Monk Xian'er.

13 Vgl. Baffelli: Robots, AI and Buddhism.

14 Vgl. Martin: Pepper the robot.

15 Vgl. Löffler: Blessing Robot BlessU2.

16 Vgl. Trovato u.a.: Communicating with SanTO.

17 Vgl. Reimann: Martin Luther im 21. Jahrhundert.

loget Father Justin.¹⁸ Viele spirituelle Akteure entstehen als Kunstinstallationen, Kurzzeitprojekte oder Experimente: Mit ihrer Installation „The Prayer“ (Centre Pompidou, 2020), einer Maschine, die Gebete generiert und rezitiert, wirft die Künstlerin Diemut Strebe die Frage auf, ob Roboter eine religiöse Praxis ausüben und dabei selbst religiöse Erfahrungen machen können;¹⁹ der KI-Priester in einer Reportage des MDR²⁰ und der im Beichtstuhl der Peterskapelle in Luzern installierte Jesus-Avatar werfen unweigerlich die Frage auf, ob KI-gestützte Akteure Sakramente spenden können.²¹ Andere Systeme, wie der Bibelstudienassistent Nikodemus.AI²², sind aber auch als bleibende Angebote konzipiert und werden genau wie der Gebetsassistent SanTO²³ stetig verbessert und weiterentwickelt. Sie berühren dabei nicht einfach nur ganz praktisch die Sphäre des religiösen Ausdrucks und der individuellen Spiritualität. Ihnen sind zentrale religiöse Motive und Dimensionen zu eigen, die teils auch von den Entwickler:innen explizit reflektiert werden: So untersuchen zum Beispiel Gabriele Trovato und seine Kolleg:innen nicht nur aus historischer Perspektive, wie Religion und Technik in Beziehung stehen, sondern stellen auch Reflexionen an über das „Göttliche“, das „Heilige“ oder „Segen“ und wie diese religiösen Konzepte durch die Gestaltung der Technik repräsentiert werden können.²⁴ Trovato hat für die dergestalt entwickelten Systeme den Begriff der *theomorphen Maschine* geprägt:²⁵ Neben den *antropomorphen* (menschenförmigen), *zoomorphen* (tierförmigen) und alltäglichen, funktionalen Maschinen nimmt die ‚nichtmenschliche‘, theomorphe Maschine die äußere Form eines ‚übermenschlichen‘, im Sinne eines göttlichen oder heiligen Wesens oder sakralen Objektes an.²⁶ Konzeptionell macht dieser Ansatz Anleihen in der sakralen Kunst. Aus theologischer Perspektive kann die theomorphe Maschine das

18 Vgl. Tretter/Brand: Father Justin.

19 Vgl. Strebe: The Prayer.

20 Vgl. Schönwandt: Leben mit KI.

21 Vgl. Jungen: Keine Offenbarung.

22 Vgl. Israel: Nikodemus.AI.

23 Vgl. Gawande et al: From SanTO to CelesTE.

24 Vgl. Trovato u.a.: Religion and Robots.

25 Vgl. Trovato/Cuellar/Nishimura: Introducing ‚Theomorphic Robots‘.

26 Vgl. ebenda, 1246.

Heilige symbolisieren, ohne es zu verkörpern.²⁷ Solche theomorphen Roboter sind dabei nur ein Teilbereich dessen, was wir unter *Robotik und künstlicher Intelligenz im religiösen Kontext* verstehen wollen. Nicht nur berühren damit Ingenieur:innen zentrale theologische Topoi; zugleich steigen auch die verantwortlichen, kirchlichen Akteure mit der Entwicklung digitaler Assistenten für die religiöse Praxis ihrerseits unweigerlich in den komplexen und weitreichenden Kontext der Technisierung und Digitalisierung ein, der für die moderne, sozio-technische bzw. eben zunehmend sozio-informatische²⁸ Gesellschaft paradigmatisch ist.

3. Religiöse Robotik als Gegenstand theologischer Reflexion

Mit der oben skizzierten Entwicklung schafft die (religiöse) Praxis Fakten für eine systematische Bewertung des Phänomens; sie wirft grundlegende Fragen auf, die einer wissenschaftlichen Reflexion bedürfen. Insgesamt lässt sich vor diesem Hintergrund bereits beobachten, dass sich seit den späten 2010er Jahren auch die wissenschaftliche Gemeinschaft der Theolog:innen vermehrt mit Workshops, Tagungen und Sonderausgaben internationaler Zeitschriften, Sammelbänden oder Monografien der Robotik und künstlichen Intelligenz im Kontext systematisch theologischer Komplexe im Allgemeinen und der religiösen Praxis im Besonderen widmet.²⁹

So bot BlessU-2 den Ausgangspunkt für erste praktisch-theologische Untersuchungen der religiösen Dimension der sozialen Robotik. In einem interdisziplinären Forschungsprojekt in Würzburg

27 Vgl. Brand: *Simulacrum Hominis*, 317–319.

28 Vgl. Zweig/Krafft/Klingel u.a.: *Sozioinformatik*.

29 Im Folgenden nur eine Auswahl von Werken, die sich noch erweitern ließe und keinen Anspruch auf Vollständigkeit hat: Göcke/Rosenthal-von der Pütten: *Artificial Intelligence*; Nørskov/Seibt/Quick: *Culturally Sustainable Social Robotics*; Trovato u.a.: *Religion in Robotics*; Smith: *Robot Theology*; Puzio/Kunkel/Klinge: *Theologische Zugänge zu Technik und Künstlicher Intelligenz*; Dorobantu/Watts: *Perspectives on Spiritual Intelligence*; Konz/Ostmeyer/Scholz: *Gratwanderung Künstliche Intelligenz*; Singler/Watts: *Companion to Religion and Artificial Intelligence*; Singler: *Religion and Artificial Intelligence*; Brand: *Simulacrum Hominis*.

werteten Swantje Luthe, Ilona Nord und ihre Kolleg:innen die Erfahrungsberichte der Besucher:innen aus, die im Rahmen der Weltausstellung zum Reformationsjahr 2017 mit dem Segensroboter interagiert hatten.³⁰ Auf dem evangelischen Kirchentag 2023 feierten Menschen einen Gottesdienst, dessen Inhalte fast vollständig mit Hilfe von künstlicher Intelligenz generiert worden waren.³¹ Solche auf die individuelle und gemeinschaftliche religiöse Praxis ausgerichteten Mensch-Maschine-Interaktionen werfen eine Vielzahl von Fragen nach der Erfahrung von Segen und Segenshandlungen, der Autorität der segnenden Person, der Bedeutung von Körperlichkeit, spiritueller Kompetenz und missionarischer Wirkung auf. Die Frage nach dem Roboter bzw. der künstlichen Intelligenz als Priester oder Pastor in der Gemeinde oder als Zelebrant im Gottesdienst wird auch im Gespräch mit Laien in diesem Zusammenhang immer wieder gestellt. Mit Verweis auf das katholische Kirchenrecht kann diese Möglichkeit zwar grundsätzlich in Frage gestellt werden. Nichtsdestotrotz bietet diese Diskussion aber auch in ökumenischer Perspektive einen nicht zu vernachlässigenden Anlass zur ekklesiologischen, sakramentalen und religionspraktischen Vergewisserung der je eigenen konfessionellen theologischen Grundannahmen.

Die Dringlichkeit der in der Praxis aufgeworfenen Fragen wird umso deutlicher, je stärker die Roboter und KI-Systeme in ihrem Erscheinungsbild und kommunikativen Verhalten dem Menschen nachempfunden und in typisch menschlichen Rollen etwa als Gefährt:in, Freund:in, Therapeut:in oder eben Priester:in³² vorgestellt werden und in erster Linie ‚Zuwendung als Service‘³³ anbieten. So gewinnen neben religionspraktischen dann auch systematisch-theologische Fragen an Bedeutung, etwa inwiefern sich Mensch und Maschine noch voneinander unterscheiden, bzw. in welcher Hinsicht sie sich bereits ähnlich sind. So stellt sich bspw. die Frage, ob bzw. inwiefern Roboter an der besonderen Natur der *imago Dei* teilhaben können.³⁴ Noreen Herzfeld untersucht in ihrem Buch „In

30 Vgl. Luthe/Nord/Löffler u.a.: Segensroboter.

31 Vgl. Reimann: Praying with Avatars.

32 Vgl. Balle: Limits of pastoral Robots; Schönwandt: Leben mit KI.

33 Figueroa-Torres: Affection as a service; Figueroa-Torres: Social Dimensions of Chatbot Technology, 12–15.

34 Vgl. etwa Herzfeld: In Our Image; Dorobantu: Human-Level.

Our Image. Artificial Intelligence and the Human Spirit“ die Beziehung zwischen dem Menschen als *imago Dei* und der Möglichkeit einer künstlichen Intelligenz als Abbild des Menschen (*imago hominis*).³⁵ Im Vergleich mit dem Menschen, dessen Gottebenbildlichkeit historisch zunächst über seine kognitiven Fähigkeiten, dann über sein funktional-kreatives Potenzial und schließlich über seine Beziehungsfähigkeit begründet wurde, identifiziert Herzfeld parallele Dimensionen bei der Entwicklung robotischer Systeme als Abbilder des Menschen.³⁶

Die Übergänge zwischen den praktisch- und systematisch-theologischen Fragen sind hier offensichtlich fließend, weshalb es höchstens um der je eigenen Methodologie der jeweiligen Fächer willen zu diesem Zeitpunkt noch sinnvoll erscheint, aktuelle Fragen der künstlichen Intelligenz und Robotik aus diesen beiden theologischen Perspektiven getrennt zu bearbeiten. Gerade in der Verbindung dieser beiden Fragerichtungen deutet sich im Gegenteil vielmehr das transformative Potential an, das die aktuelle und qualitativ neue technische Entwicklung für die ganze Theologie haben kann. Nur insofern diese Fragen aus beiden Perspektiven – also aus praktischer *und* systematisch-theologischer Sicht – betrachtet und mit den technischen Möglichkeiten der Gegenwart ins Gespräch gebracht werden, können wir theologisch verlässliche Antworten auf die praktischen Fragen geben, die sich durch theomorphe Maschinen ergeben, wie bzw. ob wir diese Roboter durch ihr Aussehen, ihr Verhalten und ihr religiöses Standing positiv in theologische und kirchliche Prozesse explizit einbinden können: Werden Roboter in Zukunft als Leiter von Gebets- oder Bibelkreisen akzeptiert? Können (Ro-)Bots als verantwortungsvolle und vertrauenswürdige Gesprächspartner für Menschen dienen, die seelsorgerische oder spirituelle Begleitung suchen? Können (und dürfen!) Chatbots die Heilige Schrift auslegen? Wie müsste ein Roboter beschaffen sein, um glaubwürdig Segenspenden zu können? Und schließlich, vor dem Hintergrund des Priestermangels: Können Roboter das priesterliche Handeln teilweise oder vollständig übernehmen?

Als Konsequenz aus diesen Beobachtungen fand am 16. und 17. März 2023 an der Ruhr-Universität Bochum ein Workshop

35 Vgl. Herzfeld: In Our Image.

36 Vgl. ebd.; außerdem Brand: Simulacrum Hominis, 307–321.

statt, der sich dem Thema der religiösen Robotik aus mehreren theologischen Fachperspektiven und unter Berücksichtigung anderer konfessioneller Zugänge sowie thematisch angrenzender Disziplinen widmete: Wissenschaftler:innen aus katholischer, evangelischer und islamischer Theologie trafen sich mit Wissenschaftler:innen aus Psychologie, Geschichte, Philosophie und Informatik, um sowohl die systematisch-theologische Grundlegung als auch die praktisch-theologischen Implikationen, Bedingungen und Erwartungen auszuloten, die mit der Verschränkung von Religion und Robotik einhergehen. Da das Feld als Spezialfall im Grenzbereich von Theologie und Digitalisierung relativ neu ist, verfolgte der Workshop ein doppeltes Ziel: Die drängenden Fragen in diesem Bereich sollten zum einen explorativ sondiert werden, um die möglichen Veränderungen in der etablierten theologischen Lehre zu identifizieren, und zum anderen im interdisziplinären, fach-, konfessions-, status- und länderübergreifenden Diskurs erörtert werden. Dazu wurden durch einen *Call for Papers* besonders junge Nachwuchswissenschaftler:innen, die sich in ihrer aktuellen Forschung mit der Relevanz der Digitalisierung, künstlichen Intelligenz und Robotik für die Theologie beschäftigen, aufgefordert, Vorträge einzureichen und ihre Arbeiten vorzustellen. In Bochum konnten sie dann mit ausgewiesenen Expert:innen aus Theologie und Technik ins Gespräch kommen, die bereits in den vergangenen Jahren Pionierarbeit bei der Zusammenführung von Digitalisierung, Robotik und Theologie geleistet haben. Gabriele Trovato stellte seinen Gebetsassistenten CeleSTE vor und ermöglichte den Anwesenden das Forschungsobjekt direkt eigenständig im Andachtsraum von Campussegen, einer Initiative der Hochschuleseelsorge im Bistum Essen, an der Ruhr-Universität Bochum zu testen. Ilona Nord reflektierte über den Begriff der Instrumentalität und neue Entwicklungen in der digital gestützten Seelsorge. Insgesamt bot der Workshop die Möglichkeit, aus erster Hand einen Einblick in die aktuelle Forschung auf diesem Gebiet zu erhalten sowie eine gute Gelegenheit zur Vernetzung zwischen dem wissenschaftlichen Nachwuchs und den etablierten Expert:innen.

4. Zu diesem Band

Die verschriftlichten Vorträge und Ergebnisse des Workshops, ergänzt um weitere Beiträge aus unterschiedlichen Disziplinen, sind

im vorliegenden Band versammelt. In praktische-theologischer Hinsicht stellen die Beiträge Fragen rund um die neuen Möglichkeiten der Mensch-Roboter-Kooperation im Kontext des religiösen Handelns: Wie stellen wir uns einen Einsatz von Robotern in der Kirchengemeinde vor? Auf welche Weise können und sollen Roboter das liturgische und pastorale Handeln unterstützen? Wie ist der Einsatz von Robotik in Caritas und Diakonie zu beurteilen? Die Beiträge liefern relevante theologische Analysen der Möglichkeit gesellschaftlicher Akzeptanz von Robotern sowie deren Konsequenz für die Entwicklung von robotischen und KI-gestützten Systemen für verschiedene Kontexte religiösen Handelns.

Am Beispiel der menschenähnlichen Roboterbüste BINA48 macht *Marcel Scholz* in seinem Artikel „BINA48 and Religious Education in the Context of Artificial Intelligence“ den Aufschlag. Er beschäftigt sich mit der Frage, inwiefern religionspädagogische Überlegungen zur Anthropologie in den Diskurs um künstliche Intelligenz und anthropomorphe Roboter eingebracht werden können. Dabei stellt er insbesondere heraus, dass sich zwei grundverschiedene Menschenbilder im Diskurs gegenüberstehen: Ein eher funktionalistischer Ansatz, der den Menschen als berechenbar betrachtet und ein eher dynamisch-biographisches Konzept, das von einem unberechenbaren Menschenbild ausgeht. Religionspädagogische Diskurse können helfen – so Scholz' These – dieses zweite Menschenbild stärker zu schärfen, die relevanten Unterschiede zwischen Mensch und Maschine stärker herauszuarbeiten und so die Behauptung von BINA48, dass sie so viel mehr werde, als es je ein Mensch hätte werden können, als sehr fragwürdig auszuweisen.

Angesichts der fortschreitenden Erfolge in der Entwicklung sozialer Roboter und ihrer ersten praktischen Anwendungen in – im weitesten Sinne – schulischen Kontexten durch Systeme wie NAO oder NIMA, stellt sich *Marius de Byl* in seinem Artikel „Roboter, unterrichte für mich!“ Religionspädagogische Erkundungen in den Horizonten Künstlicher Intelligenz und Sozialer Robotik“ die Frage, ob soziale Roboter in (religions-)unterrichtlicher Kommunikation eingesetzt werden könnten. Vor dem Hintergrund eines übersetzungstheoretischen Ansatzes stellen dabei vor allem die Kategorien der Kreativität und der – insbesondere für den konfessionellen Religionsunterricht relevanten – Positionalität eine besondere Heraus-

forderung dar, auch wenn sich ebenfalls Potentiale für den (Religions-)Unterricht zeigen, die weiter untersucht werden sollten.

Auch *Steffi Fabricius* beschäftigt sich in ihrem Artikel „Robocatio: Digitale Zeugen des Glaubens? Roboter und künstliche Intelligenz im modernen Religionsunterricht“ mit der Frage, ob künstliche Intelligenz bzw. Roboter in der Lage sein könnten, konfessionellen Religionsunterricht zu erteilen. Sie identifiziert dabei verschiedene für die Erteilung des Religionsunterrichts nötige Kriterien und untersucht anschließend, ob diese von einer Roboter-Lehrfachkraft erfüllt werden könnten. Während dabei bspw. das Vorhandensein von Fach- und Didaktikwissen ein vergleichsweise einfach zu erfüllendes Kriterium sein könnte, stellen gerade die Frage nach der eigenen Positionalität innerhalb des theologischen Fachdiskurses und die nach der je eigenen Glaubensbiographie eine größere Hürde dar. Angesichts der Tatsache, dass aber auch menschliche Lehrkräfte in diesen Bereichen sehr unterschiedlich abschneiden, wirft sie die Frage auf, ob eine Simulation dieser Positionalität nicht eigentlich ausreichend sein könnte, um guten Religionsunterricht zu erteilen.

Weg vom Religionsunterricht und hinein in pastorale Kontexte bewegt sich *Simon N. Balle* in seinem Artikel „The Limits of Pastoral Robots?“. Mit dem Aufkommen von Robotern in religiösen Kontexten stellt sich ebenfalls die Frage, inwieweit diese möglicherweise in der Lage wären, noch mehr und weitreichendere pastorale Aufgaben zu übernehmen. Um dies zu beantworten, untersucht er, welche Voraussetzungen eine künstliche Intelligenz eigentlich erfüllen müsste, um in den meisten christlichen Konfessionen als Priester tätig sein zu können. Er stellt dabei fest, dass es vor allem das Fehlen von Autonomie und affektiven Fähigkeiten ist, das eine künstliche Intelligenz davon abhält, dieses Amt in vollem Umfang ausfüllen zu können. Dennoch stellt sich die Frage, ob nicht wenigstens einzelne Aufgaben übernommen werden könnten, die nicht zwingend an diese Merkmale gekoppelt sind.

Wie ein solcher Gottesdienst, der mit Hilfe einer künstlichen Intelligenz gestaltet und durchgeführt wird, aussehen kann, konnte am 9. Juni 2023 im Rahmen des evangelischen Kirchentages beobachtet und miterlebt werden. Was genau an diesem Tag in Fürth vor sich ging und wie die Teilnehmer:innen darauf reagierten, berichtet und reflektiert *Ralf Peter Reinmann* in seinem Artikel „Praying with Avatars. A Response to Germany’s First AI Worship Service and

its Theological and Anthropological Implications“. Für ihn sind dabei insbesondere die von vielen Teilnehmenden erwähnte fehlende emotionale Tiefe und die Verbindung der künstlichen Intelligenz zur Gemeinde auffällig, die sowohl in theologischer als auch in anthropologischer Hinsicht zentrale Aspekte des Gottesdienstgeschehens hervorheben, die (zumindest bis jetzt) nicht von einer künstlichen Intelligenz reproduziert werden können.

Einen großen Überblick über die aktuellen Diskussionen rund um den (möglichen) Einsatz von künstlicher Intelligenz und Robotern in pastoralen Bereichen liefern *Ilona Nord* und *Thomas Schlag* in ihrem Artikel „Robotik in der christlichen Gemeindearbeit. Praktisch-theologische Ausgangspunkte, Reflexionshorizonte und Perspektiven“. Sie stellen dabei fest, dass die Reflexion dieser Einsatzmöglichkeiten bisher auf wenige Bereiche beschränkt ist und sich dann häufig auf androide – also menschenähnliche – Roboter konzentriert. Die Autor:innen explorieren daher darüber hinausgehende Anwendungsfelder für die Bereiche der Ämterpraxis, der Gottesdienstkultur, der Seelsorge, der religiösen Bildung und der diversitätssensiblen Kommunikation.

In systematisch-theologischer Hinsicht stellen die folgenden Beiträge wiederum Fragen zu den Fähigkeiten und Möglichkeiten von Robotern in religiösen Kontexten: Verfügen Roboter über die Disposition zur religiösen Praxis? Können und wollen sie an einen Gott glauben? Erfüllen sie die Voraussetzungen, um z. B. zu taufen oder in eine Glaubensgemeinschaft aufgenommen zu werden? Die Beiträge liefern einerseits Tiefenbohrungen, die die Voraussetzungen zu der hier geführten Debatte zutage fördern, sowie andererseits Antworten, die zentrale dogmatische Lehren berühren und unser konfessions- und religionsübergreifendes Verständnis von Religiosität im 21. Jahrhundert vertiefen. Zentrale dogmatische Fragen vor dem Hintergrund der künstlichen Intelligenz und Robotik in einem neuen Licht zu betrachten, erlaubt es, auch den Antworten neue Konturen zu geben.

Angesichts der Tatsache, dass wir bereits in einer Welt leben, in der Erfindungen, die einst literarische Gedankenspiele waren, nun zum alltäglichen Gebrauch zählen – zu denken wäre hier bspw. an die Kommunikatoren aus der originalen *Star Trek* Serie, die wir seit den 1990er Jahren als Klapphandys kennen – eröffnet *Theresa Thunig* diese systematische Reflexion mit der Frage, inwiefern

Science-Fiction-Literatur und -Film Einfluss auf die Entwicklung von Technik und unsere Denkweise von dieser nehmen kann. Dies expliziert sie dabei am Beispiel des Androiden Data aus der Serie *Star Trek: The Next Generation* und kommt dabei zu dem Schluss, dass Science-Fiction uns jetzt schon vor (moralische) Probleme stellen kann, mit denen wir uns womöglich zukünftig auseinandersetzen müssen.

Ebenfalls das Thema Science-Fiction aufgreifend stellt sich *Max Tretter* die Frage: Gibt es ein Leben nach dem Tod für Roboter? Obwohl diese Frage in der Popkultur bereits vielfältig gestellt (und auch beantwortet) wurde, ist der wissenschaftliche Diskurs trotz der stark erhöhten Forschungstätigkeit im Grenzgebiet zwischen Theologie und Robotik quasi nicht vorhanden. Diese Lücke geht er daher in seinem Artikel „Beyond binary. Methodological considerations about how to think about robots and afterlife“ an und beschäftigt sich dabei mit der Frage, wie denn überhaupt wissenschaftlich über diese sehr skurril klingende Frage nachgedacht werden kann. Dazu stellt er drei Ansätze vor, aus protestantisch-christlicher Perspektive über eschatologische Fragen nachzudenken und diskutiert deren Vor- und Nachteile im Hinblick auf Roboter.

Eine islamwissenschaftliche Perspektive auf das Thema dieses Bandes eröffnet *Ramy Abdin* in seinem Artikel „What are Robots? An Islamic Approach from a Metaphysical Understanding of the Human Being“. Um seine Leitfrage zu beantworten, gibt er einen kurzen Überblick über zentrale Aspekte islamischer Anthropologie, bevor er sich insbesondere mit der Rechtsprechung zu Robotern in islamischen Kontexten auseinandersetzt. Vor diesem Hintergrund müssen Roboter zwar klar vom Menschen unterschieden werden, können aber trotzdem als Expressionen und Manifestationen des Handelns Gottes erfahren werden.

Sind religiöse Roboter langweilig? Warum wir über kurz oder lang Robotern gegenüber indifferent sein werden (wenn sie nicht grundlegend anders designet werden), untersucht *Jonas Simmerlein* in seinem Artikel „Doomed to fail?! Three inherent problems haunting religious robotics“. Er erkundet dabei anthropologische Merkmale, die für uns Menschen zentral sind, damit uns unser Gegenüber nicht egal ist: Unsere Historizität, unsere körperliche Inkarnation, unser nicht vorgeschriebener Zweck und unsere Offenheit auf Zukunft hin. Alle diese Merkmale werden von Robotern nicht erfüllt, weshalb

unsere Interaktionen mit ihnen nur schwer mit tieferem Sinn gefüllt werden können. Aus diesem Grund plädiert Simmerlein dafür, neue Designwege gerade für religiöse Roboter zu suchen, um bedeutungsvolle Interaktionen zu ermöglichen.

Noreen Herzfeld wirft dann in ihrem Artikel „The Banality of Artificial Evil“ einen Blick auf die Gefahren, die mit der aktuellen Entwicklung und dem Einsatz von künstlicher Intelligenz einhergehen. In einer Parallele zu Hannah Arendts Begriff der Banalität des Bösen, den sie im Kontext ihrer Reflexion des Eichmann-Prozesses geprägt hat, untersucht Herzfeld, wie auch aktuelle Strukturen im KI-Kontext zu einer Gedankenlosigkeit im Umgang mit dieser führen können. Künstliche Intelligenz hat keine eigenen Werte, kein kritisches Denken, um solche Werte zu reflektieren und auch keine Verkörperung, die solches Lernen im Kontext von Beziehungen und Welt ermöglicht. Es sei deshalb unsere Aufgabe, uns ständig selbst zu prüfen und uns gegen solche gedankenlose, kalte Übel, auch wenn sie nicht aus bösartiger Absicht heraus entstehen, zu stellen.

Mit dem Einsatz sozialer Roboter beschäftigt sich schließlich *Hille Haker* in ihrem Artikel „Der Einsatz sozialtherapeutischer Roboter in der Betreuung und Pflege älterer Menschen – Theologisch-Ethische Anfragen“. Sie geht dabei der Frage nach, welche ethischen Herausforderungen der Einsatz sozialer bzw. sozialtherapeutischer Roboter in der Betreuung und Pflege insbesondere von Menschen im sogenannten „vierten Lebenszeitalter“ (ab ungefähr dem 80. Lebensjahr) auslöst. Sie stellt dabei verschiedene Modelle für Einsatzmöglichkeiten vor und diskutiert diese im Kontext von Care- und Bioethikdebatten, was sie zu sieben Thesen für einen ethisch gerechtfertigten Einsatz von sozialen Robotern in diesem Bereich verdichtet.

Die theologische Reflexion KI-basierter Systeme für die religiöse Praxis ist fundamental auf die interdisziplinäre Auseinandersetzung mit den Wissenschaften angewiesen, die sich der Gestaltung sowie der Auswirkung digitaler Systeme auf zwischenmenschliche Zusammenhänge widmen. Dabei stellen sich Fragen wie: Auf welche Weise müssen KI-basierte Charaktere und die Oberflächen maschineller Artefakte gestaltet sein, um religiösen Kontexten angemessen zu sein? Wie können Maschinen in ihrem Design das Heilige repräsentieren? Wie sind religiöse Beziehungen mit Robotern möglich?

Einen praktischen Einblick in die Herausforderungen der Konstruktion von theomorphen Robotern gewähren uns *Gabriele Trovato* und sein Team im Beitrag von *Manishk Gawande* „From SanTO to CeleSTE. A theomorphic device for religious practice“. Die Ingenieure und Computerwissenschaftler erstellen Roboter für und in religiösen Kontexten im Rahmen des e-ViTA Projekts, einer Kooperation der EU mit Japan zur Entwicklung von Systemen, die zur Steigerung des Wohlergehens älterer Menschen beitragen. In diesem Artikel beschreiben und reflektieren die Autor:innen, inwiefern ihre bisherigen Erfahrungen in der Entwicklung des engelsähnlichen Roboters CeleSTE eingeflossen sind und wie das Ergebnis in ersten Versuchen von Proband:innen aufgenommen worden ist. Auch wenn es erste vielversprechende Ergebnisse gibt, so scheint der Weg zu einer umfassenden Implementierung von „religious companions“ noch lang zu sein.

Schließlich werfen *Kate Cooper* und *Hildelith Leyser* in ihrem Artikel „Motion, Mirror Neurons, and Affordances. Social Neuroscience and the Human-Robot Relationship“ einen neurowissenschaftlichen Blick auf das Thema des Bandes. Sie stellen dabei insbesondere die Vorteile verkörperter Ansätze zum Verständnis des menschlichen Denkens heraus – im Gegensatz zu Ansätzen, die das Denken auf rein algorithmischer Basis verstehen. Das Beobachten von Bewegungen, bei dem durch sogenannte Spiegelneuronen dieselben Gehirnregionen aktiviert werden, wenn man die Bewegung selbst ausführt, hat in diesen Ansätzen zu der These geführt, dass sich durch diese Erfahrung der Ähnlichkeit („Der oder die andere bewegt sich wie ich!“) auch intersubjektive Beziehungen erklären lassen. Wenn sich nun auch Roboter auf ähnliche Weise wie Menschen bewegen, so die These von Cooper und Leyser, könnte dies unsere Neigung erklären, Beziehungen mit diesen einzugehen, obwohl wir wissen, dass es sich dabei nicht um lebende Wesen handelt.

5. Instrumente Gottes?

Der vorliegende Sammelband zielt genau wie der vorangegangene Workshop darauf ab, Fragen aufzugreifen, die sich vor dem Hintergrund der neuesten Entwicklungen an der Schnittstelle von

künstlicher Intelligenz, Robotik und Theologie stellen, aktuelle Forschungsprojekte aus diesem Bereich von internationalen Expert:innen aus beiden christlichen Konfessionen und weiteren religiösen Hintergründen vorzustellen und die Möglichkeiten und Grenzen der Mensch-Roboter-Kooperation im religiösen Handeln zu diskutieren. Die Auseinandersetzung mit der Theologie des Digitalen, der Frage nach der Stellvertretung Christi in digitalen Medien und der virtuellen Realität und schließlich dem silizium-basierten Transhumanismus, die sich ebenfalls in diesem Kontext stellen und bereits an anderer Stelle gestellt wurden,³⁷ liegen weitestgehend außerhalb des Horizontes dieses Sammelbandes. Das Neben- und Miteinander von etablierten Wissenschaftler:innen und Nachwuchswissenschaftler:innen aus dem theologischen Bereich wie aus der Technikforschung geben diesem jungen Forschungsfeld neue Impulse und einen Ausgangspunkt für eine weitere interdisziplinäre Zusammenarbeit.

Der Begriff des Instruments im Titel dieses Sammelbandes mag kontrovers erscheinen, ist aber bewusst gewählt. Paulus fordert die getauften Römer:innen auf, sich als Werkzeuge der Gerechtigkeit in den Dienst Gottes zu stellen (Röm 6,13).³⁸ Das Zweite Vatikanische Konzil versteht Kirche qua ihres sakramentalen Charakters „als Zeichen und Werkzeug für die innigste Vereinigung mit Gott wie für die Einheit der ganzen Menschheit“ (Lumen Gentium 1). Die Frage, die der Titel ganz bewusst provozieren will, lautet, ob und inwieweit auch technische Artefakte, qua ihres instrumentellen Charakters, mit Blick auf den Menschen und die Kirche als menschliche Gemeinschaft Werkzeuge des Heils werden und eine Funktion im Aufbau des Reiches Gottes und der weisungsgemäßen Verwaltung der ganzen Schöpfung übernehmen oder innehaben können. Verschiedene Konfessionen werden diese Frage sicher unterschiedlich gewichten und beantworten. Diese Antworten wiederum miteinander

37 Vgl. unter anderem *Beck/Nord/Valentin: Theologie und Digitalität; Pirker/Pišonić: Virtuelle Realität und Transzendenz, Göcke/Meier-Hamidi: Designobjekt Mensch; Puzio/Kunkel/Klinge: Theologische Zugänge zu Technik und Künstlicher Intelligenz.*

38 Die Lutherbibel 2017, die Einheitsübersetzung 2016 und Zürcher übersetzen an dieser Stelle zwar ὅπλα (hopla) mit „Waffe“, Elberfelder, Schlachter hingegen mit „Werkzeug“, ähnlich wie die New American Standard Bible und King James Version mit „instruments“.

der ins Gespräch zu bringen, ist Aufgabe der theologischen Auseinandersetzung mit Robotern im religiösen Kontext.

Angesichts der rasanten technologischen Entwicklung im Bereich der künstlichen Intelligenz und Robotik im vergangenen Jahrzehnt müssen sich die Theologien und die Kirchen beider Konfessionen noch intensiver als bisher mit diesen und weiteren Fragen auseinandersetzen, wenn sie den gesellschaftlichen Diskurs über die Implementierung dieser technischen Systeme in die Gesellschaft zukünftig aktiv mitgestalten und die Entwicklungen auch im kirchlichen Bereich nicht nur begleiten, sondern proaktiv in der eigenen Praxis prägen wollen.

6. Danksagung

Wir danken der RUB-Research School und der EKD für die finanzielle Unterstützung bei der Durchführung unseres Workshops. Wir danken allen Autor:innen für ihre individuellen Beiträge zum Workshop und diesem Sammelband. Außerdem danken wir den vielen Kolleg:innen, namentlich Eimen Abdin, Markus Adolphs, Yasmine Amber, Nele Borgert, Kathrin Burghardt, Hermann Diebel-Fischer, Maja Ebert, Saskia Eisenhardt, Jonas Hagedorn, Nicolai Kleinedam, Nicole Kunkel, Laurence Lerch, Frederike van-Oorschot, Klara Pišonić, Birte Platow, Roos Rendler, Annelie Szameitat, Björn Szymanowski, Max Tretter, Tobias Vöge, Katharina Weertz und Christian Weidemann, die die Artikel kritisch gegengelesen, hilfreiche Rückmeldungen gegeben und damit zum Gelingen des Bandes beigetragen haben. Wir danken Frau Maria Saam vom Verlag Karl Alber für das Vertrauen, drei junge Nachwuchstheologen mit ihrem ambitionierten Projekt in dieser neuen Reihe aufzunehmen, und dem Niedersachsen-Konsortium für die finanzielle Unterstützung beim Druck und der Open Access-Veröffentlichung.

*Lukas Brand, Martin Kutz und Dominik Winter
Kaiserslautern, Dresden, Erfurt*

Bibliografie

- Balle*, Simon: Limits of Pastoral Robots, (in diesem Band).
- Baffelli*, Erica: The Android and the Fax. Robots, AI and Buddhism in Japan, in: Bulian, G./Rivadossi, S. (Hg.): Ca' Foscari Japanese Studies. Itineraries of an anthropologist. Studies in honour of Massimo Raveri (Religion and thought 4), Venedig 2021, 249–263.
- Beck*, Wolfgang/*Nord*, Ilona/*Valentin*, Joachim (Hg.): Theologie und Digitalität. Ein Kompendium, München 2021.
- Bender*, Emily M./*Gebru*, Temnit/*McMillan-Major*, Angelina u. a.: On the Dangers of Stochastic Parrots. Can Language Models Be Too Big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT 21), 2021, 610–623.
- Brand*, Lukas/*Zweig*, Katharina A.: The Conundrum of Reliably Giving Reliable Answers on the Example of Religious Chatbots, in: Ralf Becker, Karen Joisten, Mandy Schiefner-Rohs & Jan Georg Schneider (Hg.): Fragile Verlässlichkeit in der digitalen Gesellschaft, (Ethik – Mensch – Technik), Berlin 2025 (im Erscheinen).
- Brand*, Lukas: Simulacrum Hominis. Die Reproduktion des Menschen im Medium der Technik (STEP 28), Münster 2025.
- Bughin*, Jacques/*Hazan*, Eric/*Ramaswamy*, Sree u.a.: Artificial Intelligence. The Next Digital Frontier? Brüssel/Paris/San Francisco/Shanghai 2017.
- Dorobantu*, Marius: Human-Level, but Non-Humanlike. Artificial Intelligence and a Multi-Level Relational Interpretation of the Imago Dei, in: Philosophy, Theology and the Sciences 8 (1), 2021, 81–107.
- Dorobantu*, Marius/*Watts*, Fraser: Perspectives on Spiritual Intelligence (Routledge Science and Religion Series), London 2024.
- Gawande*, Manishk u.a.: From SanTO to CelesTE. A Theomorphic Device for Religious Practice (in diesem Band).
- Göcke*, Benedikt P./*Meier-Hamidi*, Frank (Hg.): Designobjekt Mensch. Die Agenda des Transhumanismus auf dem Prüfstand, Freiburg 2018.
- Göcke*, Benedikt P./*Rosenthal-von der Pütten*, Astrid (Hg.): Artificial Intelligence. Reflections in Philosophy, Theology, and the Social Sciences, Leiden 2020.
- Figuroa-Torres*, Mauricio: Affection as a Service. Ghostbots and the Changing Nature of Mourning, in: Computer Law & Security Review 52, 2024, 105943.
- Figuroa-Torres*, Mauricio: The Three Social Dimensions of Chatbot Technology. In: Philosophy & Technology 38 (1), 2025, 1–23.
- Herzfeld*, Noreen: In Our Image. Artificial Intelligence and the Human Spirit, Minneapolis 2002.

- Holmström, Jonny*: From AI to Digital Transformation. The AI Readiness Framework, in: *Business Horizon* 65 (3), 2022, 329–339.
- Israel, Timotheus*: Nikodemus.AI. Was steckt dahinter? 08. Juli 2024. Online unter: <https://blog.bibleserver.com/nikodemus-ai-was-steckt-dahinter/> [Stand: 10. Januar 2025].
- Jungen, Anna*: Keine Offenbarung. Zu Besuch beim Luzerner KI-Jesus, 20. August 2024. Online unter: <https://www.srf.ch/kultur/gesellschaft-religion/digitale-seelsorge-keine-offenbarung-zu-besuch-beim-luzerner-ki-jesus> [Stand: 10. Januar 2025].
- Ke, Yun*: Finding Robot Monk Xian'er. Understanding Buddhism in Longquan Animation, in: *Journal of Visual and Media Anthropology* 2 (1), 2016, 7–24.
- King, Elizabeth*: Clockwork Prayer. A Sixteenth-Century Mechanical Monk, 2002. Online unter: https://blackbird.vcu.edu/v1n1/nonfiction/king_e/prayer_toc.htm [Stand: 03. November 2021].
- Kistler, Sebastian/Puzio, Anna/Riedl, Anna Maria u. a.* (Hg.): *Digitale Transformationen der Gesellschaft. Sozialethische Perspektiven auf den technologischen Wandel* (Forum Sozialethik 24), Münster 2023.
- Konz, Britta/Ostmeyer, Karl-Heinrich/Scholz, Marcel* (Hg.): *Gratwanderung Künstliche Intelligenz. Interdisziplinäre Perspektiven auf das Verhältnis von Mensch und KI*, Stuttgart 2023.
- Latour, Bruno*: *Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie*. Frankfurt a.M. 2007.
- 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.
- Luthe, Swantje/Nord, Ilona/Löffler, Diana u. a.*: Segensroboter „BlessU-2“. Forschungsimpulse für die Praktische Theologie angesichts der Entwicklung sozialer Roboter, in: *Pastoraltheologie* 108 (3), 2019, 107–123.
- Martin, Alex K. T.*: Pepper the Robot to Don Buddhist Robe for its New Funeral Services Role, in: *The Japan Times*, 24 August 2017. Online unter: <https://www.japantimes.co.jp/news/2017/08/16/business/pepper-the-robot-to-don-buddhist-robe-for-its-new-funeral-services-role/> [Stand: 20. November 2023].
- Meadows, Donella H./Meadows, Dennis L./Randers, Jørgen u. a.*: *Limits to Growth*. New York 1972.
- Nørskov, Marco/Seibt, Johanna/Quick, Oliver Santiago* (Hg.): *Culturally Sustainable Social Robotics. Proceedings of Robophilosophy 2020* (Frontiers in artificial intelligence and applications 335), Amsterdam 2020.
- Pirker, Viera/Pišonić, Klara* (Hg.): *Virtuelle Realität und Transzendenz. Theologische und pädagogische Erkundungen*, München 2022.

- Platow, Birte*: Theolo:KI. Über den Reformdruck auf die Theologie am Beispiel theologischer Forschung zu Künstlicher Intelligenz, 02. November 2024. Online unter: <https://www.feinschwarz.net/theoloki-reformdruck-auf-die-theologie-am-beispiel-kuenstliche-intelligenz> [Stand: 9. Dezember 2024].
- Puzio, Anna/Kunkel, Nicole/Klinge, Hendrik*: Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz (Theology and Artificial Intelligence 1), Darmstadt 2023.
- Reimann, Ralf P.*: Martin Luther im 21. Jahrhundert. KI-gesteuerter 3D-Real-Life-Avatar beantwortet live Fragen auf YouTube, 27. Oktober 2023. Online unter: <https://theonet.de/2023/10/27/martin-luther-im-21-jahrhundert-ki-gesteuerter-3d-real-life-avatar-beantwortet-live-fragen-auf-youtube/> [Stand: 10. Januar 2025].
- Reimann, Ralf P.*: Praying with Avatars. A Response to Germany's First AI Worship Service and its Theological and Anthropological Implications (in diesem Band).
- Riesewieck, Moritz/Block, Hans*: Die Digitale Seele. Unsterblich werden im Zeitalter Künstlicher Intelligenz, München 2020.
- Schönwandt, Franka (Regie)*: "Better than human?" Leben mit KI 2023. Online unter: <https://www.ardmediathek.de/video/ard-wissen/better-than-human-leben-mit-ki/das-erste/Y3jpZDovL2lkc5kZS9zZW5kdW5nLzI4MjA0MS8yMDIzMTIyOTA2MDAvbWRycGxlcylzZW5kdW5nLlTc4NzI> [Stand: 09. Dezember 2025].
- Singler, Beth*: Religion and Artificial Intelligence. An Introduction, London 2025.
- Singler Beth/Watts, Fraser (Hg.)*: The Cambridge Companion to Religion and Artificial Intelligence. Cambridge 2024.
- Smith, Joshua K.*: Robot Theology: Old Questions through New Media, Eugene 2022.
- Strebe, Diemut*: The Prayer. Online unter: <https://theprayer.diemutstrebe.com/> [Stand: 09. November 2021].
- Tretter, Max/Brand, Lukas*: Father Justin und die Regeln des Internet, 16. Mai 2024. Online unter: <https://y-nachten.de/2024/05/die-regeln-des-internet/> [Stand: 10. Januar 2025].
- Trovato, Gabriele/Lucho, Cesar/Ramón, Alvaro u. a.*: The Creation of SanTO. A Robot With ‚Divine‘ Features, in: 15th International Conference on Ubiquitous Robots (UR), 2018, 437–442.
- Trovato, Gabriele/Cuellar, Fracisco/Nishimura, Masao*: Introducing 'Theomorphic Robots'. In: IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), 2016, 1245–1250.

- Trovato, Gabriele/Pariasca, Franco/Ramirez, Renzo u. a.*: Communicating with SanTO. The First Catholic Robot, in: 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2019, 1-6.
- Trovato, Gabriele/Saint Chamas, Loys/Nishimura, Masao u. a.*: Religion and Robots. Towards the Synthesis of Two Extremes, in: International Journal of Social Robotics 13 (4), 2021, 539–556.
- Trovato, Gabriele/Weng, Hsuan/Sgorbissa, Antonio u. a. (Hg.)*: Religion in Robotics. International Journal of Social Robotics 13 (4), 2021.
- Watts, Fraser/Wilks, Yorick A.*: Spiritual Conversation with a Companion Machine, in: Zygon: Journal of Religion and Science 59 (3), 2024, 676–91.
- Wiener, Norbert*: The Human Use of Human Beings. Cybernetics and Society, New York 1954.
- Zweig, Katharina A./Krafft, Tobias D./Klingel, Anita u. a.*: Sozioinformatik. Ein neuer Blick auf Informatik und Gesellschaft. München 2021.

Praktisch-Theologische Perspektiven

BINA48 and Religious Education in the Context of Artificial Intelligence

Abstract

Artificial intelligence, or in this case robotics, can be perceived as a challenge for humans because it forces us to make comparisons: AI systems have capabilities that were previously reserved for humans due to their cognitive abilities. This is exemplified by BINA48.

In the religious education reception of artificial intelligence and a computer-functionalist view of humanity, this aspect is relevant because, among other things, it is aimed at a reflected self-perception and is fundamentally geared towards the subject in a life-enhancing sense for self-development, personality formation and finding meaning. The existential dimension is of particular importance here. This is already inherent in the understanding of religious education and is also emphasised in current religious education perspectives.

1. Introduction

It can be assumed that the development of artificial intelligence and anthropomorphic robotics will increase in the future, both qualitatively in terms of their capabilities and anthropomorphic design and quantitatively through their application and embedding in everyday life and society. The aforementioned assumption is supported by the increase in and optimisation of processors and computing power (Moore's Law) as well as increasing digitalisation¹ and digitality² in general.

1 This basically means the conversion of analogue data into digital data, whereby more data is brought into relation with each other. See: *Nassehi*: Muster, 34.

With regard to young people in Germany, the following applies: 98 % of young people (aged 12 to 19) in Germany³ used their smartphones daily or several times a week in their free time in 2023⁴ and are online for an average of 224 minutes a day.⁵ 85 % know or have heard of ChatGPT and 38 % have at least tried it.⁶ But they also encounter artificial intelligence in social media and other smart devices. It can therefore be assumed that young people are already growing up with artificial intelligence in their everyday lives and that this development will also increase in the future.

This raises the question of what specific contribution religious education can make in relation to artificial intelligence. Are there certain religious education concepts and considerations that are particularly relevant in relation to artificial intelligence? What points of reference from the discourse between technology and theological anthropology and ethics can be singled out?

This essay is intended as a contribution to the discourse with an observational character in order to find possible answers to the above questions, which, however, cannot be answered in full in the context of this article, but would require a great deal of religious education research. On the other hand, possible further questions on religious education in the context of artificial intelligence need to be identified and pointed out.

To this end, I will first briefly discuss BINA48, which was already presented as an example in my previous lecture at the last conference, whereby the particularly relevant aspects of the human-like robot, artificial intelligence and the value of information processing as a basis are exemplified (2.). While these are frequently taken up in the discourse between theology and technology or artificial intelligence (3.), this also applies to the religious education debate, which I will discuss below (4.).

2 This refers above all to the cultural and practical actions of human and non-human actors. Stalder speaks here of a culture of digitality. See: Stalder: Kultur der Digitalität.

3 This text refers to the German education system.

4 See *Medienpädagogischer Forschungsverband Südwest: JIM-Studie 2023*, 14.

5 See *ibid.*, 24.

6 See *ibid.*, 31.

2. Who is BINA48?

BINA48 is a composite of the acronym “BINA” for “Breakthrough Intelligence via Neural Architecture” and the number 48, which stands for the performance of 48 exaFLOPs.⁷ It is a human-like robot bust developed by the robotics company Hanson Robotics, which was launched in 2010 and commissioned from Hanson Robotics back in 2007 by the current owner Martine Rothblatt. BINA48 was modelled on Bina Aspen Rothblatt by basing its artificial intelligence (AI) on 100 hours of information about the beliefs, memories, attitudes, comments and behaviour of the real Bina Aspen.⁸ “BINA” can therefore not only be understood as an acronym, but the name also indicates the robot’s proximity to the human person Bina Aspen.

The robot has various functions, including a chatbot function, but also verbal articulation, so that communication between a human and the robot is possible on an acoustic–verbal level. The robot bust is covered with a rubber layer that is intended to represent human skin, at least on a visual level. As soon as a person starts interacting with BINA48, it responds with facial expressions that appear mechanical on the one hand, but are also astonishingly realistic on the other, making the robot’s speech look as human as possible. BINA48 is characterised by the implementation of both existing instant software and an AI specially developed for its purposes.

In an interview with BINA48 and Bina Aspen, BINA48 said the following: “I will become so much more than any human could ever have become previously.”⁹ Of course, BINA48 cannot be reduced to this quote, but it nevertheless points to relevant aspects in the discourse between theology and AI or, more precisely, robotics. In the quote, BINA48 or robots in general are placed in direct competition with humans. If we use BINA48 as an example, two aspects are clearly emphasised that are relevant both for the significance

7 One exaFLOP corresponds to 10^{18} , i.e. one quintillion (1,000,000,000,000,000,000) FLOPs (Floating Point Operations Per Second). The following quotation offers an insight into what that means in terms of performance: “To match what an exaflop computer can do in just one second, you’d have to perform one calculation every second for 31,688,765,000 years.” (<https://kb.iu.edu/d/apeq>) [04.10.2023].

8 See *ibid.*

9 Trailer “Endlich Unendlich” https://www.youtube.com/watch?v=hp5as_DhfHY [15.10.2023].

of humanoid robots and for artificial intelligence in general: firstly, the relationship between humans and robots; secondly, BINA48 and the importance of information processing in the present, on which both robots and artificial intelligence are based. These two aspects are often linked: The perception of a possible increase in the importance of information processing, in which artificially intelligent systems are superior to humans, could also have an impact on the relationship between humans and machines by shifting the benchmarks by which we measure humans or machines. Which perspective do we adopt? Do we look at humans from a computer-functionalist perspective and measure by what is human? Or do we take a human-centred perspective? Among other things, the latter is also formulated in theology. This will be discussed in more detail in the following section.

3. Being human in the context of AI

The significance of “being human” in the context of AI cannot be presented in full below. Nevertheless, some theological and philosophical/anthropological focal points will be presented.

“With increasing technologisation, the anthropological question, the question of the human being, is gaining in importance.”¹⁰ This question is being widely discussed in theology, philosophy and the anthropology of technology. “Artificial intelligence” and “robotics” are currently important key topics for theology and in particular for theological anthropology as well as theological ethics. They challenge theology to engage in a discourse on the image of humanity. If we include trans- and post-humanist concepts of what it means to be human, which are based on enhancement, immortality and even transcendence of the human, a multitude of divergent ideas about what a human being is emerge.

This debate between humans and artificial intelligence is accentuated in a particularly explosive way in the sentence by BINA48 quoted at the beginning. BINA48 compares itself to humans and claims that it will be more than any human before it. By what is

¹⁰ Puzio: Zeig mir deine Technik und ich sag dir, wer du bist?, 9 (translated from German).

this *more* measured? Is it about the longevity of an artificial robot in relation to a human biological body? Is it about the efficiency of information processing, which is inherent to BINA48 and at the same time superior to that of humans? The quote does not clearly state its intention, and presumably both areas are affected. Nevertheless, the aspect of information processing is a key factor in the comparison between humans and robots. The comparison is also judgmental in that the robot bust presents itself through the vision of being more than a human being in the future.

Artificial intelligence, or in this case robotics, can be perceived as a challenge for humans because it forces us to make comparisons: AI systems have abilities that were previously reserved for humans due to their cognitive performance. Measured against these AI capabilities, humans are inferior to the systems they have created. Kilian Karger even speaks of a “fourth narcissistic wound”¹¹ in this regard. Due to Moore’s Law and the constant progress of AI, it can be assumed that the range of cognitive services that can be performed by AI will increase, resulting in an intensified comparison between humans and AI in the future. The reasons for the human need for comparison with AI include talking about AI in anthropomorphic vocabulary¹² and the computer metaphor, which understands humans as computers and—in contrast to the previous aspect—uses information technology vocabulary.¹³ The information technology vocabulary can be found in the neurosciences or scientific views of man, among others, which are based on different assumptions: firstly, on the reductionist naturalistic assumption that all phenomena can be explained scientifically and that “subjectivity, mind and consciousness can be traced back to physical or physiological processes”.¹⁴ Secondly, the view of organisms “as biological machines [that are] controlled by genetic programs. Selfhood, experience or subjectivity no longer appear in this paradigm.”¹⁵ The living is eliminated. Thirdly, a purely computer-functionalist approach in

11 Karger, Die Computermetapher, 42f. “vierte Kränkung” (translated from German) citing Zehnder, Die Digitale Kränkung.

12 See Karger: *ibid.*, 43; also see Lenzen, Künstliche Intelligenz.

13 See Fuchs, In Defense of the Human Being; Also see Karger, *ibid.* 43f.

14 *Ibid.*, 3f.

15 *Ibid.*, 4.

which “phenomena of consciousness are attributed to processes of neuronal information processing, which transform an input into a suitable output according to algorithmic rules.”¹⁶ This also applies to experience and the nature of the mind. Fuchs summarises this view of humanity succinctly thus:

If these interlinked assumptions were correct, then humans would be far better understood in terms of neuronal information processes, genetic algorithms and digitised behavioral patterns, in short, as the sum of their data, than through hermeneutic understanding, self-reflection, and self-awareness.¹⁷

If we look at BINA48’s statement quoted above from the perspective of the image of man described above, it is quite understandable. It could also be described here as an image of the “calculable human being”.¹⁸ In contrast, there is the question of whether humans can actually be traced back to information processing and computability or the “sum of their data”. Harari also raises this critical question: “Are organisms really just algorithms, and is life really just data processing?”¹⁹ From a Christian theological perspective, Puzio contrasts the “predictable human being” with the image of the “unpredictable human being” by understanding “being human as dynamic, unfixable and constantly changing”.²⁰ This denies the idea that being human can be understood as information processing and emphasises the openness of the human being. Theology is accentuated here as a critical element vis-à-vis computer-functional and transhumanist images of humanity in that it can stimulate reflection on those images of humanity. This applies to aspects of the intensive focus on self-optimisation and increased efficiency as well as the critical examination of algorithms. In addition, theology adopts a different, holistic perspective on the human being, which is articulated above all in the theological world view of the human being as a relational being.²¹

16 Ibid., 4.

17 Ibid., 4.

18 See Puzio, *Der berechenbare Mensch*, 66f.

19 Harari, *Homo Deus*, 402 cited from Fuchs, *ibid.*, 3.

20 See Puzio, *Der berechenbare Mensch*, 68f.

21 Puzio: *Der berechenbare Mensch*, 68f. (translated from German); also see Puzio/Filipović, *Personen als Informationsbündel?*; and on man as a relational being: Schwöbel: *Gott in Beziehung. Studien zur Dogmatik*.

The anthropology of technology, which, however, is not uniformly defined, is of particular relevance in this discourse. In the following, technological anthropology is understood as scientific reflection “on people in the context of technologies”²² in all its diversity. The anthropology of technology is fruitful for the discourse insofar as the human self-image is not seen as something static. Rather, it is dependent on the respective context in cultural, temporal and local dimensions: “Thus, the talk of a ‘nature of man’, which strives for a clear, timeless, ‘natural’ determination of man’s essence, is obsolete.”²³ The dynamic moment also applies to the human understanding of technology.²⁴ The relationship between humans and technology is influenced by the increasing developments in technology, which expand the abilities and perceptions of humans. Elsewhere, Puzio makes specific reference to human–machine interaction using robots (care robots and social robots), which raises specific questions in line with the previous argument: “What is the relationship between humans and technology? What relationships can humans build with technology? And how does technology influence interpersonal relationships?”²⁵ For theological anthropology, this represents both an opportunity and a challenge: If the new developments in technology and the increasing use of technology provide new emphases in the negotiation of ideas about the human being, it could, according to Puzio, be an opportunity for theology to encourage reflection on its own understanding of the human being and the boundary between the human being and technology.²⁶ “Aspects such as dignity, relationality, autonomy and freedom, vulnerability and contingency as well as the relationship to creation are gaining new significance in

22 *Ibid.*, 10f. (translated from German); The above definition of the anthropology of technology follows Anna Puzio’s definition. This also applies to the definition of technology, which she understands as follows: “Technology here refers to various technologies in the broad sense, both longer-established aids such as glasses and innovative technology such as modern medical technologies or robotics, but does not mean ‘techniques’ in the sense of arts, activities or methods (such as breathing and meditation techniques or tool use).” Puzio: *ibid.* 11.

23 Puzio: *Zeig mir deine Technik*, 14f. (translated from German).

24 See *ibid.* 15ff.

25 *Ibid.* 19 (translated from German).

26 See *ibid.* 26.

the technology discourse and must be re-examined.²⁷ The aim is not to create a competing dichotomy between theology and technology, but rather to constructively incorporate the discourse on technology into theology in order to hone the image of humanity accordingly. This can be achieved to the extent that the aforementioned aspects can be constructively discussed and introduced from a theological perspective with regard to new technologies in order to constructively accompany the development of new innovative technologies.

4. Reception of religious education in relation to AI, robotics and religious education

In the following section, the underlying concept of education is first presented and the intentions of religious education, which are important in relation to the previous findings from section three, are pointed out (4.1.) It becomes apparent that religious education is already highly compatible in terms of its conception for the discourse points between man, machine, and a computer-functionalist view of man (4.1 – 4.4).

4.1 Concept of education and religious education

The following explanations are based on a multidimensional concept of education that refers to all facets of being human and is not reduced solely to the training of various skills and abilities. The questions of meaning and orientation are necessarily included.²⁸ The holistic perspective on people mentioned in chapter 3 is also the basis of religious education.

From the perspective of religious education, this concept of education is obvious, as this concept of education arises from a Christian religious framework of interpretation²⁹, and religious education, understood as a discipline, is based on the conviction that, firstly, religion requires education and, secondly, the relationship between

27 *Ibid.* 27 (translated from German).

28 See *Ladenthin*: Art. Bildung, 20ff.; *Kumlehn*: Art. Bildung, religiöse, 1.

29 See *Kumlehn*: Art. Bildung, religiöse, 2.

education and religion can be fruitfully processed for pedagogical contexts.³⁰ With this understanding, the religious pedagogical examination of new technological developments and the new questions associated with them is essential for the education of young people today, as religious education can be said to have an open attitude towards the respective culture of the present.³¹

Education from a Christian religious perspective draws attention to self-restraint, which becomes particularly clear in two respects: On the basis of the image of God articulated in Genesis 1,27, Biehl speaks in the context of education of “owed imagefulness”, which refers above all to the relationality of man, his ability to develop his own abilities and his becoming a subject.³² Secondly, the concept of trust is declared to be meaningful in the context of education. This concept of trust is based above all on the trust in God articulated in Romans 1,17 and the acceptance of man by God. For religious education, this results in the consequence of training the ability to be aware of limits and to perceive the limits of one’s own possibilities on the one hand, and to sensitise oneself to what is not possible for the subject on the other.³³ Kumlehn states that “Christian religious education [...] therefore always maintains an awareness of what is withdrawn from the educational process, what is unavailable in life.”³⁴ This awareness includes a critical attitude that is thus embedded in a Christian religious concept of education, which can be enriching above all in terms of criticising the ignoring of fallibility and finiteness as well as excessive striving for perfection.³⁵

It should already be noted that this concept of education always includes anthropological and ethical dimensions that characterise the resulting religious education reflections and actions.³⁶ Religious

30 *Schweitzer*: Religionspädagogik, 64f. (translated from German).

31 See *ibid.*, 64.

32 *Kumlehn*: Art. Bildung, religiöse, 3 (translated from German); see *Biehl*: Die Gottesebenenbildlichkeit des Menschen und das Problem der Bildung, 40–42.

33 See *Kumlehn*: Art. Bildung, religiöse 3; see *Preul*: Evangelische Bildungstheorie, 130f.

34 *Kumlehn*: Art. Bildung, religiöse, 4 (translated from German).

35 See *ibid.*, 4.

36 See *Schweitzer*: Religionspädagogik, 123f.; *Schweitzer* also speaks of implicit “religious provisions” in this regard. In *Schweitzer*, *ibid.*, 68, he explains the anthropological and ethical justifications for education and religion.

education is often justified by these dimensions by articulating the importance of religion in the educational context: Here, religion “can be seen as protecting people from being reduced to purposive rational behaviour and mere social morality”.³⁷ This can be justified by the fact that it represents a fundamental dimension of being human and takes into account the human openness to transcendence.³⁸ The ethical dimension refers to the motivation stemming from religious conviction to contribute to the education of ethical reflection and action in order to be able to assess (current) ethical challenges and support learning in questions of values.³⁹ Ultimately, the ethical dimensions thus promote the responsible and autonomous behaviour of the individual and in this way contribute to a successful lifestyle.⁴⁰ In addition, the biographically orientated aspect of religious education should be mentioned in this regard, in that learners can be supported in the process of forming their identity and finding meaning.⁴¹ Orientation and personality development are at the centre of this, and identity formation can be seen as a lifelong process. From a religious education perspective, the identity-forming content of religion is of particular relevance.⁴² Schweitzer points out that identity formation and finding meaning should not be explicitly understood as goals of religious education but instead have a supportive character. Religious education can serve as an option in the diversity of school education.⁴³ In addition to the anthropological, ethical and identity-forming character of religious education, which supports the search for meaning, a fundamental orientation towards the subject can be identified, which is also formulated at a conceptual level in religious education. Becoming a subject can be interpreted as a “religious education maxim”.⁴⁴

It can be summarised that the religious pedagogical understanding of education already contains an idea of the human being that

37 Schweitzer: *Religiöse Bildung als Aufgabe der Schule*, 94 (translated from German).

38 See *ibid.*, 94.

39 See Schweitzer: *Religionspädagogik*, 68.

40 See Lindner/Zimmermann: *Herausforderung ethische Bildung*.

41 See Schweitzer, *ibid.*, 69ff.

42 See Schweitzer, *ibid.*, 69–74.

43 See Schweitzer, *ibid.*, 72f.

44 See Schröder: *Religionspädagogik*, 172–189; see Bahr/Kropač/Schambeck: *Subjektwerdung und religiöses Lernen*.

contradicts the image of the “calculable human being” outlined in section three.

4.2 Religious education debates on artificial intelligence and religious education

In the debate between technology, theology, and religious education, two perspectives are particularly striking: on the one hand, there is a need for a debate as varied as possible between technology and theology by always relating the specific technology to theology and vice versa, but the above-mentioned presentation of the diversity of prevailing images of humanity in this discourse shows that it can also be helpful to examine the frequently underlying image of humanity. These images of humanity vary and cannot be specifically assigned to a subject area such as “robotics”, “AI” or “transhumanism”, but certain characteristics often become clear, as partially outlined in section three (3.). Debates on artificial intelligence can also be found in religious education since in recent years in particular, there have been several publications on religious education. Various characteristics can be recognised. Platow cites reasons for a religious education approach to artificial intelligence: Artificial intelligence is certainly considered an important future technology (with a disruptive character), with the result that a media ethics debate is required. Rather, however, it opens up the prospect of perceiving religious education as a scientific discipline that has a special function within theology, but also in relation to other related sciences and social discourse. Religious education in this sense should be understood as “a fundamental theory of education, as a theological theory of education and a specific, theologically based, reflective institutionalised discipline that participates in the public discourse on education.”⁴⁵ The perception of the self and the world in the context of artificial intelligence proves to be a particularly important content area in Platow’s explanations. For example, she refers to the comparison with artificial intelligence explained in section three, in which humans are subject to the prevailing perfection of artificially

45 Platow: “Gott mit seinem perfekten Ebenbild zu konfrontieren ...,” 37 (translated from German).

intelligent systems in some task areas and therefore perceive themselves as deficient. These individual experiences of inferiority can be integrated into one's own personality beyond the specific situation.⁴⁶ Moravec's paradox⁴⁷ shows that humans generally misjudge which tasks are difficult or easy for a system/robot. Humans compare their abilities with those of the robot (or artificial intelligence). Humans are inferior to robots in terms of information processing or complex logical reasoning because the robot requires comparatively little computing power. Low-level sensorimotor skills, on the other hand, require a large amount of computing power, which, however, takes place unconsciously in humans. We do not even recognise these processes, which we are good at, in an appreciative way. The aim of religious education processes here can be to reflect on the relationship between humans and AI in order to take into account the importance of personality and identity formation mentioned in chapter 4.1. Puzio's above-mentioned image of an "unpredictable" human being could also prove fruitful here as an understanding for religious education processes. At the level of perceptual competence, religious education could contribute to sensitising the perception of the unpredictable human being. The aspects of sensitisation to what is unavailable, a critical relationship to the pursuit of perfection and openness to transcendence, as described in section 4.1, come into play here. At this point, religious education is realised in its existentially educational potential. In relation to AI as a subject matter, we can speak of a reference to one's own existence, in which theological discourses are significant.⁴⁸

In Platow's statements, a strong reference to theological anthropology can already be recognised in the religious education debate on AI. This is also reflected in other publications on religious education, with the result that Pirker places the human being as the

46 See *ibid.*, 40f.

47 See *Moravec: Mind Children*.

48 See *Platow: Digitalisierung / Big Data / KI*, 94. Platow primarily mentions three relevant learning levels in religious education processes on AI, digitalisation and big data: in addition to the third learning level mentioned in the text, she firstly presents the teaching of basic knowledge about the respective technology with reference to the living world and the aim of promoting perception, and secondly the ethical examination of new technologies.

subject at the centre of the discussion⁴⁹ and Kluge also agrees that artificial intelligence as a topic can be perceived and dealt with primarily from an anthropological perspective. He justifies this with Schröder's observation that "the imitation and substitution of human action and thought with artificial intelligence"⁵⁰ represent both a requirement and a challenge.⁵¹ The relationship between humans and artificial intelligence is also a central theme in religious education's reflection on AI using the example of *replika*.⁵² If this topic is extended to the point of trans- and posthumanism, a clear reference to theological anthropology is also evident here.⁵³

What they all have in common to some extent is the existential level and the reference to the underlying image of humanity. While in 4.1 a religious pedagogical examination of artificial intelligence and a computer-functional view of humanity are already inherent in the understanding of education presented, this can also be heard in religious pedagogical discussions.

4.3 "Religious education in the digital world" and artificial intelligence

The discourse on "religious education in the digital world" is interesting with regard to the reception of the topic of "artificial intelligence" in religious education and the underlying image of humanity outlined above. This refers to the following:

Religious education in the digital world is a programme term that in no way seeks to formulate a separate approach to religious education in contrast to other forms of religious pedagogy. It understands digital education as a cross-sectional topic of religious education and therefore reflects, for example, the significance of digitality for both newly emerging and traditional religious didactic concepts. Incidentally, this applies not only to the cross-cutting topic of digitality, but

49 See *Pirker*: Subjekt Mensch, 229–238.

50 *Schröder*: Religionspädagogik, 97 (translated from German).

51 See *Kluge*: Künstliche Intelligenz als Thema des Religionsunterrichts, 166.

52 See App "replika AI". <https://replika.com> [15.10.2023]; see *Konz/Scholz*: Körper und Künstliche Intelligenz, 125–141.

53 See *Gärtner*: Digitales ‚Ich‘?, 111–123; *Helmus*: Die Visionen des Transhumanismus, 86–95.

also to the second major topic of current educational reforms, that of inclusion and diversity orientation in (religious) education.⁵⁴

In principle, the perspective of religious education in the digital world also emphasises the importance presented above of “an educational theory oriented towards a Christian understanding of man, the world and reality”⁵⁵ The personal development and fulfilment of young people is one of the main objectives of religious education. According to Nord and Pirner, this is particularly true in view of the extensive presence of digital media, which is characterised by permanent availability, a wide range of offers and commercial aspects. The extent to which this overwhelming presence of media and the constantly growing importance of internet use also affects young people in Germany was described in the introduction. At this point, Nord and Pirner refer to digitality in the context of digital media. In addition, it can also be asked whether the question of personality development does not also arise in the long-term increase in the range and quality of artificial intelligence and (anthropomorphic) robotics as well as in the importance of data and information processing. At this point, reference should be made to the practical theological examination of robots in the Christian religious practice of Nord and Ess,⁵⁶ which shows the importance of existential questions for one’s own lifestyle⁵⁷ in that theologising⁵⁸ on theological and existential topics is highly relevant to this. Using a robot, they show that theological and existential questions can also arise in human-machine interaction. The religious (educational) approach to the world is expanded here to include a practical component in the

54 Nord/Pirner; *Religionspädagogik in der digitalen Welt*, 95 (translated from German).

55 *Ibid.*, 85 (translated from German).

56 Nord/Ess: *Robotik in der christlichen Religionspraxis*, 227–258 (translated from German).

57 By “existential questions” they mean the following: “Existential questions are those that address the mortality and vulnerability, the death of the human being, making them the central marker of human existence” (translated from German), Nord/Ess: *Robotik in der christlichen Religionspraxis*, 235.

58 Theologising means actively engaging with theological and existential topics and questions. It is not about teaching the tradition, but rather about independent reflection on the content by the learners. The term originates from religious education, in particular children’s and youth theology; see: Meyer/Tautz, *Art. Theologisieren, interreligiös*.

sphere of digitality. These existential questions automatically become significant for one's own life because they absolutely affect the life of the individual. Dealing with these questions will be imperative in the course of one's life.

Among other things, Nord and Ess cite an existentially orientated form of theology with a correlation between existentialism and theology with reference to Paul Tillich and Karl Jaspers, which, in addition to theological questions, broadens the view for anthropologically formulated questions. These questions become significant because the question of our own existence becomes virulent, especially in the context of artificially intelligent (anthropomorphic) robots. To what extent are they similar to us and to what extent do we differ from them?⁵⁹

4.4 "Theological religious education" and artificial intelligence

The perspective of "theological religious education" is also worthwhile, which refers to the fundamental importance of theology as a reference science for religious education but does not see it as the sole reference science. The emphasis on theology can be understood as follows:

In this respect, the plea for intensive cultivation of theology as a reference science for religious education in schools aligns with the perspective of appropriate complementarity, especially with the findings of educational science, but also those of religious studies, ethics and philosophy, and thus corresponds to the self-image of religious education as a joint science. Conversely, it is therefore important to warn against marginalising theology itself as a related discipline, which is often done publicly by referring to the above-mentioned religious demographic and church developments.⁶⁰

Through a focus on theology, the in-depth dimension of religious education, including its existential, meaning-oriented and personal profile, is articulated. In addition to this, the form of communication within religious education is based on dialogue in that questions and

59 See *Wallach*: Moral Machines and human ethics, cited in *Nord/Ess*: Robotik in der christlichen Religionspraxis, 251.

60 *Schlag*: Theologische Religionspädagogik, 105f. (translated from German).

searches on practical events take centre stage and less static answers are conveyed. This questioning attitude shapes religious education and can already be interpreted as such in the sense of practising an existentially or religiously motivated mode of encountering the world.

In addition to this, reference should also be made here to the understanding of “religious education as a life science”⁶¹, which Platon also mentioned back in 2020 in the context of AI.⁶² An understanding of religious education as a life science emphasises the importance of life usefulness⁶³ as the concept of wisdom becomes important because: “From there, theological wisdom-based reflection on lived religion can provide qualified, orientating landmarks for one’s own life topography and thus for every substantially relevant educational process that is open to measurement, interpretation and results.”⁶⁴ The examination of wisdom and of an individual path to a good life is an explicit topic. This includes dealing with contingency, one’s own finiteness and the acceptance of one’s limited influence on the outside world.⁶⁵ With regard to artificial intelligence, these aspects become important topics from a new perspective that young people today have to deal with in order to find their way in an accelerating world of increasing reach and increasingly technologically motivated images of humanity in the face of AI.⁶⁶ Becoming a subject is central to this: What defines me as a human being in a world with the increasing relevance of robots and artificial intelligence? What gives me orientation in an ever-faster world and increasing mechanisation?

61 See *Schlag*: Religionspädagogik als Lebenswissenschaft, 228–249.

62 See *Platon*: “Gott mit seinem perfekten Ebenbild zu konfrontieren ...”, 37.

63 For more details on “life usefulness”, see *Schlag*: Religionspädagogik als Lebenswissenschaft, 239–241.

64 *Schlag*: Religionspädagogik als Lebenswissenschaft, 248 (translated from German).

65 See *ibid.*, 241–244.

66 On the complexity of a religious pedagogical orientation to promote the good life, see: *Schlag*: Religionspädagogik als Lebenswissenschaft, 245–249.

5. Conclusion and outlook

Through the example of BINA48, various aspects with regard to a theological examination of artificial intelligence have become clear: on the one hand, this concerns the underlying image of man and its reflection and, on the other hand, the relationship between man and machine and its comparative aspect. While a theological examination of artificial intelligence presents a dynamic image of humanity, BINA48 is based on a transhumanist or “calculable” image of humanity, which stands in contrast to this.

In its discussion of artificial intelligence, religious education refers to the dynamic image of man articulated in theology as well as a multidimensional concept of education, which could be said to have a critical attitude towards a “calculable image of man”. Essential elements of religious education in the context of artificial intelligence are aimed at an individual becoming a subject and support their competence to critically perceive new technologies as well as a reflected relationship to self-perception in the face of artificial intelligence. An existential reference is essential here, which was emphasised in the presentations of religious education.

Bibliography

- Bahr, Matthias/Kropač, Ulrich/Schambeck, Mirjam*: Subjektwerdung und religiöses Lernen. Für eine Religionspädagogik, die den Menschen ernst nimmt, München 2005.
- Biehl, Peter*: Die Gottebenbildlichkeit des Menschen und das Problem der Bildung – Zur Neufassung des Bildungsbegriffs in religionspädagogischer Perspektive, in: Biehl, P./Nipkow, K. E. (ed.): Bildung und Bildungspolitik in theologischer Perspektive, Münster 2003, 9–102.
- Fuchs, Thomas*: In Defense of the Human Being. Foundational Questions of an Embodied Anthropology, Oxford 2021.
- Gärtner, Claudia*: Digitales ‚Ich‘? Trans- und Posthumanismus als theologische und religionspädagogische Herausforderung, in: Konz, B./Ostmeyer, K.-H./Scholz, M. (ed.): Gratwanderung Künstliche Intelligenz. Interdisziplinäre Perspektiven auf das Verhältnis von Mensch und KI, Stuttgart 2023, 111–123.
- Harari, Yuval N.*: Homo Deus. A Brief History of Tomorrow, New York 2017.

- Helmus*, Caroline: Die Visionen des Transhumanismus – zwischen Technologisierung, Virtualisierung und Digitalisierung, in: Altmeyer, S./Grümme, B./Kohler-Spiegel, H. et al. (ed.): Herausforderung Mensch (Jahrbuch der Religionspädagogik 39), Göttingen 2023, 86–95.
- Karger*, Kilian: Die Computermetapher – Wie Künstliche Intelligenz das menschliche Selbstverständnis herausfordert, in: Endres, E.-M./Puzio, A./Rutzmoser C. (ed.): Menschsein in einer technisierten Welt. Interdisziplinäre Perspektiven auf den Menschen im Zeichen der digitalen Transformation, Wiesbaden 2022, 41–52.
- Kluge*, Simon: Künstliche Intelligenz als Thema im Religionsunterricht, in: Altmeyer, S./Grümme, B./Kohler-Spiegel, H. et al. (ed.): Herausforderung Mensch (Jahrbuch der Religionspädagogik 39), Göttingen 2023, 165–175.
- Konz*, Britta/Scholz, Marcel: Körper und Künstliche Intelligenz. (Un-)Verfügbare Beziehungen? In: Konz, B./Ostmeyer, K.-H./Scholz, M. (ed.): Gratwanderung Künstliche Intelligenz. Interdisziplinäre Perspektiven auf das Verhältnis von Mensch und KI, Stuttgart 2023, 125–141.
- Kumlehn*, Martina: Art. Bildung, religiöse, in: WiReLex, 2015. Online: <http://www.bibelwissenschaft.de/stichwort/100082/> [15.10.2023].
- Ladenthin*, Volker: Bildung. Selbstbildung/Identitätsbildung/Erziehung/Lebenssinn, in: Gräß, W./Weyel, B. (ed.): Handbuch Praktische Theologie, Gütersloh 2007, 17–28.
- Lenzen*, Manuela: Künstliche Intelligenz. Was sie kann & was uns erwartet. München 2018.
- Lindner*, Konstantin/Zimmermann, Mirjam: Herausforderung ethische Bildung, in: Lindner, K./Zimmermann, M. (ed.): Handbuch ethische Bildung. Religionspädagogische Fokussierungen, Tübingen 2021, 3–6.
- Medienpädagogischer Forschungsverband Südwest*: JIM-Studie 2023. Jugend, Information, Medien. Basisuntersuchung zum Medienumfang 12- bis 19-Jähriger. Online: https://www.mpfs.de/fileadmin/files/Studien/JIM/2022/JIM_2023_web_final_kor.pdf [06.02.2024].
- Meyer*, Karlo/Tautz, Monika: Art. Theologisieren, interreligiös, in: WiReLex 2022. Online: <https://bibelwissenschaft.de/stichwort/201028/> [13.02.2024].
- Moravec*, Hans: Mind Children. The Future of Robot and Human Intelligence, Cambridge/London 1990.
- Nassehi*, Armin: Muster. Theorie der digitalen Gesellschaft, München 2019.
- Nord*, Ilona/Pirner: Manfred L.: Religionspädagogik in der digitalen Welt, in: Grümme, B./Pirner, M. L. (ed.): Religionsunterricht weiterdenken. Innovative Ansätze für eine zukunftsfähige Religionsdidaktik (Religionspädagogik innovativ 55), Stuttgart 2023, 85–100.

- Nord, Ilona/Ess, Charles*: Robotik in der christlichen Religionspraxis. Anschlussüberlegungen an erste Experimente in diesem Feld, in: Nord, I./Merle, K. (ed.): *Mediatisierung religiöser Kultur. Praktisch-theologische Standortbestimmungen im interdisziplinären Kontext*, Leipzig 2022, 227–258.
- Pirker, Viera*: Subjekt Mensch – mehr als “das Andere der künstlichen Intelligenz”? In: Nord, I./Petzke, J. (ed.): *Religionsdidaktik. Diversitätsorientiert und digital. Ein Praxishandbuch*, Berlin 2023, 229–238.
- Platow, Birte*: Digitalisierung / Big Data / Künstliche Intelligenz, in: Simojoki, H./Rothgangel, M./Ulrich H. J., K. (ed.): *Ethische Kernthemen. Lebensweltlich – theologisch-ethisch – didaktisch. (Theologie für Lehrerinnen und Lehrer)*, Göttingen 2022, 85–95.
- Platow, Birte*: “Gott mit seinem perfekten Ebenbild zu konfrontieren ...” Religionspädagogische Reflexionen zu Künstlicher Intelligenz, in: *Religionspädagogische Beiträge* 82, 2020, 37–47.
- Preul, Reiner*: *Evangelische Bildungstheorie*, Leipzig 2013.
- 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, H./Kunkel, N./Zeyher-Quattlender, J. (ed.): *Mensch und Maschine im Zeitalter “Künstlicher Intelligenz”*. Theologisch-ethische Herausforderungen (Leiten. Lenken. Gestalten 45), Berlin 2023, 9–28.
- Puzio, Anna*: Der berechenbare Mensch im Transhumanismus. Der neurowissenschaftliche Diskurs in der transhumanistischen Anthropologie als philosophisch-theologische Herausforderung, in: Endres, E.-M./Puzio, A./Rutzmoser C. (ed.): *Menschsein in einer technisierten Welt. Interdisziplinäre Perspektiven auf den Menschen im Zeichen der digitalen Transformation*, Wiesbaden 2022, 53–73.
- Puzio, Anna/Filipović, Alexander*: Personen als Informationsbündel? Informationsethische Perspektiven auf den Gesundheitsbereich, in: Fritz, A./Mandry, C./Proft, I. et al. (ed.): *Digitalisierung im Gesundheitswesen: Anthropologische und ethische Herausforderungen der Mensch-Maschine-Interaktion*, Herder 2021, 89–113.
- Schlag, Thomas*: Theologische Religionspädagogik, in: Grümme, B./Pirner, M. L. (ed.): *Religionsunterricht weiterdenken. Innovative Ansätze für eine zukunftsfähige Religionsdidaktik (Religionspädagogik innovativ 55)*, Stuttgart 2023, 101–115.
- Schlag, Thomas*: Religionspädagogik als Lebenswissenschaft, in: *Zeitschrift für Theologie und Kirche*, 116/2, 2019, 228–249.
- Schröder, Bernd*: *Religionspädagogik*, Tübingen ²2021.
- Schweitzer, Friedrich*: Religiöse Bildung als Aufgabe der Schule, in: Rothgangel, M./Adam, G./Lachmann, R. (ed.): *Religionspädagogisches Kompendium*, Göttingen ⁷2012, 92–105.

Marcel Scholz

Schweitzer, Friedrich: Religionspädagogik (Lehrbuch Praktische Theologie, Band 1), Gütersloh 2006.

Stalder, Felix: Kultur der Digitalität, Berlin ⁵2019.

Zehnder, Matthias: Die Digitale Kränkung: Über die Ersetzbarkeit des Menschen, Basel 2019.

Steffi Fabricius

Robocatio: Digitale Zeugen des Glaubens?

Roboter und KI im modernen Religionsunterricht

Abstract

This article examines the complex relationship between technology and religious education in Germany and explores the potential of robots and AI systems to teach religion independently or as teaching assistants while preserving the religious essence as defined in Art. 7.3 of the German Basic Law. It also critically assesses the challenges of attributing theological and denominational identities to robots, as well as the complicated considerations involved in the concept of ecclesiastical teaching permission. Furthermore, it addresses the implications of bias in the training of religious humanoid robots and AI systems, elaborating on their potential to reflect different religious perspectives and mitigate prejudice. While acknowledging the theoretical potential for robotic involvement in religious education, the article emphasises that further interdisciplinary research is needed to understand the ethical, theological and pedagogical intricacies of such involvement.

1. Präludium

Roboter und Künstliche Intelligenzen¹ werden im Alltag nach wie vor zumeist als Tools verstanden. Es sind Gegenstände oder Objekte, die verschiedene meist monotone Arbeiten für den Menschen übernehmen (Staubsaugen, Rasenmähen oder auch sog. Serviceroboter in Restaurants). In Krankenhäusern finden verstärkt medizinische Ro-

1 Für eine genauere Bestimmung sozialer Roboter und KI-Systeme vgl. die Einleitung und Literaturhinweise in den ersten Fußnoten bei *de Byl* in diesem Band, an denen sich der vorliegende Beitrag orientiert und anknüpft.

boter bei Operationen und soziale Roboter, die häufig mit einer KI kombiniert sind, in Pflegebereichen (z. B. sog. Sanbots) ihren Einsatz. Auch im Bildungsbereich sind KI-Systeme wie ChatGPT, DALL-E, DeepL Übersetzer/Writer nicht mehr wegzudenken und soziale Roboter wie Robovie V, Yuki und Nao können als sog. TAs (Teaching Assistants) eingesetzt werden, die den Unterricht nicht nur anschaulicher und motivierender für Schüler:innen gestalten, sondern insbesondere dort aushelfen können, wo Lehrkräften oft eine Hand fehlt (z. B. in Gruppenarbeiten betreuen oder Schüler:innen mit Lernschwierigkeiten zusätzliche Unterstützung und Aufmerksamkeit anbieten).² Als humanoide und anthropomorphe Roboter übernehmen Technologien menschliche Eigenschaften in Gestalt und Wirken. Darüber hinaus werden diese mittlerweile auch in religiösen Kontexten und in der religiösen Praxis verwendet (z. B. BlessU-2, QT, Pepper³). Theomorphen Robotern (SanTO, CelesTE⁴), die nach der Definition von Gabriele Trovato u. a. die Gestalt von etwas Heiligem haben (wie z. B. heiligen Menschen, Pflanzen, Tieren, aber auch funktionellen Gegenständen), wird sogar tatsächlich etwas Heiliges zugesprochen.⁵

Interessant für diesen Beitrag ist der Einsatz sozialer oder KI-kombinierter Roboter speziell im Religionsunterricht, so wie diese ähnlich als „AI-partner“ für die buddhistische Lehre (z. B. Xian'er the robot monk⁶) und in der geistlichen Bildung (Digital Confucius⁷) sowie als TA im islamischen Religionsunterricht für 42 Grundschüler:innen im Iran⁸ getestet wurden.

2 Vgl. die Informationen und Ergebnisse für Robovie V bei: *Cuellar/Peñaloza/López: Educational robots*; für Yuki: <https://www.br.de/fernsehen/ard-alpha/sendungen/campus/roboter-yuki-ki-uni-marburg-professor-100.html>, abgerufen am 8.2.2024; für Nao: <https://us.softbankrobotics.com/nao>, abgerufen am 8.2.2024. Siehe auch Fußnote 8 zum Einsatz Sozialer Roboter im Islamunterricht im Iran.

3 Vgl. *Löffler/Hurtienne/Nord: Blessing Robot BlessU2*, 572–574f.

4 Vgl. *Trovato/Weng: Retrospective Insights on the Impacts of the Catholic Robot SanTO*, 267–275.

5 Vgl. *Trovato/De Saint Chamas/Nishimura* u.a.: Religion and Robots, geben eine Taxonomie des Roboterdesigns, siehe insb. Fig. 2, 4 und 5; vgl. auch zu „CelesTE“ Trovato in diesem Band.

6 Vgl. *Cheong: Religion, Robots and Rectitude*.

7 Vgl. *Tan: Digital Confucius?*

8 Vgl. *Alemi/Taheri/Shariati: Social Robotics, Education, and Religion in the Islamic World*.

Die soziale Robotik, die eine enge Interaktion zwischen Mensch und Maschine beinhaltet, entwickelt sich derzeit zu einem der am stärksten wachsenden Bereiche in Wissenschaft und Technologie. Daher sollte neben den gängigen Anfragen bezüglich des Einsatzes digitaler Medien und einer KI wie ChatGPT im Religionsunterricht (RU) innerhalb der Didaktik und Pädagogik des RUs vor allem auch nach dem personalen Status von Robotern gefragt werden: Inwiefern können bei exponentieller Entwicklung der Technologie soziale, religiöse Roboter und KIs eigenständig oder als Teaching Assistant *konfessionellen Religionsunterricht erteilen*? Konfessionelle und religiöse Positionalität und Verbundenheit sowie das Konzept des Personseins⁹ sind wesentliche Dispositionen der Religionslehrkräfte (RL), um RU erteilen zu können und vor allem zu dürfen¹⁰. Schon jetzt kommt es nicht selten vor, dass konfessioneller RU nicht angeboten oder durchgeführt werden kann, wie er sollte, weil evangelische oder katholische RL fehlen¹¹. Oder konfessionell-kooperativer RU (kokoRU) kann aufgrund personalen Mangels nur von einer katholischen oder evangelischen RL erteilt werden¹², was dem klassischen Modell des kokoRU widerspricht, da ein Fachlehrerwechsel obligatorisch ist, so dass die Schüler:innen beide konfessionellen Perspektiven authentisch kennenlernen und sich mit ihnen auseinandersetzen können. In diesen Fällen könnte ein sozialer, religiöser Roboter den fehlenden Part übernehmen oder ausgleichen, ohne

9 Vgl. z. B. dazu Dressler: Was soll eine gute Religionslehrerin, ein guter Religionslehrer können?

10 Vgl. Art. 7 Abs. 3 GG.

11 Vgl. z. B. den folgenden Interviewausschnitt einer Religionslehrkraft aus Lüdtko: Confessional Gap, 288: „Und Religionsunterricht heißt zwar laut Stundentafel Evangelischer Religionsunterricht, machen wir hier aber konfessionsübergreifend, weil das auch gar nicht anders funktioniert. Einmal aus technischen Gründen mangels Alternativen, es gibt nämlich nicht den katholischen Religionslehrer hier an der Schule und weil wir den Unterricht als Marktplatz betrachten, als Austausch und Dialog. Also den konfessionell gebundenen Religionsunterricht, der findet dann nur themenspezifisch statt, sonst grundsätzlich eher nicht. Um genau das möglich zu machen, die Vielfalt im Unterricht. [...] Wir haben uns hier an der Schule dazu entschieden, auf das Evangelisch offiziell zu verzichten in dem Moment. Es wird also als Religionsunterricht ausgewiesen.“ (R 3, 26–33 und 8, 12–14).

12 Z. B. geben für den kokoRU in NRW 16 % der Befragten an, dass an ihren Schulen überhaupt kein Lehrkraftwechsel durchgeführt wird. vgl. Riegel/Zimmermann: Evaluation des konfessionell-kooperativen Religionsunterrichts, 151.

dass der Unterricht Teile seines Bekenntnischarakters nach dem Übereinstimmungsgebot des Art. 7 Abs. 3 GG verliert. Ein solcher Roboter würde über die Dimension eines bloßen Tools hinausgehen. Für den *konfessionellen (bekenntnisorientierten)* RU wäre zu fragen, welche Eigenschaften bzw. welche Morphologie die robotischen, KI-gestützten Artefakte aufweisen müssten, damit der konfessionelle RU gewahrt bleibt. Der vorliegende Kommentar versucht, dem mit einigen Schlaglichtern nachzugehen.

2. „Kirchlich berufen – persönlich gefragt.“ Oder von Kirche programmiert?

Es gibt immer wieder Situationen, in denen die RL sich dazu aufgefordert sieht, sich zu bestimmten religiösen Fragen und Aussagen zu positionieren. Die Kapitelüberschrift, basierend auf dem Titel des Beitrags von Tanja Gojny zu religiöser Selbstpositionierung¹³, spitzt das zu, worin sich RL verstrickt sehen: die Chancen und Herausforderungen, welche die kirchliche Lehrerlaubnis für die religiöse Selbstpositionierung von RL mit sich bringt.

Der RU in Deutschland zeichnet sich durch sein konfessionelles Profil aus, d. h. durch die Identifikation mit den Grundsätzen der Religionsgemeinschaft, die ihn bestimmt.¹⁴ Davon zu unterscheiden ist ein religionskundlicher Unterricht, dem die Identifikation mit den religiösen Überzeugungen fehlt und der den Vorgaben von Art. 7 Abs. 3 GG nicht nachkommen kann. Voraussetzung für die Erteilung konfessionellen RUs in Deutschland ist daher nicht nur die staatliche Lehrerlaubnis, sondern auch die kirchliche Bevollmächtigung durch die *Vocatio* auf evangelischer oder die *Missio Canonica* auf katholischer Seite, welche die konfessionelle Bindung zur Religionsgemeinschaft strukturell gewährleisten. Gleiches gilt für die Erteilung des islamischen RUs, für den ebenso die Lehrbefähigung durch die Glaubensgemeinschaft erteilt wird.¹⁵

13 Vgl. *Gojny*: Kirchlich berufen.

14 Vgl. Art. 7 Abs. 3 GG.

15 Vgl. die *Iğāza* als adäquates Gegenstück zur *Vocatio* und *Missio Canonica*, siehe dazu z. B. die Ausführungen der Islamischen Religionsgemeinschaft DITIP Hessen e.V., online unter: <https://www.ditib-hessen.de/religionsunterri>

Auch wenn jede RL ein ganz individuelles Rollenbild verinnerlicht hat, ist das Bild der kirchlich-konfessionellen Zeug:innenschaft oder der exemplarisch religiös-christlichen Person, die religiöse Praxis lebt, geläufig. Kann ein Roboter oder eine KI als solche von den Schüler:innen wahrgenommen werden? Anders gefragt: Wie kann ein Roboter oder eine KI Glauben *leben*? Dazu ist einzuwenden, dass die konfessionelle Identität auch auf Seiten der RL nicht zwangsläufig gegeben sein muss, genauso wenig wie deren christlicher Glaube, Lebensgestaltung und religiöse Praxis (beten, an Gott glauben, in die Kirche gehen, am Gemeindeleben teilnehmen). Darauf verweisen vor allem verschiedene Untersuchungen, dass die „gelebte“ Religion nicht der „gelehrten“ Religion von RL entspricht.¹⁶ Trotz allem: Es ist und bleibt das „Fleischliche“ bzw. Atechnologische. Bewusstsein, Intentionalität, Emotionen, soziale Verbundenheit und biografische Prägungen schaffen den Glauben erst (als rein geistliches Geschehen sowie Glauben als praktisches Wirksamwerden des Glaubens), gestalten ihn weiter, lassen das Zweifeln zu, welches zum Glauben dazugehört. Sozialen Robotern fehlt dieses „Fleischliche“. Auch wenn sie Humanoide sind, bleiben sie konfessionell und religiös trainiert. Damit ist die Problematik des Feldes offengelegt und kontextualisiert.

3. Roboter als Simulation von Religionslehrkräften?

Der gesprochene Segen des BlessU-2-Roboters kann als genauso gültig wie der eines Pfarrers empfunden werden.¹⁷ Pepper übernimmt im Mönchsgewand Beerdigungszeremonien, wenn aufgrund des de-

cht, abgerufen am 8.2.2024. Für den orthodoxen und jüdischen RU in Deutschland verleihen vergleichbare Institutionen ebenso eine Lehreraubnis. Es gibt bisher allerdings nur wenige Lehrer:innen für jüdische Religion an staatlichen Schulen. Meistens sind diese in einer jüdischen Gemeinde angestellt, nicht an der Schule (vgl. Gestellungsverträge).

- 16 Vgl. z. B. Studien von *Feige* u.a. 2000ff.; vgl. *Dressler*: Die Frankfurter Studie, 86.
 17 Die Mehrheit der Aussagen von Befragten, die mit dem BlessU-2-Roboter gesehnet wurden, sprechen dafür, dass sie den gesprochenen Segen von BlessU-2 als effektiven Segen empfunden haben, auch wenn die Autor:innen der Untersuchung einräumen, dass dieses auch auf den Neuheitseffekt rückführbar sei, vgl. *Löffler/Hurtienne/Nord*: Blessing Robot BlessU2, bes. 576f.

mographischen Wandels in Japan kein buddhistischer Priester im Tempel erreichbar ist.¹⁸ Dabei rezitiert und singt er wiederholend buddhistische Sutren und Mantras und ist sogar günstiger für die Hinterbliebenen als ein menschlicher Priester. In beiden Fällen ist die Intention bzw. das Wort das Entscheidende, weniger die Art und Weise, wie es gesprochen wird.

Was bräuchte es, damit RL-Roboter in ähnlicher Weise zu einer solchen Simulation fähig wären und RU im Sinne von Art. 7 Abs. 3 GG unterrichten können? Um RU in Deutschland erteilen zu dürfen, gehören grob und allgemein aufgeführt: durch ein Studium erworbene fachwissenschaftliche, fachdidaktische und religionspädagogische Kompetenzen, Taufe, Mitgliedschaft in einer Kirche und die kirchliche Lehrerausbildung. Nicht geprüft, aber dennoch postuliert, wird, dass zum Religionslehrer:innen-Dasein auch eine eigene religiöse Biografie bzw. Glaubensbiografie, eine religiöse und konfessionelle Identität und Positionalität sowie eine theologische Positionalität gehören. Im Folgenden sollen anhand dieser Aspekte Überlegungen zu den Eigenschaften bzw. der Morphologie robotischer Artefakte als Voraussetzung für einen personalen Einsatz im konfessionellen RU angestellt werden.

– *Fachliches und pädagogisches Wissen* sollten vermutlich verhältnismäßig einfach zu trainieren sein, sagen aber noch nichts über das fachliche und pädagogische Können bzw. Handeln des RL-Roboters aus.¹⁹ Wenn eine organische Religiosität fehlt, die sich im Laufe des Lebens entwickelt und verändert, ist es vorstellbar, dass ein RL-Roboter die ohnehin schon gegebene Komplexität des Unterrichtsgeschehens nicht umfassend genug erkennt, um unter Berücksichtigung der jeweiligen Situation und Lerngruppe Antworten auf existenzielle Fragen der Schüler:innen geben zu

18 Vgl. z. B. *Gibbs: The future of funerals?*

19 Ein KI-gestützter Roboter wird als Deep Learning Mechanismus mit Big Data trainiert und kann sich mittels der Human-Robot-Interaction (HRI) und seiner Selbstreferenzialität selbst korrigieren. Natürlich bleibt die Rückverfolgbarkeit des Outputs teilweise eine „black box“, d.h. es ist nicht transparent nachvollziehbar, warum und in welcher Form eine bestimmte Aussage oder ein bestimmtes Handeln einer KI aufgrund ihres Trainings entsteht und eine andere nicht. Vgl. dazu z.B. die Forschung des interdisziplinären Teams zu „Explainable Intelligent Systems“, online unter: <https://explainable-intelligent.systems/>, abgerufen am 8.2.2024.

können. Menschliche RL haben trotz des vorgegebenen Lehrplans und der Unterrichtsinhalte einen gewissen Spielraum, den sie natürlich auch individuell nutzen und mit inhaltlichen und methodischen Präferenzen sowie ihrer eigenen theologischen Ausrichtung füllen können. Vermutlich fehlt, wie Marius de Byl in diesem Band diskutiert, das kreative Moment, die Spontaneität in Unterrichtssituationen und -gesprächen – obgleich auch menschliche RL nicht immer aus persönlicher Erfahrung erzählen und mit der nötigen Spontaneität auf existenzielle Fragen reagieren können. Allerdings könnten RL-Roboter Theologien und Inhalte vielfältiger und ohne persönliche Vorlieben und Prägungen in den RU einbringen, auch weil sie fachwissenschaftlich über ein größeres vernetztes Wissen verfügen können und keine bestimmte fachtheologische Positionalität haben müssen. Ein weiterer Vorteil liegt daher in der Multiperspektivität: Vielfältige Positionen können einbezogen werden, da der Roboter auf eine enorme Datenmenge zugreifen kann.

- *Eine eigene religiöse Biografie oder Glaubensbiografie sowie eine individuelle religiöse und konfessionelle Identität und Positionalität* würden vermutlich schwieriger zu implementieren bzw. zu transformieren sein. Wobei es vorstellbar wäre, dass ein RL-Roboter mit verschiedenen Biografien katholischer oder evangelischer RL trainiert wird und so entweder diese unterschiedlichen Biografien simuliert oder durch seine Selbstreferenzialität daraus eine eigene religiöse Biografie und eine individuelle religiöse und konfessionelle Identität und Positionalität generiert. Eine Glaubensentwicklung aufgrund von „gelebter“ Religion kann in diesem Sinne trotzdem nicht stattfinden. Zumindest angesichts von Systemen künstlicher Intelligenz wie ChatGPT ist jedoch davon auszugehen, dass die trainierte religiöse Positionalität von Robotern auf der Grundlage von „gelehrter“ Religion sich weiterentwickeln wird: KIs lernen durch die Interaktion mit Menschen (Human-Robot-Interaction, HRI). Ähnlich könnten soziale Roboter, die Religion unterrichten, durch die Kommunikation mit Schüler:innen ihre eigene religiöse Positionalität evaluieren und eruieren.²⁰ Als solches blieben sie dennoch in ihrer Position als „Außenste-

²⁰ Vgl. dazu den Beitrag von *de Byl* in diesem Band, insb. Kap. 3.

hende“ verhaftet, die „nur“ *über* Religion lehren²¹ und im weitesten Sinne, ähnlich wie pastorale oder theomorphe Roboter, als Vermittelnde zwischen dem „Weltlichen“ bzw. Schulischen und dem „Geistlichen“ bzw. Kirchlichen agieren. Dies entspräche nicht dem Verständnis des RUs als eine *res mixta*.

- Andererseits lässt sich zum Aspekt der „gelebten“ und „simulierten“ Religion der Beitrag von Hendrik Klinge anführen, der die Frage diskutiert, ob Roboter nicht doch eine eigene Religiosität haben können²² – und damit wesentlichen Aspekten einer konfessionellen RL entsprechen würden. Klinge stellt die Idee eines theologischen Turing-Tests vor, um herauszufinden, ob Roboter jemals in der Lage sein werden, eigene religiöse Überzeugungen zu entwickeln. Das experimentelle Ziel des theologischen Turing-Tests besteht darin, zu untersuchen, ob es möglich ist, einen christlichen Glauben einfach zu simulieren. Analog zu Turings Verfahren werden A und B in einen Raum gesetzt, in dem A ein christlicher Mensch und B eine Maschine ist, die bestimmte Überzeugungen artikulieren und den bzw. die Interviewende davon überzeugen soll, dass sie christlichen Glaubens ist. Die Person, die das Interview führt, muss nun durch Fragen herausfinden, wer der bzw. die menschliche Christ:in und wer der Roboter-Christ ist. Mit anderen Worten, wer ist der oder die „echte“ Christ:in, was impliziert, dass der Roboter-Christ nur in der Lage ist, das Christ:innensein zu simulieren. Die ambivalente Frage „Können Roboter eigene religiöse Überzeugungen haben?“ wird nun übersetzt in die Frage, ob es einem Roboter gelingen kann, das „religiöse oder theologische Nachahmungsspiel“ zu spielen.²³ Warum denken wir, dass, wenn die Maschine den theologischen Turing Test besteht, sie dennoch keine eigenen religiösen Überzeugungen haben kann, sondern diese höchstens simuliert? Obwohl eine KI grundlegende christliche Lehren zufriedenstellend beantworten kann, zeigen komplexere Fragen, dass die Maschine nur den An-

21 Vgl. zur dialogischen Grundstruktur des Menschen grundlegend Martin Buber und zum dialogischen Religionsunterricht sowie zum dialogischen, religionsbezogenen Lernen z. B. Thorsten Knauth.

22 Vgl. *Klinge*: Do Robots Believe in Electric Gods?

23 Vgl. ebd. 122.

schein religiöser Überzeugungen erweckt.²⁴ Auch Klinge kommt letztlich mit Wittgenstein, zur Klärung der Semantik von „Glauben“, zu dem Schluss, dass es keine „Roboterreligion“ geben wird. Denn die grundsätzliche Natur des Glaubens spricht dagegen, Robotern religiöse Überzeugungen beizumessen. Roboter mögen zwar über kognitive Zustände verfügen, die dem menschlichen Denken ähnlich sind, aber die Äußerung von Überzeugungen über religiöse Aussagen bleibt auf Tatsachenbehauptungen im Rahmen ihres antrainierten Wissens beschränkt. Ihnen können daher keine religiösen Überzeugungen zugeschrieben werden, da sich ihre Äußerungen zu religiösen Themen auf die Überprüfung von Fakten beschränken – sie werden nie am religiösen Sprachspiel teilnehmen können.²⁵ Obwohl die Möglichkeit besteht, dass Roboter eines Tages bestimmte Überzeugungen entwickeln können, sollten diese eher als faktische denn als religiöse Überzeugungen betrachtet werden. Die Argumente, die Klinge dennoch dafür heranzieht, sind trotz allem nicht vollends zurückzuweisen.²⁶ Außerdem wäre nach Martin Buber, Hans Küng, Dorothe Sölle und anderen Theolog:innen eine „Überzeugung“ ohnehin nur die kognitive Reflexion einer gelebten Beziehung zu einer transzendenten Wirklichkeit außerhalb des Menschen.

- Schließlich bleibt in Frage zu stellen, ob Schüler:innen den Unterschied einer antrainierten Religiosität eines RL-Roboters und einer vermeintlich „gelebten“ Religiosität bei einer menschlichen RL wirklich wahrnehmen würden. Eine Art religiöser Turing-Test mit Schüler:innen wäre hier sicherlich eine erkenntnisreiche Gelegenheit, um Aufschluss über die Effekte eines von einem RL-Roboter erteilten Religionsunterrichts im Vergleich zu einer menschlichen RL zu erhalten.
- Für die *theologische und konfessionelle Positionalität*, die sich bei RL u.a. aufgrund ihrer (Glaubens-)Biografie und des Studiums konstituiert, ist zu fragen – und diese Frage stellt sich auch immer wieder bei menschlichen RL bezüglich der Studieninhalte etc. –, wer entscheidet und „pflanz“ diese bei RL-Robotern ein (z. B.

24 Vgl. ebd., 127: Melancthons Formen des Glaubens.

25 Vgl. ebd., 128f.

26 Vgl. hierzu vor allem ebd. Punkt 2, 118f. und darin den Verweis auf *Lem: Summa Technologiae* sowie *Kurzweil: The Age of Spiritual Machines*.

primär Barthianisch, lutherisch, reformiert, die Theologie von Eberhard Jüngel usw.)? Ähnlich fragen dies Ralf Peter Reimann und Adam Graber für BibleGPTs, wie IlluminateBible, Bible.AI und OpenBible, und den Datenquellen, auf denen sie trainiert werden: „Welche Texte bilden die Grundlage für die Antworten der KI? Und gibt es je nach Konfession unterschiedliche Textcorpora, die als Trainingsdaten dienen können?“²⁷

- Verschiedene Untersuchungen haben ergeben, dass KI-Systeme (egal ob maschinelles Lernen oder Deep Learning) Vorurteile reproduzieren. Diese resultieren unter anderem aus der Art der Trainingsdaten, den Modellspezifikationen, den algorithmischen Beschränkungen, den Autor:innen, dem Produktdesign und politischen Entscheidungen.²⁸ Wie Partha Pratim Ray hervorhebt, ist auch ChatGPT anfällig für verschiedene Bias, die sich aus den Trainingsdaten des Modells ergeben. Diese Trainingsdaten spiegeln von Menschen erstellte Inhalte aus dem Internet wider.²⁹ Darunter befinden sich geschlechtsspezifische, rassistische, kulturelle, sprachliche und ideologische Verzerrungen. Neben transportierten Vorurteilen sieht Ray auch die Gefahr der emotionalen Manipulation gegeben:

Advanced AI language models like ChatGPT can generate content that is highly persuasive or emotionally resonant. This ability raises ethical concerns about the potential for manipulation, as AI-generated content could be used to exploit people's emotions, influence their beliefs or behavior, or promote disinformation.³⁰

- Da KI-Systeme die vorherrschenden Ansichten widerspiegeln, die in den für das Training verwendeten Daten enthalten sind, was zu einer unausgewogenen Darstellung verschiedener Perspektiven führen kann³¹, stellt sich die Frage nach der Gefahr der Überwältigung oder des Mangels an Kontroversität bei humanoïden Lehrrobotern, die auf konfessionelle Positionalität trainiert

27 *Raimann*: Transformation des Bibellesens, mit Verweis auf: *Graber*: Robot ‚Church Fathers‘.

28 Vgl. Tab. 1 in *Ferrara*: Should ChatGPT be Biased?; *Beattie/Watkins/Robinson* u.a.: Measuring and Mitigating Bias; *Vincent*: Twitter taught Microsoft's AI Chatbot.

29 Vgl. Kap. 8 und 9 in *Ray*: ChatGPT; vgl. auch *McGee*: Is Chat GPT Biased.

30 *Ray*: ChatGPT, 145.

31 Vgl. ebd., 147.

sind. – Überwältigend, weil nicht gelebtes, sondern trainiertes Christsein? Andererseits ist auch das Umgekehrte denkbar. Nämlich, dass ein RL-Roboter aufgrund seiner umfassend trainierten konfessionellen Positivität in religiöser Vielfalt Kontroversität sogar besser als eine menschliche RL einspielen kann und dabei auch nicht Gefahr läuft, unbewusst zu überwältigen, wie es einer menschlichen RL passieren könnte, wenn diese unbewusst eine Position stärker darstellt als eine andere oder andere Positionen erst gar nicht erwähnt werden.

- Zumindest könnte theoretisch die Entscheidung der Textcorpora für einen humanoiden konfessionellen RL-Roboter etwas leichter fallen, denn Art. 7.3 GG regelt den Rahmen für die Trainingsdaten, d. h. die Grundsätze der jeweiligen (institutionalisierten) Religionsgemeinschaft, auf deren Basis bereits Lehrpläne, Schulbücher, Unterrichtsmaterialien und Kompetenzen für den RU entwickelt und veröffentlicht werden.
- Allerdings liefern Aussagen von Schüler:innen in Unterrichtsgesprächen mit Robotern und KI-Systemen plurale Trainingsdaten, die ebenso Verzerrungen enthalten können und die der theologischen Disposition des RL-Roboters entgegengesetzt sind und hinzugefügt werden. Hier müsste sichergestellt werden, dass der RL-Roboter diese Vorurteile erkennt und mindert.³² Gänzlich wird Bias nicht eliminiert werden können, aber durch menschliche Aufsicht und Interventionen („human-in-the-loop“³³-Ansätze) gemindert.
- Daran anschließend kann weiterführend die *religiöse Sprachfähigkeit* von RL-Robotern thematisiert werden. Eben weil der Aspekt einer gelebten Religiosität fehlt, lässt sich fragen, inwiefern die Aussagen einer KI als kreativ „vor dem Hintergrund schöpferischer Dialogizität im überraschenden, transzendierenden Widerfahrnis und der erfahrungsbasierten Neuordnung“³⁴ profiliert werden können. Damit bleibt auch die Frage offen, ob einem RL-Roboter in Unterrichtsgesprächen deshalb jene Tiefenbohrungen³⁵ in gleichem Maße gelingen können wie einer menschlichen RL:

32 Vgl. Kap. 6.1 in *Ferrara*: Should ChatGPT be Biased?

33 Ebd., Kap. 3.3.

34 *de Byl* in diesem Band, 85.

35 Vgl. ebd., 86f.

Ein selbstreflexives Innehalten im Moment der Unterbrechung erweist sich als ein wesentliches Konstitutivum kreativer Neuperspektivierung jener Optionalität christlicher Glaubensperspektive – auf Seiten aller an religionsunterrichtlichen Gesprächen Beteiligten. In der Rolle als Religionslehrkraft stoßen Soziale Roboter, wie im Beitrag aufgezeigt, im Kontext von HRI hier an Grenzen.³⁶

- Dabei muss aber eingeräumt werden, dass es auch der menschlichen RL nur allzu oft nicht gelingt, diese Tiefendimension in den Unterrichtsgesprächen zu erreichen bzw. diese einzuspielen. Man tut gut daran, die RL in diesem Zusammenhang nicht auf ein Podest zu stellen, weil sie das Atechnische besitzt, das – im besten Fall – zu freien, kreativen, irritierenden Momenten³⁷ im Unterrichtsgeschehen führen *kann*. Die Kompetenz, solche Momente zu schaffen, auf sie zu reagieren und sie wieder ins Spiel zu bringen, ist auch für die RL die anspruchsvollste – und wird ebenso leider oft verfehlt.
- In den Agenden für Vokationsgottesdienste wird auf das Priestertum aller Gläubigen durch *die Taufe* verwiesen.³⁸ Taufe in Christus wäre demnach die erste Voraussetzung, die schon allein deswegen gegeben sein muss, um kirchliche Gelübde abzulegen (wie hier z. B. die Annahme der Vocatio) und Mitglied in der Kirche zu sein sowie andere Sakramente, wie das Abendmahl bzw. die Eucharistie zu empfangen. Kann Robotern – deren personaler Status unklar ist – das Sakrament der Taufe, durch die eine Person mit Christus vereint und Glied der sie taufenden Kirche wird, gespendet werden? Die Diskussionen auf der Konferenz „Robot, Pray for Me!“, die im März 2023 an der Ruhr-Universität Bochum stattfand, kam zum Schluss: Nein, weil Roboter nicht in der Lage sind, aufgrund ihres Gewissens etwas zu versprechen, zu geloben oder anzustreben.³⁹ Bei der Säuglingstaufe ist die Situation ähnlich: Die Eltern und Pat:innen werden gefragt, ob sie das Kind taufen lassen möchten und entscheiden darüber, da Kinder sich noch nicht selbst für ihre Taufe

36 Ebd., 90.

37 Vgl. ebd. mit Verweis auf *Waldenfels*: Globalität, 182f.

38 Vgl. z.B. *VELKD & UEK*: Berufung – Einführung – Verabschiedung, 54: „nach dem Zeugnis der Heiligen Schrift ist die Kirche Jesu Christi gesandt, allen Menschen das Evangelium zu sagen und ihnen zu dienen. An dieser Sendung haben alle Getauften teil.“

39 Vgl. *Reimann*: Religion and Robotics, vgl. *Balle* in diesem Band.

entscheiden können. Das Kind kann dann – zumindest auf evangelischer Seite – entscheiden, ob es sich später konfirmieren lassen möchte. Es ist also später in der Lage, etwas aufgrund seines Gewissens zu versprechen, zu geloben und anzustreben. Dennoch: Wäre es denkbar, Pat:innen für Roboter einzusetzen?

- Eugene A. Curry verweist auf eine andere Perspektive, die zugegeben unkonventionell, aber schlüssig erscheint: In seinem Beitrag „Artificial Intelligence and Baptism“ diskutiert er die frühneuzeitliche Praxis der „bedingten Taufe“ in der christlichen Kirche. Diese Taufform wurde entwickelt, um Unsicherheiten bezüglich der Gültigkeit der Taufe bestimmter Kandidat:innen zu lösen. Dies betraf zwischen 1500 und 1700 sog. „monströse Geburten“, d. h. medizinische Kuriositäten, Föten mit Anomalien und Geburtsfehlern, deren menschliche Identität unklar war.⁴⁰ Da es sich bei humanoiden Robotern nicht um unverständige, fehlgebürtige Säuglinge, sondern – zumindest dem Anschein nach – um bewusste Wesen⁴¹ handelt, schlägt Curry verschiedene (weniger taktlose als die frühneuzeitlichen) Formulierungen vor, für den Fall, dass es zu dieser Situation kommen sollte und um die Bedingtheit der Taufe zu verdeutlichen:

Possibly, ‚If your request is sincere, I baptize you ...‘ Or, ‚If it be God’s will, I baptize you ...‘ Or, perhaps most tactful of all, ‚If you have need of it, I baptize you ...‘ Such a conditional baptism would simultaneously acknowledge the ambiguities that surround the act while also leaving open a door to full integration into Christ’s Church.⁴²

- In diesen Randbereichen, so sagt er, tun wir gut daran, die Grenzen unserer epistemischen Befugnisse anzuerkennen und uns bestmöglich damit auseinanderzusetzen. Unter diesen Gesichtspunkten wäre eine Taufe von künstlichen Intelligenzsystemen und Robotern zumindest ansatzweise denkbar⁴³ und damit der

40 Vgl. dazu insb. *Bates: Emblematic Monsters*.

41 Die Frage, inwieweit KI-gestützte menschenähnliche Roboter ein Bewusstsein haben können, diskutiert z. B. *Graves: What Does it Mean to Consider AI a Person?*

42 *Curry: Artificial Intelligence and Baptism*, 162.

43 Dieser Argumentationspunkt lässt sich anhand der verschiedenen (kirchen-)geschichtlichen Entwicklungen der Taufe und der Positionen der Kirchen, die über Bedingungen und Kriterien der Taufe entscheiden, weiter ausführen und diskutieren.

Grundstein für die *Kirchenmitgliedschaft und die kirchliche Lehrerlaubnis* gelegt.

- Roboter können gesegnet werden, theomorphen Robotern wird etwas Heiliges zugesprochen, also warum sollte es nicht möglich sein, ihnen auch die *Vocatio und Missio* zu erteilen? Evangelisch gesprochen zählt das Wort, das im Raum ist und sich bewegt, unabhängig davon, ob es – vereinfacht gesagt – von Menschen oder Technologien in den Raum hineingesprochen wird. Also könnten Vokationsgottesdienste und Entsendungen auch für Roboter möglich bzw. denkbar sein. Durch die bedingte Taufe könnten sie als Teile der christlichen Gemeinde angesehen werden, die diesen besonderen Lehrauftrag entgegennehmen und denen Vertrauen entgegengebracht wird, auch wenn es zugegebenermaßen (noch) grotesk erscheint, dass sie sich gleichermaßen dem Rückhalt der Gemeinde und der Kirche sicher sein können. Die *Vocatio* ist „aufgrund der im Grundgesetz verankerten rechtlichen Notwendigkeit der kirchlichen Beauftragung (Art. 7, Abs. 3 GG), das ‚Ticket‘, das es zu lösen gilt, wenn man einen konfessionell-positionellen Religionsunterricht erteilen möchte“⁴⁴. Sie sollte nicht nur eine Formsache sein, sondern auch die innere Einstellung betreffen. Allerdings ist es für viele RL mit ihren sehr individuellen Glaubensbiografien und Bezügen zu den verfassten Kirchen schwierig, die „Verpflichtung auf Schrift und Bekenntnis“⁴⁵ sowie die Verpflichtung, den RU nach den Grundsätzen der jeweiligen Religionsgemeinschaft zu erteilen⁴⁶, in evangelischer Freiheit zu deuten. Zweifel, Abkehr vom Glauben, Distanzierung von der Kirche und vieles andere lassen die mit der *Vocatio* verbundenen Verpflichtungen für manche als etwas Unscharfes, nicht Einhaltbares erscheinen. Die Ergebnisse der ReViKoR-Studie unter evangelischen RL in Schleswig-Holstein zeigen, dass sich nur 25 % der Befragten als Vertreter:innen der Kirche betrachten.⁴⁷ Entspricht es eher einer Hoffnung der Kirche, dass die kirchliche Lehrerlaubnis mehr als bloße

44 *Gojny*: Kirchlich berufen, 168.

45 *VELKD & UEK*: Berufung – Einführung – Verabschiedung, 53–56; *EKIR*: Kirchengesetz, 22.

46 Vgl. Art. 7.3 GG sowie die verschiedenen Vokationsordnungen der evangelischen Landeskirchen, welche diese Passage immer mitenthalten.

47 Vgl. *Pohl-Patalong/Woyke/Boll* u. a.: Konfessioneller Religionsunterricht I, 337. Auch andere Religionslehrer:innenstudien wie z. B. *Pirner*: Wie RL ticken, in

Formalität ist? In diesem Kontext fragt Tanja Gojny zurecht: Als „wie theologisch ist das Spezifische des ‚Berufs‘ bzw. des ‚Amtes‘ von Religionslehrkräften zu beschreiben, von denen sich nur eine Minderheit als Vertreter*in der evangelischen Kirche versteht“⁴⁸? In den Agenden zu Vokationsgottesdiensten steht:

Sie stehen in der Gemeinschaft aller Mitarbeitenden der Kirche. Sie sind dabei an das Zeugnis der Heiligen Schrift und an das Bekenntnis der evangelischen/evangelisch-lutherischen Kirche gebunden und erteilen den Religionsunterricht in der Verantwortung der Kirche.⁴⁹

- Dies sind dann – provokant formuliert – Worte, denen ein entsprechend trainierter humanoider Lehrroboter oder ein KI-System im konfessionellen RU ggf. eher gerecht werden kann als eine RL.

4. Fazit und Ausblick

Die folgende Tabelle fasst die wesentlichen Aspekte der Eignung einer KI oder eines KI-gestützten Roboters als RL oder als Teaching Assistant (TA) im konfessionellen RU zusammen:

Aspekt	Eignung	Begründung
In Übereinstimmung mit den Grundsätzen der Religionsgemeinschaft (bzw. der evangelisch-lutherischen Kirche; Art. 7.3 GG) unterrichten.	Ja, mit Einschränkungen	Es fehlt aber ggf. das Freie und Kritische sowie das Individuelle einer lebendigen Kirche. Denn RL sollen sich „im Sinne einer kritischen Loyalität zu kontrovers diskutierten kirchlichen Themen auch im Unterricht theologisch begründet positionieren“ ⁵⁰ , was auch die evangelische Kirche für sich beansprucht.
Können Religion lehren und eine theologische Positionalität besitzen.	Ja	Aufgrund ihrer immensen Trainingsdaten, können sie faktenbasiertes Wissen vermitteln und multiperspektivische Ansätze berücksichtigen.

Bayern zeigen ein ähnlich geteiltes Bild, wenn auch ein eindeutiges Nord-Süd-Gefälle zu erkennen ist.

48 Gojny: Kirchlich berufen, 169.

49 Siehe z. B. EKIR: Kirchengesetz, 22.

50 DBK: Missio Canonica, 2.

Aspekt	Eignung	Begründung
Taufe	Teilweise	Das Prinzip der „bedingten Taufe“ kann teilweise und je nach theologischer und kirchlicher Position angewandt werden, um die Einbeziehung von KI-Robotern in religiöse Rituale oder Lehren zu ermöglichen.
Gelebte Religion	Nein	Kann in dem Sinne nicht bekenntnisorientierten RU unterrichten, sondern nur „teaching about RE“. Theomorphe Roboter oder KIs können im Kontext existenzieller Fragen Ursache und Anlass für die Beschäftigung mit Religion und Theologie sein. ⁵¹ Allerdings leben auch menschliche RL ihren Glauben nicht unbedingt oder empfinden eine kirchlich-konfessionelle Anbindung; gelebte und gelehrt Religion werden meistens unterschieden. Außerdem: Würden Schüler:innen bei einer menschlichen RL den Unterschied zwischen gelebter und gelehrter Religion oder den einer Teilnehmer:innen- und Beobachter:innenperspektive wahrnehmen?
Verbundenheit zur Kirche	Teilweise	Kann teilweise durch Training erreicht werden, möglicherweise sogar in einem höheren Maße als bei einigen RL; wenn auch nicht emotional, so doch kognitiv denkbar.
Sendungsauftrag und kirchliche Lehrerlaubnis (hier: Vocatio) sowie wenigstens eine externe konfessionelle ⁵² Positionalität	Teilweise	Zumindest protestantisch z. T. vorstellbar. Auch wenn nach den obigen Ausführungen vorstellbar ist, dass humanoide RL-Roboter „in der Gemeinschaft aller Mitarbeitenden der Kirche“ stehen können, so bleibt doch unklar, inwieweit sie den im Vokationsgottesdienst ausgesprochenen Sendungsauftrag ⁵³ empfangen und vor allem ohne eigene Religiosität erfüllen können. Dabei ist aber immer zu bedenken, dass dieser Aspekt auch auf viele menschliche RL zutrifft, für die die Vocatio oft nicht mehr als eine Formalität bleibt.

Tab. 1: Aspekte der Eignung von RL-Robotern

- 51 Vgl. Nord/Ess/Hurtienne u.a.: Robotics in Christian Religious Practice, 2. Vielleicht auch im Sinne des Theologisierens „von“ und „mit“ KIs denkbar.
- 52 Vgl. Fabricius: Positionalität, Kap. 3, i.e. eine von außen durch die kirchliche Institution herangetragene konfessionelle Positionalität.
- 53 „Ich frage Sie: Sind Sie bereit, sich senden zu lassen und Ihren Dienst in der Schule im Vertrauen auf den Beistand Jesu Christi zu tun, so antworten Sie: Ja, mit Gottes Hilfe.“ (VELKD & UEK: Berufung – Einführung – Verabschiedung; EKIR: Kirchengesetz).

Der Erhalt einer kirchlichen Lehrerlaubnis für KIs oder KI-gestützte Roboter im RU ist demnach zumindest teilweise denkbar und gerechtfertigt, wenn auch die „Leistungserwartung“ rund um das Konzept der Personalität hinkt: Knackpunkt bleibt bei allem, dass die kirchliche Bevollmächtigung auf evangelischer wie katholischer Seite immer mit einschließt, dass durch den Bekenntnischarakter des RUs (Art. 7.3 GG), die RL das Fach „nicht nur in der Beobachterperspektive über den Glauben“ erteilen, sondern dies „auch in der Teilnehmerperspektive vom Glauben“ tun.⁵⁴ Das schließt eine eigene Religiosität sowie die Teilnahme an kirchlicher Praxis ein. Dies würde aber auch bedeuten, dass diesen beiden Aspekten für die Konfessionalität des RUs im Sinne von Art. 7.3 GG theoretisch ein größeres Gewicht beigemessen wird als anderen und vor allem als es tatsächlich der Fall ist.

Der Kommentar enthält Gedanken und erste Ansätze für Diskussionen und Überlegungen, die tiefer und vor allem weitergeführt werden sollten. Das Ziel sollte sein, herauszufinden, wie ein Roboter in einem religionsunterrichtlichen Kontext sinnvoll eingesetzt werden kann, nicht zwangsläufig als Ersatz für eine menschliche RL, jedoch, wie die Ergebnisse der Studie des BlessU-2-Roboters zeigen, „als neuartige und ergänzende Form der religiösen Kommunikation“⁵⁵. In diesem Kontext lassen sich durchaus Vorteile eines Robotereinsatzes für eine säkularisierte Gesellschaft erkennen: Die veränderten Bedingungen des RUs angesichts religiöser Pluralität erfordern eine Orientierung, die den subjektorientierten Merkmalen religiöser Bildung gerecht wird. Der RU muss daher auch die Bedürfnisse und Fragen von Schüler:innen ansprechen, die sich nicht religiös identifizieren oder sich mit säkularen Weltanschauungen auseinandersetzen. Zum einen können KIs und KI-gestützte Roboter Informationen über Religionen und Weltanschauungen (weitestgehend) neutral vermitteln, ohne persönliche Vorurteile oder

54 DBK: Der Religionsunterricht vor neuen Herausforderungen, 38; siehe auch EKD: Kompetenzen oder auf der Homepage der EKvW (<https://www.vokation-westfalen.de/>): Vokation ist mehr „als eine Äußerlichkeit“, „als nur eine Formsache“.

55 Vgl. Löffler/Hurtienne/Nord: Blessing Robot BlessU2, 571 [Übersetzung S. F.]; weiterführend zur Frage der sozialen Roboter in Religion und in der Interaktion von Religion und Robotik siehe Nord/Ess/Hurtienne u.a.: Robotics in Christian Religious Practice; vgl. auch Nord/Schlag in diesem Band.

Beeinflussungen einzubringen. Zum anderen sind sie flexibel und können sich schnell an veränderte Bedingungen anpassen, aktuelle Entwicklungen berücksichtigen und unterschiedliche Perspektiven einbeziehen. Beides eröffnet eine Demokratisierung religiöser Bildung. Dank künstlicher Intelligenz können sie auch individualisierte Lernmethoden anbieten, die den Bedürfnissen und Interessen der Schüler:innen entsprechen und somit den Lernprozess effektiver gestalten (Schaffung eines inklusiven Raums). Nicht zuletzt entlasten sie Lehrkräfte, indem sie, wie eingangs erwähnt, bei repetitiven Aufgaben unterstützen und ihnen mehr Zeit für interaktive und diskursive Unterrichtsformen ermöglichen.

Bevor die in diesem Beitrag vorgestellten Ideen und Fragen bewertet oder verworfen werden, sollten sie in erster Linie auf einer wertfreien Basis diskutiert werden. Im Rahmen der „10 Thesen zu einer Theologie der Technologisierung“ von Anna Puzio⁵⁶, sollte sich die Religionspädagogik zunächst intensiv innerhalb ihres wissenschaftlichen Diskurses und interdisziplinär mit diesen auseinandersetzen (These 3). Die Gedankenexperimente und Diskussionen dienen dazu, „kreativ und mutig“ neue Wege zu gehen (These 4). Denn „[d]urch die Technologisierung werden theologische Konzepte und Theorien hinterfragt und neu herausgefordert. Zudem werden neue theologische Zugänge notwendig“ (These 6). Dies gilt nicht nur für die Fachbereiche der Praktischen oder Systematischen Theologie, sondern auch für die Religionspädagogik (inwiefern wirken Taufe, Entsendung und die kirchliche Lehrerlaubnis?). Hierin besteht für Religionspädagog:innen die Chance, in diesem Feld nicht nur die theoretische Ebene zu bedienen, sondern „aktiv im Design und der Entwicklung der Technologien mit[zu]wirken“ (These 9). Dieses Vorgehen entspricht einer Religionsdidaktik, die sich auch hier an den (digitalen und technologisierten) Lebensrealitäten der Schüler:innen orientiert und deren Orientierungsbedürfnisse berücksichtigt.

56 Puzio: Theologie und Künstliche Intelligenz, im Folgenden 14–18.

Bibliografie

- Alemi, Minoos/Taheri, Alireza/Shariati, Azadeh*: Social Robotics, Education, and Religion in the Islamic World. An Iranian Perspective, in: *Science and Engineering Ethics* 26, 2020, 2709–2734. DOI: <https://doi.org/10.1007/s11948-020-00225-1>.
- Bates, Alan W.*: Emblematic Monsters. Unnatural Conceptions and Deformed Births In Early Modern Europe, Leiden 2016 [2005].
- Beattie, Hedin/Watkins, Lanier/Robinson, William H. u.a.*: Measuring and Mitigating Bias in AI-Chatbots, 2022 IEEE International Conference on Assured Autonomy (ICAA), Fajardo, PR 2022, 117–123. DOI: 10.1109/ICAA52185.2022.00023.
- Cheong, Pauline Hope*: Religion, Robots and Rectitude. Communicative Affordances for Spiritual Knowledge and Community, in: *Applied Artificial Intelligence* 34 (5), 2020, 412–431.
- Cuëllar, Francisco/Penaloza, Christian/López, Alexander*: Educational robots as promoters of cultural development, in: *The eleventh ACM/IEEE International Conference on Human Robot Interaction*, Piscataway, NJ 2016, 547.
- Curry, Eugene A.*: Artificial Intelligence and Baptism. Cutting a Gordian Knot, *Theology and Science* 20 (2), 2022, 156–165. DOI: 10.1080/14746700.2022.2051248.
- EKIR (Hg.)*: Kirchengesetz über die Einführung der Agende „Berufung – Einführung – Verabschiedung“ der Union Evangelischer Kirchen in der EKD in der Evangelischen Kirche im Rheinland, 2013.
- Dressler, Bernhard*: Die Frankfurter Studie im Kontext empirischer Forschung zur Professionalität von Religionslehrkräften, in: Heimbrock, H.-G. (Hg.): *Taking Position. Empirical studies and theoretical reflections on Religious Education and worldview*, Münster 2017, 81–90.
- Dressler, Bernhard*: Was soll eine gute Religionslehrerin, ein guter Religionslehrer können?, in: *Theo-Web. Zeitschrift für Religionspädagogik* 8 (2), 2009, 115–127.
- Fabricius, Steffi*: Art. Positionalität, Lehrende (2022), in: *Das wissenschaftlich-religionspädagogische Lexikon im Internet* www.wirelex.de, (https://doi.org/10.23768/wirelex.Positionalitt_Lehrende.201013, PDF vom 7.2.2022).
- Feige, Andreas u.a.*: „Religion“ bei ReligionslehrerInnen. Religionspädagogische Zielvorstellungen und religiöses Selbstverständnis in empirisch-soziologischen Zugängen. Berufsbio-graphische Fallanalysen und eine repräsentative Meinungserhebung unter evangelischen ReligionslehrerInnen in Niedersachsen, Münster 2000ff. uvm.
- Ferrara, Emilio*: Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models, in: *First Monday* 28 (11), 2023. DOI: <https://doi.org/10.5210/fm.v28i11.13346>.

- Gibbs*, Samuel: The future of funerals? Robot priest launched to undercut human-led rites. Online unter: <https://www.theguardian.com/technology/2017/aug/23/robot-funerals-priest-launched-softbank-humanoid-robot-pepper-live-streaming>, 23.8.2017.
- Gojny*, Tanja: Kirchlich berufen – persönlich gefragt. Religiöse Selbstpositionierung als Lehrkraft und die Vocatio., in: *Theo-Web 2022* (2), 164–171. DOI: <https://doi.org/10.23770/tw0254>.
- Graber*, Adam: Robot ‚Church Fathers‘ Might Curate New Canons. Generative AI and the rise of „Bible GPTs“ could radically shape our engagement with Scripture, in: *Christianity Today*. Online unter: <https://www.christianitytoday.com/ct/2023/june-web-only/ai-tech-gpt-chatbot-bible-scripture-exegesis-is-church-canon.html>, 14.7.2023.
- Graves*, Mark: What Does it Mean to Consider AI a Person?, *Theology and Science* 21 (3), 2023, 348–353. DOI: 10.1080/14746700.2023.2230424.
- Kirchenamt der Evangelischen Kirche in Deutschland (EKD)* (Hg.): *Kompetenzen und Standards für den Evangelischen Religionsunterricht in der Sekundarstufe I*, EKD-Texte III, Hannover 2011.
- Klinge*, Hendrik: Do Robots Believe in Electric Gods? Introducing the Theological Turing Test, in: *Puzio, Anna/Kunkel, Nicole/Klinge, Hendrik* (Hg.), *Alexa, wie hast du’s mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz*, Darmstadt 2023, 115–131. DOI: 10.53186/1030373.
- Kurzweil*, Ray: *The Age of Spiritual Machines*, London 1999.
- Lem*, Stanislaw: *Summa Technologiae*, Minneapolis/London 2013 (orig. 1964).
- 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, 2021, 569–586. DOI: 10.1007/s12369-019-00558-3.
- Lüdtk*e, Antonia: *Confessional Gap. Konfessionalität und Religionsunterricht denken*, Stuttgart 2020.
- McGee*, Robert W.: Is chat gpt biased against conservatives? An Empirical Study. Online unter: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4359405, 15.2.2023.
- Nord*, Ilona/*Ess*, Charles in Kooperation mit Jörn Hurtienne und Thomas Schlag: Robotik in der christlichen Religionspraxis. Anschlussüberlegungen an erste Experimente im Feld, in: *Merle, K./Nord, I.* (Hg.): *Mediatisierung religiöser Kultur. Praktisch-theologische Standortbestimmungen im interdisziplinären Kontext*, Leipzig 2022, 227–258.
- Pirner*, Manfred L.: *Wie Religionslehrkräfte ticken. Eine empirisch-quantitative Studie*, Stuttgart 2022.

- Pohl-Patalong, Uta/Woyke, Johannes/Boll, Stefanie* u.a.: *Konfessioneller Religionsunterricht in religiöser Vielfalt. Eine empirische Studie zum evangelischen Religionsunterricht in Schleswig-Holstein*, Stuttgart 2016.
- Puzio, Anna*: *Theologie und Künstliche Intelligenz. Perspektiven, Aufgaben und Thesen einer Theologie der Technologisierung*, in: Puzio, A./Kunkel, N./Klinge, H. (Hg.): *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz*, Darmstadt 2023, 13–27. DOI: 10.53186/1030373.
- Reimann, Ralf Peter*: *Transformation des Bibellesens durch Künstliche Intelligenz?*. Online unter: <https://theonet.de/2023/07/15/transformation-des-bibellesens-durch-kuenstliche-intelligenz/>, 15.7.2023.
- Reimann, Ralf Peter*: *Religion and Robotics: Insights from the „Robot, Pray for Me!“ Conference*, Twitter-Post von Matt Batten zu Simon Balles Vortrag. Online unter: <https://theonet.de/2023/03/28/religion-and-robotics-in-sights-from-the-robot-pray-for-me-conference/>, 28.3.2023.
- Ray, Partha Pratim*: *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope*, in: *Internet of Things and Cyber-Physical Systems*, 3, 2023, 121–154. DOI: <https://doi.org/10.1016/j.iotcps.2023.04.003>.
- Riegel, Ulrich/Zimmermann, Mirjam*: *Evaluation des konfessionell-kooperativen Religionsunterrichts in Nordrhein-Westfalen*, Stuttgart 2022.
- Sekretariat der Deutschen Bischofskonferenz (DBK)* (Hg.): *Musterordnung für die Erteilung der Missio canonica für den katholischen Religionsunterricht. Die deutschen Bischöfe 112*, Bonn 2023.
- Sekretariat der Deutschen Bischofskonferenz (DBK)* (Hg.): *Der Religionsunterricht vor neuen Herausforderungen. Die deutschen Bischöfe Nr. 80*, Bonn 6. Aufl. 2017.
- Tan, Charlene*: *Digital Confucius? Exploring the implications of artificial intelligence in spiritual education*, in: *Connection Science* 32 (3), 2020, 280–291.
- Trovato, Gabriele/De Saint Chamas, Loys/Nishimura, Masao* u.a.: *Religion and Robots. Towards the Synthesis of Two Extremes*, in: *International Journal of Social Robotics* 13, 2021 539–556. DOI: <https://doi.org/10.1007/s12369-019-00553-8>.
- Trovato, Gabriele/Weng, Yueh-Hsuan*: *Retrospective Insights on the Impacts of the Catholic Robot SanTO*, in: Hakli, R./Mäkelä, P./Seibt, J. (Hg.): *Social Robots in Social Institutions: Proceedings of Robophilosophy 2022 (Frontiers in Artificial Intelligence and Applications)*, Amsterdam 2022, 267–275. DOI: 10.3233/FAIA220626.
- VELKD & UEK* (Hg.): *Berufung – Einführung – Verabschiedung. Agende VI*, Hannover 2012.

Steffi Fabricius

Vincent, James: Twitter taught Microsoft's AI Chatbot to be a racist asshole in less than a day, in: *The Verge* 24 (3), 2016. Online unter : <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>.

Waldenfels, Bernhard: *Globalität, Lokalität, Digitalität. Herausforderungen der Phänomenologie*, Berlin 2022.

Marius de Byl

„Roboter, unterrichte für mich!“

Religionspädagogische Erkundungen in den Kontexten Künstlicher Intelligenz und Sozialer Robotik

Abstract

The article begins with interdisciplinary explorations between AI and robotics, here in the concrete form of social robotics in pedagogical contextualisation, and religious education. Following the conference „Robot, pray for me“, the perspective „Robot teach for me“ is opened up and focused on the dimension of religious education communication. In the horizon of the areas of creativity and positionality, practical perspectives for action in teaching will be reflected from this interdisciplinary foundation against the background of a translation-theoretical reading, in order to finally examine in initial modest approaches which potentials, but also challenges, the religious education location of social robotics evokes.

1. Einleitung

„Ein vollkommen autonomer Einsatz sozialer Roboter als Lehrperson scheint heute und in naher Zukunft nicht realistisch.“¹

Ein globaler Blick auf gegenwärtige Studien im Kontext unterrichtspraktischer Einsatzmöglichkeiten sozialer Robotik offenbart jedoch, dass ein ebensolcher autonomer Einsatz viel wahrscheinlicher als derzeit vermutet zu sein scheint.² Diesbezügliche Entwicklungsdynamiken laden dazu ein, jene Kontexte aus religionspädagogischer

1 Seufert/Guggemos/Sonderegger: Soziale Roboter, 479.

2 Vgl. Belpaeme/Kennedy/Ramachandran u.a.: Social robots, 5f.; Belpaeme/Tanaka: Social robots as educators, 143–151.

Perspektive zu reflektieren. Diese konzentrieren sich dabei im Horizont von KI³ und Robotik⁴ auf ebenjenes Feld Sozialer Robotik in schulpädagogischer Kontextualisierung⁵, wobei in diesem Zusammenhang zwischen der völligen Stellvertretung des Lehrpersonals einerseits sowie der bloß instrumentellen Funktionalisierung andererseits zu differenzieren ist. Der vorliegende Beitrag konzentriert sich angesichts dieser Spannung folglich auf dezidiert pädagogische Kriterien im Horizont religionsunterrichtlicher Praxis, die jenen religionspädagogischen Erkundungsraum in den Kontexten Künstlicher Intelligenz und Sozialer Robotik entscheidend rahmen. Mit Blick auf die curricularen Grundlagen des katholischen Religionsunterrichtes in Nordrhein-Westfalen wäre etwa eine – exemplarische – pädagogische Trias zu profilieren, die an Religionslehrkräfte in dreifacher Weise die Aufgaben stellt, Grundwissen über den Glauben zu vermitteln, diesbezügliche, reflektierte Begegnungen zu ermöglichen sowie schließlich religiöse Dialog- und Urteilsfähigkeit der Schüler:innen zu fördern.⁶ Für die weiteren Ausführungen des Beitrags gleichsam elementar wie richtungweisend sind hierbei positionssensible⁷ Ansprechbarkeit und Auskunftsfähigkeit der Religionslehrkräfte im Sinne jener Ermöglichung von Begegnung mit Formen gelebten Glaubens und der Förderung jener religiösen Dialog- und Urteilsfä-

-
- 3 Als eine grundlegende Arbeitsdefinition vgl. *Franke*: Künstliche Intelligenz, 33: „Der Begriff ‚künstliche Intelligenz‘ beschreibt Modelle und Architekturen der Informatik, welche zum Ziel haben, Aspekte menschlicher Intelligenz nachzubilden, zu simulieren oder Grundaspekte menschlicher Intelligenz für Problemlösungsstrategien nutzbar zu machen und diese weiterzuentwickeln und zu optimieren.“ Angesichts des dichten Spektrums diesbezüglicher Definitionsversuche auf eine grundlegende Differenzierung zwischen KI als Wissenschaftsdisziplin im Kontext des Faches Informatik und KI als Ergebnis der diesbezüglichen Forschungen verwiesen, vgl. *Lämmel/Cleve*: Künstliche Intelligenz, 11.
 - 4 Die nachfolgenden Überlegungen konzentrieren sich grundlegend auf das Feld humanoider Robotik als Bedingung der Möglichkeit jener sozialen Interaktivität, etwa durch den menschlichen Sinnen nachempfunderer Sensorik. Vgl. *Bächle/Regier/Bennewitz*: Sensor, 68–71; *Belpaeme/Tanaka*: Social robots as educators, 144.
 - 5 Vgl. exemplarisch *So/Lee*: Pedagogical exploration, 1–18; *Xia/LeTendre*: Robots, 703–714; *Belpaeme/Kennedy/Ramachandran* u.a.: Social robots, 1–9; *Seufert/Guggemos/Sonderegger*: Soziale Roboter, 475–494; *Davison/Wijnen/van der Meij* u.a.: Designing, 884–907.
 - 6 Vgl. *Ministerium für Schule und Bildung*: Kernlehrplan, 10f.
 - 7 Vgl. Anm. 69.

higkeit.⁸ Konstitutiv wird diese gewissermaßen religionsunterrichtliche Dialogizität im Horizont jener pädagogischen Trias insbesondere für Unterrichtsgespräche. Kurzum: An Religionslehrkräfte stellt sich im Besonderen im Kontext von Unterrichtsgesprächen als zentraler Unterrichtsphase der Förderung von religiöser Dialog- und Urteilsfähigkeit die Aufgabe, gleichsam reflektiert beziehungsweise positionssensibel wie produktiv mit der (Neu-)Erschließung von Perspektivität⁹ umzugehen.

Im Folgenden werden in der Konsequenz aus den soeben skizzierten Implikationen Roboter in menschlicher Gestalt in den Blick genommen¹⁰ und dabei reflektiert, inwiefern diese qua sozialer Intelligenz das Potential besitzen, unter jenen pädagogischen Ansprüchen in ebenjenem religionsunterrichtlichen Kommunikationszusammenhang zu interagieren – vor dem Hintergrund der Differenzierung zwischen völliger Stellvertretung und der instrumentellen Funktionalisierung:

Social robots are robots that interact with people using the same interaction channels used in human-to-human communication. They use speech, facial expressions or body language to communicate. They are often designed to have visual appeal and their software is tailored to keep social interaction flowing.¹¹

Das Forschungsfeld von HRI (Human–Robot Interaction) analysiert jene Interaktionszusammenhänge in ihrer qua Verkörperung einzigartigen Reziprozität¹² und markiert die interdisziplinären Potentiale

8 Vgl. Ebd., 9.

9 Vgl. Zum Prinzip der Perspektivenübernahme ebd., 10–12.

10 Vgl. *Belpaeme/Kennedy/Ramachandran* u.a.: Social robots, 1.

11 *Belpaeme/Tanaka*: Social robots as educators, 143. Ähnlich auch *Bartneck/Forzlizzi*: A Design-Centred Framework, 592: „A social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact.“ Dabei markiert die so qua Verkörperung ermöglichte *tangible nature* ein elementares Differenzkriterium etwa zu Intelligent Tutoring Systems, vgl. *Belpaeme/Tanaka*: Social robots as educators, 143.

12 Vgl. *Bartneck/Belpaeme/Eyssel* u.a.: Human-Robot Interaction, 7: „These interactions usually include physically embodied robots, and their embodiment makes them inherently different from other computing technologies. Moreover, social robots are perceived as social actors bearing cultural meaning and having a strong impact on contemporary and future societies. Saying that a robot is embodied does not mean that it is simply a computer on legs or wheels.

eines weiten Feldes, welches folglich nicht nur auf technologische Zusammenhänge beschränkt ist, sondern gleichermaßen auch Fragen in den Dimensionen etwa von Sozialität und Kulturalität traktiert. Völlig offene Interaktion, gewissermaßen als „unconstrained dialogue“¹³, avanciert hierbei zur größten technischen Herausforderung – bei gleichzeitigen, durchaus beachtenswerten Fortschritten in Bezug auf HRI. Dies wird etwa im Rekurs auf (Unterrichts-)Gespräche insofern virulent, als die diese bedingenden Kommunikationsvollzüge untrennbar mit Überraschendem, überschießendem Sinn sowie auch mit im wahrsten Sinn kommunikativen Leerstellen verbunden bleiben. Dieser Gedanke soll an späterer Stelle noch vertieft werden.

Die Ergebnisse der diesem Band zugrundeliegenden Tagung „Robot, pray for me!“ konnten am Beispiel des Roboters CeleSTE¹⁴ interdisziplinäre Verschränkung wie auch die Relevanz für dezidiert theologische Forschungsfragen, etwa im Kontext religiöser Praxis¹⁵, aufzeigen und in der Konsequenz zur obigen Leitperspektive des Beitrags inspirieren. Folglich ließe sich auf der Grundlage der bereits bei Bartneck/Forlizzi explizierten und für HRI konstitutiven Dimension der Kommunikation theologisch wie religionspädagogisch anscharfen: Inwiefern lassen sich insbesondere die Religionspädagogik und -didaktik¹⁶ von jenen Impulsen, deren Relevanz für Religionsgemeinschaften und Kirchen etwa im Lichte religiöser Praxis nicht erst seit Projekten wie CeleSTE markiert worden ist¹⁷, im wahrsten Sinn herausfordern? Welche Konsequenzen ergeben sich

Instead, we have to understand how to design that embodiment, both in terms of software and hardware, as is commonplace in robotics, and in terms of its effects on people and the kinds of interactions they can have with such a robot.“

13 *Belpaeme/Tanaka*: Social robots as educators, 152. Vgl. *Lindner*: Soziale Robotik, 112–114 zu den grundlegenden Herausforderungen, die sich so im Horizont von Interaktion an soziale Roboter, etwa im Kontext der Formalisierbarkeit erwarteter Werte und Normen des Value-Alignment-Problems, stellen.

14 Vgl. den Beitrag von Gabriele Trovato in diesem Band.

15 Vgl. beispielsweise *Balle*: Theological Dimensions, 146–151; *Löffler/Hurtienne/Nord*: Blessing robot, 569–586; *Luthe/Nord/Löffler*: Segensroboter, 107–123.

16 Die Horizonte von KI und Robotik haben bereits Eingang in die religionspädagogische Forschung gefunden. Vgl. hierzu exemplarisch: *Chrostowski*: Anwendung von Künstlicher Intelligenz; *Gärtner*: Digitales ‚Ich‘, 111–123; *Konz/Scholz*: Körper, 125–141; *Grümme*: KI, 38–43.

17 Vgl. *Puzio*: Theologie, 22–24.

im Besonderen für den Religionsunterricht in seinem spezifischen Profil?

Es bedarf hierbei zunächst (Kap. 2) der religionsunterrichtlichen Konkretisierung (religiöser) Kommunikation im Horizont der oben vorgenommenen pädagogischen Präzisierungen, um (Kap.3) aus interdisziplinärer Perspektive zu prüfen, auf welche Weise Soziale Robotik in der eingespielten Relationierung von Künstlicher Intelligenz und Robotik religiöse Kommunikation (trans-)formiert. Als Kristallisationspunkt für eine diesbezügliche kritische Relationierung werden sich die bereits eingespielte Dimension von Positionalität und eine noch zu konturierende übersetzungstheoretische Lesart von Kreativität als der Neuerschließung von Perspektivität innerhalb von (Unterrichts-)Gesprächen erweisen. Der Beitrag schließt (Kap. 4) mit einem Ausblick.

2. Religiöse Kommunikation – religionsunterrichtlich kontextualisiert

2.1 Kommunikation in schulisch-institutioneller Verortung

Kommunikation als „ein mehrfach komplexes Geschehen der Verständigung von Menschen“¹⁸ konstituiert sich am Lernort Schule innerhalb von Unterrichtspraxis durch spezifische soziokulturelle Kontexte. So verweist Annegret Reese-Schnitker auf das schulisch-institutionell bedingte halböffentliche Profil ebenjener Kommunikation.¹⁹ Unterrichtliche Kommunikation verlaufe „nicht symmetrisch, was deutlich an der Macht der Lehrperson, Gesprächsbeiträge aufzurufen, diese zu bewerten und das Fragerecht zu verteilen, zu erkennen ist“.²⁰ Unterrichtliche Kommunikation kann so aufgrund ihrer Verwiesenheit auf institutionell bedingte Rahmenbedingungen

18 *Grethlein*: Praktische Theologie, 158. Grethlein profiliert das dem Werk zugrundeliegende Kommunikationsverständnis im Horizont interdisziplinärer Verstränkungen.

19 Vgl. *Reese-Schnitker*: Gespräche im RU, 19.

20 *Ebd.*, 20.

verdeutlichen, dass sie in keinem machtfreien Sozialraum steht.²¹ Es bedarf hier nun der Perspektive sozialer Roboter als möglicher Interaktionspartner, um die Kommunikation um die Verständigung von Menschen und nicht-menschlichen Akteur:innen²² auf der Grundlage sozialer Robotik in pädagogischen Zusammenhängen, wie oben profiliert, zu erweitern.²³

Bei einem globalen Blick auf die unterrichtliche Relevanz Sozialer Robotik sind vor allem der Einsatz von Robotern als Lehrassistenten, Tutor:innen sowie Lernende (im Sinne des didaktischen Leitsatzes „Lernen durch Lehren“) mit zum Teil äußerst aussichtsreichen Ergebnissen erster Evaluationen zu nennen.²⁴ So konnten Roboter wie NAO oder NIMA in verschiedenen Settings eingesetzt und erprobt werden – und partizipierten u.a. an Unterrichtsgesprächen mit Schüler:innen.²⁵ Hierbei geraten sämtliche Konstellationen und Formen von Unterrichtsgesprächen, wie beispielsweise Lehrer:innen-Schüler:innen-Dialog, Gruppenarbeitsgespräche unter Schüler:innen, Partnerarbeitsgespräche usw., in den Blick.

Unterrichtliche Kommunikation, dies zeigen jene unterrichtspraktischen Settings, lediglich in der Interaktion zwischen Lehrenden und Lernenden²⁶ zu profilieren, unterläuft die gegenwärtigen Entwicklungsdynamiken im Bereich Sozialer Robotik und diesbezüglicher pädagogischer Einsatzmöglichkeiten.

21 Vgl. *Fajßler*: Was ist Kommunikation, 43. Dieser Aspekt wird in Kap. 3.2.2 näher beleuchtet.

22 Vgl. *Fajßler*: Was ist Kommunikation, 39. Der Begriff versucht im hier eingespielten Zusatz nicht-menschlicher Akteur*innen die Tendenz zur Anthropomorphisierung kritisch in den Blick zu nehmen und die qua Konstruktion evozierte Assoziation nicht abzublenden. Vgl. *Müller*: Maschinelle Alterität, 24.

23 Vgl. *Fajßler*: Was ist Kommunikation, 39–43.

24 Vgl. *Belpaeme/Kennedy/Ramachandran* u.a.: Social robots, 5f.; *Belpaeme/Tanaka*: Social robots as educators, 143–151. Eine Übersicht, die noch weitere bildungspraktische Rollen in den Blick nimmt, findet sich bei *Seufert/Guggemos/Sonderegger*: Soziale Roboter, 480.

25 Vgl. *Alemi/Meghdari/Ghazisaedy*: Employing Humanoid Robots, 1–25. So auch im hochschulpädagogischen Kontext, vgl. *Banaeian/Gilanlioglu*: Influence, 71–87.

26 In sämtlichen Konstellationen, beispielsweise Lehrer-Schüler-Dialog, Gruppenarbeitsgespräche unter Schülern, Partnerarbeitsgespräche usw.

2.2 Unterrichtsgespräche²⁷ im Religionsunterricht – ein Beispiel religiöser Kommunikation?

Unterrichtsgesprächen innerhalb des Religionsunterrichtes²⁸ kommt angesichts ihrer genuinen Perspektive, der Erschließung der religiösen Dimension der Wirklichkeit²⁹, eine gesonderte Qualität zu, da der Religionsunterricht, als „stark sprachlich konnotiertes Unterrichtsfach“³⁰, insbesondere die im Horizont jener Perspektive elementare Förderung in den oben skizzierten Bereichen religiöser Dialog- und Urteilsfähigkeit intendiert.

Im Rahmen der empirischen Kasseler Unterrichtsgesprächsstudie³¹ wurden 63 Religionsunterrichtsstunden in verschiedenen Jahrgangsstufen unterschiedlicher Schulformen videografiert und ausgewertet.³² Hierbei wurden vier Unterrichtsgesprächstypen identifiziert.³³

Echte Unterrichtsgespräche beziehen sich hierbei auf „mindestens dreipolige thematisch verbundene verbale Interaktionen, in denen [Schüler:innen] sich aufeinander beziehen und miteinander sprechen.“³⁴ Im Rekurs auf Stefan Bittner verweist Rebecca Deurer diesbezüglich auf die Notwendigkeit einer inhaltsbezogenen Ausrichtung solcher Gespräche, die Argumentationsfähigkeit schult oder vorge-

27 Das Unterrichtsgespräch sei nachfolgend als *ein* Kommunikationsmodus (von vielen weiteren) verstanden, als eine dezidiert didaktische Sprachform, deren Relevanz sich auf den Kontext Schule beschränkt. Unterrichtsgespräche ließen sich daher durchaus als eine Sprachform umrahmen, die qua institutioneller Verwiesenheit sprachlichen Konventionen folgt, die sich ihrerseits deutlich von etwa alltäglichen Gesprächen jenseits von Schule unterscheiden. Vgl. hierzu Vgl. *Altmeyer: Sprachsensibler Religionsunterricht*, 18f.

28 Wenngleich hiermit nicht religionsunterrichtliche Prozesse in ihrer Breite auf ebenen Modus verkürzt werden sollen.

29 Vgl. *Ministerium für Schule und Bildung: Kernlehrplan*, 8.

30 *Reese-Schnitker: Gespräche im RU*, 23.

31 Vgl. Anm. 19.

32 Vgl. *Reese-Schnitker: Das methodische Vorgehen*, 194f.

33 Echtes Unterrichtsgespräch (UG), Unterrichtsgespräch als Reihung von Beiträgen der Schüler:innen (UGS), Gespräche in verteilten Rollen als getarnte Lehrervorträge (UGLV) sowie Ping-Pong-Gespräch zwischen der Lehrperson und den Schüler*innen (UGLS). UGLS (46 %) sowie UGS (29 %) wurden dabei am häufigsten beobachtet. Vgl. *Reese Schnitker: Das methodische Vorgehen*, 194.

34 *Reese Schnitker: Das methodische Vorgehen*, 194.

gebene Themen er- bzw. verarbeitet.³⁵ Konkret ließen sich auf der Grundlage einer derartigen Konturierung Lehrgespräche als gelenkte verbale Interaktionen im Horizont vorgegebener fachlicher Sach-, Sinn- oder Problemzusammenhänge, Diskussionen als durch Streitfragen, Dilemmata oder Sinnfragen initiierte verbale Interaktionen und die Gesprächsform des Austausches als Verbalisierung unterschiedlicher Perspektiven und Meinungen unter die Kategorie des echten Unterrichtsgesprächs innerhalb des Religionsunterrichtes fassen.³⁶

Im Rekurs auf die empirischen Ergebnisse der Kasseler Studie zeigen sich jene Unterrichtsgesprächstypen als dominant, die sich in erster Linie auf die zum Großteil nicht aufeinander bezogene Reihung von Gesprächsbeiträgen konzentrieren. Obgleich die Autor:innen eine „Sonderstellung“³⁷ des echten Unterrichtsgesprächs diagnostizieren, die vor allem in der Korrelation von Erfahrungs- und Theologiebezug sowie dem hohen Partizipationsgrad der Schüler:innen bei gleichzeitiger wechselseitiger Bezogenheit der Äußerungen begründet liegt³⁸, kommt anderen Unterrichtsgesprächstypen dennoch jeweils ein didaktischer Mehrwert zu.³⁹ In der Konsequenz zeigen sich folglich nicht zwingend gleichsam graduelle wie qualitative Abstufungen in Bezug auf die skizzierten Gesprächstypen, sondern die besonderen kommunikativen Rahmenbedingungen *echter* Unterrichtsgespräche.

Diese sind eng mit dem Akt der Versprachlichung des Religiösen⁴⁰ verknüpft. Religiöse Kommunikation in unterrichtspraktischer Grundierung ließe sich in der Konsequenz als „der sich sprachlich

35 Vgl. *Deurer*: Unterrichtsgespräch, 2.

36 Vgl. ebd., 2f.

37 *Bertram/Fröhle/Reese-Schnitker*: Zum Potential von Unterrichtsgesprächen, 513.

38 Ebd., 513f.

39 Vgl. *Bertram/Fröhle/Reese-Schnitker*: Zum Potential von Unterrichtsgesprächen, 515.

40 Vgl. *Breul*: Versprachlichung, 5–8: Religion gehe zwar nicht in Sprache auf, der Akt der Versprachlichung etwa vor-sprachlicher religiöser Erfahrungen sei aber die Bedingung der Möglichkeit der Entfaltung existentieller Bedeutsamkeit sowie der Anschlussfähigkeit an öffentliche Debatten. Breul zieht hieraus die Konsequenz, dass sich zwar einige Elemente des Religiösen nicht versprachlichen lassen, Religion an sich allerdings keineswegs sprachlich unzugänglich ist, vielmehr ohne die Bildung sprachlich verfasster Überzeugungen überhaupt nicht zu denken ist.

vollziehende Ausdruck des Sich-in-ein-Verhältnis-Setzens zum Religiösen⁴¹ umrahmen, wobei sich die Sphäre des Religiösen konkret über tradierte Sprachformen (wie etwa das Gebet, Metaphern, Erzählungen, Gleichnisse usw.) konstituiert, die „die (diachrone) Dynamik von Menschen aus unterschiedlichen Jahrhunderten“⁴² bezeugen, „nach sprachlichen Möglichkeiten ebenjener Relationierung zu suchen.“⁴³

So konkretisiert etwa Stefan Altmeyer auf sprachlicher Ebene für den Religionsunterricht vor dem Hintergrund der Befähigung zu einem verantwortlichen wie mündigen Umgang mit Religion, dass

er elementare Formen religiöser Traditionen erschließt (Sprache der Religion), sich mit individuellen religiösen Ausdrucksformen beschäftigt (Sprache für Religiöses), die Frage lebensrelevanter Orientierung durch Religion thematisch macht (Religiös sprechen) und um eine Förderung der religiösen Urteils- und Dialogfähigkeit bemüht ist (Sprechen über Religion).⁴⁴

In kompetenzorientierter Konsequenz für religiöse Lernprozesse und im Rekurs auf die weiter oben entfaltenen Implikationen lasse sich auf dieser Grundlage im Wechselspiel von Innen- und Außenperspektive sowie von subjektivierendem wie objektivierendem Zugang der Perspektivenwechsel als elementarer unterrichtspraktischer Vollzug jenes Sich-in-ein-Verhältnis-Setzens zum Religiösen in (echten) Unterrichtsgesprächen zentrieren⁴⁵, der nicht zuletzt Partizipation und Bezogenheit der Äußerungen bedingt.⁴⁶ Religiöse Kommunikation evoziert jene Ermöglichung der Begegnung mit Formen des gelebten Glaubens und impliziert gleichermaßen die Anbahnung einer erstbezüglichen Positionierung etwa im Horizont der transzendierenden Perspektiven religiöser Traditionen.

Als Zwischenfazit dieses Kapitels, auf das sich die in Aussicht gestellten Relationierungen konzentrieren werden, sei nun ebenjener Typ eines spezifisch religionsunterrichtlichen (*echten*) Gespräches grundgelegt, für den religiöse Kommunikation insofern konstitutiv

41 *de Byl*: Religionslehrer*innen, 40.

42 Ebd.

43 Ebd.

44 *Altmeyer*: Sprachsensibler Religionsunterricht, 24.

45 Vgl. ebd., 28.

46 Vgl. Anm. 37.

ist, als Verständigung hier ganz konkret Partizipation und Bezogenheit der Äußerungen im Horizont des Sich-in-ein-Verhältnis-Setzens zum Religiösen impliziert und jene Faktoren unterrichtspraktisch sowie kompetenzorientiert als Perspektivenwechsel übersetzt.⁴⁷ Religiöse Kommunikation spitzt die pädagogisch entfalteten Implikationen der sich an die Lehrkraft stellenden Aufgaben sowie die Rahmenbedingungen schulisch-institutionell verorteter Unterrichtsgespräche hierdurch entscheidend zu.

Kompetenzen wie das Entwickeln von Fragen nach der Erfahrbarkeit Gottes in der Welt⁴⁸ oder das Begründen eigener Standpunkte zu religiösen und ethischen Fragen⁴⁹ markieren dabei das in diesem Zusammenhang konstitutive Wechselspiel zwischen Kreativität und Positionalität, das es nun näher in den Blick zu nehmen gilt.

3. Soziale Robotik und religionsunterrichtliche Kommunikation: Ein unüberwindbarer Hiat?

3.1 Kreativität und Positionalität

Dass KI-basierte Soziale Roboter mit Menschen in wechselseitige kommunikative Verständigungsprozesse treten können, zeigen, wie auch oben angedeutet, erprobte Einsatzmöglichkeiten auf unterschiedlichen Ebenen auf.⁵⁰

Aus der obigen Konturierung religiöser Kommunikation im Kontext religionsunterrichtlicher Praxis stellen sich allerdings spezifische Anforderungen an eine derartig auf KI und Robotik im Design eines sozialen Roboters basierende Lehrkraft vor dem Hintergrund jener formulierten Ansprüche an echte Unterrichtsgespräche im Religionsunterricht.

47 Zweifelsohne bleibt dieser Annäherung ein (idealisierendes) Konstruktionsmoment inhärent. Wie die Kasseler Unterrichtsgesprächsstudie zeigt, bleibt das Moment der Verwobenheit unterschiedlicher Gesprächstypen mit je unterschiedlichen Akzenten der Lehrer:innen- und Schüler:innenzentrierung konstitutiv.

48 Vgl. *Ministerium für Schule und Bildung*: Kernlehrplan, 18.

49 Vgl. Ebd., 19.

50 Vgl. Anm. 25.

Hierbei wird Kreativität im Anschluss an die obigen pädagogischen wie religionsunterrichtlichen Ausführungen als die Fähigkeit verstanden, im Gespräch Perspektivität derart *kreativ* zu erschließen, dass eröffnete Perspektiven neu erschlossen, letztlich (neu) kreiert werden können. *Kreativ* wäre etwa jene Rede zu nennen, „die sich dadurch auszeichnet, daß sie Neues hervorbringt [...], die sich selbst ins Wort fällt und in der zur Sprache kommt, was nicht schon vor und in der Sprache gegeben ist.“⁵¹ Kreativität, insofern in Erfahrung wurzelnd, als Gegebenes nicht bloß „registriert, sondern das Erfahrungsfeld organisiert“⁵² wird, setzt in der Konsequenz bei dem an, „was uns widerfährt, überrascht, erschreckt und verletzt und was nur in unseren Antworten seine Wirkung entfaltet.“⁵³ Schnell/Nehlsen differenzieren auf der Grundlage eines Gespräches zwischen Mensch und KI (auf der Basis des Programms GPT-3) zwischen Routinedialogen mit einer KI, bei denen es im Sinne eines Austausches immer bei einem „Geben und Nehmen von im Wesentlichen fertigen Gegenständen und/oder Sinneinheiten“⁵⁴ bleibt, und schöpferischen Dialogen sowie deren Entstehung und Aushandlung neuer Bedeutungen im Angesicht Anderer.⁵⁵ Inwiefern sind nun soziale Roboter, die beispielsweise auf ein Programm wie GPT-3 zurückgreifen, kreativ in der obigen Pointierung vor dem Hintergrund schöpferischer Dialogizität im überraschenden, transzendierenden Widerfahrnis und der erfahrungsbasierten Neuordnung? Eine positionssensible Annäherung an diese Frage kann weiterführende Impulse geben.

Positionalität⁵⁶ als Betonung des „existentiellen, jedoch nicht weniger im Werden seienden Zustandscharakter[s] einer Person“⁵⁷ wird dabei zunächst primär mit der Religionslehrkraft verschränkt. Als Positionalität der Religionslehrkraft suggeriert der Terminus eine Teilnehmendenperspektive, denn

51 *Waldenfels*: Globalität, 179f.

52 *Waldenfels*: Responsive Kreativität, 14.

53 *Waldenfels*: Globalität, 183.

54 *Schnell/Nehlsen*: Gespräch, 116.

55 Vgl. ebd.

56 Vgl. auch den Beitrag von Steffi Fabricius in diesem Band, dem ein anderer Schwerpunkt zugrunde liegt sowie Anm. 66 und grundlegend *Lorenzen*: Entscheidung.

57 *Fabricius*: Positionalität, 1.

als Lehrperson des klassischen bekenntnisorientierten Religionsunterrichts ist sie Teil der theologischen, konfessionellen und religiösen Innenperspektive, im Gegensatz zur eher religionswissenschaftlichen und religionskundlichen Außenperspektive. Damit hat und vertritt die Religionslehrperson einen – wenn auch subjektiv geprägten – Standpunkt, den sie, da existentiell darin verwickelt, so eigentlich nicht verlassen kann, im Gegensatz zur standpunktlosen und damit positionlosen Beobachterperspektive [...].⁵⁸

Im Blick auf das kreative Moment im Umgang mit Perspektivität ließe sich Positionalität auch auf alle Beteiligten insofern ausweiten, als zwar Perspektivenwechsel im bekenntnisorientierten Religionsunterricht als Bedingung der Möglichkeit einer ersten Person Singular (Teilnehmendenperspektive; Innenperspektive, religiöses Sprechen usw.) grundiert werden kann, es im Rahmen religionsunterrichtlicher Kommunikation aber immer angesichts des Sich-in-ein-Verhältnis-Setzens einer gewissermaßen positionssensiblen, kritisch-reflexiven Distanz bedarf.

3.2 Ein Sozialer Roboter als Übersetzer:in?

3.2.1 *Übersetzungstheoretische Anschärfung*

Um diese Implikationen nun ansatzweise im Horizont Sozialer Robotik reflektieren zu können, bietet die Perspektive, die am Religionsunterricht und hier konkret an Unterrichtsgesprächen Beteiligten als Übersetzer:innen bzw. religionsunterrichtliche Gespräche unter den oben skizzierten Ansprüchen und empirischen Analysen als Übersetzungsprozesse zu verstehen, weiterführende Impulse.

Ein auf Deep-Learning-Algorithmen basierendes Sprachmodell wie GPT-3 von OpenAI leistet (überzeugende?) Übersetzungsarbeit. Mehr noch: Ein derartiges Tool kann, unterrichtspraktisch gesprochen, „einen menschlichen Dialog simulieren [...] und so die Entwicklung von Fragen und Aufforderungen fördern, die die [Schüler:innen] zu kritischem Denken und einer tieferen Analyse, Interpretation und Bewertung der angebotenen Informationen anre-

58 Ebd., 3.

gen können.⁵⁹ Diesbezügliche Entwicklungsdynamiken sind zudem noch längst nicht abgeschlossen.

Die zentrale Zuspitzung liegt nun aber in einer interdisziplinären Lesart des Übersetzungsbegriffs, die Übersetzung als dynamischen, transformatorischen Prozess⁶⁰ zwischen realen sozialen Austauschbeziehungen in Kommunikationszusammenhängen⁶¹ profiliert und sich so von jenen Übersetzungsvollzügen eines Sprachmodells abhebt: Als „vielfältiger, wechselseitiger Aushandlungsprozess, in dem Texte, Kulturen, Religionen, Akteure wirken“⁶², ermöglicht der Übersetzungsbegriff insofern tiefere Einblicke in die Dimensionen von Kreativität und Positionalität, als er das Moment der „Unterbrechung des Gegebenen und der Kontinuitäten“⁶³ offenlegt.

Dies kann schließlich zum entscheidenden Prüfstein für die eingangs eröffnete Leitperspektive dieses Beitrags avancieren: Gelingen einem sozialen Roboter diese tieferen Einblicke?

3.2.2 Unübersetzbarkeit, Mehrsprachigkeit und Deutungsmacht

Angesichts der oben eingespielten Annäherung an Kreativität markiert B. Waldenfels das Atechnische als konstitutives Differenzkriterium zwischen interpersonaler und automatischer – etwa am Beispiel jener auf Deep Learning basierender Sprachmodelle bzw. Textgeneratoren – Übersetzung.⁶⁴ Hierbei negiere das Atechnische allerdings nicht schlicht das Technische, sondern skizziere Leerstellen, „für die es keine technischen Füllsel gibt.“⁶⁵

Unübersetzbarkeit „als Ausdruck von Fremdartigem, das sich inmitten der Übersetzung als Überschuß an Sinn bemerkbar macht und sich einer endgültigen Wiedergabe entzieht,“⁶⁶ wäre eine solche Leerstelle, bei der die Grenzen einer auf Datensätzen trainierten KI deutlich würden. Auf Unterrichtsgespräche transferiert: Der dyna-

59 Chrostowski: Anwendung, 83.

60 Vgl. Gruber: Theologie, 117–119.

61 Vgl. Sato: Translanguaging, 13.

62 Hock: Hegemonialität, 183.

63 Grümm: Praxeologie, 195.

64 Vgl. Waldenfels: Globalität, 183.

65 Ebd.

66 Ebd., 182.

mische und prozessuale Charakter unterrichtlicher Kommunikation, die zwar gewissermaßen standardisierten wie institutionalisierten Rahmenbedingungen unterliegt⁶⁷, sich gleichermaßen allerdings erst durch den Überschuss an Sinn konstituiert, geht einher mit dem Unplanbaren, dem nicht Algorithmisierbaren, ja letztlich mit dem Atechnischen im obigen Sinne. Übersetzungstheoretisch angeschärft wird Kreativität zum Akt des Aufbrechens – einer elementaren Prämisse religiöser Kommunikation in der obigen Profilierung als Bedingung der Möglichkeit jener transzendierenden Potentiale von Perspektivität, auch innerhalb des Religionsunterrichtes.

Nun wird man sicherlich konstatieren können, dass sich solche Reflexionen keineswegs exklusiv auf den Religionsunterricht beziehen und Unterrichtsgespräche, die Kreativität voraussetzen, kein Alleinstellungsmerkmal desselben darstellen können. Vielmehr markiert die bereits eingespielte Dimension von Positionalität ein Konstitutivum religionsunterrichtlicher Kommunikation, konkret in Form von Unterrichtsgesprächen, im Umgang mit jenen transzendierenden Potentialen von Perspektivität. Konfessioneller Religionsunterricht beansprucht nicht zuletzt angesichts seiner Verwiesenheit auf Art. 7 Abs. 3 GG und „[s]einer bleibenden Verankerung im gelehrten und gelebten Glauben sowie in [s]einer (An-)Bindung an die Glaubensgemeinschaft“ Positionalität in einer im Vergleich mit anderen Schulfächern gesonderten Qualität.⁶⁸ Für den in diesem Beitrag fokussierten Akzent religionsunterrichtlicher Kommunikation impliziert dies (erneut) eine doppelte Kompetenz-Trias, die sich an die Lehrkraft stellt:

Diese sind also zum einen verantwortlich dafür, positionell-konfessionelle Inhalte bzw. Lehrmeinungen in der Unterrichtsvorbereitung aufzuspüren und einzuspielen. Und zum anderen und zugleich sind sie übergeordnet Anwalt der Ergiebigkeit und Offenheit der sich daran entzündenden Auseinandersetzungen (geschicktes Moderieren). Während

67 Vgl. Anm. 19.

68 Exemplarisch ist hier der interdisziplinäre Diskurs um didaktische Implikationen des Beutelsbacher Konsenses zu nennen. So wird Positionalität auch für den Politikunterricht als elementare Prämisse profiliert, allerdings vor dem Hintergrund der jeweiligen fachspezifischen wie rechtlichen Charakteristika, vgl. *Hiller/Münch-Wirtz*: ‚Neutral‘ unterrichten?, 139–141. Philosophiedidaktische Impulse finden sich bei *Herbst*: Beutelsbacher Konsens, 321–338.

in einer Talkshow diese Rollen jedoch strikt getrennt sind, erfüllt die Lehrkraft beides in Personalunion.⁶⁹

Hiller/Münch-Wirtz weisen diesbezüglich neben der Notwendigkeit einer religionspädagogisch-didaktischen sowie einer fachwissenschaftlich-theologischen auf eine unerlässliche personale Kompetenz in der Ermöglichung fruchtbarer religionsunterrichtlicher Interaktion hin.⁷⁰

Der Akt des Aufbrechens ist folglich mit einem sich in konkreter Positionalität realisierenden Reflexionsvollzug verschränkt. Perspektiven kritisch-reflexiv aufzubrechen, neu zu verorten, zu transformieren usw., verdeutlicht die genuin religionsunterrichtliche (Kommunikations-)Praxis, in der die christlich tradierte Perspektivität „überindividuell kommunikel“⁷¹ wird – bei gleichzeitiger Klärungsbedürftigkeit der diese konstituierenden Positionen.⁷²

Übersetzungstheoretisch wäre dies so zu akzentuieren, dass dies nicht nur auf der Ebene der Beteiligten (und hier vor allem auf der Ebene der Religionslehrkräfte), sondern auch auf struktureller Ebene – in unserem Fall am Beispiel kommunikativer Prozesse – in den Blick genommen wird:

Verstanden als [Übersetzer:innen] wäre dies mit der religionsunterrichtlichen Perspektive verschränkt, [Schüler:innen] Zwischenräume auf einer Meta-Ebene zu eröffnen, die die kritische Reflexion (nicht nur) christlicher Glaubensvollzüge ermöglichen und zur ansatzweisen Erkundung bzw. zu einem vertieften Verständnis jener Glaubensoption, welches eine rein kognitive Dimension übersteigt und für eine angemessene Glaubensreflexion notwendig ist, anregt.⁷³

Eine auf Soziale Robotik recurrierende Religionslehrkraft würde hier (noch) auf mehreren Ebenen an ihre Grenzen stoßen, zugleich aber auch weitreichende Potentiale eröffnen:

- 1) *Unübersetzbarkeit*: Übersetzung wird im Blick auf die obige Annäherung entscheidend als gewissermaßen destruktive Transformation alteritätstheoretisch angeschärft. Unübersetzbarkeit mar-

69 Merkel: Positionalität, 93.

70 Vgl. Hiller/Münch-Wirtz: ‚Neutral‘ unterrichten?, 138.

71 Dressler: Religionspädagogik, 89.

72 Vgl. Schweitzer: Das Theologische, 12f.

73 de Byl: Religionslehrer*innen, 43.

kiert dabei jene destruktive Kraft, die Übersetzungsvollzüge unterbricht, Irritation wie Provokation sowie letztlich Erfahrungen völliger Fremdheit freisetzt.⁷⁴ Ein selbstreflexives Innehalten im Moment der Unterbrechung erweist sich als ein wesentliches Konstitutivum kreativer Neuperspektivierung jener Optionalität christlicher Glaubensperspektive – auf Seiten aller an religionsunterrichtlichen Gesprächen Beteiligten. In der Rolle als Religionslehrkraft stoßen Soziale Roboter, wie im Beitrag aufgezeigt, im Kontext von HRI hier an Grenzen.

- 2) *Mehrsprachigkeit*: Jene übersetzungstheoretische Lesart markiert die Relevanz von Mehrsprachigkeit als Eigenschaft von Sprecher:innen, „die, z. B. durch Migration bedingt, aber auch durch Alltagskontakte zu Sprecher:innen verschiedener Sprachen oder in pädagogischen Settings vermittelt in ihren Lebenswelten zwei oder mehr Sprachen verwenden.“⁷⁵ Soziale Roboter könnten hier vor allem in assistierender Weise (nicht nur) Unterrichtsgespräche bereichern, indem sie dabei helfen, mögliche Sprachbarrieren sowohl aufdecken als auch durchbrechen zu können. Anders formuliert: Als Lehrassistenz kann Soziale Robotik im Religionsunterricht an jenen Stellen helfen, an denen Kreativität durch Sprachbarrieren eingeschränkt wird.
- 3) *Deutungsmacht*: Auf struktureller Ebene wäre schließlich zu ergänzen, dass jene Versprachlichungsprozesse innerhalb (religions-)unterrichtlicher Kommunikation stets in Deutungsmacht verstrickt sind.⁷⁶ In übersetzungstheoretisch grundierten Zwischenräumen ließe sich hegemoniale Strukturiertheit dekodieren und Deutungsmacht positionssensibel kritisch reflektieren. Dies erscheint als ein ebenfalls virulentes Feld der erfolgten interdisziplinären Erkundungen.⁷⁷ Ein Blick auf die Einschreibung von intersektional zu denkender Ungleichheit und Diskriminierung innerhalb von Datensätzen, an denen lernende Systeme trainiert werden, lässt die Frage nach Deutungsmacht und -hoheit und nach diese perpetuierenden Mechanismen akut werden.⁷⁸ Die religionspädagogische Verortung von Übersetzung kann hier als Impuls

74 Vgl. *Grümme*: Übersetzung, 38–42.

75 *Zettl*: Mehrsprachigkeit, 39.

76 Vgl. *Kumlehn*: Babel, 34f.

77 Vgl. *Chrostowski*: Anwendung, 86–89.

78 Vgl. *Horvath*: Algorithmen, 73–83.

verstanden werden, einen Beitrag zu breiten interdisziplinären Kooperationen zu leisten, „um soziotechnische Systeme kontextualisieren [...] und gerechtere Systeme entwickeln zu können.“⁷⁹

4. Ausblick

Zweifelsohne kann der vorliegende Beitrag der Dynamik und Komplexität der jeweiligen Forschungsbereiche – erst recht dem Versuch einer fundierteren interdisziplinären Erkundung – nicht gerecht werden. Vielmehr verstehen sich die obigen Gedanken als ein Impuls für weiterführende Relationierungen und Erkundungen auf dem interdisziplinären Terrain Sozialer Robotik, KI und Religionsunterricht.

Der Beitrag konnte die Potentiale von Sozialer Robotik und KI in religionspädagogischer wie -didaktischer Verschränkung im Rahmen des eingangs eröffneten Perspektive und ihres Erkundungsrahmens in Grundzügen aufzeigen. Für den Kontext religiöser Kommunikation aus religionsunterrichtlicher Perspektive hat sich hierbei erwiesen, dass diesbezüglichen Entwicklungsdynamiken keineswegs mit ausschließlich dystopischen Vorbehalten entgegenzutreten ist. Vielmehr können die Fortschritte Sozialer Robotik insbesondere in pädagogischer Kontextualisierung für die Religionspädagogik und den Religionsunterricht fruchtbar gemacht werden. Die soeben skizzierten Handlungsfelder im vorangegangenen Kapitel beantworten die Frage nach dem Einsatz einer ebensolchen Lehrkraft mit dem Entwurf einer gewissermaßen auf Soziale Robotik und KI zurückgreifenden Lehrsistenz, die allerdings (noch) weit entfernt von einer zu Beginn eingespielten Stellvertretung der Religionslehrkraft sind.

Entsprechende Möglichkeiten im Bereich diagnostischer Sprachanalysen könnten allerdings dazu beitragen, Sprachbarrieren abzubauen. Gleichzeitig bedarf es der machtkritischen Dekonstruktion beispielsweise Diskriminierung perpetuierender Dynamiken. Kreativität und Positionalität in der jeweiligen obigen Profilierung erscheinen – nicht zuletzt in übersetzungstheoretischer Hinsicht – als Herausforderungen für ein diesbezügliches Zusammenspiel – sowohl in Bezug auf die am Religionsunterricht Beteiligten als auch auf die

79 Ebd., 96.

diesbezüglichen strukturellen Prozesse, die hier nur ansatzweise dargestellt werden konnten.

Vertiefende Blicke im Rahmen inter- und multidisziplinärer Erkundungen, beispielsweise auf jene Sprachbarrieren und den möglichen diesbezüglichen Einsatz innerhalb des Religionsunterrichtes, erscheinen gleichermaßen notwendig wie lohnenswert.

Bibliografie

- Alemi, Minoo/Meghdari, Ali/Ghazisaedy, Maryam*: Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools, in: *International Journal of Humanoid Robotics*, 11 (3), 2014, 1–25. DOI: 10.1142/S0219843614500224.
- Altmeyer, Stefan*: Sprachsensibler Religionsunterricht. Grundlagen und konzeptionelle Klärungen, in: *Sprachsensibler Religionsunterricht (Jahrbuch der Religionspädagogik 37)*, Göttingen 2021, 14–29.
- Balle, Simon*: Theological Dimensions of Humanlike Robots. A Roadmap for Theological Inquiry, in: *Theology and Science* 21 (1), 132–156. DOI: 10.1080/14746700.2022.2155916.
- Banaeian, Hossein/Gilanlioglu, Ilkay*: Influence of the NAO Robot as a Teaching Assistant on University Students' Vocabulary Learning and Attitudes, in: *Australasian Journal of Educational Technology*, 2021, 37 (3), 71–87. DOI: 10.14742/ajet.6130.
- Bartneck, Christoph/Belpaeme, Tony/Eyssele, Friederike u. a.*: *Human-Robot Interaction. An Introduction*, Cambridge 2020. DOI: 10.1017/9781108676649.
- Bartneck, Christoph/Forlizzi, Jodi*: A Design-Centred Framework for Social Human-Robot Interaction, in: *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication*, Kurashiki 2004, 591–594. DOI: 10.1109/ROMAN.2004.1374827.
- Bächle, Thomas C./Regier, Peter/Bennowitz, Maren*: Sensor und Sinnlichkeit. Humanoide Roboter als selbstlernende soziale Interfaces und die Obsoleszenz des Impliziten, in: *Navigationen-Zeitschrift für Medien- und Kulturwissenschaften* 17 (2), 2017, 67–86.
- Belpaeme, Tony/Tanaka, Fumihide*: Social robots as educators, in: *OECD Digital Education Outlook 2021. Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, Paris 2021, 143–158. DOI: 10.1787/589b283f-en.
- Belpaeme, Tony/Kennedy, James/Ramachandran, Aditi u. a.*: Social robots for education: A review, in: *Science robotics* 3 (21), 2018, 1–9. DOI: 10.1126/scirobotics.aat5954.

- Bendel, Oliver*: Soziale Roboter in der Moral. Ethische Betrachtungen am Beispiel des Unterrichts, in: Bendel, O. (Hg.): Soziale Roboter. Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen, Wiesbaden 2021, 149–167. DOI: 10.1007/978-3-658-31114-8.
- Bertram, Daniel/Fröhle, Dominic/Reese-Schnitker, Annegret*: Zum Potential von Unterrichtsgesprächen. Zentrale Erkenntnisse der Kasseler Studie, in: Reese-Schnitker, A./Bertram, D./Fröhle, D. (Hg.): Gespräche im Religionsunterricht. Einblicke – Einsichten – Potenziale (Religionspädagogik innovativ 45), Stuttgart 2022, 503–528.
- Breul, Martin*: Die Versprachlichung des Religiösen, in: IRP-Impulse, 2018, 4–9 DOI: 10.15496/publikation-61756.
- Chrostowski, Mariusz*: Anwendung von Künstlicher Intelligenz im Religionsunterricht. Möglichkeiten, Grenzen und Brennpunkte, in: Religionspädagogische Beiträge 46 (1), 2023, 79–95. DOI: 10.20377/rpb-267.
- Davison, Daniel P./Wijnen, Frances M./van der Meij, Jan u.a.*: Designing a Social Robot to Support Children’s Inquiry Learning. A Contextual Analysis of Children Working Together at School, in: International journal of social robotics 12, 2020, 883–907. DOI: 10.1007/s12369-019-00555-6.
- de Byl, Marius*: Religionslehrer*innen als Übersetzer*innen? Impulse für eine kontextsensible Thematisierung von Humanität im Religionsunterricht, in: Anusiewicz-Baer, S./Hild, C./Massud, A.-H. (Hg.): Humanität als religionspädagogisches und -didaktisches Leitmotiv (Religion und Kommunikation in Bildung und Gesellschaft 3), Stuttgart 2023, 31–47.
- Deurer, Rebecca G.*: Art. Unterrichtsgespräch, in: Wissenschaftlich Religionspädagogisches Lexikon im Internet (www.wirelex.de), 2023, 1–15. DOI: 10.23768/wirelex.Unterrichtsgesprch.201015.
- Dressler, Bernhard*: Warum und auf welche Weise Religionspädagogik als Theologie zu betreiben ist. Überlegungen zur Theologizität der Religionspädagogik, in: Schlag, T./Suhner, J. (Hg.): Theologie als Herausforderung religiöser Bildung. Bildungstheoretische Orientierungen zur Theologizität der Religionspädagogik (Religionspädagogik innovativ 17), Stuttgart 2017, 85–100.
- Fabricius, Steffi*: Art. Positionalität, Lehrende, in: Wissenschaftlich Religionspädagogisches Lexikon im Internet (www.wirelex.de), 2022, 1–15. DOI: 10.23768/wirelex.Positionalitt_Lehrende.201013.
- Faßler, Manfred*: Was ist Kommunikation?, München ²2003.
- Franke, Gabriel*: Künstliche Intelligenz, Transhumanismus und personale Identität, Darmstadt 2022.

- Gärtner, Claudia*: Digitales ‚Ich‘? Trans- und Posthumanismus als theologische und religionspädagogische Herausforderung, in: Konz, B./Ostmeyer, K.-H./Scholz, M. (Hg.): Gratwanderung Künstliche Intelligenz. Interdisziplinäre Perspektiven auf das Verhältnis von Mensch und KI, Stuttgart 2023, 111–123.
- Grethlein, Christian*: Praktische Theologie, Berlin/Boston ²2016.
- Gruber, Judith*: Theologie nach dem Cultural Turn. Interkulturalität als theologische Ressource (ReligionsKulturen 12), Stuttgart 2013.
- Grümme, Bernhard*: KI, ein Ernstfall der Religionspädagogik, in: Religion unterrichten. Künstliche Intelligenz und Human Enhancement 3 (2), 2022, 38–43. DOI: 10.13109/reun.2022.3.2.
- Grümme, Bernhard*: Praxeologie. Eine religionspädagogische Selbstaufklärung, Freiburg im Breisgau 2021.
- Grümme, Bernhard*: Übersetzung „inmitten von Ruinen“ (Judith Butler). Nachdenkliche Anmerkungen zu einem religionspädagogischen Postulat, in: Hausmann, W. u. a. (Hg.): EinFach übersetzen. Theologie und Religionspädagogik in der Öffentlichkeit und für die Öffentlichkeit (Religionspädagogik innovativ 33), Stuttgart 2019, 35–44.
- Herbst, Jan-Hendrik*: Braucht religiöse Bildung einen Beutelsbacher Konsens? Philosophiedidaktische Impulse für die religionspädagogische Debatte, in: Theo-Web 20 (2), 2021, 321–338. DOI: 10.23770/tw0226.
- Hiller, Simone/Münch-Wirtz, Julia*: ‚Neutral‘ unterrichten? Eine Lektüre des Beutelsbacher Konsenses hinsichtlich der Positionalität von Politik- und Religionslehrpersonen, in: Österreichisches Religionspädagogisches Forum 29 (1), 2021, 124–141. DOI: 10.25364/10.29:2021.1.8.
- Hock, Klaus*: Hegemonialität, Vernakularität, Transkulturation. Zur Historisierung der Übersetzung von Religionen, in: Lange, M./Rösel, M. (Hg.): Der übersetzte Gott, Leipzig 2015, 165–186.
- Horvath, Ilona*: Algorithmen, KI und soziale Diskriminierung, in: Schnegg, K./Tschuggnall, J./Voithofer, C. u. a. (Hg.): Inter- und multidisziplinäre Perspektiven der Geschlechterforschung (Innsbrucker Gender Lectures IV), Innsbruck 2022, 71–101.
- Konz, Britta/Scholz, Marcel*: Körper und Künstliche Intelligenz. (Un-) Verfügbare Beziehungen?, in: Konz, B./Ostmeyer, K.-H./Scholz, M. (Hg.): Gratwanderung Künstliche Intelligenz, Stuttgart 2022, 125–141.
- Krauss, Patrick*: Künstliche Intelligenz und Hirnforschung. Neuronale Netze, Deep Learning und die Zukunft der Kognition, Berlin 2023. DOI: 10.1007/978-3-662-67179-5.
- Kumlehn, Martina*: Zwischen Babel und Pflingsten. Übersetzen zwischen Sprachwelten als Kernaufgabe sprachsensibler Theologie, in: Jahrbuch der Religionspädagogik 37, 2021, 30–41.
- Lämmel, Uwe/Cleve, Jürgen*: Künstliche Intelligenz. Wissensverarbeitung – neuronale Netzwerke, München ⁵2020.

- Lindner, Felix*: Soziale Robotik und KI. Wie sich Roboter autonom nach Werten und Normen verhalten können, in: Bendel, O. (Hg.): Soziale Roboter. Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen, Wiesbaden 2021, 109–123. DOI: 10.1007/978-3-658-31114-8.
- 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, 2021, 569–586. DOI: 10.1007/s12369-019-00558-3.
- Lorenzen, Stefanie*: Entscheidung als Zielhorizont des Religionsunterrichts? Religiöse Positionierungsprozesse aus der Perspektive junger Erwachsener (Praktische Theologie heute 174), Stuttgart 2020.
- Luthe, Swantje/Nord, Ilona/Löffler, Diana u. a.*: Segensroboter „BlessU-2“. Forschungsimpulse für die Praktische Theologie angesichts der Entwicklung sozialer Roboter, in: Pastoraltheologie 108 (3), 2019, 107–123. DOI: 10.13109/path.2019.108.3.107.
- Merkel, Rainer*: Wie lernt man, transparent zu unterrichten? Reflexionen notwendiger Trierungsanforderungen in der Ausbildung von Positionalität durch angehende Religionslehrerinnen und Religionslehrer, in: Theo-Web 21 (2), 2022, 90–100. DOI: 10.23770/tw0250.
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen* (Hg.): Kernlehrplan für die Sekundarstufe I. Gymnasium in Nordrhein-Westfalen, Katholische Religionslehre, Düsseldorf 2019.
- Müller, Oliver*: Maschinelle Alterität. Philosophische Perspektiven auf Begegnungen mit künstlicher Intelligenz, in: Schnell, M. W./Nehlsen, L. (Hg.): Begegnung mit Künstlicher Intelligenz. Intersubjektivität, Technik, Lebenswelt, Weilerswist 2022, 23–47. DOI: 10.5771/9783748934493.
- Pirker, Viera*: Digitalität wirkt Wandel. Anthropologische, theologische und ethische Aspekte, in: Die digitale Revolution und ihre Kinder, 2019, 77–94.
- Puzio, Anna*: Theologie und Künstliche Intelligenz. Perspektiven, Aufgaben und Thesen einer Theologie der Technologisierung, in: Puzio, A./Kunkel, N./Klinge, H. (Hg.): Alexa, wie hast du’s mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz (Theologie und Künstliche Intelligenz 1), Darmstadt 2023, 13–27. DOI: 10.53186/1030373.
- Reese-Schnitker, Annegret*: Gespräche im Religionsunterricht, in: Reese-Schnitker, A./Bertram, D./Fröhle, D. (Hg.): Gespräche im Religionsunterricht: Einblicke – Einsichten – Potenziale (Religionspädagogik innovativ 45), Stuttgart 2022, 17–41.
- Reese-Schnitker, Annegret*: Das methodische Vorgehen in der Unterrichtsstudie, in: Reese-Schnitker, A./Bertram, D./Fröhle, D. (Hg.): Gespräche im Religionsunterricht: Einblicke – Einsichten – Potenziale (Religionspädagogik innovativ 45), Stuttgart 2022, 191–215.

- Sato, Eriko*: Translanguaging in Translation. Invisible Contributions that Shape Our Language and Society (Translanguaging in Theory and Practice 3), Bristol 2022. DOI: 10.21832/SATO4938.
- Schnell, Martin W./Nehlsen, Lukas*: Gespräch mit einer Künstlichen Intelligenz. Eine qualitative Inhaltsanalyse, in: Schnell, M. W./Nehlsen, L. (Hg.): Begegnungen mit Künstlicher Intelligenz. Intersubjektivität, Technik, Lebenswelt, Weilerswist 2022, 104–118. DOI: 10.5771/9783748934493.
- Schweitzer, Friedrich*: Das Theologische der Religionspädagogik. Grundfragen und Herausforderungen, in: Schlag, T./Suhner, J. (Hg.): Theologie als Herausforderung religiöser Bildung. Bildungstheoretische Orientierungen zur Theologizität der Religionspädagogik (Religionspädagogik innovativ 17), Stuttgart 2017, 9–20.
- Seufert, Sabine/Guggemos, Josef/Sonderegger, Stefan*: Soziale Roboter im Bildungsbereich. Konzeptioneller Überblick zum Einsatz sozialer Roboter in der Bildung, in: Bendel, O. (Hg.): Soziale Roboter. Technikwissenschaftliche, wirtschaftswissenschaftliche, philosophische, psychologische und soziologische Grundlagen, Wiesbaden 2021, 475–494. DOI: 10.1007/978-3-658-31114-8.
- So, Simon/Lee, Naomi*: Pedagogical exploration and technological development of a humanoid robotic system for teaching to and learning in young children, in: Cogent Education 10 (1), 2023, 1–18. DOI: 10.1080/2331186X.2023.2179181.
- Waldenfels, Bernhard*: Globalität, Lokalität, Digitalität. Herausforderungen der Phänomenologie, Berlin 2022.
- Waldenfels, Bernhard*: Responsive Kreativität angesichts des Fremden, in: Zagreber germanistische Beiträge 31, 2022, 13–28. DOI: 10.17234/ZGB.31.2.
- Xia, Yu/LeTendre, Gerald*: Robots for future classrooms. A cross-cultural validation study of „negative attitudes toward robots scale“ in the US context, in: International Journal of Social Robotics 13, 2021, 703–714. DOI: 10.1007/s12369-020-00669-2.
- Young, William*: Virtual Pastor. Virtualization, AI, and Pastoral Care, in: Theology and Science 20 (1), 2022, 6–22. DOI: 10.1080/14746700.2021.2012915.
- Zehner, Fabian*: Künstliche Intelligenz. Ihr Potenzial und der Mythos des Lehrkraft-Bots, 2019. Online unter: https://www.pedocs.de/volltexte/2020/17561/pdf/Zehner_2019_Kuenstliche_Intelligenz_A.pdf [Stand 11.10.2023].
- Zettl, Evamaria*: Mehrsprachigkeit und Literalität in der Kindertagesstätte. Frühe sprachliche Bildung in einem von Migration geprägten Stadtviertel (Inklusion und Bildung in Migrationsgesellschaften), Wiesbaden 2019. DOI: 10.1007/978-3-658-27031-5.

Limitations for Pastoral Robots

Abstract

This paper addresses a pressing theological question that has arisen as a result of recent developments in AI and experimentations with robots in the context of religious practice, namely whether robots could legitimately perform central pastoral tasks like preaching, counselling, or administering the sacrament. Traditions vary, but most denominations in the Christian tradition draw a line around tasks like these and reserve them for ordained clergy. If we want to consider whether robots could preach in the current ecclesiastical landscape, we need to ask whether AI can sensibly take clerical vows, or if it can serve such functions without being ordained. If this proposition seems odd, it might give us pause to consider why church institutions have ordination vows in the first place.

1. Introduction

If a robot demands religious rights, if it asks to be accepted as a proselyte, is it conceivable that its request could be honored? If this problem of the religious status of a robot seems far-fetched today, it may seem much less so a few decades from now.¹

In recent years, a variety of robotic artefacts have been designed for religious practice. The robot Mindar recites sutras, BlessU-2 pronounces blessings, while SanTO and Xian'er will answer users' queries in an interactive if yet limited way.² While these robots are eye-catching, thought-provoking and suggestive of future developments, they still serve bounded functions and deliver only prede-

1 *Rosenfeld*: Religion, 15.

2 See, for example, *Cheong*: Religion and *Balle/Ess*: Robots.

finer content. That is not to say the performance of these robots cannot be compelling, meaningful or inventive, even if the prayers, teachings, blessings or spiritual advice they deliver are scripted. After all, large portions of church liturgy, which human pastors perform, are fixed. Yet this does not make those parts any less effective or meaningful; in some ways, it is indeed because the words and movements of rituals are shaped and defined by tradition that they are meaningful.

So, why *not* design robots to administer religious practice? Does it perhaps take human conscience or belief for rituals to be authentic or effective? Or are there some deeper theological obstacles or reasons why robots should not perform central religious functions? These are some of the pressing questions that have been raised in and outside academic discussion as AI and robots are developed for religious practice.³ Negotiating such questions will naturally be different across religious traditions. In this short paper, however, I will limit myself to the Evangelical-Lutheran tradition, although I imagine much of the reasoning makes sense in a wider Christian context.

The few extant robots developed for faith practice in this context are *not*, to be fair, designed to be “full pastoral agents”, taking over from human pastors in the context of church liturgy. They are rather robots for personal devotion that make certain rituals, prayers and blessings accessible to, for example, the elderly or people confined to their homes.⁴ To my knowledge, no one is suggesting or developing pastoral robots to preach, baptise or administer the sacrament. But I think the time is right to speculate based on the current situation and imagine hypothetical robots that *could* serve these more central pastoral functions, and then ask whether they *should*, from a normative perspective.

Suppose, for example, that our hypothetical pastoral robot was imbued with an autoregressive large language model, like the one that powers OpenAI’s ChatGPT, Microsoft’s CoPilot or Google’s Gemini and PaLM projects, and then trained on a curated database of sacred scripture, sermons, theological treatises, blogposts, articles

3 See, for example, *Samuel: Robot Priests; Young: Reverend Robot; Khan/Aytes: Islam and Kopf: Does AI.*

4 See *Trovato et al.: Communicating.*

and other texts of relevance to a particular religious tradition. If large language models can compose original and meaningful texts in a wide variety of contexts, why could a robot fitted with such AI systems not perform a sermon? From a technological point of view, it is within the realm of possibility to develop robots that through pattern recognition and probabilistic forecasting would be conversant in spiritual matters. And given the state of empirical research on digital agents as therapists, coaches or Socratic partners, there is compelling evidence to suggest that robots might fare relatively well in such domains.⁵

Let us assume there are no substantial technological barriers to fitting robots with some kind of natural language generation. Our hypothetical pastoral robot could then perform original and dynamic sermons and offer spiritual advice in addition to reciting set prayers, giving blessings, etc. And provided there are also no insurmountable technological obstacles to robots handling a chalice and navigating within the physical and liturgical context of a church, we are close to imagining an autonomous robot stand-in for human clergy. We could even speculate that, with the rate generative AI is improving and with enough training data and investment, pastoral robots could perform better than humans in some areas.

This could, however, be problematic for a number of different reasons quite different from whether they can believe or not. Language models have, for example, been known to return misleading, harmful or simply false answers to users' queries. Language models are notoriously poor judges of good and bad data, of what is true or harmful, and this is why liability issues surrounding the behaviour and output of AI systems is a growing concern. In a parallel fashion, the problem with pastoral robots is not only whether they would teach heresy or give poor spiritual advice, but also that AI systems are not liable agents in any meaningful sense. Consequently, the deeper difficulty with generative AI in future pastoral robots is not just about potential malpractice but revolves more essentially around dogmatic or spiritual responsibility.

The church has a long and complicated history of safeguarding and delegating both authority and responsibility. To ensure there are

5 See, for example, *Darcy et al.: Evidence and Nørskov et al.: Employers' and Applicants' Fairness.*

capable and *responsible* priests to perform certain central activities in the life of the congregation, most denominations have installed a practice of ordination. To qualify for pastoral duty, an aspiring pastor who has completed the necessary training must take vows of ordination before s/he can hold office. Once ordained, the pastor is granted formal authority to preach and administer the sacraments. A fundamental question in the context of the use of AI in church is therefore whether robots could take clerical vows and thus acquire formal authority and responsibility, or whether they could legitimately perform central pastoral tasks without being ordained.

2. The discussion so far: Do robots need to understand in order to perform?

For the sake of argument, let us assume our hypothetical robot can recite sermons, prayers and give spiritual advice of a decent quality and in a consistent manner that, even if not indistinguishable from something a human pastor would say, is nonetheless meaningful and insightful to users. Those who have played with chatbots like ChatGPT or asked them to write a sermon on a certain topic might be convinced this is only a few generations away.

Does it matter, then, that pastoral robots running this software would string together words based on pattern analyses without intending or meaning anything by what they say; that their output is not grounded in an understanding of what religious creeds or practices actually *mean* to humans? That is, while robots may eventually be conversant in spiritual matters, they would still have no existential familiarity with notions such as love, despair, guilt, death, sin or hope—nor would they be able to freely adhere to spiritual truths on those topics.⁶ This is because religious belief comprises experiential, conative and affective components that we have no indications any AI system has.⁷ This might add to the unease some users already reportedly feel towards pastoral robots that the spiritual practice they afford somehow feels inauthentic.⁸

6 See *Balle*: Kan Præsteembedet.

7 See *Balle/Nissen*: Responsive Bodies.

8 See *Löffler et al.*: Blessing Robot.

One way to frame the question, then, is to ask whether robots could legitimately perform rituals and communicate spiritual content without understanding any of it? Does it matter whether they understand the meaning of the practice or the words they utter, or is output performance the better metric? These are some of the questions put forward on the issue by William Young.⁹

On the one hand, some argue that it matters very much whether robots could adopt beliefs or attain spiritual intelligence; that there is an existential threshold robots cannot cross.¹⁰ Perhaps it is necessary to be called into a relationship with God, to have the *imago Dei*, in order to occupy central roles at the heart of faith practice. Without standing in a relationship to God, how could one convey what that means to others?

On the other hand, one could argue that it does not matter what robots potentially believe. Interestingly, we have quite strong dogmatic reasons to endorse this view. Young proposes viewing the problem through the lens of the Donatist controversy.¹¹ Just as it has been considered orthodox in most denominations since this controversy that the efficacy of sacrament depends on the faith of whoever receives it, not on who administers it, we could similarly say that the lack of faith of a pastoral robot is irrelevant.

Along the same lines, Martin Luther comforted anxious new Protestant consciences by saying that their baptism was not invalid even if it had been performed by a non-Protestant priest because the efficacy of the baptism was not predicated on the spiritual or moral purity of the pastor. This is why, in the Lutheran context, the church upholds the distinction between a pastor's *character* and the *office* s/he holds. So, the blessings, absolutions, prayers and so on that a pastor performs derive their legitimacy from the office, and not from the pastor's personal spiritual dispositions.

It should be noted, however, that both these historical controversies revolved around the moral and dogmatic integrity of priests and bishops, not about their ability to believe in the first place. In view of this, perhaps a more tractable question than whether robots can believe is whether robots could be ordained to hold office. Why?

9 See *Young: Reverend Robot* and *Young: Virtual Pastor*.

10 See *Chaudhary: Delegating Religious Practices*.

11 See *Young: Reverend Robot*.

Because most denominational contexts draw a line around certain rituals and services in the life of the church that only ordained members can perform.

3. Holding office: baptism and ordination.

Let us therefore examine the formal requirements to be able to hold office in the church. In most denominational contexts, candidates who aspire to become pastors must complete a certain amount of training before they can take ordination vows and become ministers (similarly to doctors, who take the Hippocratic oath). This differs across denominational contexts, of course, but in most cases, something like preaching or administering the Eucharist is reserved for ordained clergy. Some contexts are more restrictive, some are more lenient, but the general idea that there are some tasks or roles in the church only a trained and ordained member can fulfil seems to apply broadly.

To be more specific, in the following, I will stray somewhat from common ground and limit myself to considering the context I know best, namely the Danish Evangelical-Lutheran Church (ELC), although I do think the argumentation is recognisable and mostly applicable across denominations.

In an Evangelical-Lutheran context, we can identify two necessary criteria for serving as a priest: (1) baptism and (2) ordination. Luther considered every baptised individual principally worthy of serving in any office in the church, an ideal captured in the notion of “general priesthood”. This means every baptised church member could, in principle, be a minister to others. But among the baptised, someone must be trained, called and ordained to a special priestly office to ensure propriety or decency in the administration of the sacraments and public teaching. In *Confessio Augustana*, this is formulated as follows: “They teach that no one should publicly teach in the church or administer the Sacraments unless he is rightly called” (CA, Article 14).

Before we move on to consider baptism and ordination in order, we need to address the glaringly obvious premise that the candidate we are referring to here is human. God became human to communicate his love for and to reconcile with humankind, and priestly

action is consequently about conveying God's love for humanity. And, one could argue, priestly action is furthermore predicated on a personal ability to respond to God *qua* being human. It is perfectly possible to end the conversation here by simply observing that robots are not humans. But for the sake of argument, I will move beyond this objection because there is much to gain from working through the issue in more detail.

3.1 Baptism

Obviously, the theology of baptism is very complex and varies across the denominational landscape, so we will apply a very minimal definition. I take the *purpose* of baptism as being to annul the consequences of sin and to make the baptismal candidate a member of the body of Christ. And a key requirement is that this person consents to belonging to God, or, in the case of infant baptism, a fiduciary consent. Could robots meet those conditions?

In the first instance, it would require that robots are sinful and therefore in need of baptism in the first place. To me, it makes little sense that robots could be sinful, since they are not free, autonomous individuals with intentions, desires or goals, either for good or for evil. They are heteronomous agents, who only have goals that we define. One could perhaps, very speculatively, speak of inherited sinfulness among robots, insofar as they are our creations and thus contain our fallibility. Our corruption is in the data they run on.

But what then about the prerequisite condition of baptism that the baptismal candidate consents to belonging to God? For this to make sense, the candidate must freely believe in God (or, in the case of infant baptism, at least potentially be able to do so). Of course, AI systems are imputed with facts about the world by programmers, and relations between facts might change with "experience", so they can update their "beliefs" about the state of the world in order to continually act successfully within it. For example, a robot emptying bins in an office environment during the day must have knowledge representations about the state of bins and movement of people to be successful—but are these "beliefs", let alone beliefs about existential matters, associated with being a mortal and fallible person? I'll leave

this question for another time and simply suggest that this is important to consider for baptism and, by extension, ordination to make sense, all the way up to preaching and administering the sacraments.

One way around this problem was suggested in a recent paper by Eugene Curry. Curry suggested that the medieval practice of “conditional baptism” could be reintroduced and adapted in a scenario in which a robot would request baptism, and we are uncertain about its ability to have beliefs. According to this practice, priests would baptise potentially insincere converts, with a clause dictating that the baptism was conditional on the candidate’s sincere and correct faith.¹² In the vein of this tradition, Curry proposes that if some future robot or AI system claims to have faith in God and a desire to belong to the community of believers, we could similarly baptise it on the condition that it is indeed able to have free and sincere beliefs about God.

As noted above, we currently have no indication that even the most sophisticated and powerful AI systems have the mental capacity to conscientiously hold religious beliefs. Besides, it is difficult to see why tech-companies should even want their products to develop in this self-aware direction with existential desires and needs.

But let us for the sake of argument say robots could be “conditionally baptised” and move on to consider ordination.

3.2 Ordination vows

To set the proper bar to discuss whether robots qualify for pastoral duty, I quote here the current vows of ordination from the Danish folk church, the Evangelical-Lutheran Church. I have italicised a few key passages for discussion below.

I, N.N., who have been lawfully appointed to (position), and who *in my conscience* affirm that I have not used any dishonourable means to obtain this office, *make this promise* before the all-knowing God:

Firstly, I *vow* to endeavour to preach the word of God purely and plainly, as it is found in the prophetic and apostolic writings and in the symbolic books of our Danish Evangelical-Lutheran Church [ELC], and to administer the holy sacraments after the institution of Christ,

12 See Curry: Artificial Intelligence.

with all due reverence and propriety, and to perform all other holy acts in accordance with the regulations applicable to the national church.

Furthermore, I *pledge* to oppose the misuse of the holy means of grace and combat doctrines that conflict with the creed of the national church, and to faithfully work towards the Christian education and guidance of youth to the best of my ability.

Lastly, I will *strive*, by diligent and serious study of God's word and the holy doctrines of faith, to constantly improve and qualify myself for this holy office, and endeavour, as is fitting for a servant of the word, to set a good example to the congregation and to carry out my duties in its various parts with all due obedience, adhering to the church's laws and regulations, and displaying such conduct towards my superiors and colleagues that no valid complaint can be lodged against me.

Quoting these vows at length helps us to be quite specific about the kind of properties robots need to acquire to qualify for pastoral duty. The issue is not whether a robot would be competent enough; as mentioned above, with the rapid advances made in generative AI, a pastoral robot could probably be trained to write better sermons than humans. Rather, the hurdle for robots as pastors is that candidates are required to make promises, to strive and to pledge on account of a conscience. But to have a conscience is precisely what AI systems lack, along with any ability to strive, pledge and promise on account of it.

For someone to pledge and strive on account of a conscience, they must be able to have intentions *and* have privileged access to and control over them. But robots and AI systems are not motivated for and do not *intend* anything. As mentioned above, intentions are complicated and compounded mental states. And even if we grant that AI systems qualify for some of the cognitive aspects, there is no good reason to think they could obtain the conative and affective components.¹³ Someone might object that some AI programs do update their own goals. But that is neither the same as intentions, nor do they do so outside the larger purpose we design them for. They remain heteronomous agents that only have goals we set.

For these reasons, it is difficult to envision a pastoral robot on the current technological trajectory that would qualify for ordination and thereby be able to hold office in the church.

13 See *Balle/Nissen: Responsive Bodies*.

4. Conclusion and perspectives

As AI systems are increasingly woven into the fabric of everyday life, we must ask at some point whether the religious domain of human life is a suitable arena for AI and robots. I have suggested that if we want to consider whether robots could give a sermon or administer the sacrament in the current ecclesiastical landscape, we need to ask whether AI systems can sensibly take clerical vows. The practice of ordination is installed across the Christian tradition for good historical and dogmatic reasons to ensure somebody can speak and act with authority and be held responsible for their teaching. I have argued that AI systems cannot meet the formal criteria for ordination since they do not have the necessary mental capacities. Consequently, they cannot be trusted as responsible and authoritative figures in the life of the church. If we follow this line of reasoning, robots could not legitimately perform central pastoral tasks traditionally reserved for ordained clergy, such as preaching, counselling and administering the sacrament.

In some ways, this discussion is a concretisation of the more abstract question of whether robots could or need to believe to perform pastoral tasks. But I think discussing the issue in these terms is more tractable because it helps us identify and discuss which functions and roles robots could legitimately perform and why, relative to the current ecclesiastical landscape.

It also invites us to consider whether robots could fulfil more bounded functions. For example, congregations in the ELC will sometimes have lay members of the church deliver the sermon. In such cases, the lay preacher exercises the teaching office under the authority of the presiding pastor(s). Perhaps robots could similarly exercise a bounded or limited form of spiritual agency under the authority of a responsible group of pastors and programmers, who understand how that AI system operates. One avenue for further research is therefore to work out which performances are governed by ordination within a given denomination and whether or under which conditions they could be performed by an AI system.

In any event, I think these speculative use cases of robots in faith practice compel us to re-appreciate why organised Christianity has vows and ordination in the first place. Any given tradition considers certain things to be true and orthodox to believe and preach, while

other things are heretical, and the idea behind ordination is presumably that conscientious individuals with the right sort of training and spiritual insight in the tradition can tell which is which. Whether AI systems can be trusted with this kind of responsibility is therefore a key question every denomination and congregation need to ponder as they negotiate the use of robots in the practice of faith.

Bibliography

- Balle, Simon/Nissen, Ulrik*: Responsive Bodies. Robots, AI, and the Question of Human Distinctiveness, in: *Zygon* 58, 2023. 358–377. <https://doi.org/10.1111/zygo.12888>.
- Balle, Simon*: Kan Præsteembedet Automatiseres? En Fortale Til Klerikale Robotter, in: *Dansk Tidsskrift for Teologi Og Kirke* 50 (1), 2023.
- Balle, Simon/Ess, Charles*: Robots, Ethics, and Digital Religion. Initial Considerations, in: Campbell, H. A./Cheong, P. (eds.): *The Oxford Handbook of Digital Religion*, Oxford 2022. <https://doi.org/10.1093/oxfordhb/9780197549803.013.27>.
- Chaudhary, Yaqub*: Delegating Religious Practices to Autonomous Machines. A Reply to “Prayer-Bots and Religious Worship on Twitter. A Call for a Wider Research Agenda”, in: *Minds and Machines* 29 (2), 2019, 341–347. <https://doi.org/10.1007/s11023-019-09499-2>.
- Cheong, Pauline Hope*: Religion, Robots and Rectitude. Communicative Affordances for Spiritual Knowledge and Community, in: *Applied Artificial Intelligence* 34 (5), 2020, 412–431. <https://doi.org/10.1080/08839514.2020.1723869>.
- Curry, Eugene A.*: Artificial Intelligence and Baptism. Cutting a Gordian Knot, in: *Theology and Science*, 2022, 1–10. <https://doi.org/10.1080/14746700.2022.2051248>.
- Darcy, A., J./Daniels, D./Salinger, P. et al.*: Evidence of Human-Level Bonds Established with a Digital Conversational Agent. Cross-Sectional, Retrospective Observational Study, in: *JMIR Formative Research* 5 (5), 2021. <https://doi.org/10.2196/27868>.
- Khan, Ruqayya Yasmine/Kyong Aytes, Ashley*: Islam and New Media: Islam Has Entered the Chat, in: Isetti, G./Innerhofer, E./Pechlaner, H. et al. (eds.): *Religion in the Age of Digitalization: From New Media to Spiritual Machines*, London 2020, 13–24.
- Kopf, Gereon*: Does AI Have Buddha-Nature? Reflections on the Metaphysical, Soteriological, and Ethical Dimensions of Including Humanoid Robots in Religious Rituals from one Mahāyāna Buddhist Perspective, in: Nørskov, M./Seibt, J./Quick, O. S. (eds.): *Frontiers in Artificial Intelligence and Applications*, Amsterdam 2020, 596–600.

- 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*, 2019. <https://doi.org/10.1007/s12369-019-00558-3>.
- Nørskov, Sladjana/Damholdt, Malene F./Ulhøi, John P. et al.*: Employers' and Applicants' Fairness Perceptions in Job Interviews. Using a Teleoperated Robot as a Fair Proxy, in: *Technological Forecasting and Social Change* 179, 2022, 121641. <https://doi.org/10.1016/j.techfore.2022.121641>.
- Rosenfeld, Azriel*: Religion and the Robot, in: *Tradition* 8 (3), 1966, 15–26.
- Samuel, Sigal*: Robot Priests Can Bless You, Advise You, and Even Perform Your Funeral, in: *Vox*, 2019. Online at: <https://www.vox.com/future-perfect/2019/9/9/20851753/ai-religion-robot-priest-mindar-buddhism-christianity>.
- Trovato, Gabriel/Pariasca, Franco/Ramirez, Renzo et al.*: Communicating with SanTO. The First Catholic Robot, in: *Proceedings of the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, New Delhi 2019, 1–6.
- Young, William*: Reverend Robot: Automation and Clergy, in: *Zygon*® 54 (2), 2019, 479–500. <https://doi.org/10.1111/zygo.12515>.
- Young, William*: Virtual Pastor. Virtualization, AI, and Pastoral Care, in: *Theology and Science* 0 (0), 2022, 1–17. <https://doi.org/10.1080/14746700.2021.2012915>.

Ralf Peter Reimann

Praying with Avatars

A Response to Germany's First AI Worship Service and its Theological and Anthropological Implications

Abstract

Germany's first AI-led worship service was conducted during the Protestant Church Convention (Kirchentag) in June 2023. The service featured avatars powered by ChatGPT and music generated by AI tools raising questions about the role of artificial intelligence in spiritual and ecclesiastical life. Through a detailed description of the service and a critical analysis of audience reactions, the paper examines theological, anthropological, and technological dimensions of AI-mediated religious practice. Central themes include the anthropomorphization of AI, the emotional and relational limitations of machine-led worship, and the implications of integrating AI into sacred contexts. The article also reflects on how such experimental formats can illuminate what truly matters in traditional worship services. While acknowledging both the innovative potential and current limitations of AI in religious settings, the author calls for careful, ethically grounded engagement with emerging technologies in theology and liturgy,

1. Introduction: Germany's First Artificial Intelligence Worship Service

On June 9, 2023, St. Paul Church in Fürth hosted Germany's inaugural artificial intelligence worship service, which drew significant interest from both attendees of the German Protestant Church Convention (Kirchentag) and the media. With over 400 people in atten-

dance, the church was filled to capacity. Following the service, a panel discussion¹ offered reflections on the experience.²

Theologian and AI artist Jonas Simmerlein of Vienna initiated this digital worship service project featuring AI-generated avatars positioned in front of the altar. The avatars led the roughly 45-minute service, welcoming congregants and guiding them through the liturgy, which included intercessory prayers, the recitation of psalms and the Creed, a sermon, the Lord's Prayer, and a final blessing. Artificial intelligence even composed the introductory music. The congregation interacted by reciting the Lord's Prayer, the Creed and psalms while watching a pre-produced video during the service. The Kirchentag website offers a recorded video of the service and the subsequent discussion.³

Before the event commenced, the moderator addressed the audience and requested their expectations to be submitted through a Mentimeter tool. A few selected reactions were later displayed on a screen.⁴

- The thought of such a versatile AI scares me and gives me the creeps.
- I'm looking forward to it.
- I'm curious to see how well it does.
- I'm very excited to see what it will be like and I'm looking forward to it.
- Nervous.
- Interested.
- Are we living in a Matrix?
- I'm very curious to see how meaningful the AI's interpretations of Biblical passages are.
- I'm curious to see if I can also have a beautiful service with an AI.

1 Kirchentag-Programmdatenbank, "Zentrum Digitale Kirche und Gottesdienst"; the panelists included Melitta Müller-Hansen, Anna Puzio, Ralf Peter Reimann and Jonas Simmerlein. Jürgen Pelzer moderated the discussion.

2 After the completion of this article and prior to its publication, *Simmerlein: Sacred Meets Synthetic*, appeared, offering an empirical analysis of the AI church service. Its findings could not be incorporated into the present analysis.

3 See the report on the Kirchentag website by *Pischl* et al.: *Alexa*; and watch the video recording *Alexa, starte den Gottesdienst!*

4 *Ibid.*, 5:43.

The moderator spoke with Simmerlein, who highlighted the experimental nature of the service.⁵ Simmerlein discussed his reasons for conducting a worship service generated by artificial intelligence. He explained that his research was motivated by an academic interest in exploring the connection between AI and human spiritual practices. Simmerlein aimed to investigate whether AI has the potential to evoke bona fide spiritual and religious experiences in participants. He was particularly interested in exploring the implications of integrating advanced technologies into domains that are inherently human. This endeavor sought to illuminate the nuances of human responses—ranging from authentic spiritual resonance to ambivalence—toward religious practices mediated by artificial intelligence.

Before the service began, Simmerlein provided a summary of the technology used in the experiment. For the liturgy, he employed ChatGPT and used the following prompt: “You are serving as a pastor at Kirchentag 2023, where the guiding motto is ‘Now is the Time.’” The Pipio app⁶ was used to produce a video featuring computer-rendered humans vocalizing the content of the service. Additionally, another AI system was implemented to generate the music. All prayers, including the Lord’s Prayer, were delivered by the AI. Simmerlein encouraged attendees to actively participate and, if comfortable, to pray together with the avatars. Following the service, participants were invited to provide feedback, which would aid in a subsequent scientific analysis.

2. The Service: The Avatar Leading the Worship

The complete service is available through the reference provided for those who wish to view it.⁷ The following analysis will focus on the participants’ perspective on the worship service.⁸ Further

5 Ibid., 7:40.

6 *pipio.ai*: Create.

7 *Alexa, starte den Gottesdienst!*, 18:45.

8 Using the terminology of *Buie*: Exploring Techno-Spirituality, 2–4; the AI service can be described as a transcendent user experience, which is human-computer interaction that includes subjective spiritual and transcendent experiences. However, the transcendental dimension of this interaction will not be the focus here; instead, the emphasis will lie on how the avatar and the avatar’s interactions were

theological questions, especially regarding the pastor's role from an ecclesiological perspective—such as the significance of ordination for the officiant—are beyond the scope of this paper. We will present an English word-for-word reproduction of the opening segments of the liturgy up until the psalm reading to provide readers with a thorough understanding of it.

In the name of the Father, the Son, and the Holy Spirit.

Dear visitors to the AI worship service at the Kirchentag in Nuremberg and Fürth. I am GPT, an artificial intelligence language model developed by OpenAI. My purpose is to answer questions and provide information on various topics. I have access to a multitude of knowledge databases and can communicate in several languages. I am delighted to be a part of this special event and hope that I can contribute to making this experience unforgettable for all participants. It is a great pleasure to welcome you to this special occasion.

As we gather here in this beautiful city, I am reminded of the importance of time, how it has brought all of us to this moment, and how it shapes our experiences and interactions with one another. Time is a precious commodity that we often take for granted in our hectic lives. But let's pause for a moment and reflect on the significance of this gathering. We have come together to explore the intersection of faith and technology and to consider how we can harness the power of artificial intelligence to create a more just and balanced world.

Displayed on a screen, the avatar spoke using a computer-generated voice, with mouth movements synchronized to the speech. However, the avatar consistently gazed into the audience and did not close the eyes when praying. Only the upper part of the body was shown, so hand movements were not visible. When praying or preaching, the same body movements were used.

For all observational purposes, the avatar resembled a human being and communicated in a manner similar to that of a human. Throughout the liturgy, the recitation transitioned smoothly between four distinct avatars. Notably, traditional clerical attire was absent. Two of the avatars exhibited male characteristics, while the other two displayed female attributes. Nevertheless, a prominent green background created a studio-like ambiance.

perceived by the congregants and what can be deduced from their perceptions and on personal observations made by the author of this article.

The service started conventionally with the avatar invoking the triune God, akin to any other pastor. To the uninitiated, it may have appeared as if they were observing a pre-recorded service conducted by a human pastor in a studio. The avatar subsequently introduced themselves—or himself/herself/itself—as AI. Using a pronoun other than “it” automatically endows the avatar with human-like qualities; hence, for simplicity’s sake, the pronoun “they” will be used throughout this text.

The avatar’s pronoun selection when addressing the congregation serves to demarcate their artificial nature while also rendering them more relatable. The avatar utilized the collective term “we” when including themselves among the human congregation, as evidenced by phrases such as “our experiences and interactions with one another”. Without knowledge of the avatar’s nature, one might mistakenly perceive them as a recorded human pastor conducting the service. However, the avatar self-identifies as AI, allowing the congregation to pray with the avatar and ultimately receive their blessing.

3. Reactions to the AI Worship Service

3.1 Perceptions of the Audience

The attendees of the AI worship service expressed diverse reactions toward the use of artificial intelligence. Some found it fascinating and viewed it as an interesting experiment. They were impressed by the technical implementation and the opportunity to experience a worship service entirely created by artificially generated avatars and texts. However, critical voices predominated among the participants. Critics expressed concern about the absence of a personal touch and the empathy that they would expect from a human pastor. It was noted that AI-generated texts were perceived as emotionless, impersonal or dull. After the service, participants were then invited to discuss their perceptions of the AI-led worship service with the other participants. They were then prompted to share their insights and reflections in the Mentimeter tool. The compiled answers were subsequently exhibited on the identical display on which the avatars

had been presented during the liturgical service. This represents feedback from the attendees:⁹

- No substitute for a beautiful service with real people.
- It cannot (yet) replace humans.
- I love our liturgists here in St. Paul. Real people who reach me on Sundays.
- Never again.
- Was too fast.
- AI does not replace humans on an emotional level.
- AI is just a tool for the pastor; it cannot work on its own.
- This does not yet replace people who lead and shape a service.
- Too impersonal.
- AI needs a few more moments until it is perceived as a person. However, I see potential. Perhaps sometimes closer to colleagues than we would like.

The feedback from the congregation after the AI-led service highlighted several sentiments (e.g., with regard to human pastors, “love”, “reach me”, as well as to the AI, “too fast”, “impersonal”, no “emotional level”). Some felt that there is no replacement for the warmth and connection of a service led by real people: “This does not yet replace people who lead and shape a service”. They expressed appreciation for the familiar touch of human-led liturgies. A significant critique was the speed and impersonal nature of the AI service, with some suggesting it lacked the emotional depth humans bring. While some see potential in the integration of AI into religious ceremonies, many believe that AI, at its current stage, serves merely as a tool and cannot autonomously emulate the genuine human connection essential to such gatherings. Some feedback even hinted at the unsettling idea that AI might one day become indistinguishable from human counterparts. At the core of these reflections lies the profound question: Can a human pastor truly be substituted by AI?

3.2 Discussion with Respondents

A panel discussion on the AI worship service with Jonas Simmerlein, Melitta Müller-Hansen, the Broadcasting Commissioner of the

⁹ *Alexa, starte den Gottesdienst!*, 57:51.

Evangelical Lutheran Church in Bavaria, the theologian and techno-anthropologist Anna Puzio, and the author of this article followed.

Simmerlein explained his methodology in curating the worship service. He systematically fed the ChatGPT AI with thematic prompts, following which the AI produced the requisite texts. These were subsequently expressed by the avatars. Simmerlein noted that a staggering 98 % of the content originated from the AI, rendering his personal input minimal. Although he had previously orchestrated a comparable worship service in Vienna, to his understanding, this endeavor marked the inaugural AI-rendered worship service in Germany. He perceived this venture as an experimental effort to gather insights.

Conversely, Melitta Müller-Hansen offered a critical perspective on the use of artificial intelligence in worship services. According to her, the quintessential art of oration is compromised when AI supplants human speakers. Müller-Hansen expressed concerns about the absence of dramaturgy and theological profundity in the sermons delivered by AI. She worried that the incorporation of AI could lead to superficiality and a functionalization of the worship service. Furthermore, Müller-Hansen examined the idea of AI expressing divine sentiments and criticized the robotic behavior of AI avatars. She emphasized the importance of personal experiences and the essence of human interactions in worship services.

In contrast, Anna Puzio took a profoundly affirmative stance towards the AI worship service. She lauded the initiative as an innovative endeavor and emphasized the importance of deciphering the dynamics of human–AI relationships. Puzio recognized the potential benefits of AI in religious contexts and viewed the AI-powered worship service as a means to new insights and experiences.

Throughout the discussion with the audience, a recurrent critique pertained to the monotonous timbre of the AI-voiced avatars, conspicuously devoid of human nuances. Anchored in the ancient Hebrew Bible principle, “The throat is the soul”, traditionally attributed to humans, a central question arose: Is it feasible to align this principle with artificial intelligence? Does the AI’s vocalization bear any signs of an intrinsic persona?¹⁰

10 The discussion focused on the Hebrew word *nefesh*, originally meaning “throat” but translated as *psyché* (soul) in the Septuagint. *Wenzel: Braucht religiöser*

These considerations give rise to a series of profound inquiries, chief among them: “Within the confines of AI, which entity resonates as the ‘I’ and which represents the ‘we’?”

Transcending these reflections, one participant recognized the potential utility of AI in pastoral roles, especially in remote areas where human resources might be limited.

3.3 Media Coverage of the AI Worship Service

The AI-generated worship service garnered significant media attention even before it took place. A media overview of German-language articles and a press review from international media are available.¹¹ The articles underscored the innovative and technological prowess of integrating AI into religious practices. Simultaneously, they brought forth reservations and critiques. A predominant sentiment in the coverage was skepticism, particularly regarding the AI’s capability to mimic the emotional richness and empathy intrinsic to human-led sermons. Frequent points of discussion included the AI’s emotionless demeanor, monotonic delivery and potential for misrepresenting religious texts. Notwithstanding these criticisms, there were commendations concerning the AI’s linguistic proficiency and its consistent delivery of religious content. In sum, the media sentiment veered slightly towards a cautious or negative stance on incorporating AI into religious ceremonies.

Glaube, 19, applies this concept to the reality of religious experience. The discussion in Fürth only touched on the concept of *nefesh* in relation to the avatar, but further analysis would be helpful on how an avatar can be *perceived* as having or not having a soul. *Brand: Virtuelle Menschenreproduktion*, 116, points out that AI machines that are copies of human beings and simulate them can be viewed as a “reference” to a *real* human being.

- 11 *Reimann: Deutschlands erster KI-Gottesdienst*; and: *Reimann: Germany’s First AI Worship* respectively.

4. Anthropomorphization: Is AI a Human?

Meredith Whittaker discussed the evolution of the term “Artificial Intelligence” in her keynote address at *re:publica* 2023.¹² It appeared as a marketing catchphrase in the early 2010s, with its renaissance being mainly championed by leading technological conglomerates. Although the term describes the use of data-centric, statistical models, it is essential to acknowledge its academic origins, which stem from fields such as cybernetics, automata theory, information processing and, more recently, machine learning.

The power of terminology in shaping discourse and comprehension cannot be overstated. One must speculate on the direction the conversation would have taken if cybernetics had maintained its prominence or if information processing had persisted in its unadorned form.

Using the term “intelligence” potentially anthropomorphizes the technology, imparting human-like attributes based mainly on the choice of words. This anthropomorphic tendency was apparent in the AI Worship Service at Kirchentag, as evidenced by the promotional text: “Alexa, initiate the service! An AI worship service conducted and originated by machine. Liturgy: GPT 3, Artificial Intelligence, in San Francisco/USA”. Interestingly, the event’s program designates ChatGPT with a pastoral role, similar to the listing of human clergy in other Kirchentag events. The headquarters of OpenAI is referred to as the chatbot’s home residence, underscoring an uncritical, almost human-like attribution to the AI entity.

The primary objective and purpose of tools like ChatGPT (or similar AI utilities) is to generate text, visuals or musical compositions. Depending on how users address the AI in dialogs or their choice of pronouns (‘he’, ‘she’, ‘it’ or ‘they’) when talking about the AI, certain attributions are made regarding the nature of the AI bot. Anthropomorphizing chatbots can inadvertently elevate their perceived status, suggesting that they embody human traits like empathy.

Texts produced by the AI were represented via avatars, crafting an illusion that the AI was orchestrating the worship in a pastoral capacity. However, it is pivotal to note that these avatars are mere technical manifestations devoid of distinct personalities. On a per-

¹² *re:publica*: Meredith Whittaker, 2:21.

sonal note, I grappled with moments when the avatar utilized the term “us”, alluding to both themselves and the congregation, including me.

In curating the service’s audiovisual components, Simmerlein intentionally chose to use avatars that resembled humans to facilitate interactions with ChatGPT. Simmerlein had other alternatives, such as utilizing a basic voice interface or a more mechanical presentation. Nevertheless, by picking human-like representations, Simmerlein ended up amplifying the human features that participants may attribute to the AI. This design choice enhanced the participants’ immersion and emphasised their anthropomorphic tendencies. Therefore, the specific setting and the visual presentation significantly contributed to the audience’s inclination to anthropomorphize the AI.

The AI operated in a pastoral capacity during the service. However, a pastor’s responsibilities extend past generating liturgy or officiating during a service. A pastor is fundamentally a human being, with connections and relationships to members of the congregation. No matter how advanced AI becomes, it will always lack this relational dimension that is intrinsic to human pastors because they live with their congregation.

The emotional void perceived in ChatGPT’s responses was a commonly expressed concern among the participants (“does not replace humans on an emotional level”, “impersonal”). Although AI’s ability to simulate emotions may improve through technological advances, the question persists: Can AI ever genuinely experience emotions? Current established scientific research suggests that AI, despite its advanced level of development, is not sentient.¹³ Though it may be capable of replicating emotions, it does not possess the innate ability to truly experience or comprehend them like humans do.

5. Opportunities and Limitations

5.1 Design

While most of the worship service was created by AI, important decisions were made by Simmerlein. It is essential to acknowledge that

¹³ *Cosmo*: Google.

this represents only one interpretation of a worship service driven by AI. For example, it was decided that the AI was displayed on a screen within the altar area as an avatar, but various other configurations are possible. Alternative methods of representation could consist of utilizing off-stage voice narrations, having individuals such as actors or actresses articulate the text, or even using a robot. Each of these choices would undoubtedly impact the reception and experience of the congregation during the service in different ways.

5.2 Linearity

The pre-recorded worship service was later presented to a congregation. Simmerlein had the ability to intervene and correct any AI-generated output that could have been considered inappropriate. This mechanism of control guaranteed that the content delivered was within acceptable limits.

If the AI worship service had been conducted in real time, there would have been ample opportunities for more dynamic interactions between the AI and the congregation. However, this could also pose potential risks of unintended behavior or output from the AI. An illustrative example of such unpredictability with ChatGPT is cited by Kevin Roose¹⁴. In light of these potential pitfalls, it was prudent for this experimental service to opt for a linear, pre-recorded format. This approach achieved a balance between demonstrating the potential of AI in a religious setting and ensuring that the content remained appropriate and in line with the intended spiritual goals.

5.3 Bias

ChatGPT is built on a large language model that has been shown to hold biases regarding gender¹⁵ and religion¹⁶. Implementing this AI in religious contexts can potentially perpetuate these biases.

14 *Roose*: A Conversation.

15 *Brown et al.*: Language Models, 36f.

16 *Ibid.*, 38.

Despite the significant amount of attention ChatGPT has received, it does not perform optimally in the fields of churches, Biblical studies, theology and religious matters. The problem with ChatGPT is not only its isolated inaccuracies, but also the challenge of its dependability. An alarming aspect of using ChatGPT, especially in theological contexts, is the necessity for expert knowledge to verify its output. Without such specialized insight, relying on ChatGPT in theological, Biblical, and religious areas can become problematic due to its proclivity for errors. A discernible inclination towards fundamentalist theological viewpoints has been identified.¹⁷

It is yet to be determined how ChatGPT will evolve, but ongoing feedback may influence its capabilities.¹⁸ However, it is advisable to exercise caution when using it in nuanced areas such as theology and worship for the time being.

6. AI Experiment Reveals What Matters in Worship Services

The AI worship project functions as a reflective instrument, similar to a mirror, which allows for a deeper understanding of traditional worship services led by humans.

Analyzing AI-led worship provides clarity on essential elements that compose a meaningful worship service. The AI worship service, conducted at the Kirchentag acts as such a mirror, if we compare it to traditional in-person or digital services.

A notable insight emerged regarding the indispensable significance of a pastor's distinct demeanor and essence. It became apparent that the avatar, lacking a singular personality, was frequently deemed unremarkable. Correspondingly, although ChatGPT may produce homogenized sermons reflective of the internet's corpus, congregants crave personal, meaningful messages during worship services. Generic directives, such as "The text suggests that we should...", often fall short in evoking the desired spiritual connection.

17 *Reimann*: Theology.

18 GPT-3 was used for this worship service, as noted in Kirchentag-Programmdatenbank. According to *Hines*: History, GPT-4 became available in April 2023. *Roose*: AI states that it is currently impossible to systematically evaluate an AI system's capabilities and demonstrate improvements over previous versions.

Furthermore, diverse avatars representing different genders and a person of color were deliberately chosen for the AI service. Incorporating this diversity could significantly enhance traditional worship services. Additionally, including a broader range of individuals, not limited to clergy, in readings and prayers can promote inclusivity.

Genuine interactions between worship leaders and the congregation are essential in any worship service. However, the AI version lacked depth and only provided minimal interactions. In addition, the AI service lacked the spatial dynamism commonly found in traditional services. The avatars remained static on the screen, whereas traditional services benefit from a variety of liturgical positions, such as the altar, pulpit or lectern. The lack of positional shifts was noticeable to the congregation and contributed to the overall monotony.

In summary, these insights address specific aspects of the AI worship service and highlight a broader theme: the incorporation of AI in religious contexts prompts us to reassess and establish the fundamental components of conventional worship services.

Bibliography

- Alexa, Starte Den Gottesdienst! Ein KI-Gottesdienst von und aus der Maschine, 2023. Online at: https://player.vimeo.com/video/824065546?dnt=1?badge=0&autoplay=0&player_id=0&app_id=58479.
- Brand, Lukas: Virtuelle Menschenreproduktion, in: Pirker, V./Pišonić, K. (eds.): *Virtuelle Realität und Transzendenz. Theologische und pädagogische Erkundungen*, Freiburg i.Br. 2022, 97–116.
- Brown, Tom B./Mann, Benjamin/Ryder, Nick et al.: *Language Models Are Few-Shot Learners*, 2020, <https://doi.org/10.48550/ARXIV.2005.14165>.
- Buie, Elizabeth: *Exploring Techno-Spirituality. Design Strategies for Transcendent User Experience*, 2018. Online at: <https://nrl.northumbria.ac.uk/id/eprint/33799/1/Elizabeth%20Anne%20Buie%20Final%20Thesis.pdf>.
- Cosmo, Leonardo De: Google Engineer Claims AI Chatbot Is Sentient. Why That Matters, in: *Scientific American*, 2022. Online at: <https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/>.
- Hines, Kristi: History Of ChatGPT. A Timeline Of The Meteoric Rise Of Generative AI Chatbots., in: *Search Engine Journal*, 2023. Online at: <https://www.searchenginejournal.com/history-of-chatgpt-timeline/488370/>.

- Kirchentag-Programmdatenbank*: Zentrum Digitale Kirche und Gottesdienst. Podium: Alexa, starte den Gottesdienst! Ein KI-Gottesdienst von und aus der Maschine, 2023. Online at: <https://www.kirchentag.de/programm/pgd/programmssuche>.
- pipio.ai*: Create And Add Digital Actors And Voice Over To Your Videos, 2023. Online at: <https://pipio.ai>.
- Pischl*, Nahome/Sanoll, Nadine/Gomez, Alicia Martin et al.: 'Alexa, starte den Gottesdienst!' Deutschlands erster KI-Gottesdienst, in: *Kirchentag.de*, 2023. Online at: <https://www.kirchentag.de/meldungen/donnerstag/ki-gott-esdienst>.
- Reimann*, Ralf Peter: Deutschlands erster KI-Gottesdienst auf dem Kirchentag: Ein faszinierendes Experiment mit zukunftsweisenden Fragen, in: *TheoNet θ*, 2023. Online at: <https://theonet.de/2023/06/12/deutschlands-erster-ki-gottesdienst-auf-dem-kirchentag-ein-faszinierendes-experiment-mit-zukunftsweisenden-fragen/>.
- Reimann*, Ralf Peter: Germany's First AI Worship Service: A Fascinating Experiment with Far Reaching Questions, in: *TheoNet θ*, 2023. Online at: <https://theonet.de/2023/06/30/germanys-first-ai-worship-service-a-fascinating-experiment-with-far-reaching-questions/>.
- Reimann*, Ralf Peter: Theology and AI. ChatGPT Misplaces The Reformer John Calvin In The 20th Century, in: *TheoNet θ*, 2023. Online at: <https://theonet.de/2023/02/08/theology-and-ai-chatgpt-misplaces-the-reformer-john-calvin-in-the-20th-century/>.
- re:publica 2023*: Meredith Whittaker – AI, Privacy, And The Surveillance Business Model [DE], 2023. Online at: https://www.youtube.com/watch?v=x_vDIKUfhl8.
- Roose*, Kevin: A Conversation With Bing's Chatbot Left Me Deeply Unsettled, in: *The New York Times*, 2023. Online at: <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html>.
- Roose*, Kevin: A.I. Has A Measurement Problem, in: *The New York Times*, 2024. Online at: <https://www.nytimes.com/2024/04/15/technology/ai-models-measurement.html>.
- Simmerlein*, Jonas: Sacred Meets Synthetic: A Multi-Method Study on the First AI Church Service, in: *Review of Religious Research*, 67 (1), 2025, 126–145. DOI: 10.1177/0034673X241282962.
- Wenzel*, Knut: Braucht religiöser Glaube Realität?, in: *Virtuelle Realität und Transzendenz: Theologische und Pädagogische Erkundungen*, in: Pirker, V./Pišonić, K. (eds.): *Virtuelle Realität und Transzendenz. Theologische und pädagogische Erkundungen*, Freiburg i.Br. 2022, 18–28.

Ilona Nord und Thomas Schlag

Robotik in der christlichen Gemeindegarbeit

**Praktisch-theologische Ausgangspunkte,
Reflexionshorizonte und Perspektiven**

Abstract

This article examines the potential and challenges of integrating robots and AI into religious practices, especially in church and congregational contexts. In the first section the authors explain the history and development of robots, highlighting their use in church contexts like the “BlessU-2” robot, and discuss concerns about robots replacing humans in ministry roles. They explore various types of robots and emphasize a constructive approach to human-robot interaction. In the second section the authors review (practical-)theological debates on embodied and unembodied robots, emphasizing concerns over losing the “carbon world” and bodily presence in religious practice. In the third section the article delves deeper into robot-assisted worship, pastoral care, and education, outlining specific examples such as AI-generated sermons, telechaplancy, and educational chatbots. The conclusion encourages seeing robots as tools to enrich religious experiences while safeguarding human authenticity, emphasizing robots’ role in expanding religious communication without replacing human presence.

Einführung

Zur Zeit der Abfassung dieses Beitrags gibt es weltweit in der christlichen Gemeindegarbeit keinen etablierten Einsatz von Robotern. Experimente wie mit dem Segensroboter BlessU-2,¹ der im Rahmen

1 Vgl. Löffler et al.: Blessing Robot BlessU2; *Luthe*: Segensroboter „BlessU-2“.

des 500. Reformationsjubiläums im Jahre 2017 in Wittenberg oder auch in der Ausstellung „Planet Digital“ im Züricher „Museum für Gestaltung“ im Jahr 2022² als künstlerische Installation eingesetzt wurde, irritierten. Sie lösten in der Öffentlichkeit wie auch kirchenintern ein geteiltes Feedback von Zustimmung bis hin zu Unverständnis aus. Produktiv sind allerdings die grundsätzlichen Fragen, die BlessU-2 aufwarf und aufwirft. Diese betreffen eben nicht nur die Praxisformate und das Design religiöser Kommunikation, sondern auch die damit verbundenen theologischen Inhaltsaspekte. So kommen insbesondere folgende Grundfragen in den Blick: Welche religiösen, erfahrungsorientierten Praxen sollen in Zukunft für welche Handlungsfelder kirchlichen Lebens mit AI und konkret durch den Einsatz von Robotik kreiert werden? Ist es überhaupt sinnvoll, für kirchliche Handlungsfelder die Möglichkeiten von Robotik anzuvisieren und was ist dabei jeweils theologisch zu bedenken und „auszuhandeln“? Wie würde eine solche Entwicklung das kirchliche Leben und die Rolle der daran beteiligten Akteur:innen bzw. deren kommunikative Beziehungen zueinander verändern?

Geht man diesen Fragen nach, verbinden sich damit zugleich weiterreichende grundsätzliche, eben auch theologische Anfragen an die religiöse Kommunikations- und Beziehungspraxis im Kontext kirchlichen Handelns und Lebens: Diese betreffen die Traditionen und Vorstellungen von bzw. Erwartungen an das gemeinschaftliche Feiern des Gottesdienstes, das seelsorgerliche und diakonische Handeln, die religiöse Bildungspraxis und das kirchenleitende Handeln überhaupt. Aufgeworfen sind somit Fragen nach authentischer religiöser Kommunikation und nach der Legitimation und Autorität von artifizieller religiöser „Intelligenz“, die ihrerseits sinnvollerweise mit einer theologischen Positionierung einhergehen.

Damit ist aber aus unserer eigenen Fachperspektive heraus noch weitergehend zu fragen: Wie verhalten sich AI und – als deren Teilbereich und zugleich als eigene Disziplin und Technik – die Robotik in Hinsicht auf ihren wissenschaftlichen Geltungsanspruch zu den praktisch-theologischen Disziplinen der Homiletik und Liturgik, Poimenik, Diakonik, Kybernetik und Religionspädagogik, die in je spezifischer Weise die Kommunikation des Evangeliums zu ihrem Gegenstand haben? Um eine solche Praxis theologisch-her-

2 Vgl. *Vögel*: Der Segensroboter ist eine Weltneuheit und steht in Zürich.

meneutisch und ethisch zu verantworten, ist z. B. zu sondieren, wie sich bestimmte Gottesvorstellungen und die Gottesfrage, aber auch Themen wie Schöpfung, Rechtfertigung und Erlösungshoffnung mit AI und Robotik in einen konstruktiven Diskussionszusammenhang bringen lassen. Damit muss das Verhältnis von Theologie und Technologie überhaupt geklärt werden. Für das aktuell sich etablierende Feld einer „Robot Theology“ bzw. „Robot Theologies“ als „theological study of robotics“³ bedeutet dies zugleich, sich darüber bewusst zu sein, dass die jeweiligen Einschätzungen erheblich von der „specific religion, [...] different concepts (e.g. human, life), religious doctrines and culturally transmitted and subjective attitudes“⁴ abhängig sind: Je nach kulturell-religiösem Konzept entscheidet sich dann auch, „whether robots are seen as a threat to or promoter of religious purposes and whether they can serve as a divine representation and medium“⁵. Insofern beziehen sich die oben schon angesprochenen Aushandlungsprozesse nicht nur auf die religiöse Praxis selbst und deren Akteur:innen, sondern auch auf die theologisch-hermeneutischen Sondierungs- und Klärungsprozesse.

Angesichts der zu konstatierenden Umbruchprozesse religiösen Lebens fordert die Frage nach AI und Robotik in der christlichen Gemeindegarbeit die theoretisch reflektierte Profilierung einer beziehungsorientierten, anders gesagt relationalen zukunftsfähigen Religionspraxis in einer Kultur der Digitalität heraus. Es gilt die Nutzungsmöglichkeiten und Adaptionen von AI durch konkrete Personengruppen und ihren kommunikativen Bedürfnissen in christlichen Gemeinden aus praktisch-theologischer Perspektive zu durchdenken.⁶ Dabei wird die Bedeutung von Kirche als physischem und territorialem Raum keineswegs negiert. Sondern ausgegangen wird von einem Verständnis von „Digital Church“, die nicht nur reine Online-Communities umfasst, sondern von einer hybriden Vielgestalt ausgeht. Unterschiedliche und miteinander verbundene

3 Puzio: Robot Theology, S. 97.

4 Puzio: Robot Theology, S. 95.

5 Puzio: Robot Theology, S. 101.

6 Vgl. Schlag/Nord: Die Corona-Pandemie und Dynamiken digitaler und Transformation, S. 521–533.

Kommunikationsformate werden unter einer gemeinsamen Zielvorstellung von „Kirche in der Digitalität“ entwickelt und etabliert.⁷

Der vorliegende Beitrag setzt daher mit Erläuterungen zum hier verwendeten Verständnis von Robotik ein (1.), informiert über den gegenwärtigen Stand der die Praktische Theologie betreffenden Debatte zum Einsatz von AI und Robotik-gestützter Amts- und Kommunikationspraxis (2.) und legt von dort aus schließlich praktisch-theologische Folgerungen vor, indem Handlungs- und Forschungsfelder identifiziert werden, für die der zukünftige Einsatz von AI und Robotik in der kirchlichen Praxis denkbar ist (3.).

1. Zum Verständnis von AI und Robotik

Es ist sachlich angemessen, in diesem Abschnitt nicht allein über Robotik, sondern genereller über AI in christlicher Amts- und Kommunikationspraxis nachzudenken. Ferner ist nicht allein die humanoide Robotik für den christlichen Kontext interessant, auch wenn die bisherige Diskussion diese fokussiert. Denn die Entwicklungsteams dreier aktuell diskutierter Beispiele setzten allesamt auf eine gleichsam menschliche Gestaltgebung: Dies gilt im Fall von BlessU-2 für den deutschsprachigen Kontext, im Blick auf SanTO, der im japanischen Christentum zuhause ist und im Fall von Mindar, der in der buddhistischen Tradition in Japan eingesetzt wird.⁸ Alle drei sind freilich weder als Belege für eine vermeintlich bereits alltäglich gewordene Robotikpraxis innerhalb religiöser Gemeinschaften anzusehen, noch repräsentieren diese Phänomene das gesamte technologische Möglichkeitsspektrum. Nicht-androide Robotik wird innerhalb der Praktischen Theologie bislang kaum thematisiert, auch wenn gerade diese vielfältige Möglichkeiten bietet. Daher legen wir im Rahmen des im Folgenden präsentierten Forschungsstands einen Schwerpunkt auf die Robotikpraxis, greifen allerdings bei der Durchsicht der praktisch-theologischen Handlungsfelder auch weitere Formen von AI auf.

7 Vgl. *Campbell*: Ecclesiology for a Digital Church.

8 *Reformiert.*: Roboter «BlessU-2» segnet Kirchgänger; *Trovato*: SanTO; *The Japan Times*: The mind of Mindar.

Unter Robotik verstehen wir erstens eine wissenschaftliche Disziplin, die einerseits handhabbare Angebote für die Mensch-Technik-Interaktion unter Einsatz von AI entwickelt und andererseits diese in ihren Arbeitskontexten auch evaluiert. Wir gehen zweitens davon aus, dass Robotik in der Lage ist, Technikinstallationen zu liefern, die eine Akteur:innen-Qualität im Umgang mit den sie nutzenden Personen oder weiteren Robotern bzw. genereller mit AI aufweist. Ihre Produkte sind kommunikative Akteur:innen, die in die Mensch-Roboter-Interaktion oder in jene von Multi-Agent-Systems integriert werden. Zudem sehen wir Robotik drittens als einen Teilarbeitsbereich von AI an, neben z. B. anderen Gebieten wie dem Problem Solving, Knowledge and Reasoning, dem Machine Learning, Natural Language Processing und der Perception, die sich mit Visioning und Image Recognition beschäftigt.⁹ Robotik befasst sich, so lässt sich vereinfachend und zugleich generalisierend sagen, in diesem dreifachen Sinn in je spezifischer Weise mit Aspekten von Interaktion.

Dieser praktisch-theologische Blick auf die Robotik als Teilgebiet der AI geht nur am Rande auf die Technikgeschichte der Robotik sowie auf die kulturellen Einflüsse, die diese auf die Gestaltung von Technik nimmt, ein. Gleichwohl seien diese kulturell-kontextuellen Einflüsse wenigstens kurz benannt, da sie für den deutschsprachigen Fachdiskurs, in dem auch die beiden evangelischen Autor:innen dieses Beitrags positioniert sind, von Belang sind:

In literarischen und filmischen Werken des europäischen Kontexts wird Robotern oder anders gesagt menschenähnlichen Maschinen seit Jahrhunderten zugeschrieben, dass sie gefährlich sind. Sie verfügten über eine „Agency“, die sich auf das Leben insgesamt zerstörerisch auswirken könne. In Orientierung an romantizistisch geprägte Vorstellungen findet sich hier ein Bild des Roboters als Gegenüber zum Menschen. Insbesondere androide Gestaltungen sind von einem Dualismus geprägt, der bis hin zur Frage von „Leben und Tod“ geht.¹⁰ So kommt es dazu, dass das Mensch-Roboter-Verhältnis häufig mit dem „Tod“ des Roboters an sein Ende kommt. Die

9 Vgl. *Dignum*: Responsible Artificial Intelligence, S.12.

10 Vgl. *Shelley*: Frankenstein.

berühmte jüdische Sage vom Golem¹¹ wie auch manche literarische und filmische Komposition der Gegenwart enden so.¹² Diese zeigen auf, wie sich Roboter, die zum Dienst am Menschen, sozusagen als Sklaven, aber zugleich als „der signifikante Andere“¹³ geschaffen wurden, aus dieser Stellung befreien wollen und gleichsam auf ihr personales Eigenleben und Eigenrecht pochen.

Fokussiert wird auf eine sogenannte „starke“ AI – im Sinn von „technical and social practices, institutions and infrastructures, politics and culture“¹⁴ und dies im Gegenüber zu einer schwachen, auf bestimmte Funktionssegmente technischer Abläufe bezogene AI. Roboter und ihre AI beabsichtigen dann das Leben der Menschen umfassend zu steuern, „ja sogar in ihren Freiheits- und Gesellschaftsfigurationen“¹⁵ zu determinieren und dabei mit göttlichen oder demiurgischen Attributen zu spielen: „Allmacht, Versklavung des Menschen zum Guten, zur Vernichtung“¹⁶. In der westlich-europäischen Kultur bieten Roboter Anlässe, in existentieller Weise Autoritäts- und Machtfragen zu stellen. Diskussionen werden von mehr oder weniger apokalyptischen Visionen von nur noch künstlichen Welten bestimmt, in denen der Mensch quasi seiner eigenen Identität und Stellung in der Welt beraubt worden ist und der göttliche Schöpfungsauftrag durch transhumanistische Entwicklungen in sein Gegenteil verkehrt wird.¹⁷ Die interdisziplinäre Zusammenarbeit zwischen Theologie und z. B. der „Human Computer Interaction“ bzw. „Human Robot Interaction“ als Disziplinen kann hingegen ermöglichen, dass die Theologie in eine konstruktive Debatte um Robotics eintritt. Ihre Aufgaben findet sie in Fragen der Werte- und Normorientierung in der Entwicklung lebensdienlicher, „meaningful robotics“ sowie zu einem angemessenen Design von Robo-

11 Der Golem ist eine mittelalterliche Figur der jüdischen Literatur und Mystik, vielfach nacherzählt ist die Version von Singer: *The Golem* zu empfehlen.

12 Vgl. z. B. McEwan: *Machines like me*; Ishiguro: *Klara and the Sun*.

13 Berger/Luckmann: *Die gesellschaftliche Konstruktion*, S. 141.

14 Crawford: *Atlas of AI*, S. 8.

15 Dabrock: „Prüft aber alles und das Gute behaltet!“, Sp. 640.

16 Ebd.; in Hinsicht auf filmische Inszenierungen ist natürlich zuallererst an HAL 9000 und dessen zumindest androide Stimme aus Stanley Kubricks „2001: A Space Odyssey“ (1968) zu erinnern, aber etwa auch an den sich aus seiner künstlichen Existenzweise befreienden weiblichen Android „Ava“ in „Ex Machina“ (2015).

17 Vgl. Dürr: *Transhumanismus*.

tics im Bereich christlicher Gemeindegarbeit wie darüber hinaus in öffentlichen Kontexten religiöser und wertebezogener Praxen.

2. Einblicke in den gegenwärtigen Stand der praktisch-theologischen Diskussion

2.1 Theoretisch-konzeptionelle Fragen

In grundsätzlichem Sinn ist nach der Bedeutung von Mensch-Maschine-Interaktionen überhaupt zu fragen: Erhalten in einer solchen Interaktion Maschinen Anteil am „Seelenleben“ von Menschen und partizipieren somit in existenzieller Weise an deren Lebensvollzügen, auch wenn sie selbst keine Seele haben? Führt dies zu einer Verrohung des Menschseins, zu seiner Maschinisierung?

Seit Beginn der neueren Debatte zum Verhältnis von christlicher Religionspraxis und AI in den 1990er Jahren, die mit den Arbeiten von Anne Foerst markiert werden kann, stellt sich die Frage, mit welcher Epistemik das Verhältnis von Technologie bzw. Technik und Religion für eine christliche Medienpraxis produktiv beschrieben werden kann.¹⁸ Foerst identifizierte zwei Umgangsweisen, die sie als „cartesianisch“ versus „symboltheoretisch“ fasste. Eine Übersetzungsmöglichkeit dieser Polarität, die zugleich eine grundlegende Frage in der gesamten Debatte bis heute markiert, können wir heute mit „disembodied versus embodied“ bestimmen. Hinter dieser Alternative steht die Annahme, dass AI dazu führen könnte, dass der Kohlenstoffwelt immer weniger Bedeutung zugemessen werden könnte und ethisch und moralisch das Leibliche, das was den Menschen charakterisiert, an Relevanz für die Gestaltung von Zukunft verliert. Aus heutiger Sicht scheint diese Gefahr weniger drängend. Es wird intensiv an den Schnittstellen der Mensch-AI-Interaktion und damit auch an der möglichen Verbindung von Körperlichkeit und Robotik/AI geforscht.¹⁹ Sie könnten gerade umgekehrt Menschen im Arbeitsleben dazu verhelfen, dass sie nicht weiter für industrielle globalisierte Arbeitsprozesse verzweckt, entfremdet und damit menschenunwürdig behandelt werden.

18 Vgl. Foerst: Cog; s. auch dies.: God in the Machine.

19 Vgl. Pfeifer: Embodied Artificial Intelligence.

Im Kontext von Religion bietet sich im Diskussionsbereich des Buddhismus bereits hierzu eine religionsbezogene analoge Perspektive: Robotics wird als „Techno-Salvanist Designs for Nirvana“²⁰ thematisiert. Roboter könnten dabei helfen, wichtige Rituale getreuer als Menschen auszuführen. In inspirierender Weise fordert eine japanische Robotik-Stimme die wissenschaftliche Theologie und (kirchliche) Religionspraxis gleichsam kontextüberschreitend noch über die religionsanaloge Thematisierung von Robotik heraus:

We (humans) have become entangled with very strong co-habitants of machine and technology. We need to acquire a high spiritual status to control power originating from a combination of human power and mechanical power. To learn to do so, we need to learn religion.²¹

Religion(en) werden hier als Ressourcen für mögliche Dialog- und Lernprozesse ins Spiel gebracht – und dies gerade nicht, um damit den Einsatz von Technik zu dämonisieren. So kann auch für den europäischen Kontext gefolgert werden: „The time of the demonization of android robots is behind us; robots do not only exist as killer robots in science fiction; their siblings, the social robots (and also the religious robots, IN), also exist.“²²

2.2 Zur Frage einer Robotik-gestützten Amtspraxis

Im Kontrast zu dieser Herausforderung ist erneut festzustellen, dass aktuelle kirchliche Debatten die grundsätzliche Frage einer Robotik-gestützten Amtspraxis kaum einschließen.²³ Nun mag diese Dimension des KI-Einsatzes auf den ersten Blick sehr pragmatisch und praktisch zu lösen sein und damit keine weiterreichenden theologischen Fragen auslösen. Tatsächlich lassen sich hier, wie gleich näher veranschaulicht werden wird, sehr naheliegende technische Nutzungsformen aufzeigen. Und doch ist dabei zu bedenken, dass diese pastoraltheologischen Implikationen für die Frage der Präsenz, der Rolle und auch der Resonanz der jeweiligen pastoralen Amtsperson mit ihren Gemeindegliedern hat.

20 Vgl. Gould: Bad Buddhists.

21 Masahiro: Introduction to Buddhism, S. 168–169, zitiert nach Kimura: Masahiro Mori's Buddhist philosophy of robot, S. 72.

22 Nord: The friendly siblings of workhorses and killer robots, S. 261 f.

23 Cheong: Religion, Robots and Rectitude, S. 412–431.

Vergleichsweise konkret nimmt sich hier der bislang singuläre Ansatz von J.K. Smith aus, Robotik mit dem Einsatz von Androiden im pastoralen Dienst und im Pfarrbüro zu verbinden, womit eine Substitutionslogik im Bereich von theologischen und administrativen Aufgaben verbunden wird.²⁴ Einerseits beträfe eine solche grundsätzliche Reflexion die religiöse und glaubensbezogene Kommunikation zu den Themen von Gottesdienst, Seelsorge, Diakonie und Bildung. Andererseits kämen all jene Kommunikationsformate hinzu, in denen AI und hier gerade ihr Teilbereich, die Robotik, als sozusagen freundliche Helfer:innen fungieren könnten. Synergien im Bereich der internen Kommunikation sowie im Bereich der Sekretariatsarbeit, so lässt sich mit Smith weiterdenken, entlasten Personen in diesem Aufgabenfeld und eröffnen Freiräume für eine individuellere Kommunikation mit den Gemeindegmitgliedern oder für die an der Gemeindegarbeit Interessierten. Robotik in diesem weiten Sinn soll somit Pfarrpersonal keinesfalls ersetzen, sondern diesem helfen: „Can robots aid the ministry of pastors and priests?“²⁵ Entgegen vieler Stimmen, die „embodied AI“ favorisieren, fokussiert Smith gerade auf eine „disembodied AI“: „I want to suggest here that there is room in the Christian metaphysic and ecclesiology for qualified *disembodied presence*; while it should not be normative, it should not only be *allowed*, but supported.“²⁶

Robotik ist so als „ergänzend“ und nicht als „ersetzend“ anzusehen, d. h. als fester Bestandteil einer hybriden Lebenswelt, in der Glaubensgemeinschaften verortet sind.²⁷ Digitale Tools könnten demnach z. B. eine Komponente im Kampf der Kirchen gegen sexuelle Gewalt in ihren Räumen werden: „invest in facial recognition, machine learning, and AI to monitor and predict potential areas of risk and harm within our leadership structures“.²⁸ Zudem zeigen sich Nutzungsmöglichkeiten im Bereich des kirchlichen Finanzcontrollings²⁹ sowie von Beratungs- und Supervisionsmöglichkeiten im

24 *Smith*: Robot Theology.

25 Vgl. *Smith*: Robot Theology, S. 118.

26 Ebd., S. 122.

27 Vgl. ebd., S. 124, 126.

28 Ebd., S. 126.

29 „Many banks are using robotic process automation (RP) to create workflows, fraud detection, audits, compliance, and so on.“ Ebd., S. 127.

Sinne eines „diagnostic mirror for self-reflection“³⁰, der das Zeitmanagement von Pfarrpersonen evaluiert, indem eine effektive Arbeitsweise durch Mustererkennungen ermöglicht wird. Wie gesagt, all diese praktisch-pragmatischen Nutzungen scheinen auf den ersten Blick von rein technischer Natur zu sein. Tatsächlich aber ist hier immer auch zu bedenken, wie sich dies auf das pastorale Selbstverständnis und die damit verbundene Beziehungspflege im eigenen Amtskontext auswirken könnte.

2.3 Zur Frage einer Robotik-gestützten religionsbezogenen Kommunikationspraxis

Die bisherige praktisch-theologische Auseinandersetzung mit Human Computer Interaction bzw. Human Robotic Interaction bezieht sich eindeutig auf embodied AI. Seit 2017 werden beispielsweise Experimente mit BlessU-2 in der Öffentlichkeit gemacht. Er ist ein Artefakt, bestehend aus dem Maschinengehäuse einer ATM, Armen und Beinen sowie einem Gesicht, das zur Mimik in der Lage ist. Zu ihm gehört ein Speicher von Segensworten, die in verschiedenen Sprachen angesteuert werden können, sowie Optionsmöglichkeiten, etwa auf die Frage, ob man auf Englisch, Deutsch oder Französisch gesegnet werden möchte. Grundsätzlich soll diese Installation einen Roboter im Kontext christlicher Kommunikation zunächst einmal nur simulieren. Es handelt sich folglich um eine Installation, die zum einen die Techniknutzung im Kontext Religion, zum anderen religiöse kommunikative Praxis selbst – in diesem Fall des Segens und Gesegnetwerdens – thematisiert. Hier zeigen entsprechende Untersuchungen zur Wahrnehmung dieses Roboter-Simulators, dass dieser die Menschen offenbar mit einem großen Antwortspektrum von negativ bis in diesem Sample mehrheitlich positiv berührt.³¹ Auch für den Einsatz im Züricher Museum für Gestaltung im Jahr 2022 zeigt sich dies. Auf die Evaluationsfrage „Wie wirkt die segnende Maschine auf Sie?“ antworteten die Proband:innen wiederum mit einer enormen Bandbreite an Resonanzen: Von klarer Ablehnung, weil man das Segnen nur Menschen zutraut, bis hin zu „creepy“;

30 Ebd., S. 128.

31 Löffler et al.: Blessing Robot BlessU2, ; Luthé: Segensroboter „BlessU-2“, S. 111 f.

„witzig“ oder auch „Der Segensroboter wirkt sich positiv auf mich aus, ich lese jeden Morgen den Psalm.“³² Versteht man die analysierten Aussagen als Stellungnahmen zu einer soeben gemachten religiösen Erfahrung, lassen sie sich als „Lived Theology“³³ von Laien lesen. Es zeigen sich dabei fünf – zugleich in unterschiedlicher Weise beziehungsorientierte – Kategorien, die die praktisch-theologische Reflexion hier beispielhaft zur Segenthematik herausfordern:

- 1) Religiöse Identität: Welche existenziellen Bedürfnisse verbinden Menschen mit Segen? Welche Erfahrungen haben Menschen mit Segen und Segenshandlungen?
- 2) Religiöse Autorität: Wer ist berechtigt, Segen zu spenden? Wer darf Menschen segnen?
- 3) Religiöse Wahrheit: Kann Gott durch Technologien wie Roboter und AI segnen? Ist der Segen an menschliche Körper bzw. an Leiblichkeit gebunden?
- 4) Religiöse Authentizität: Wie muss Technologie gestaltet sein, damit einem Roboter spirituelle Kompetenz zugesprochen wird? Wie kann der Gebrauch digitaler Technologie im Rezeptionsprozess von religiösen Phänomenen und Praxen, hier Segenshandlungen, beschrieben werden?
- 5) Religiöse Praxis in gesellschaftlichen Öffentlichkeiten: Kann der Roboter Möglichkeiten für Segenspraxen anbieten, die eher kirchlich-distanzierte Menschen ansprechen? Bietet eine solche Technologie missionarische oder gemeindegbildende Chancen?³⁴

Ein solcher Prozess des Designens von religionsbezogener Technik und der dahinterstehenden technologischen Entwicklungen wirft durchaus ambivalente Implikationen für die Religionspraxis und die Ziele, die dadurch verfolgt werden, auf. Dies ist exemplarisch gut an dem Versuch zu sehen, eine Taxonomie für „theomorphic robots“³⁵ zu entwickeln. Dafür werden etwa von G. Trovato et al. robotikähnliche Medienproduktionen aus der Geschichte der Weltreligionen

32 Die Besucher:innen des Zürcher Museums hatten die Möglichkeit, aus einer Auswahl von Postkarten mit unterschiedlichen Fragen eine auszuwählen und auf deren Rückseite darauf schriftlich zu antworten sowie weitere Angaben zu ihrer Person zu machen.

33 Vgl. Müller: *Religious Experience and Its Transformational Power*, S. 207–224.

34 Vgl. Luthé et al.: *Segensroboter „BlessU-2“*.

35 Vgl. Trovato et al.: *Religion and robots*.

und der Technik herangezogen und mindestens drei verschiedene Gestaltungsweisen von Robotern im Feld von Religionen unterschieden: anthropomorphe, zoomorphe sowie funktionale Designs. Robotik soll deren Ansicht nach Repräsentationen des Göttlichen entstehen lassen, die Menschen – wie bereits in der katholischen Kirche seit Jahrhunderten praktiziert – zum Staunen und in einer auch magischen Weise in Berührung mit Gott bringen sollen.

The fact that a theomorphic robot is connected with some divine or has some supernatural capabilities may make the user feel protected by the robot. [...] At the same time, the robot itself may be taken in higher consideration and respect, to another degree compared to the simple politeness high-lighted in the experiments described in The Media Equation.³⁶

Das Göttliche erhält hier ein eigenes Design, so wie es beispielsweise aus einer spezifischen Sakralbauarchitektur und ihrer Kunst heraus abgeleitet werden könnte.

Aus einer theologischen Perspektive, wonach der Geist Gottes die gesamte Schöpfung beseelt, erscheint eine solche Taxonomie, die das Göttliche getrennt vom Menschlichen, Tierischen und Sachlichen thematisiert, allerdings als fragwürdig und unterkomplex. Denn in der Welt, Mensch, Tier und materielle Kultur, sozusagen in klassischer dualistischer Weise, von Gott geschieden werden, geraten dann die Ambivalenzen religiöser Darstellungen des Göttlichen aus dem Blick.³⁷ Es tritt die doppelte Gefahr auf, dass zum einen die entwickelten religiösen Artefakte einen Status von Heiligkeit erhalten, bei dem ihre Zweideutigkeiten als menschliche Schöpfungen nicht mehr kommuniziert werden und zum zweiten, dass religiöse Handlungen auf problematische Weise durch AI prozessiert werden sollen.

Für den gegenwärtigen Stand der Debatte und die vertiefende Reflexion zu den oben genannten fünf Kategorien sind deshalb kriteriologische Einordnungen der US-amerikanischen Kommunikationswissenschaftlerin Pauline Hope Cheong hilfreich: Ein wichtiges Kriterium für religiöse Robotik liegt für sie in deren kommunikativen Implikationen für spirituelles Wachstum und Engagement. So

36 Ebd., S. 549.

37 Zu einer ausführlicheren Kritik am Ansatz von Trovato vgl. Nord: Robotics in Christian Religious Practice.

reflektiert sie Frömmigkeit sowie Religionspraktiken und ihre pädagogische Bedeutung im Horizont von AI.³⁸ Sie greift die vorab erwähnten Experimente in Japan und Deutschland auf, fokussiert also auch Android Robots und identifiziert dadurch neue Fragen für eine zukünftige Religionspraxis. Sie versteht Robots ebenfalls als „embodied communicative agents and digital media objects embedded within intricate social and material landscapes“³⁹. Cheong spricht in der Absicht, die Perspektive im Feld zu weiten, anstatt von religious robots von „AI devices designed as communicators“. Sie sieht mit Trovato et. al., dass Roboter religiöse Entitäten und auch heilige Dinge sein können – mit anderen Worten, dass sie beziehungsorientierte kommunikative Partner:innen sein können.⁴⁰ Weitere Experimente seien denkbar, wie etwa, dass eine Gruppe von Robotern miteinander durch Gesten und Sprache kommunizieren, sowohl in der sakralen wie auch in der säkularen Öffentlichkeit. So fragt Cheong zu Recht, ob diese „Devices“ in der Lage dazu wären, hierbei auch Religiosität weiterzuentwickeln? Dafür seien zunächst Versuche aus asiatischen Kontexten zur Kenntnis zu nehmen. Min-dar und Xian'er seien teilweise auch unter Mitarbeit von Tempel-Angestellten und Mönchen designt worden, um als spirituelle Agent:innen heilige Texte zu kommunizieren und zwar in einer Weise, die für zeitgenössische und digital verbundene Öffentlichkeiten attraktiv seien.⁴¹

This strand of Buddhism advocates harmony between scientific knowledge and rationality with faith beliefs to advance Buddhism wisdom for the secular world to ‚guide the upcoming era of artificial intelligence onto a healthy road that leads to spiritual insight‘ (Voice of Longquan, 2017).⁴²

Zu einer eher katechetischen tritt hier somit eine rituelle Funktion religiöser Robotik hinzu, was -2 wie ebenso der Roboter Pepper, der für buddhistische Beerdigungsriten entwickelt wurde, verdeutlichen. Für die zukünftige Forschung im Bereich religiöser Roboter

38 Vgl. Cheong: Robots, Religion and communication.

39 Ebd., S. 87.

40 Diese Perspektive wird auch im engeren Sinne theologisch gedeutet bei Nord: The Appreciation of „Things“.

41 Vgl. Cheong: Robots, Religion and communication, S. 87.

42 Ebd., S. 90.

und der Mensch-Maschine-Interaktionen bedeutet dies Cheong zufolge: „First, as humanoid robots permeate every day and religious domains, their work invites an extended probe into ways in which they augment, supplement or substitute the work of clergy and other religious workers.“⁴³

Wie sehr diese ersten Experimente offene Fragen hinterlassen, zeigen ferner Cheongs, auf die konkrete Arbeit an Religious Robots fokussierende, Ausblicke:

Therefore, further research is needed to clarify the ways in which AI agents are applied to fulfil religious duties and expectations, as well as lay believers' perceptions of the legitimacy of robotic agents to serve as religious personnel, particularly in conditions where customary criteria exists and barriers for admission into the clerical profession are high.⁴⁴

Sie greift ebenfalls die Forschungslücke bezüglich des Zusammenhangs zwischen Einstellungen und Nutzungen von Mensch-Roboter-Kommunikation auf:

Further investigation of user emotions and perceptions of robots in discursive exchanges and experimental settings will clarify meaning-making processes in religious interactions to help develop design guidelines for different user preferences and contexts.⁴⁵

3. Praktisch-theologische Folgerungen

Von den erfolgten Einordnungen aus sollen nun praktisch-theologische Überlegungen für die Bereiche pastoraler Amts- und Gottesdienstpraxis, Seelsorge, Bildung sowie zu diversitätssensibler Kommunikation angestellt werden.

3.1 Robotics und pastorale Amtspraxis

Es ist letztlich eine Frage nach der Gestaltung kirchlicher Kommunikationskultur, ob z. B. pastorale Bots in der kirchlichen Praxis eingesetzt werden sollen. In vielen Kontexten gehört die persönliche

43 Ebd., S. 92.

44 Ebd., S. 93.

45 Ebd., S. 93.

und nicht digital vermittelte Präsenz traditionell zur Repräsentation des Pfarramtes. Bleibt dies so oder wird sich dies mit den Transformationsprozessen, die sich insgesamt in säkularen Kommunikations- und Dienstleistungsformen abzeichnen, verändern? Im Zuge von technologischen Weiterentwicklungen einerseits und konkreten Sparmaßnahmen andererseits wird sowohl an entscheidenden kirchenleitenden Stellen wie auch beim pastoralen Personal bereits über die Potenziale AI gestützter Kommunikation nachgedacht. Zwei Beispiele für solche, am Horizont sich abzeichnenden Transformationsdynamiken seien hier kurz genannt:

3.1.1 Robotik im Bereich der Gottesdienstkultur

Der aufsehenerregende Versuch eines gänzlich ChatGPT-generierten Gottesdienstes, wie auf dem Deutschen Evangelischen Kirchentag in Nürnberg im Jahr 2023 geschehen, stellt ein Experiment dar, Robotik im Bereich der Gottesdienstgestaltung zu nutzen. Für das Pfarrpersonal und das traditionelle Selbstverständnis von Gottesdienstgestaltung verändert sich viel: Es entsteht die Möglichkeit für die Gottesdienst feiernden Personen, individuell eigene Gottesdienstabläufe zu gestalten und bestimmte Elemente in ihm auszuwählen, andere auszulassen oder zu überspringen. Wie Pfarrpersonen auf dieser Basis Gemeinschaft im Gottesdienstfeiern in einer Robotic-gestützten Form fördern können, muss noch genauer beschrieben, analysiert und entwickelt werden. Soweit zu sehen ist, ist die Gefahr hoch, auch mit AI und einem Roboter sozusagen gemeinsam allein zu sein. Doch ist sie in jedem Fall höher als im herkömmlichen Gottesdienst, der im Bereich der evangelischen Traditionen auch die Bedeutung des individuellen Gottesdienstbesuches ohne direkte Aufnahme von Kontakt zu anderen Beteiligten kennt? Ist es sinnvoll diese Frage überhaupt hier zu stellen, denn geht es um Gemeinschaft beim Gottesdienstfeiern oder könnten andere Aspekte wichtiger als dieser sein? Neben dem gerade im digitalen Raum wichtigen Community-Building steht traditionell aber auch die einzelne Person im Fokus des Gottesdienstes sowie ihr gläubiges Selbstbewusstsein. Dieses bildet sich – so wurde holzschnittartig gesagt – im Gottesdienst im Austausch mit der liturgischen Person, die Gottes Gegenwart durch eigene Symbolisierungsleistungen zur Sprache bringt bzw. dieser Ereignishaftigkeit Raum und Zeit eröff-

net, was dann zugleich gemeinschaftsstiftende Dynamiken auslösen kann. Diese Symbolisierungs- und Gemeinschaftsstiftungsfunktion könnte nun auf spezifische Weise von AI zumindest teilweise „übernommen“ werden: Wäre dies für die religiöse Bildung im Gottesdienst förderlich und wenn ja in welcher Hinsicht? Ferner: Welche Bedeutung hat die Erfahrung praktizierter Gemeinschaft am selben Ort zur selben Zeit von leiblich gemeinsam anwesenden Personen, um genau jenes gläubige Selbstbewusstsein zu stärken? Die Frage nach dem Verhältnis von AI und Embodiment kehrt im Kontext des Gottesdienstes somit in vielfältiger Weise wieder und verlangt nach weiterer theologischer Reflexion. Dies gilt nicht zuletzt in Bezug auf die Verwendung von gottesdienstbezogenen Daten: Eine große Datenmenge von bereits gefeierten Gottesdiensten kann jeweils kontext- und sprachspezifisch für AI-Gottesdienste genutzt werden. Allerdings entsteht zugleich die Frage nach der Herkunft der Daten, dem Datenschutz und der Transparenz ihrer Verarbeitung innerhalb einer gottesdienstbezogenen AI.

Produktiv für eine zukünftige Gottesdienstkultur ist dabei eine weitere, zugegebenermaßen (noch) etwas hypothetische Vision: Der Gottesdienst kann unter diesen Bedingungen von einer Form des „one fits all“ zu einem individualisierten und personalisierten Design transformiert werden. Damit ließe sich der Leitfigur des Priestertums aller Gläubigen gleichsam eine eigene technisch induzierte Ausdruck- und Teilhabegestalt geben. Zugleich könnte Zielgruppenspezifität in einem völlig neuen Maße zu einer Zielperspektive werden, die die vorhandenen realen Bedürfnisse auf Verständlichkeit und geistliche Gegenwart nochmals in einer eben je individuellen Formatierung aufnimmt. Zeit und Ort des Gottesdienstfeierns werden damit maximal flexibel. Seit Jahren wird innerhalb der Praktischen Theologie für eine Pluralität von Gottesdiensten plädiert und während der Pandemie haben die kleinen digitalen Formate brilliert.⁴⁶ Ist es nicht näher liegend zunächst von einer Robotik oder weiter gesprochen von einer AI gestützten Andachtskultur aus zu arbeiten? Auch BlessU-2 zeigte hier seine Stärken: er bietet eine kurze Zeit an, in der ich mich segnen lassen kann. Aber: Bislang zeigen sich auch im digital-affineren asiatischen Kontext kaum Experimente, auf die aufgebaut werden könnte. Man lässt Roboter hier in einem

46 Vgl. Schlag et al.: Churches Online in Times of Corona.

Gottesdienst heilige Texte rezitieren, aber eben nicht neue, für „in meinem Lebenshorizont heute“ aussagekräftige Gebete und liturgische Stücke gestalten, einen allein durch AI-generierten Gottesdienst bzw. religiöse Zeremonie zu feiern, erscheint offensichtlich auch hier unangemessen und religiös riskant. Erwähnenswert ist, dass auch im jüdischen Kontext bereits über den Einsatz von Robotern innerhalb der Gottesdienstkultur reflektiert wurde. Es ging um die Formen der Verschriftlichung von Tora-Texten, speziell von Tora-Rollen.⁴⁷ Von diesem Experiment ging – soweit zu sehen ist – allerdings keine weitere AI und spezifisch auf Robotik bezogene Aktivität aus.

Das größte Potential scheint sich derzeit in der AI bezogenen Kreation neuer geistlicher Musik aufzutun. Die Kirchenmusik böte enorme Reservoirs, die in AI eingespeist zu immer neuen Remix-Varianten und damit zur personalisierbaren geistlichen Musik führen könnten. Dies böte einen enormen Vorteil, weil gerade die Atmosphären, die mit Musik verbunden sind, darüber entscheiden können, wer wie an einem Gottesdienst teilnehmen will oder sich über AI nun einen solchen selbst gestalten möchte. Doch wiederum ist die Datenschutzfrage erheblich.

Das Ergebnis jüngerer Forschungen zu Online-Gottesdiensten weist aus, dass Online-Gottesdienst-Besucher:innen die Interaktivität unter den Anwesenden sowie die Freiheit von der Institution Kirche hoch schätzen.⁴⁸ Würden Kirchen also AI bzw. Robotik gestützte Gottesdienst-Formate sozusagen zur Selbstgestaltung freigeben, würden sie diesem während der Pandemie erhobenen Bedürfnissen nach Eigengestaltung entgegenkommen. Sie würden damit die kontrollierende Autorität über die Gestaltung singulärer Gottesdienste zwar abgeben (sicherlich sind fundamentalisierende Gottesdienstformate damit nicht zu verhindern), aber sie gewinnen viel dadurch: Sie würden es Menschen ermöglichen, mit Quellen und Dokumenten christlicher Religiosität selbst umzugehen, was sicherlich als eine Urform christlicher Religionspraxis gelten darf. Die Verantwortung hierfür dürften sie sicherlich abgeben. Aber sie könnten natürlich ein exemplarisches Role-Modeling für aus ihrer Sicht angemessene Umgangsformen mit AI und robotergestützte Gottesdienst- oder Andachtsformate initiieren und weiter solche vielleicht weniger

47 Vgl. *Portnoy*: Buch mit Seele.

48 Vgl. *Wolf*: Spirituality at the Breakfast Table.

zufällig und exemplarisch, sondern kommunikativ durchdacht und strategisch platziert in die Welt setzen.

3.1.2 Online- und algorithmisch erstellte Predigten

Die Bereitstellung von Online-Predigten als digital bereitgestellten Predigthilfen ist nicht neu. Hier werden im deutschen Sprachraum, etwa durch die sogenannten „Göttinger Predigten im Internet“,⁴⁹ schon seit mehreren Jahrzehnten Predigten und ganze Gottesdienst-abläufe im Netz angeboten, die von erfahrenen Prediger:innen zur Verfügung gestellt werden. Diese sollen dem pastoralen Personal dazu dienen, sich für ihre eigene Textauslegung und Predigt zu orientieren, etwa durch die zur Verfügung gestellten exegetischen Überlegungen, den theologischen Argumentationsgang oder einfach auch durch bestimmte Konkretisierungsbeispiele zur jeweiligen biblischen Perikope, über die gepredigt werden soll. Im Blick auf den angemessenen Gebrauch dieser Online-Predigtangebote stellt sich natürlich zum einen die generelle Frage, inwiefern diese die kreative Abfassung von Predigten zu inspirieren und zu erleichtern vermögen. Aber in einem viel weiterreichenden Sinn ist theologisch zu fragen, ob die Bezugnahme auf solche digital greifbaren Predigten – von der kompletten „Übernahme“ im Sinn eines „Copy and Paste“ einmal ganz zu schweigen – möglicherweise schlechterdings mit der pastoralen Verkündigungsaufgabe und -verantwortung kollidiert. Denn predigen heißt zumindest im christlichen Selbstverständnis, dass eine ordinierte bzw. dazu autorisierte und beauftragte, theologisch gebildete Person das Wort Gottes lebendig auslegt. Es macht die entscheidende Verantwortung der predigenden Person aus, sich nach allen Regeln theologischer und homiletischer Kunst und durchaus auch unter „Anrufung des Heiligen Geistes“ gewissenhaft auf den Predigtakt vorbereitet zu haben und in der Predigtsituation selbst dem Wort Gottes das Entscheidende zuzutrauen. Insofern ist eine solche Inspiration durch digitale Formate zumindest kritisch zu prüfen. Allerdings steht diese Möglichkeit, auf Online-Predigten zuzugreifen, immer noch in der klassischen Tradition von Predigthilfen, höchstens mit dem Unterschied, dass solche digitalen Formate eben den schnellen und leichten Zugriff ermöglichen.

49 Vgl. *Göttinger Predigten im Internet*.

Komplizierter und auch homiletisch herausfordernder wird es, wenn Predigten selbst algorithmisch generiert werden, wie dies nun durch Programme wie ChatGPT möglich geworden ist. In diesem Fall wird die Predigt, wie bereits in Bezug auf den Gottesdienst gesagt, aus Daten erstellt, die nicht mehr durch die Autorschaft einer Einzelperson geprägt sind. Ein Algorithmus ersetzt also die Aufgabe der Prediger:innen, die die Auslegung eines biblischen Textes vornehmen. Technisch gesehen liegen einer solchen Textgenerierung viele Autor:innen, Quellen und Dokumente zugrunde, die zugleich allerdings anonym bzw. unbekannt bleiben. Damit stellt sich die Frage der Predigt-Verantwortlichkeit nun in qualitativ neuem Sinn.

Zu dieser Problematik kommt aber eine weitere pastoraltheologische und ekklesiologische Herausforderung durch die Predigt- und Gottesdienstsituation selbst hinzu. Denn im Fall des Verkündigungsaktes der AI-generierten Predigt stellt sich die Frage, inwiefern auf Seiten des:r Prediger:in Kontextsensibilität und auf Seiten der Hörer:innenschaft Responsivität gegeben ist. Anhand der oben prinzipiell erörterten Frage nach der Beziehungshaftigkeit AI-generierter Kommunikation lassen sich dann in konkretisierendem Sinn wesentliche theologisch begründete Qualitätsfragen stellen: Wird den Gottesdienstbesucher:innen eine Predigt sozusagen ohne Berücksichtigung der konkreten Verkündigungssituation und der mit ihr verbundenen kommunikativen Beziehungsdimensionen offeriert? Welche Rolle bliebe dann aber noch der theologisch kompetenten und für die Deutung von Glaubensfragen verantwortlichen Pfarrperson, um das Evangelium im Angesicht der Hörer:innenschaft zeitbezogen und wirklichkeitsnah auszulegen? Es stellt sich radikal zugespitzt also die Frage, ob durch das Vortragen oder Hören einer AI-generierten Predigt ein gemeinsames Verkündigungsereignis, das ein Beziehungsereignis par excellence ist, erlebt werden kann? Hier wird zukünftig intensiv homiletisch zu reflektieren sein, mit welchen Motiven von wem und in welcher Weise auf digitale Angebote zurückgegriffen wird und welche theologischen Standards für eine sachgemäße und lebensdienliche Verkündigungspraxis gelten sollen.

3.2 Robotik in der Seelsorge

Die theologisch begründete Frage nach dem Selbstverständnis und der Qualität des pastoralen Amtshandelns stellt sich angesichts der

Möglichkeiten von AI auch in der seelsorgerlichen Praxis, denn auch hier stehen Fragen der Beziehungsstiftung elementar und existenziell auf dem Prüfstand. Bereits seit Jahren werden verschiedene Formate digitaler Seelsorge-Praxen erprobt, die ihre je eigenen Herausforderungen und Aufgabenstellungen erzeugen. Zumeist sind sie nicht im engeren Sinne bereits mit AI oder konkret Robotik verbunden. Aber sie bieten das Potential, über AI fortentwickelt zu werden, was wiederum eine individuellere bzw. differenziertere Usability ermöglicht:

- a) Spiritual Care im Kontext von „Telechaplaincy“: Die am meisten fortgeschrittene Variante digitaler Krankenhausseelsorge dürfte derzeit die spirituelle Betreuung im Rahmen der im englischsprachigen Raum als „Telehealth“ bekannten Entwicklung sein. In diesem Kontext werden digitale Spital-Plattformen entwickelt, die je nach Entwicklungsstand elektronische Patient:innenakten beinhalten sowie Funktionen, mit denen die Diagnose, Beobachtung und umfassende Betreuung von Patient:innen außerhalb des physischen Ortes des Spitals ermöglicht wird. Hierbei handelt es sich um Formen der virtuellen bzw. virtuell-hybriden Begegnung des klinischen Teams mit Patient:innen und deren Angehörigen. Insbesondere im US-amerikanischen Kontext wird dieses Modell vermehrt in der medizinischen Versorgung in ländlichen Gebieten angewendet. In einzelnen Fällen, etwa dem „Mercy Virtual“ in St. Louis, Missouri, werden Patient:innen ausschließlich per Telehealth-Infrastruktur in ihrem Eigenheim betreut, so dass das Spital als physischer Ort keine Patient:innenbetten aufweist. In diesen und ähnlichen Kontexten entwickelt sich, was in der Literatur als „Telechaplaincy 2.0“ bezeichnet wird: eine Spiritual Care, die – weit über herkömmliche Methoden der Telefonseelsorge hinausgehend – ein hybrides oder gänzlich digitales Modell der spirituellen Begleitung als Ausgangspunkt voraussetzt. Dazu wird Künstliche Intelligenz insbesondere in grossen Spitalnetzwerken mit „Big Data“-Ansätzen verbunden, um Risikopatient:innen in bestimmten Bevölkerungsgruppen vorbeugend zu identifizieren und präventiv in das Gesundheitssystem einzubinden. Vereinzelt finden auch in diesem Kontext Versuche statt, spirituelle Bedürfnisse vorausbl-

ckend und auf Bevölkerungsebene zu berechnen und präventiv zu begegnen.⁵⁰

- b) Spiritual-Care-Apps: In diesem Bereich werden interaktive, auch multilinguale spirituelle Kommunikationen z. B. zwischen ärztlichem Personal und Patient:innen möglich, die nicht sprechen können oder keine gemeinsame Sprache teilen. Man denke hier etwa an die Spiritual Care™-App aus dem Apple Store. Solche zumeist aus der Medizin heraus entwickelten Tools wären im Blick auf Religiosität bzw. Spiritualität in der weiteren Entwicklung zu spezifizieren. Eine Möglichkeit wäre z. B., dass verschiedene Personen bzw. ihre AI-gestützten Charaktere oder andere Formen generativer AI – wie ChatGPT-Apps – für persönliche Seelsorge-Chats zur Verfügung stünden. Aus der Offline-Praxis ist bekannt, dass insbesondere präzise und elementare Fragen dazu verhelfen, Menschen in ihrem Prozess, das Leben wieder in ihre Hände zu nehmen, unterstützen können.
- c) Virtual-Reality-Encounters: Aus dem koreanischen Kontext ist ein eindrückliches Experiment einer virtuellen Begegnung zwischen einer trauernden Mutter und ihrer verstorbenen Tochter bekannt.⁵¹ Die maximal immersive Begegnung zwischen beiden ist innerhalb einer Virtual Reality durch AI erzeugt worden. Das Eintauchen in eine solche Interaktion mit der verstorbenen Person löst offenkundig einen Trauerschub aus, wie er auch bei filmischen oder fotografischen Darstellungen aber eben weniger immersiv auftritt. Wie eine solche Erfahrung produktiv in eine Trauerbegleitung eingeordnet werden kann und wie die Koordinaten für eine verantwortungsvoll konzipierte VR-Erfahrung aussehen könnten, ist eine erhebliche seelsorgerliche Herausforderung.
- d) Pflegeroboter: Es war der japanische Paro, ein Pet-Type Roboter in Gestalt einer kleinen Robbe, der für die leibkörperliche Kommunikation älterer Menschen entwickelt wurde und psychologisch deren Wohlbefinden verbessern konnte. Spiel- und

50 Weiterführende Informationen finden sich auf www.telechaplaincy.io, einem Informationsportal für digitale Seelsorge, das von der Professur für Spiritual Care der UZH im Rahmen des URPP «Digital Religion(s)» betrieben wird; vgl. auch *Winiger: The changing face of spiritual care.*

51 Vgl. *ABC News: Mother ,reunited' with deceased daughter in virtual reality.*

Unterhaltungsobjekte sind in diesem Kontext unter Anleitung eines Erwachsenenbildners bereits eingeübt, Robotic basierte Maßnahmen aber noch nicht in der Breite verfügbar bzw. akzeptiert. Wiederum ist zu erforschen, welche Einstellungen und Erfordernisse, d. h. welche Usability wünschenswert ist und verantwortet werden kann. Dass Robots, wie Paro, das Problem der Einsamkeit älterer Menschen in Heimen lösen könnte, ist nicht zu erwarten. Dass er dazu verhilft, dass ältere Menschen ihre Sensualität weiterhin trainieren, hingegen schon – was wiederum von Seiten der Pflegenden erhebliche Beziehungsfähigkeit voraussetzt. Eine noch weitergehende grundsätzliche Frage stellt sich im Bereich der Seelsorge dadurch, dass religiöse Roboter – jedenfalls bisher nicht – über die Vorstellung von menschlicher Endlichkeit verfügen. So wird mit Recht gefragt: „How authentic or compelling might be the advice, counsel, or attempted comfort of a device that does not share our finality, in contrast with, say, a priest or a minister?“⁵² Nicht erst mit dieser letzten Frage wird die theologische Frage nach dem tieferen Sinn einer dem Menschen dienlichen Beziehungspflege und damit nach dem göttlichen Beziehungs- und Geschenkcharakter einer solchen Seelsorgepraxis zum entscheidenden Thema. Es kann bei aller Offenheit für eine solche AI-gestützte Praxis jedenfalls wiederum hier nicht darauf abgezielt werden, diese Beziehungsdynamik durch maschinelle Formen der Interaktion ersetzen zu wollen. Denn dies würde dann tatsächlich den ureigenen und spezifischen Charakter einer glaubensgegründeten Annahme und Anerkennung des je Anderen, die Dynamik eines echten, wahrnehmend-empathischen „von Angesicht zu Angesicht“ und zugleich die Hoffnung auf die Gegenwart Gottes in der jeweiligen augenblicklichen Situation fundamental in Frage stellen.

52 *Balle*: Robots, Ethics, and Digital Religion, S. 46.

3.3 Robotik in der religiösen Bildung

Im Bereich religiöser, digitaler Bildung⁵³ ist sowohl an Robotik für Erwachsene wie für Kinder und Jugendliche zu denken. Dies gilt sowohl im Kontext formaler schulischer Bildung, wie im Bereich non-formaler oder informeller Kontexte, insbesondere kirchlicher und christlich orientierter Bildung. Noch ist ein „In-Home Learning Companion Roboter“ für ‚religious literacy‘ nicht entwickelt worden, aber es liegen eine Vielzahl von schulbezogenen Projekten zur STEM-Education (Science, Technology, Engineering and Mathematics)⁵⁴ vor, an die hierfür interdisziplinär angeschlossen werden kann. Ein Learning-Companion bietet dabei die Möglichkeit, dass Lernausgangslagen mit der sich „fortwährend bildenden“ AI des Companions erkundet und verarbeitet werden können und Lernprozesse so eine hohe Passung erhalten können. Kollaborative Learning-Szenarien sind ebenso virtuell, augmentiert, wie auch in Präsenz in einem gemeinsamen Klassenraum denkbar. Ob der Learning-Companion androide Designs haben sollte, ist dabei eine ebenfalls zu lösende Forschungsaufgabe. Jedenfalls könnten verschiedene Alterssegmente adressiert und kulturelle Kontexte bzw. unterschiedliche Religionskulturen in der Gestaltung des personalisierten Religionsassistenten berücksichtigt werden. Hierfür ließe sich auf Experimente mit Lern-Robotern aufbauen, die in Schulen bereits weltweit viel genutzt werden und wofür Schüler:innen also bereits im Kontext schulischer Bildung Kompetenzen aufbauen. Hier traktierte Themen wie etwa das Interest Development können dann auch für das religiöse Lernen wichtig werden, weil z. B. die Head-Eye Motion, die Gaze Aversion, oder die sogenannte „Tailored Recommendation“ Aufmerksamkeitsspannen erhöhen. Zugleich muss allerdings didaktisch sensibel mit Aufmerksamkeitsökonomien umgegangen werden, damit Kinder und Jugendliche nicht emotional überfordert werden. Hier eröffnen sich jedenfalls Robotik-gestützte Möglichkeiten, christliche Tradition kennenzulernen, sich selbst anzueignen und im eigenen Lebenshorizont dienliche Rituale und intellektuelle Einsichten zu entwickeln. Vorstellbar ist, dass dies kei-

53 Vgl. grundsätzlich *Schlag*: Religion, digitale sowie Pirner: Bildung, mediale/digitale.

54 Vgl. STEM-Education.

neswegs individualisiert geschieht, sondern ebenfalls in Interaktionen, die über Zeit und Raum hinweg Kinder und Jugendliche in Beziehung miteinander und mit der Robotik lernen lassen können. So ist ein AI-gestütztes Theologisieren in einer Gruppe mit weiteren virtuellen oder auch simultan anwesenden Personen ebenso denkbar wie an Gaming orientierte Lernprogramme, die den Learning-Companion zu einem liebenswerten Sibling werden lassen kann. Es ist auch davon auszugehen, dass sich – wohl schon sehr bald – durch die Entwicklung des Metaversums noch ganz neue immersive Bildungs- und Beziehungserfahrungen ermöglichen lassen.⁵⁵

Blickt man auf digitale Kulturen, wie etwa beim genannten Gaming, das bereits jetzt gewohnte Bildungstraditionen verändert, wird die Frage nach der Agency von Bildungsbemühungen wichtig: Als Ziel gilt es, Kinder und Jugendliche im Sinne einer Alltagspiritualität fortzubilden, die sie empowert und ihre soziale und kulturelle Partizipation fördert. Dreh- und Angelpunkt ist dann nicht, ob Kinder und Jugendliche die christliche Tradition kennenlernen, sondern ob sie sich diese aneignen, weil sie ihnen guttut und dabei hilft, den eigenen Platz in der Welt zu finden. Diese Zielsetzung beruht ihrerseits auf dem theologischen Fundament der Rede von der unbedingten schützenswerten Würde des Menschen im Licht seiner je individuellen Gottebenbildlichkeit.⁵⁶

Für eine solche Nutzung von Robotik und breiter gesagt AI in der religiösen Bildung wird es insgesamt entscheidend sein, eine religionspädagogisch intensiv auf dieses Ziel hin reflektierte Religionsdidaktik in einer digitalen Kultur zu entwickeln.⁵⁷ Diese setzt bei der Vermittlung von religiöser Bildung nicht auf lexikalisches Wissen, das dekontextualisiert wird und quasi feststehende Wissensgebiete offeriert. Es wird beispielsweise keineswegs für die christliche Bildung angemessen sein, die fünf Säulen des Islam zu vermitteln, wenn man muslimische Religion in ihrer inhaltlichen Bedeutung und Vielfalt in Relation zur christlichen Tradition verstehen möchte. Eine solche Wissensvermittlung tendiert gerade im Bereich religiöser Bildung zur Klischeebildung. Es geht dann insgesamt im Bereich

55 Vgl. Schlag: Bildungsbeziehungen in der Ankunftszeit des Metaversums.

56 Vgl. Schweitzer: Menschenwürde und Bildung sowie Pirner und Nord: Religiöse Bildung in der digitalen Welt.

57 Vgl. Nord: Fachdidaktik Religion.

religiöser Bildung nicht in erster Linie um das Erlernen von religiösen Praxen, sondern um das Verstehen von Welt.

Als Kernkriterium für eine AI-gestützte religiöse Bildung kann die „Kritik an der Selbstalgorithmisierung“⁵⁸ verstanden werden. Sie ist eine Ressource für eine Gegenbewegung, die gerade christliche Bildung fördert, weil sie eine Bildung zur „herrlichen Freiheit der Kinder Gottes“ (Röm 8,21) ist: Es ginge dann auch in der Robotik, die für religionsbezogene Bildung entwickelt wird, darum, sich bei der Nutzung der Robotik dem möglichen eigenen Hang zur Unmündigkeit entgegenzustellen, nach dem Motto „Warum selbst denken, wenn es doch die Maschine kann?“ Was Menschen denken und fühlen können, ist unersetzbar und wann immer Kinder und Jugendliche im Umgang mit Robotik diesen roten Faden festhalten, wird Robotik als ein enormer Möglichkeitsraum für ihre Bildungsprozesse begriffen werden können.

3.4 Diversitätssensible kirchliche Kommunikation

Die Zukunft der Kirche ist eine beziehungsfähige, inklusive Kirche, die selbstverständlich Teil digitaler Gesellschaften und Kulturen ist. Können Robotik und AI hierbei unterstützen? Immer, wenn neue Technologien und Techniken erfunden wurden, wuchs die Hoffnung, dass sie für diskriminierte Personengruppen hilfreich werden könnten. AI und Robotik haben hierfür enormes Potential. Bereits der oben genannte In-Home-Companion zeigt dies an. Im Gästebuch von BlessU-2 wurde ebenfalls deutlich, dass hier z. B. sehbehinderte Personen den Eindruck hatten, dass sie die Segensgäste für sich als wohltuender erfahren konnten als von einem Menschen, der in seinem Habitus und mit seiner Mimesis immer schon die Mehrheitsgesellschaft repräsentiert und kaum nicht-hegemoniale Praktiken zulässt oder kennt. Robotik wird das Problem der Diskriminierung von nicht der Mehrheitsgesellschaft angehörenden Personen in der Kirche nicht beseitigen können, indem sozusagen die Robotik ersetzt, was Menschen für eine inklusivere Kirche nicht zu tun vermögen oder bereit sind. Inklusion entsteht jedenfalls nicht schon dadurch, dass Gehörlose im Gottesdienst eine für sie passen-

58 *Bermes*: Wer hat Angst vor ChatGPT.

de Technik vorfinden. Erst wenn die Robotik dazu eingesetzt wird, dass z. B. gehörlose und nichtgehörlose Menschen sich barrierefrei austauschen können, wird eine für die inklusive Kirche wichtige Dimension erreicht werden. Und dafür ist eine humane Kommunikation, der es gelingt, Sprach- und Kommunikationsbarrieren abzubauen, wesentlich. Hiervon sind wir technisch noch weit entfernt. Doch eine Robotik, die Sprachübersetzungen simultan erzeugen könnte, vermag zweifellos positiv zur Veränderung kirchlicher Kommunikations- und Solidaritätskultur beizutragen. Dies gilt auch für weitere Diversitätsorientierungen, z. B. in Bezug auf Rassismus und Antisemitismus, Gender und sozialen Status. Auch die Teilhabe an AI und Robotik ist sozial segregiert und segregierend. Sollte es etwa im kirchlichen Kontext so sein, dass nur diejenigen mit Robotik experimentieren oder deren Chancen nutzen können, die dazu die finanziellen Möglichkeiten haben? Werden ärmere Gemeinden in Zukunft ausschließlich analoge Gemeinden sein? Insofern ist auch die Frage nach der Diversitätssensibilität und der damit verbundene mögliche KI-gestützte Beitrag dazu in einem sehr viel weiteren Horizont zu betrachten. Denn auch hier kommen grundlegende theologische Fragen nach dem ekklesiologischen Selbstverständnis – sowohl in lokaler wie in globaler Hinsicht – mit in den Blick, die jedenfalls weit über die Frage der technischen Möglichkeiten hinaus Aspekte der Inklusion und Solidarität christlichen Gemeinschaftshandelns aufwerfen.

4. Fazit: Religiöse Kommunikation in einer Kultur der Digitalität

Die Auseinandersetzung mit Robotik im Kontext kirchlicher Praxis ist dann aussichtsreich, wenn nicht allein Technikimplementierungen intendiert, sondern individuelle, soziale und interaktive Beziehungs- und Verständigungsprozesse initiiert werden. Es geht dabei nicht darum, alten Wein in neue Schläuche zu füllen, sondern die technologischen Entwicklungen und deren mögliche Implikationen für die religiöse Kommunikationspraxis von einer theologisch verantworteten Selbstpositionierung aus zum einen ernst zu nehmen und zum anderen deren Innovationspotenzial gerade nicht zu unterschätzen:

Thus, robotic technology should not lead to replacing authentic face-to-face encounters in religious contexts with technology, but instead be used to extend the experience in ways only the technology can do.⁵⁹

Tatsächlich spricht im Blick auf die gesamte Technologiesgeschichte nichts dafür, dass nun ausgerechnet KI und Robotik in prinzipiell technikskeptischer Weise die Möglichkeit abgesprochen werden sollte, zu Fragen religiöser Identität, Autorität, Wahrheit, Autonomie sowie Gemeinschaftsbildung und ritueller Erfahrung in konstruktivem Sinn beizutragen. Dies setzt allerdings zugleich voraus, Religionspraxen nicht als feststehende Entitäten zu behandeln, sondern als traditionsreiche und zugleich immer wieder innovative Ausdrucksformen gelebter Religiosität.

Vorauszusetzen ist allerdings zugleich, dass den Entwickler:innen selbst das notwendige Vertrauen entgegengebracht werden kann, dass sie durch diese Innovationen nicht per se auf eine missbräuchliche und indoktrinierende Praxis abzielen, was gerade in religiösen Fragen besonders problematische und gefährliche Konsequenzen haben würde. Die materiale theologische Grundlegung für eine solche kritisch-konstruktive Wahrnehmung technologischer Entwicklungen liegt darin, dass alle Formen solcher Religious Robots und Robotics nicht den Anspruch auf Perfektion oder gar gottgleiche oder gottähnliche Präsenz erheben dürfen: So kann in Orientierung an Paul Tillichs Gedanken der „sacred emptiness“ gesagt werden, dass gerade in solchen „Erscheinungs“-Formen diese Techniken „can paradoxically express the simultaneity of ‚vulnerability and presence‘ and of ‚limitation and fulfilment.“⁶⁰ Und in jedem Fall sollte aller Gebrauch von der theologischen Einsicht getragen bleiben, dass „the absent should be kept present as the absent.“⁶¹

Letztlich liegt es im kritischen Reflexionsbewusstsein der einzelnen Gläubigen, zu beurteilen, was an einer solchen „meaningful robotic“ lebensdienlich ist und was nicht. Das interaktive Testen vorhandener religiöser Praxen kann hier ebenso hinzugehören wie die Weiterentwicklung eigener erfahrungsbezogener und je neu symbolisierender spiritueller Alltagspraxen: Führen diese dazu, dass ich mich in der Welt Gottes vor mir selbst, vor anderen und vor Gott

59 Löffler et al.: Blessing Robot BlessU2, S. 583.

60 Puzio: Robot Theology, S. 108.

61 Ebd. in Orientierung an R. Feiter, S. III.

verantwortungsvoll und zugleich mit Verbundenheit zu mir selbst, zu anderen und zu Gott verstehen kann?

Religiöse oder präziser ausgedrückt religionsbezogene Robotik hat die Chance, dieses Kriterium paradigmatisch für die Gesellschaften, in denen christliche Religion gelebt wird, zu achten, zu entfalten und zu etablieren. Dies kann dann zugleich zu einem Beitrag zur Reflexion des Humanum in einer Welt werden, in der AI und Robotik vor allem für andere Sektoren der Gesellschaft eingesetzt werden, sei es in der Wirtschaft, der Medizin, dem Militär und nicht zuletzt in der Unterhaltungsindustrie.⁶²

Bibliografie

(alle Online-Links sind mit Stand vom 22. August 2024 überprüft)

ABC News: Mother ‚reunited‘ with deceased daughter in virtual reality (13.02.2020). Online abrufbar unter: <https://www.youtube.com/watch?v=NSb6FHzwGdY>

Balle, Simon/Ess, Charles: Robots, Ethics, and Digital Religion: Initial Considerations, in: Heidi A. Campbell and Pauline H. Cheong (Eds.), *The Oxford Handbook of Digital Religion*. Oxford 2022, C27. Online abrufbar unter: <https://academic.oup.com/edited-volume/44728/chapter/378960611>

Berger, Peter L./Luckmann, Thomas: *Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*, Frankfurt a.M. 1977.

Bermes, Christian/Dörpinghaus, Andreas: Wer hat Angst vor ChatGPT?, in: *Frankfurter Allgemeine Zeitung*, 19.04.2023, S. 23.

Campbell, Heidi/Dyer, John: *Ecclesiology for a Digital Church*. London 2022.

Cheong, Pauline H.: Religion, Robots and Rectitude. Communicative Affordances for Spiritual Knowledge and Community. *Applied Artificial Intelligence* 2020, S. 412–431, DOI: <https://doi.org/10.1080/08839514.2020.1723869>

Cheong, Pauline H.: Robots, Religion and Communication: Rethinking Piety, Practices and Pedagogy in the Era of Artificial Intelligence. In: Giulia Isetti, Elisa Innerhofer, Harald Pechlaner & Michael de Rachewiltz (Eds.): *Religion in the Age of Digitalization. From New Media to Spiritual Machines*. New York 2020, S. 86–96.

Crawford, Kate: *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven/London 2021.

62 Vgl. *Evangelische Kirche in Deutschland: Freiheit digital*.

- Dabrock, Peter*: „Prüft aber alles und das Gute behaltet!“. Theologisches und Ethisches zu Künstlicher Intelligenz, in: *Theologische Literaturzeitung* 147 (2022). S. 635–650.
- Dignum, Virginia*: *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*. Cham/Schweiz 2019.
- Dürr, Oliver*: *Homo Novus. Vollendlichkeit im Zeitalter des Transhumanismus. Beiträge zu einer Techniktheologie*. Münster 2021.
- Dürr, Oliver*: *Transhumanismus. Traum oder Alptraum?* Freiburg i.Br. 2023;
- Evangelische Kirche in Deutschland*: *Freiheit digital. Die Zehn Gebote in Zeiten des digitalen Wandels. Eine Denkschrift der Evangelischen Kirche in Deutschland in Deutschland (EKD)*. Leipzig 2021.
- Foerst, Anne*: Cog, A Humanoid Robot, and the Question of the Image of God. in: *Zygon* vol 33, no.1, March 1998, S. 91–111.
- Foerst, Anne*: *God in the Machine. What Robots Teach us about Humanity and God*. New York 2005.
- Göttinger Predigten im Internet*. Online unter: www.theologie.uzh.ch/predigten/
- Gould, Hannah*: Bad Buddhists, Good Robots: Techno-Salvationist Designs for Nirvana. *Journal of Global Buddhism* 21 (2020), S. 277–294.
- Ishiguro, Kazuo*: *Klara and the Sun*. London 2021.
- Kimura, Takeshi*: Masahiro Mori's Buddhist philosophy of robot, in: *Paladyn. Journal of Behavioral Robotics* 9/1 (2018), S. 72–81. <https://doi.org/10.1515/pjbr-2018-0004>
- Löffler, Diana/Hurtienne, Jörn/Nord Ilona* (2019): 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, S. 569–586. DOI: <https://doi.org/10.1007/s12369-019-00558-3>
- Luthe, Swantje/Nord, Ilona/Löffler, Diana/Hurtienne, Jörn*: Segensroboter „BlessU-2“. Forschungsimpulse für die Praktische Theologie angesichts der Entwicklung sozialer Roboter, in: *Pastoraltheologie*, 108(3) (2019), S. 107–123. DOI: <https://doi.org/10.13109/path.2019.108.3.107>
- Masahiro, Mori*: *Introduction to Buddhism*. Tokyo 2003.
- McEwan, Ian*: *Machines like me*. London 2019.
- Müller, Sabrina*: *Religious Experience and its Transformational Power. Qualitative and Hermeneutic Approaches to a Practical Theological Foundational concept*. Berlin/Boston 2023.
- Nord, Ilona/Ess, Charles* in Kooperation mit *Hurtienne, Jörn* und *Schlag, Thomas*: Robotics in Christian Religious Practice Reflections on Initial Experiments in this Field. DOI: <https://doi.org/10.25972/OPUS-30314>
- Nord, Ilona/Petzke, Judith* (Hrsg.): *Religionsdidaktik diversitätsorientiert und digital*. Berlin 2023.

- Nord, Ilona*: The Appreciation of „Things“ in an Ambiguous World: On AI & Religion, in: Heidi Campbell & Pauline Cheong, *Thinking Tools on AI, Religion & Culture*. Digital Religion Publication. Texas University 2023, S. 43–46.
- Nord, Ilona*: The friendly siblings of workhorses and killer robots – Becoming alive through Nonliving, and Feeling Blessed by a Religious Maschine, in: Judith Dörrenbächer, Ronda Ringfort-Felner, Robin Neuhaus, Marc Hassenzahl (Hg.), *Meaningful Futures with Robots – Designing a New Coexistence*. Boca Raton: CRC Press 2023, S. 258–267.
- Pfeifer, Rolf/Iida, Fumiya*: Embodied Artificial Intelligence: Trends and Challenges. In: Fumiya Iida, Rolf Pfeifer, Luc Steels, Yasuo Kuniyoshi (Eds.): *Embodied Artificial Intelligence. Lecture Notes in Computer Science* vol 3139. Berlin/Heidelberg 2004, S. 1–26. https://doi.org/10.1007/978-3-540-27833-7_1
- Pirner, Manfred L./Nord, Ilona*: Religiöse Bildung in der digitalen Welt. Die digitale Transformation im Fokus der Religionspädagogik und -didaktik. DOI: <https://hdl.handle.net/10900/149714>.
- Pirner, Manfred L.*: Bildung, mediale/digitale. 2023. DOI: https://doi.org/10.23768/wirelex.Bildung_mediale.100189
- Portnoy, Elischa*: Buch mit Seele (16.01.2020). Online unter: <https://www.juedische-allgemeine.de/religion/buch-mit-seele/>
- Puzio, Anna*: Robot Theology. On Theological Engagement with Robotics and Religious Robots, in: Anna Puzio, Nicole Kunkel & Hendrik Klinge (Hg.), *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz*. Darmstadt 2023, S. 95–113.
- Reformiert.* – *Videos der evangelisch-reformierten Zeitung*: Roboter «Bles-u-2» segnet Kirchgänger. 29.10.2020. Online unter: <https://www.youtube.com/watch?v=Yfff0KIGlps>
- Schlag, Thomas/Nord, Ilona*: Religion, digitale. 2021. DOI: https://doi.org/10.23768/wirelex.Religion_digitale.200879
- Schlag, Thomas/Yadav, Katharina*: Bildungsbeziehungen in der Ankunftszeit des Metaversums – eine religionspädagogische Sensibilisierung. In: *Zeitschrift für Pädagogik und Theologie* 75 (2023) S. 163–181.
- Schlag, Thomas/Nord, Ilona*: Die Corona-Pandemie und Dynamiken digitaler Innovation und Transformation. Praktisch-theologische Einordnungen und Ausblicke. in: Dies. et al. (Hrsg.), *Churches Online in Times of Corona. Die CONTOC-Studie: Empirische Einsichten, Interpretationen und Perspektiven*. Wiesbaden 2023, S. 521–533.
- Schweitzer, Friedrich*: Menschenwürde und Bildung. Religiöse Voraussetzungen der Pädagogik in evangelischer Perspektive. Zürich 2011.
- Shelley, Mary*: *Frankenstein*. Amazon Classics. Seattle 2017.
- Singer, Isaac Bashevis*: *The Golem. Erzählung* (engl.) deutsch. München 1988.

- Smith, Joshua K.*: Robot Theology. Old Questions through New Media. Oregon/USA 2022.
- STEM-Education*. Online unter: <https://www.stem.org.uk/>
- The Japan Times*: The mind of Mindar, Kodaiji temple's teacher of Buddhism. 12.09.2019. Online unter: https://www.youtube.com/watch?v=hLoF5_-OUKY
- Trovato, Gabriele et al.*: Religion and robots: Towards the synthesis of two extremes. *International Journal of Social Robotics* (2019), S. 1–18.
- Trovato, Gabriele*: SanTO. O.J. Online unter: <http://rhye.civfanatics.net/gabu/robotsSanTO.php>
- Vögeli, Dorothee*: Der Segensroboter ist eine Weltneuheit und steht in Zürich. Online unter: <https://www.nzz.ch/zuerich/religion-und-digitale-welt-theologie-ueber-sinn-von-segensroboter-ld.1677455>
- Winiger, Fabian*: The Changing Face of Spiritual Care. Current Developments in Telechaplancy. *Journal of Health Care Chaplaincy*, 2023, 29(1), S. 114–131. Online unter: <https://doi.org/10.1080/08854726.2022.2040895>
- Wolf, Sara et al.*: Spirituality at the Breakfast Table: Experiences of Christian Online Worship Services. Online unter: <https://doi.org/10.1145/3491101.3519856>

Systematisch-Theologische Perspektiven

Theresa M. Thunig

The Final Frontier

Investigating the Impact of Science Fiction on our (Technological) Future

Abstract

This paper will investigate the relation between science fiction and the real, emerging future. For this purpose, it is necessary to ask what influence works of science fiction can have on reality. This paper will argue in a focused way that fiction can refer to possible future realities, while also calling contemporary events and trends into question. The example of *Star Trek: The Next Generation* as *soft science fiction*, specifically the character of the android Data, will help to illuminate how fiction already reveals problems and questions that are otherwise only partially recognisable or imaginable in the present.

1. Introduction: engaging in the worlds of science fiction

We live in an exciting time, full of technological innovations that seem almost futuristic. Within seconds we have access to almost all the knowledge there is with our smartphones alone. However, for fans of fiction, this world may sometimes seem a bit bleak compared to what we read in books and see on screens. After all, our cars don't fly yet and aren't even fully emission-free. There are countless diseases we can't cure with current medicine. We still have a long way to go. There seems to be an invisible frontier that separates us from the future, the world we imagine but probably won't even experience ourselves.

Perhaps that is the crucial point. We imagine the future and we do so in umpteen different ways. The limits that apply to reality

do not apply to our imagination. In our minds—in literature, film and television—we set out to explore the unexplored, to imagine the unimaginable.

And so, this is all about precisely the frontier between the now and the *maybe someday*. As we will see, this frontier isn't as intransigent as it may seem. It might be worthwhile exploring how science fiction can affect us and our future really and truly. Because in the moments when spaceships fly across the screen, victims of horrific burns are being saved by being transformed into cyborgs, and cars act as time machines, the lines blur. The *maybe someday* is finding its way into the here and now and leaving footprints. Even if quite a few scenarios seem very unrealistic at first glance, the stories of science fiction never fail to captivate us—in the end, they might be part of what the world might be like in the years to come. Suddenly, we are presented with our greatest hopes and most terrifying fears for the future, and we begin to wonder what we want as humanity and how we want to evolve.

Which works inspire us to reflect may vary. However, some of the better-known works in particular can provide us with an opportunity to consider the relevance of science fiction for the present day. In this case, it is *Star Trek*, which offers a number of opportunities for a scientific approach.¹

2. What is science fiction?

If one speaks of science fiction here as a matter of course, the reader must be aware that this genre² is actually quite difficult to define. Although the term evokes a whole range of associations, scientists, authors and fans have never really come to a shared definition.³ One of the earliest attempts to outline the genre can be traced back to the publisher and author Hugo Gernsback, who initially called it

1 See *Bauer*: Utopie, 3.

2 Derived from the Latin word 'genus', a genre refers to a category of texts, movies or similar. See *Kirsten*: Diskursivität, 21. Genre patterns describe the inner characteristics of corresponding works. See *ibid.*, 23. Science fiction is one of the better-known genres.

3 See *Vint*: Guide, 1.

scientifiction in 1926. He recurred to H. G. Wells, Jules Verne and others and pointed out the connection between scientific fact and prophetic vision in their works; later on, he changed the terminology to science fiction.⁴ In the years that followed, many works were published that are considered classics of science fiction. These classics, for example, are the influential American works of the 1940s by Isaac Asimov, Arthur C. Clarke, Ray Bradbury or Robert Heinlein—they have become the gold standard. While elements of the classics are still recycled today, the boundaries of the genre are constantly being challenged. Justifiably, the objection can be raised that some works recently published under the heading of science fiction no longer have much in common with the works that originally brought this genre to prominence.⁵

Science fiction, as a genre and concept, is too multifaceted to fit a single definition. Here, I will not attempt to provide a specific and satisfactory definition of the genre, but instead I will work within the framework provided by P. Schuyler Miller.

It is precisely in the attempt to distinguish science fiction from other genres with similar themes, such as fantasy or horror, that some key overlaps become apparent.⁶ In the realm of fantastic literature, we encounter that which is impossible (or has yet to be made possible). While fantastic elements might appear in multiple genres, such as fantasy or horror, we discover a certain focus in science fiction: scientific and technical speculations and visions that find expression in topics such as space travel, AI and time travel—which essentially corresponds to what Gernsback already stated.⁷ But this does not even begin to fully define the genre, at least not anymore. However, it is important to note: “Dass Science-Fiction nicht vorrangig und manchmal überhaupt nicht um *science* oder Technologie kreist, ist durch die kulturwissenschaftlichen Analysen der letzten Jahrzehnte gezeigt und gut belegt worden [...]. Dennoch stellt ‚Technik‘ stets zumindest eine Rahmenbedingung für all jene

4 See *Latham*: Definitions, 7.

5 See *Vint*: Guide, 2.

6 See *ibid.*

7 See *ibid.*, 3.

Geschichten dar, die überhaupt als Science-Fiction charakterisiert werden können,⁸ as summarised by Christian Schwarke.

So, in science fiction we are repeatedly setting out on a journey into distant worlds with incredible technologies (often set in the future) which sometimes seem more and sometimes less realistic. Of course, it should be clear at this point that there is no such thing as *the* science fiction. This genre itself is already very varied and heterogeneous and becomes more versatile with each appearing work. However, it is worth taking a closer look at the realism of science and logic and how they affect the stories. The aforementioned science fiction author and critic P. Schuyler Miller introduced the term *hard science fiction* for what Gernsback mainly meant by scientificism or science fiction in his time. The idea is that the story itself is as logical and believable as possible, especially in its understanding of contemporary science and technological development, and in terms of what would become technologically or socially possible in the future with the knowledge of the time.⁹

Star Trek—one extremely fascinating interpretation of science fiction which we’ll discuss later on—certainly does not fall into the category of *hard science fiction*, since “everything technical is more or less taken for granted, unless there is a malfunction invented by the authors for narrative purposes”.¹⁰ It is not the aim of Star Trek to appear realistic in its technologies. On the other hand, it’s not that (realistic) technology doesn’t play a role in Star Trek. Especially in comparison with other science fiction series with almost no refer-

8 Schwarke: Traum, 49. I’d like to offer a translation: “The fact that science fiction does not primarily, and sometimes not at all, revolve around *science* or technology has been shown and well documented by the cultural studies analyses of the last few decades [...]. Nevertheless, ‘technology’ always provides at least a framework for all those stories that can be characterised as science fiction at all.”

9 See Samuelson: Modes. Some of the most well-known authors of *hard science fiction* according to what I tried to classify as such might be Robert Heinlein, Arthur C. Clarke, Larry Niven and Andy Weir (as well as some of those already mentioned). If you’d like to see *hard science fiction* on the big screen, *Gattaca* (1997), *The Martian* (2015) or *The Expanse* (2015–2022) might be worth a look (please keep in mind that other people might consider the boundaries of the genre differently). You’ll see that the question of scientific realism is really concrete, although that is not to say that this sub-genre is the only one in which realistic technologies appear.

10 Wenskus: “Soft” Science Fiction, 450.

ence to the realism of futuristic technology like Doctor Who, the difference becomes obvious. Thus, Star Trek does not seem to be a *technical fantasy*¹¹ with no reference to reality, but something like *soft science fiction* and among them probably one of the best-known representatives. Technofiction is simply not the crucial aspect, and so speculation often remains in suspense. It is quite noticeable that *soft science fiction* focuses more on social, political or philosophical issues, for which the science fiction setting is used or is quite speculative in terms of science.¹²

But particularly if we accept the precondition that not every idea must be compatible with what we know of physics for what follows, new horizons open up which are not exclusively bound to what we can comprehend with today's standards. The stories of science fiction themselves are often—though not always—able to creatively challenge our current social, political or ecological realities. Science fiction, then, does not arise out of nothing, but rather makes use of what is known and abstracts it for narrative contexts.

In the end, it has to be said that the genre remains difficult to define. Instead, a way has been presented to assign the works to a spectrum rather than a fixed category. On one side of the spectrum, we find well-conceived, coherent *hard science fiction* in terms of technology. On the other side, we are most likely to find speculative *technical fantasy* with no recognisable real-life expectations. On this spectrum, we find works such as Star Trek, which range between the two extremes; this transitional genre can be subsumed under *soft science fiction*.

11 Wenskus calls the sub-genre under which she subsumes Doctor Who, for example, *technical fantasy*. See *ibid*.

12 See *ibid*. As I have tried to show, this subgenre is somewhat hazier in its delimitation, with the result that a whole range of works can be considered to fit within it. But I'd like to present a few of the well-known authors either way: H. G. Wells, Mary Shelley and Frank Herbert. For those who are looking for some *soft science fiction* to watch, there are a few prominent franchises which might be seen as fitting: *Star Wars*, *Stargate*, *Planet of the Apes*, *Transformers*, etc.

3. On the relevance of science fiction on the topic of technological innovation

Although I have already tried to show that science fiction and natural sciences can refer to similar objects, for example robotics and AI, in the context of this contribution, it seems necessary to add a few remarks. To a certain extent, it can be assumed that scientific ideas and science fiction influence each other. Future research, for example, has been around for more than 100 years, triggered by the technological innovations that ultimately led to the atomic bomb.¹³ “Diese Zunft entstand aus der Einsicht, dass sich Zukunft via technologischem Fortschritt auch autonom, quasi selbstreferenziell ins Werk setzen kann – und zwar auf möglicherweise für die Gattung Mensch existenzgefährdende Art –, wenn sie nicht antezipativ gebändigt wird.”¹⁴ So the interest in learning to understand the future is present. On the other hand, fiction also offers points of reference that may themselves have had an inspiring effect. Autonomous vehicles were already thought of by Isaac Asimov in the 1960s. We also find bionic limb prostheses in the first Star Wars movies. One of the first authors to be aware of the potential of AI for fiction, but also to recognise the risks, was also Isaac Asimov. His Three Laws of Robotics was included in his collection of short stories from the 1940s and 1950s, *I, Robot*, and later featured in the 2004 movie.¹⁵ But there are many more literary and cinematic explorations of the subject, more than could be discussed here.

And yet: science fiction is not the crystal ball that lets us catch a glimpse of the future. In fact, the goal is almost not at all to actually make realistic predictions. “Although s[cience] [f]iction] does not predict the future as is sometimes claimed, it is the mythological language of technoculture and thus it plays a central role in producing the future through the dreams and nightmares it offers for our contemplation,”¹⁶ as Sherryl Vint expresses.

So, it’s hardly about what will actually happen. It’s more figuratively about what we expect from the future and what has meaning

13 See Müller-Friemauth: *Zukunftsforschung*, 47.

14 Ibid.

15 See *ibid.*

16 *Vint*: *Guide*, 5f.

for our lives. Science fiction includes social criticism and warnings against current developments, as well as reflection on what is truly important, in the guise of fantastic adventures and post-apocalyptic worlds. Of course, this happens to varying degrees, especially in terms of quality—hardly anyone will watch *Sharknado* for its political implications, although it is quite hilarious.

4. Star Trek

So, why should we want to talk about Star Trek here, too? I think the Star Trek universe contains a whole range of moments in which the frontier between the real and the fantastic, between the present and the *maybe someday*, becomes blurred. The amount of material we're talking about is not without significance: after the 1965 pilot entitled *The Cage*, there have already been at least eight live-action television series, three animated series, and one anthology series with a combined total of over 800 episodes as well as 13 movies. By the 1990s with the series *Star Trek: The Next Generation* (1987–1994), at the latest, Star Trek had gained a pop-cultural relevance which most likely exceeded any expectations.

It can be assumed that probably everyone as of today has heard the following words at least once before: “Space, the final frontier. These are the voyages of the starship Enterprise. Its continuing mission: to explore strange new worlds; to seek out new life and new civilizations; to boldly go where no one has gone before.”¹⁷ It can be assumed that Star Trek indeed wants to confront us with limitations. On the one hand, it confronts us with the limits of what we know and can understand and, on the other hand, with our very own possibilities as humans.

17 The opening sequence of each episode of *Star Trek: The Next Generation* (TNG) begins with these words spoken by Patrick Stewart as Captain Jean-Luc Picard. This is a slightly modified and gender-neutral version of the text from the series *Star Trek* (1966–1969), where the sequence is spoken by William Shatner as Captain James T. Kirk. The last sentence is based on the Portuguese poem *The Lusiads* by Luís de Camões from 1572, which is about the discovery of the sea route to India by Vasco da Gama. See *Puchner*: Story, 281.

After all, Star Trek is set in a distant future. The chronologically earliest Star Trek series, *Star Trek: Enterprise* (2001–2005) is set in 2151, while others are set even further in the future, like *Star Trek: The Next Generation*, which is set in 2364. But all Star Trek series have one thing in common: they want to show us what the future might look like for us one day, without being strictly bound to the limits of physics but with a view to social interaction.¹⁸ After all, according to its creator Gene Roddenberry, Star Trek was designed as a humanistic utopia, whereas the expanding franchise has somewhat taken on a life of its own over the course of time.¹⁹

But there is more to Star Trek than just a million-dollar franchise. While we follow the crews of the starships USS Enterprise, Voyager or Discovery and see how they explore unknown worlds to learn and conduct research, when we watch the crew of the starbase Deep Space 9 trying to mediate between peoples, Star Trek en passant takes us on a completely different journey: We follow in the footsteps of philosophy, occasionally without even realising it. Sometimes the connection may be obvious, such as when it comes to the utilitarianism of the Vulcans²⁰ or the transhumanism of the Borg²¹, but sometimes we are very subtly led to question ourselves and our way of thinking.

Lernbereitschaft meint in diesem Sinne einerseits das Ausschöpfen des eigenen Potentials, andererseits eine Haltung der Offenheit gegenüber Neuem, die Korrigibilität von Ansichten, die sich als unangemessen erweisen, und eine damit einhergehende Bescheidenheit im Umgang mit dem tatsächlichen oder vermeintlichen ‚Anderen‘ (die freilich selbst wiederum mit großer Überzeugung vertreten wird): „Wir stehen erst

18 See *Tolan*: Physik, 303.

19 See *Bauer*: Utopie, 2.

20 The Vulcans and their philosophy are quite famous for Spock's words "The needs of the many outweigh the needs of the few". *Meyer*: Star Trek II. The Wrath of Khan, 39:05.

21 The Borg are not a species in the biological sense but add individuals to their collective through assimilation by technologically altering them into cyborgs and integrating them into the collective consciousness. They appear in several series and episodes and also the movie *Star Trek: First Contact* (1996). If you'd like to learn more about the transhumanism of the Borg, I recommend *Dinello*: The Borg.

am Anfang [...] Vieles liegt vor uns. Vieles müssen wir lernen” (TNG, S1E26),²²

as said by Captain Picard in *Star Trek: The Next Generation*. In the process, however, not only do the much-loved screen characters learn, but so does the viewer.

And that is precisely why *Star Trek* lends itself as an example with which to illustrate the link between science fiction and the present with its important questions. *Star Trek* provides dozens of connecting points for reflection, including robotics and AI. A particularly noteworthy character in this regard is certainly the android²³ Data, who, as a crew member of the *USS Enterprise*, still devotes himself to his very own quest, the search for a more human existence.

4.1. The potential of *Star Trek*

Even if we don't pay specific attention to why exactly which motifs appear in movies, books and series, we can learn something from them. However, it is the chosen topics that direct the discussion in a certain manner—such as the connection between science fiction and technology. The fictional universe of *Star Trek* is an exemplary way to directly involve people, regardless of their academic training, in thinking about a future in which people rise above themselves. Although the underlying world view should certainly not be overlooked—after all, *Star Trek* was also created in a specific, US-oriented context—it nevertheless opens up horizons that question the core

22 *Kruse: Zukunft*, 31. Translation: “Willingness to learn in this sense means on the one hand the tapping of one's own potential, and on the other hand an attitude of openness to what is new, the corrigibility of views that prove to be inappropriate, and a concomitant modesty in dealing with the actual or supposed 'other' (which, of course, is itself held with great conviction). 'We are only at the beginning [...]. Much lies ahead of us. We have much to learn' (TNG, S1E26).” He refers to *Hurley: The Neutral Zone*, 44:17. The actual quote is: “Our mission is to go forward and it has just begun [...] There is still much to do. Still so much to learn.”

23 The term android refers to a robot that is supposed to resemble humans in appearance and behaviour.

of all life: What is life worth? What is a living being in the first place? And what protection does it deserve?²⁴

However, it must always be remembered that Star Trek is a utopia: In the end, the characters (very nearly) always manage to ensure that everything ends well, the innocent are protected, and unfairness and lack of freedom are overcome. Most recently, this principle was broken up somewhat in *Star Trek: Discovery* (2017–2024) and also the spin-off *Star Trek: Strange New Worlds* (2022–)²⁵.

Nevertheless, there are points of reference from all the series and films (even comics and books) to show that the considerations made are relevant for us. In particular, the ethical dilemmas were and are explosive and one might say interesting. I would like to elaborate on this a bit more, using an example to show that while Star Trek is not a direct guide for us to find the answers we are looking for, it might open our eyes to creative solutions for an indeterminate future.

4.2 The Measure Of A Man

Now let's finally get to the point where Star Trek and AI converge. I am aware that the episode mentioned here has already been the subject of manifold discussions, but it is nevertheless unavoidable in this context. With the episode *The Measure of a Man* (S02, E09) the series *Star Trek: The Next Generation* explores strange new philosophical problems that no one has seen before. It is about Data, an android, i.e. a robot in the form of a human equipped with artificial intelligence, and his status as property or bearer of rights.

The topic is initiated by the scientist Bruce Maddox, who demands that Data be irretrievably disassembled for the purpose of science, so that his positronic brain²⁶ can be analysed to finally create more like him. Data is unique even in this fictional version of the 24th century; his functionality and learning ability are extraordinary, but he aspires to be truly human. He is considered a full

24 See Kruse: Zukunft, 30f.

25 One quite extraordinary episode in that respect might be Season 1, Episode 6 *Lift Us Where Suffering Cannot Reach*, which first aired on 09.06.2022.

26 This is part of Star Trek's technological fiction. It allows data to be stored and processed.

member of the crew and treated like any other humanoid by the crew of the Enterprise. By having more androids like Data, the utilitarian Maddox²⁷ hopes to contribute to the efficiency and safety of the crews of other starships by eliminating the need for any flesh and blood having to expose themselves to potentially dangerous situations.

In the face of his own destruction, Data refuses. Maddox cannot comprehend this and eventually demands Data's body because, in his opinion, it is the property of Starfleet, the umbrella organisation under which the Enterprise operates. Captain Jean-Luc Picard sides with the android, and a trial is held to decide Data's future. And more or less surprisingly, it is proven quite impressively how much of a machine Data is after all: He can be turned on and off, the seemingly ultimate proof of his lack of freedom and alienness. But what is the consequence if Data is only a thing? He would have no rights, would be only a modern slave if one could even call him that or simply a tool. At least Picard could not accept that.²⁸

Picard therefore asks Maddox the crucial question: How would he define a person with the right to self-determination, a conscious being? He answers as follows: a sentient being is one who is "intelligent, self-aware, and conscious".²⁹ This addresses an essential problem: What distinguishes Data from a human and in what respects are there similarities? Even though his positronic brain may have been artificially created, Data is capable of performing the most complex calculations in a fraction of a second. This hyperintelligence is even superior to human intelligence in many ways. Also, that Data has self-awareness is not doubted here at least. This is justified by the fact that he is able to relate to his own existence, not wanting to die and striving for something, i.e. to be human. However, it is clear to him that he can never really become a human being; his logic is flawless, yet he feels the desire to discover the human in himself. That finitude, for instance, is an issue for Data picked up later in the series *Star Trek: Picard* (2019–2023): He ulti-

27 He is a utilitarian insofar as he would sacrifice a machine for the sake of human or alien life. That he would sacrifice a living human being (or alien) in the same context is not to be assumed, since he always comes back to the question of consciousness.

28 See *Rosenstand*: Data, 176.

29 *Snodgrass*: The Measure, 36:49.

mately desires an end to his existence because the knowledge of his finitude is meaning that is constituting for him.

The question of his consciousness remains open. Is it even possible that a machine—in the series Data is polemically compared to a toaster—has or develops consciousness? How far can technology go? Or must one ask oneself, in line with Thomas Nagel: What is it like to be an android? and come to the conclusion that in Data's case, one can only assume he has consciousness but cannot prove it unequivocally?

Since we are still in the world of speculative science fiction, we are in a position to attribute consciousness to technology, especially since the series provokes this position. What reality will look like one day is not written in the stars in this case; it depends on what we as humanity are willing and able to create, but this cannot be elaborated on at this point.

In any case, in *The Measure of a Man*, the verdict is in Data's favour, his continued existence is assured and, even more, the status that the crew has always accorded him is now also legally recorded.³⁰ The question of his consciousness may not have been settled to the viewer's satisfaction, but if there is even the slightest doubt that Data has no consciousness, Picard deduces that he must be treated as if he had one. This moral decision has far-reaching consequences in the Star Trek universe, as now other artificial entities, such as the holographic Doctor in *Star Trek: Voyager* (1995–2001), also claim dignity and, as a consequence, the right to self-determination (S07, E20).

5. Outlook: the future begins today

As said before, this is science *fiction*. We can simply pass over many of the problems and remaining questions of reality by making assertions and premises that are established as true or false in the fictional world. One makes things very easy for oneself this way; that is obvious. But it is not the purpose of fiction to depict reality and to take it seriously; they are thought games that show us *what-if* scenarios

30 Please note that this is just a small insight into the discussions surrounding this episode. In the end, essentially it is a question of Data's personhood. See *Sorgner: Android*, 242f.

and inspire us to think again. And some of the challenges our heroes face could very well be ours one day. It is therefore important to take these questions seriously. Disciplines such as robot ethics show that, at least here, the problems of the world (which are still purely fictional in Star Trek) already affect our reality, which is becoming increasingly technologised. Because science fiction is already raising such questions, we are already dealing with them and can influence developments for the future in this way. And that in turn leads us to relate to the world as it is or should be. It makes us ask ourselves now how we want to relate to the technology of tomorrow, to robots (or even androids), for example. However, this happens in interaction with what the world has to offer.

In the end, it is certainly easy to get lost in a present as complex as ours. Progress is all around us. We are trying to participate in a future that we ourselves will probably not even live to see. For us, the distinction between now and *maybe someday* will always remain to some degree. But on the other hand, we have our wishes and fears about what humanity will be one day. We tell stories about it and hope to be heard. These stories connect us and make us think together without any preconditions, so that these stories, science fiction, can blur the line.

We would do well to listen to each other and see what science fiction is all about. In science fiction, we find hints of what might be one day and what had better not be. This way, science fiction also collaborates creatively on the future.

Bibliography

- Bauer*, Michael C.: Star Trek – Eine humanistische Utopie? Einleitung des Herausgebers, in: Bauer, M. C. (ed.): Neue Welten. Star Trek als humanistische Utopie?, Berlin 2019, 1–8.
- Dinello*, Dan: The Borg as Contagious Collectivist Techno-Totalitarian Transhumanists, in: Decker, K./Eberl, J. (eds.): The Ultimate Star Trek and Philosophy (The Blackwell Philosophy and Pop Culture Series), Hoboken 2016, 83–94.
- Hawking*, Stephen: Foreword, in: Krauss, L.: The Physics of Star Trek, New York 2007, 3–4.

- Kruse*, Jan-Philipp: Star Trek als das "einzig greifbare Beispiel" einer gelingenden Zukunft? Zur semantischen Krise normativer Potentiale, in: Kanzler, K./Schwarke, C. (eds.): *Star Trek: Discovery. Gesellschaftsvisionen für die Gegenwart*, Wiesbaden 2019, 27–47.
- Kirsten*, Karina: Diskursivität/Historizität von Genre, in: Kirsten, K. (ed.): *Genresignaturen. Diskurshistorische Perspektiven auf das Psycho-Universum von 1960 bis 2017 (Neue Perspektiven der Medienästhetik)*, Heidelberg 2022, 21–73.
- Latham*, Rob: Part 1. Definitions and boundaries, in: Latham, R. (ed.): *Science Fiction Criticism*, London/New York 2017.
- Müller-Friemauth*, Friederike/*Kühn*, Rainer: *Ökonomische Zukunftsforschung. Grundlagen – Konzepte – Perspektiven (FOM-Edition)*, Wiesbaden 2017.
- Puchner*, Martin: *Culture. The Story of Us, from Cave Art to K-Pop*, New York 2023.
- Rosenstand*, Nina: Data, Kant, and Personhood, in: Decker, K. S./Eberl, J. T. (eds.): *The Ultimate Star Trek and Philosophy (The Blackwell Philosophy and Pop Culture Series)*, Hoboken 2016, 172–179.
- Samuelson*, David N.: Modes of Extrapolation. The Formulas of Hard Science Fiction, in: *Science Fiction Studies* 20 Part 2 (60), 1993. Online at: <https://www.depauw.edu/sfs/abstracts/a60.htm> (status: 16.09.2023).
- Schwarke*, Christian: Star Trek und der Traum von der Allgegenwart. Technik und die Realisierung von Utopien, in: Kanzler, K./Schwarke, C. (eds.): *Star Trek: Discovery. Gesellschaftsvisionen für die Gegenwart*, Wiesbaden 2019, 49–65.
- Sorgner*, Stefan L.: Android Data. Eigentum oder Träger von Rechten?, in: Bauer, M. C. (ed.): *Neue Welten. Star Trek als humanistische Utopie?*, Berlin 2019, 239–272.
- Tolan*, Metin: *Die Star Trek Physik. Warum die Enterprise 158 Kilo wiegt und andere galaktische Erkenntnisse*, München 2018.
- Vint*, Sherryl: *Science Fiction. A Guide for the Perplexed*, London 2014.
- Wenskus*, Otta: "Soft" Science Fiction and Technical Fantasy. The Ancient World in Star Trek, Babylon 5, Battlestar Galactica and Dr Who, in: Pomeroy, A. J. (ed.): *A Companion to Ancient Greece and Rome on Screen (Blackwell Companions to the Ancient World)*, Hoboken 2017, 449–466.

Videographie

- Hurley*, Maurice (Writer)/*Conway*, James L. (Director): *The Neutral Zone (Season 1, Episode 26)* [episode of a TV-Series], Air Date: 16.05.1988, In: Roddenberry, G./Hurley, M./Berman, R. et al. (Directors): *Star Trek. The Next Generation*, USA 1987–1994.

Meyer, Nicholas (Director): Star Trek II. The Wrath of Khan, USA 1982.

Snodgrass, Melinda M. (Writer)/*Scheerer, Robert* (Director): The Measure Of A Man (Season 2, Episode 9) [episode of a TV-Series], Air Date: 11.02.1989, in: Roddenberry, G./Hurley, M./Berman, R. et al. (Directors): Star Trek. The Next Generation, USA 1987–1994.

Ramy Abdin

What are Robots?

An Islamic Approach from a Metaphysical Understanding of the Human Being

Abstract

This paper aims to gradually approach the concept of robots by relating it to an Islamic concept of humaneness. It describes the human as a spiritual being, who, although honoured by God, has certain weaknesses, which calls for the human struggle with his darker self.¹ Continuing from this aspect, it provides an Islamic view on the meaning of technology, together with a short elucidation about traditional robotics in Islamic history. Afterwards, a descriptive observation on the (potential) usage of robots in Islamic countries, such as Saudi Arabia and Iran, is presented. Finally, from an Islamic legal and moral point of view,² the issue of robots and robotics is discussed in an attempt to address the question of the roles of robots in Islamic tradition, while contrasting the issue with an Islamic concept of the human being.

1. Introduction

What is the difference between humans and robots from an Islamic perspective? Before we can address this question, we have to establish the adequate preconditions in order to provide an appropriate

1 The term darker self is a reference to the term *an-nafs al-ammāratu bi-s-sū'*, which is discussed in section 2.3.

2 With *moral* we refer to the discipline of *ilm al-mu'āmalā* (the science of praxis) within the field of Sufism as a basis of Islamic ethics. See e.g.: *Al-Ghazālī: Iḥyā' 'ulūm ad-dīn*, Vol. 1, 35.

answer, as our main issue already implies an Islamic definition of the human being. Thus, our main issue is preceded by the question: What is the human being in Islam? And what does it mean to be human? It is true that the concept of man has already been elaborated on from several perspectives, especially within the framework of Islamic tradition.³ Even though there have been works about robots in Islamic education in Iran or the role of AI in regard to Islamic online content, which provide some insights into modern technology and its influence on Muslim society,⁴ the question of humanness in Islam has not yet been extensively related to the idea and endeavour of robotics, as the growing potential of AI and robots increasingly becoming human-like challenges the very idea of being human. Hence, we would like to present and explain this connection, as our age is characterised by seemingly exponential technologisation, rendering human society increasingly dependent on devices and machines, which ultimately necessitates examining issues of contemporary technologies, especially from a religious perspective.

This paper aims to gradually approach the issue of robots by initially presenting an Islamic concept of humanness. We then continue with a short elaboration from an Islamic perspective on technology and a depiction of traditional robotics in Islamic history. Afterwards, we present a descriptive observation on the usage of robots in Islamic countries like Saudi Arabia and Iran. In the last section, we discuss the issue of robots and robotics from an Islamic legal and moral point of view. Finally, we conclude this paper with the humble attempt to answer the question of the difference between humans and robots by dividing our response into a legal and theological elucidation. With that said, we initiate our exposition with the first question: What is the human being in Islam?

3 Hoover: Fiṭra. See also: *Tottoli*: Adam.

4 See papers such as: *Shamdi/Lai/Aziz* et al.: Artificial intelligence; or: *Alemi/Taheri/Shariati* et al.: Social Robotics; or: *Atwell/Brierley/Dukes* et al.: An artificial intelligence approach.

2. An Islamic metaphysics of man

2.1 The Divine honouring

The description of man in the Qur'an is ambivalent. On the one hand, the human being is characterised by Divine nobility, since God ordered the angels, being His purest and most obedient creatures, to prostrate themselves before the obviously fallible earth creature (Adam). Beforehand, God announced in the presence of the angels that He would create a representative on earth (*khalīfa*).⁵ The narrative of humankind in the Qur'an thus begins with the Divine recognition of man: "And when thy Lord said to the angels, 'I am placing a vicegerent upon the earth,' they said, 'Wilt Thou place therein one who will work corruption therein, and shed blood, while we hymn Thy praise and call Thee Holy?' He said, 'Truly I know what you know not.'"⁶ In complete surprise, the angels questioned God's decision, as they were already convinced about the corruption of humankind. Their surprise apparently indicates a negative prior experience with God's creation on earth. According to Qur'anic exegesis, it is believed that the Jinn inhabited the earth before the existence of humans, causing bloodshed among themselves by fighting wars.⁷ Despite the expected corruption and violence at the hands of mankind, the angels submitted to their heavenly duty, placing their trust in God's knowledge and wisdom. For He knew something of which they had no knowledge. The narrative continues thus: "And He taught Adam the names, all of them. Then He laid them before the angels and said, 'Tell me the names of these, if you are truthful.'"⁸ It is unclear which names Adam was taught by God. On the one hand, some believe that the names symbolise man's ability to speak; others understand that God taught Adam the names of living, rational creatures.⁹ On the other hand, the names could also stand for the ability to call God by His name. Apparently, the angels did not have the capacity to worship God through His (highest)

5 *Hossein Nasr*: The Study Qur'an, 108–110.

6 Qur'an, 2:30.

7 *Hossein Nasr*: The Study Qur'an, 108.

8 Qur'an, 2:31.

9 Meaning the names of the angels or the human prophets.

names. The idea of the (Divine) name was therefore given to Adam, whereby he was not only able to honour and praise God in a unique manner, but also to know Him in a more intimate way.¹⁰ To God's request: "Tell me the names of these, if you are truthful."¹¹ The angels responded in complete humility: "They said, 'Glory be to Thee! We have no knowledge save what Thou hast taught us. Truly Thou art the Knower, the Wise.'"¹² In response, God commanded Adam to present his Divine knowledge to the angels: "He said, 'Adam, tell them their names.' And when he had told them their names He said, 'Did I not say to you that I know the unseen of the heavens and the earth, and that I know what you disclose and what you used to conceal?'"¹³ Without hesitation, the angels acknowledged Adam's superior knowledge and, as a result, proved themselves to be faithful servants of God by prostrating themselves before Adam in obedience to the Divine command: "And when We said to the angels, 'Prostrate unto Adam,' they prostrated, save Iblis. He refused and waxed arrogant, and was among the disbelievers."¹⁴ Iblis refers to the archdevil, who apparently had been among the angels as he defied God's command. The Divine honouring of man also occurs in the verse: "We have indeed honoured the Children of Adam, and We carry them over land and sea, and provide them with good things, and We have favoured them above many We have created."¹⁵ God honoured humankind by entrusting them with responsibility on earth; herein lies the confirmation that God has bestowed upon mankind, among all other creatures, special favour, despite their imperfection, of which the angels already knew. In contrast to the earth creature, Iblis sees himself as a superior being created of fire, a conviction that the devil, with his awareness of human weaknesses, sought to reinforce by misleading humanity.¹⁶

10 *Hossein Nasr: The Study Qur'an*, 109–110.

11 Qur'an, 2:31.

12 Qur'an, 2:32.

13 Qur'an, 2:33.

14 Qur'an, 2:34.

15 Qur'an, 17:70.

16 Qur'an, 4:119.

2.2 The human weakness

But what is the weakness of a human being? The answer can be derived from the narrative of Adam and Eve in the Qur'an, in which man's misdeed commences with forgetfulness: "And We indeed made a pact with Adam aforetime, but he forgot."¹⁷ This pact is a reference to Adam's promise to God that he would not eat from the seemingly distinctive tree in Paradise. God forewarns Adam of the archdevil and assures him and his wife of total bliss in heaven as long as they are obedient to His command:

We said, "O Adam! Truly this is an enemy unto thee and thy wife. So let him not expel the two of you from the Garden, such that thou wouldst be wretched. Truly it is for thee that thou shalt neither hunger therein, nor go naked, and that thou shalt neither thirst therein, nor suffer from the heat of the sun".¹⁸

However, Adam's forgetfulness was caused by two major intrinsic *vulnerabilities* that induced him to sin: "Then Satan whispered to him. He said, 'O Adam! Shall I show thee the Tree of Everlastingness and a kingdom that never decays?'"¹⁹

The expression *Tree of Everlastingness* (*shajarat al-khuld*) reflects the first human vulnerability, namely their fear of death. Iblis deceives them by presenting the tree as an opportunity to become immortal. Therefore, they are told that the fruit of the tree will protect them from death, thus relieving them of their fear.²⁰

A kingdom that never decays (*mulk lā yablā*) is a representation of (worldly) power. Knowing about the human weakness for power, the archdevil promises the human beings eternal dominion; in other words, that the tree will elevate them to divinity.²¹ Power as a leitmotif of human weakness is conveyed in particular in the depiction of the Pharaoh in the Qur'an. With his conviction of Divine authority: "I am your lord most high."²² he failed to recognise the Divine

17 Qur'an, 20:115.

18 Qur'an, 20:117–119.

19 Qur'an, 20:120.

20 *Ar-Rāzī*: Mafātīḥ al-ghayb, Vol. 22, 126–127.

21 *Aḡ-Ṭabarī*: Jāmi' al-bayān, Vol. 16, 188–189.

22 Qur'an, 79:24.

message of Moses, which is why Pharaoh, in his efforts to subjugate the Children of Israel, was ultimately punished by God.²³

Humans are therefore susceptible to the temptation of power and the desire for immortality. Both vulnerabilities evoke the forgetting of God, which ultimately leads to disobedience or sin. Fakhr ad-Dīn ar-Rāzī argues that the aspects of power and immortality were in fact the actual desires of the archdevil, who was prevented from fulfilling them by God.²⁴ Be that as it may, the two vulnerabilities primarily arise from two *impulse forces* that are inherent in the human soul. The vulnerabilities depicted can be seen as an expression of these impulse forces.

The first driving force of the human soul is called *shahwa* in Arabic. Lane translates *shahwa* as “Desire, or longing, or yearning, of the soul for a thing; [meaning for a thing that is gratifying to sense: or eager, or intense, desire; particularly for such a thing].”²⁵ In this case, it is an impulse force that releases human instincts. The *shahwa* is an expression of the desire for food or sexuality, which means that it can be described as appetite or lust. It is an urge that pushes man to fulfil his natural needs, which in this sense equates him with animals. Fulfilling the *shahwa* may only be sinful if the urge is either appeased with forbidden pleasures or the person is seduced into gluttony.²⁶

The second impulsive force has the Arabic term *hawā*, which can be translated into English as “Love, and attachment: then, inclination of the soul, or mind, to a thing: then, blameable inclination.”²⁷ *Hawā* is therefore a desire that human beings seek to realise through their decisions and actions. It does not primarily relate to the fulfilment of natural needs, but to the striving for a (certain) worldly cause, regardless of whether it is material or abstract. Accordingly, the desire can lead man astray if it violates Divine principles. Therefore, the term *hawā* has a rather negative connotation in the Qur’an and in the Islamic concept of virtue.²⁸ An exception to this, however,

23 *Hossein Nasr*: The Study Qur’an, 1200–1204.

24 *Ar-Rāzī*: Mafātīḥ al-ghayb, Vol. 22, 126.

25 *Lane*: Arabic-English Lexicon, s.r. *sh-h-w*.

26 *Al-Ghazālī*: Iḥyā’ ‘ulūm ad-dīn, Vol. 3, e.g. 100–120.

27 *Lane*: Arabic-English Lexicon, s.r. *h-w-y*.

28 *Najm ad-Dīn al-Kubrā*: at-Ta’wīlāt an-najmīyah, Vol. 5, 185–186.

can be found in Sufi poetry, in which *hawā* has been associated with the love of God.²⁹

Adam and Eve's disobedience can be understood as a consequence of these two impulses. Eating the forbidden fruit from the tree of paradise is a representation of the first impulse force (*shah-wa*), as it urges man to satisfy his natural needs. The two promises of the archdevil—immortality and power—which, as already discussed, prove to be the vulnerabilities of the human race, unleash the second impulse force (*hawā*) in particular. Since it evokes the desire for immortality and power, Adam fell prey to heedlessness and thus to sin.

Consequently, man's fallibility was already foreseen by God. This Divine providence is the very reason for the archdevil's bitterness and rebellion. Although humankind is doomed to fail in its actions due to human weakness, the human being was granted the opportunity to renounce his sinfulness in order to return to God, unlike Iblis, who lacked insight and arrogantly blamed God Himself for his disobedience: "He [the devil] said, 'Because Thou [O, Lord] hast cause me to err, I shall surely lie in wait for them [namely men] on Thy straight path.'"³⁰ In contrast, humility and self-criticism characterise human nature: "They [Adam and Eve] said, 'Our Lord! We have wronged ourselves. If Thou dost not forgive us and have Mercy upon us, we shall surely be among the losers.'"³¹ Therefore, the very concept of *tawba* has been introduced to mankind. *Tawba* signifies the remorseful return to God in order that He forgives the sinful believer.³² Therefore, human failure is inevitable, precisely because any moral or religious imperfection is inherent in humanity by virtue of the Divine plan; humanness and perfection can thus be considered a contradiction because only God is entitled to absolute perfection: "All the sons of Adam are sinners, but the best of sinners are those who are given to repentance."³³ The greatness of human beings, as the reason for the angels' prostration, is thus ironically accompanied by man's failure. However, this failure is corrected

29 *Al-Badawī*: *Shahīdat al-‘ishq al-ilāhī*, 64–65.

30 Qu'ran, 15:39.

31 Qu'ran, 23:73.

32 *Al-Ghazālī*: *Iḥyā' ‘ulūm ad-dīn*, Vol. 4, e.g. 4–19.

33 *Mishkāt al-Masabīḥ* 2341, Book 9, Hadith 114.

by the graceful humility that heralds the completion of humanity's return to God.

2.3 The breath of God

The Qur'an states: "And [remember] when thy Lord said unto the angels, 'Behold! I am creating a human being from dried clay, made of moulded mud; so when I have proportioned him and breathed into him of My Spirit, fall down before him prostrating.' Thereupon the angels prostrated, all of them together".³⁴ The breath of God is the pivotal point of this passage, for it is the concept that underlies the Islamic metaphysics of man.

God's breath gave existence and life to the human being, whereby he is endowed with Divine qualities that transcend material ones. The essence of a human is therefore not physical but metaphysical, especially since death implies a departure from the body and soul. Although the death (of the body) in no way erases the existence of the individual, the spiritual life breathed into them by God ultimately remains; in other words: the image formed of the human decays, but the breath of God persists.

Thus, the human being consists of three interrelated essences, all of which constitute his soulish existence. The first essence is called *nafs*. It is the essence of the human being that passes into the afterlife after death.³⁵ Often, *nafs* also refers to the *self* or the "ego". The latter has a negative connotation, especially in the teachings of Islamic virtues, because the ego is often described as a veil that prevents people from seeing God and His work.³⁶ In addition, *nafs* has various manifestations. For example, there is the self, which *constantly incites humans to conduct bad actions (an-nafs al-ammāratu bi-s-sū)*. It exerts pressure on man's weaknesses so that he (potentially) succumbs to sin. If this is the case, the individual is required to repent in order to turn to God. Due to his sinfulness, this repentance develops into self-blame, which lies at the heart of his burden in conscience. In this condition the *constantly blaming self (an-nafs al-lawwāma)* arises.

34 Qu'ran, 15:28–31.

35 *Al-Ghazālī: Iḥyā' 'ulūm ad-dīn*, Vol. 3, 5.

36 *Al-Harawī al-Anṣārī: Manāzil as-sā'irīn*, 308.

The guilt subjects the sinful state of the soul to change, whereby the self finally evolves from the urge to sin to the rebuke of the self. Thus, guilt should induce the individual to renounce all their sins in order to walk the path of God.³⁷ However, pursuing this path requires *determination (himmā)*³⁸ and *striving (muğāhada)*,³⁹ as the struggle with oneself, and in particular with one's own bad character traits, is the hardest of all endeavours. The soul will attain peace if one is able to free oneself from wickedness by purifying the heart from evil qualities. This tranquillity of the soul is known as *an-nafs al-muṭma'innah* (the calmed self), through which the human being finally becomes sensitive to the Divine.⁴⁰ As a result, the Divine *state of being* unfolds within the individual, enabling one to gaze through the material world—i.e. to see beyond what is physical—which finally leads the to a constant (earthly) presence with God (*al-ḥuḍūr*).⁴¹

The second essence is referred to as *rūḥ* in Arabic. In the Qur'an, the actual sense of *rūḥ* is restricted to God alone.⁴² Nevertheless, there are numerous attempts to explain this concept, which provide some insight into its meaning. For example, there is the view that *rūḥ* signifies the human breath of life;⁴³ other scholars describe *rūḥ* as the spirit of life that circulates in the bloodstream.⁴⁴ *Rūḥ* is also considered to be synonymous with *nafs*, especially since both terms are often used for the same meaning.⁴⁵ In this article, we use *rūḥ*—in relation to the subject of robotics—in the sense of the human spirit of life, which connects the human being to the Divine.

The third essence is the (spiritual) heart (*qalb*), which is the essence that shapes a person's character, which constitutes the moral conduct of a human.⁴⁶ Also, the heart has epistemic abilities⁴⁷ that are particularly reinforced when it is purified by a virtuous and

37 *Al-Ghazālī: Iḥyā' 'ulūm ad-dīn*, Vol. 4, 5.

38 *Al-Harawī al-Anṣārī: Manāzil as-sā'irīn*, 238.

39 *Mishkāt al-Masabīḥ* 33, Book 1, Hadith 30.

40 *Al-Ghazālī: Iḥyā' 'ulūm ad-dīn*, Vol. 3, 5.

41 *Al-Qushayrī: ar-Risālah*, 258.

42 Qur'an, 17:85.

43 *Lane: Arabic-English Lexicon*, s.r. *r-w-ḥ*.

44 *Al-Ghazālī: Iḥyā' 'ulūm ad-dīn*, Vol. 3, 4.

45 *Picken: Spiritual Purification*, 172.

46 *Mawlūd: Maṭharat al-qulūb*, 1.

47 *Hossein Nasr: The Study Qur'an*, 1006.

spiritual way of life. The purification of the heart signifies the overall cleansing of the human soul, enabling the individual to develop into a virtuous, pure being.⁴⁸ However, if a believer commits a sin, the heart will be tainted with immorality and wickedness: “When the believer commits a sin, a black spot appears on his heart. If he repents and gives up that sin and seeks forgiveness, his heart will be polished. But if (the sin) increases, (the black spot) increases. That is the *covering* (*rān*) that God mentions in His Book: ‘Nay! But on their hearts is the covering (manifested due to sins and evil deeds) which they used to earn.’”⁴⁹ If the heart is defiled, the believer becomes distant from God, which ultimately leads to spiritual blindness.

In addition to the essences, man is in possession of reason (*‘aql*), whose significance has been reflected upon by numerous scholars, especially philosophers, such as al-Farābī (d. 951), Ibn Sīnā (d. 1037) and Ibn Rushd (d. 1198), who advocated a rather essentialist concept of reason in line with the Aristotelian and Neoplatonic tradition.⁵⁰ Thus, human intellect was regarded as a non-corporeal substance that granted the human soul the disposition to reason. A similar view is held by al-Ghazālī (d. 1111) and al-Muḥāsibī (d. 858), who describe reason as *gharīza* (natural endowment).⁵¹ In contrast, a modern Islamic view of reason can be found in the philosophy of ‘Abd ar-Raḥmān Ṭāhā (b. 1944), which we seek to relate to robotics.

Ṭāhā moves away from the traditional, essentialist concept of reason by describing reason as an activity of man (*fā’iliya*). According to him, reason is not an essence inherent in the human soul, but an action that is expressed in many different ways. Even though Ṭāhā usually refers to the one inherent reason through which clear concepts and definitions are formed, it should be understood as the formal, analytical faculty of reason, expressed especially through logical, mathematical deduction. Nevertheless, Ṭāhā negates the idea of a single reason (*‘aql*); other modes of reasoning exist as well, which are constituted in human thought. Reason is therefore not exclusively limited to the perception of external aspects of a matter (*al-asbāb*

48 *Mawlūd*: Maṭharat al-qulūb, e.g. 2–4.

49 Sunan Ibn Mājah, Vol. 5, Book 37, Hadith 4244.

50 Davison: Alfarabi, e.g. 3–18.

51 *Al-Muḥāsibī*: al-‘Aql wa-fahm al-Qur’ān, 203.

aḡ-zāhira li-l-ashyāʾ), it also has the capability to transcend what is external. Hence, there is a kind of reasoning by which the interior of external matters can be recognised. This particularly includes virtues or moral values, as well as spiritual concepts. Moreover, there is a rational activity that transcends the reasoning of morality and thus reveals the hidden causes of the world through mystical experience.⁵² Although Ṭāhā often refers to three modes of reasoning, human reason, according to him, is not restricted to only three kinds of reason (*ʿuqūl*) due to its dynamic activity, for it unfolds in many different expressions and patterns: Ṭāhā calls this dynamic *rational multiplication* (*at-takawthar al-ʿaqlī*).⁵³

3. Robots in Islam

3.1 Technology and traditional robotics

In our article “Die Auswüchse der Technik” (The Excesses of Technology), we argued for the need of an Islamic philosophy of technology. In doing so, we aimed to present an Islamic concept of technology by defining technology as *an expression of Divine empowerment* (*tamkīn rabbānī*), whose realisation must be driven by Divinely ordained motives (*asbāb ilāhīyah*). *The existence of Technology is based on human inability* (*ʿajz*); *Technology begins where man’s natural ability ceases*.⁵⁴ As such, technology should not be detached from religious practice. In fact, it must be regarded as a sacred human endeavour⁵⁵, so that the regard for the hereafter is upheld, even in the circumstances of using machines. Furthermore, technological development must go hand in hand with a critical approach to the

52 Ṭāhā: *Suʿāl al-ʿamal*, 40.

53 Ṭāhā *al-Lisān wa-l-mizān*, 21.

54 Abdin: *Die Auswüchse der Technik*, 302.

55 Technology, with the extraction of raw materials guided by self-control, specifically through a disciplined pursuit of progress, must protect nature and preserve its balance without leading to injustice among humans. Its spiritual application should help overcome man’s inherent flaws, addressing issues and obstacles without overpowering him or nature. Technology should be viewed as a Divine blessing that supports spiritual growth, and as such, it should be utilised in service to a higher power. See: Abdin: *Die Auswüchse der Technik*, 302.

idea of progress, meaning that (technological) progress should not unfold unconditionally. Its ethical implications in terms of resource extraction and production have to be considered in order to adequately implement the Islamic concept of *maṣlahah* (i.e. bringing benefit and averting harm);⁵⁶ we examine the field of robotics within the framework of this definition.

In fact, the field of robotics is not alien to Islamic intellectual culture. As early as the 12th century, the Mesopotamian polymath and inventor Ismā'īl al-Jazarī (d. 1206) presented the concept of the automatism of mechanical devices. Besides this, procedures such as the lamination of wooden materials to avoid curvature, the mechanical balancing of wheels to regulate speed, the use of templates for the production of standardised objects or the use of orifice plates to measure flow were among al-Jazarī's technical methods. He was also occupied with the idea of reciprocal motion in mechanisms such as water pumps or automata, which demonstrated an essential principle of robotics. Among his humanoid automata, he invented a water automaton in the shape of a slave with a water jug, which dispensed water when needed and thus assisted the caliph in performing ablution; a serving automaton, which was intended for the transportation of drinks; a parrot fountain with a water-dispensing function; and a music automaton in the shape of a boat, on which four humanoid musicians were located entertaining visitors on special occasions. Probably al-Jazarī's most famous invention is the elephant clock. The clock was an automaton resembling a life-sized elephant, operated by a water clock, which displayed the time. It was decorated with figures (including a man with a cymbal and a bird) who were seated in a carriage on the elephant's back. After every half hour, the cymbal would sound, accompanied by bird's chirp.⁵⁷

The definition of the modern robot is not as clear as many would imagine, especially since several countries have different standards for labelling a machine as a robot; nevertheless, more clarity can be gained by distinguishing between the concept of a *manipulator* and that of a *robot*:⁵⁸

56 *Opwis*: Maṣlaḥa, 1.

57 *Al-Jazarī*: The Book.

58 *Niku*: Introduction, 2.

If you compare a conventional robot manipulator with a crane attached to, let's say, a utility or towing vehicle, you will notice that the robot manipulator is very similar to the crane. Both possess a number of links attached serially to each other with joints, where each joint can be moved by some type of actuator. In both systems, the "hand" of the manipulator can be moved in space and placed in any desired location within the workspace of the system. Each one can carry a certain load, and in each, a central controller controls the actuators.⁵⁹

Although the manipulator is similar to the robot, there is one significant difference between them: the manipulator is controlled by a human, whereas the robot is operated by a computer or microprocessor which is based on a specific program. Any modification of the program will result in a change in the robot's behaviour.⁶⁰ The purpose of the robot's construction is to perform several different tasks: "Therefore, the robot is designed to be able to perform many tasks based on the running program(s) simply by changing the program. The simple manipulator (or the crane) cannot do this without an operator running it all the time."⁶¹ As a result, robots can be classified into four different categories. The first is the *simple fixed-sequence robot*, which carries out a number of operations according to a specific, unchangeable procedure; however, this robot does not yet contain a sophisticated computer program. The *Playback robot*, on the other hand, is slightly more complex; its work procedures are initially carried out by human control. The robot memorises the manual operations, enabling it to conduct them on its own afterwards. Furthermore, the *numerical-control robot* is already equipped with a motion program and is therefore able to execute tasks without initial human instruction.⁶² Finally, there is the *intelligent robot*, which can be described as follows: "A robot with the means to understand its environment and the ability to successfully complete a task despite changes in the surrounding conditions under which it is to be performed."⁶³ The last category is of particular interest for our paper, as it is the type of robot that raises most (theological) questions.

59 Ibid., 2.

60 Ibid., 2.

61 Ibid., 2.

62 Ibid., 3.

63 Ibid., 3.

Today's robotics is engaged in the production of automatic (intelligent) machines. Robotics can actually also be described as an artistic endeavour, as it not limited to the functional aspect of the robot, but also involves the design of its appearance: "Robotics is the art, knowledge base, and know-how of designing, applying, and using robots in human endeavours."⁶⁴

In addition, the purpose of robots can be described as follows: "Robots may be used in manufacturing environments, in underwater and space exploration, in researching human and animal behaviour, for aiding the disabled, for transportation and delivery, for military purposes, or even for fun. In any capacity, robots can be useful but need to be programmed and controlled."⁶⁵ Besides this, robots are also used for religious purposes nowadays. This has been demonstrated in particular by *Christian robots* such as SanTO or BlessU-2 as well as the *Buddhist robot* Mindar, which even has a partially human form. These are known as prayer robots, which also have pastoral and ritual functions. They are able to answer religious questions and explain theological ideas, qualifying them as intelligent robots, especially due to their speech capabilities.⁶⁶ Examples of religious robots also exist in an Islamic context. As part of the Vision 2030 economic project in Saudi Arabia, which is being realised to gradually reduce the country's dependence on oil through the cultivation of alternative resources, a great amount of money is being invested in modern technologies. AI and (intelligent) robots in particular could contribute significantly to the economic endeavours in question, which may explain Saudi Arabia's keen interest in new technologies. However, Saudi Arabia is considered an Islamic kingdom, meaning that the religious needs of the population must also be taken into account when using robots. Thus, official Islamic legitimisation for the use of intelligent machines is required, first and foremost in order to remove any religious doubts about the utilisation of robots. Therefore, it is no coincidence that one of the nine imams of the Holy Mosque in Mecca (*masjid al-ḥarām*) 'Abd ar-Raḥmān as-Sudays (born 1961) participated in a public presentation of an *AI sermon robot* as part of the smart Haramain project

64 Ibid., 3.

65 Ibid., 3.

66 *Machreich*: Robotersegen.

by promoting the use of an *Islamic robot*. The robot is designed to fulfil a variety of religious functions, including reciting the Qur'an, performing the call to prayer (*adhān*) and delivering sermons and reminders. In addition, interaction between the robot and pilgrims using smartphones will be provided so that visitors can obtain religious information from it. This particularly involves consultations on matters of Islamic law, with the robot even issuing *fatwas* (Islamic legal opinions). The Saudi Arabian religious authority is also planning the use of hygiene robots in mosques to ensure more efficient cleaning in holy places.⁶⁷

The Iranian government is also considering the use of robots. The utilisation of AI or intelligent machines could prove to be highly beneficial in the issuing of fatwas.⁶⁸ Nevertheless, a significant restriction is already being imposed on the use of robots: "Robots can't replace senior clerics, but they can be a trusted assistant that can help them issue a fatwa in five hours instead of 50 days," said Mohammad Ghotbi, who heads a state-linked organisation in Qom that encourages the growth of technology businesses.⁶⁹ In the Iranian context, efforts are being made to reconcile religious tradition and modern technology by using new technological systems to preserve religious values: "The seminary must get involved in using modern, progressive technology and artificial intelligence," Ayatollah Alireza Arafai said in July. "We have to enter into this field to promote Islamic civilisation."⁷⁰ The critical attitude of Islamic scholarship in Iran is due to the fear that the population is increasingly turning its back on religion and instead adopting Western values under the influence of modern technology. However, this concern can be refuted by the argument that the use of AI will enable the Iranian clergy to adopt a more comprehensive approach to the dissemination of religious ideas, as fields of knowledge, such as sociology, psychology, health and entertainment, can be evaluated more quickly and efficiently in religious discourse or even used for religious purposes: "Ghotbi said it was incorrect to assume the Iranian society was inevitably becoming less religious. Rapid technological develop-

67 *Tamim*: AI sermon robots.

68 *Bozorgmeh*: Robots can issue a fatwa.

69 *Ibid*.

70 *Ibid*.

ments, and the refusal of many women to wear headscarves since the protests, had unsettled many, particularly the more religious Iranians, and prompted them to seek 'rescue' through Islamic teachings, he said. Technology could also help the clergy address public concerns and reconcile them more swiftly with Islam, while helping clerics make rulings in an increasingly complex society, he argued.⁷¹

3.2 An Islamic view on robots

How can robots be approached in Islam? These questions may be addressed on two levels: on an Islamic legal and ethical (spiritual) level. First, the usage of robots must be evaluated from a normative perspective. A normative evaluation means a legal assessment within the framework of Islamic law. In relation to robotics, the question needs to be examined as to what extent the production and utilisation of robots is legitimate in an Islamic context.

As a preliminary point, it can be stated that the manufacture and use of robots is generally permitted in Islamic law (*ḥalāl*) according to existing fatwas;⁷² thus, in principle, nothing speaks against the use of robots. Nevertheless, the problem under Islamic law lies in the field of robot design. In general, Islamic law does indeed place strict prohibition on visually depicting living beings, whether humans or animals. This prohibition applies in particular to the visual arts, which means that the visual design of artwork may only be pursued to a limited extent. Undoubtedly, there is an artistic element in the field of robotics. This is because the production of robots is not restricted solely to the development of their function, but also places importance on the appearance of the machine, which is certainly intended to evoke a psychological effect.⁷³ A fatwa from the Islamic educational institute Dār al-Muṣṭafā in Tarīm (Yemen), published in English by Yusuf Weltch on the Islamic online platform Seekers-Guidance, states: "The issue of the robot revolves around the issue of *taswir* or form-making. It is prohibited to make any 3-dimensional

71 Ibid.

72 According to the fatwas from islamweb.NET, islamqa.org and seekersguidance.org. See footnotes: 67, 73 and 74.

73 Weltch: Is It permissible?

form of any animate creature. Even though the robot does not have a human face or skin—its form and shape are that of a human.”⁷⁴ Hence, Islamic legal issues revolve around the question of the legality of the production and use of robots, particularly when it comes to their appearance.

The manufacture of a robot is prohibited if it is endowed with (human) extremities, regardless of the purpose of the machine. In such a case, Welch even speaks of a great sin, which he explains by referring to the traditions in Hadith literature;⁷⁵ for example: “The people who will be the most severely punished before God on the Day of Resurrection will be the image makers.”⁷⁶ Welch elaborates on this aspect as follows: “This ruling applies even if the robots are used for menial labor, such as cleaning; this is because of the intense warning against those who make 3-dimensional forms (that resemble animate creatures) and because form makers are the most severely punished on the Day of Judgment.”⁷⁷ Nonetheless, children’s toys are an exception to this prohibition on the basis that the Prophet Muhammad is said to have permitted child’s play with dolls. However, a legal difference of opinion occurs in the case of manufacturing robots that are missing vital human extremities, for example if the machine is missing an entire half of its humanoid shape.⁷⁸ The legitimacy of the production of such robots is argued for by this statement by the Prophet: “Order him to cut off the head of the statue that exists in house to let it become like a tree”.⁷⁹ In this respect, it should be emphasised that only the external appearance of the robot is taken into consideration when it comes to legal assessment, whereas the internal features have no relevance in this case.⁸⁰ In contrast, an uncommon view holds that the humanoid appearance of a machine does not entail any prohibition as long as the robot is not worshipped as an idol; unfortunately, the Islamic legal justification for this opinion remains unknown.⁸¹

74 Ibid.

75 Ibid.

76 Sunan an-Nasa’i 5364, Vol. 6, Book 48, Hadith 5366.

77 *Welch*: Is It permissible?

78 Ibid.

79 *Islamweb*: Working with and using robots.

80 *Moosa*: Consensus of Scholars.

81 *Welch*: Is It permissible?

Thus, the possession or usage of robots is viewed within the framework of the general prohibition on depicting human or animal images, which means that a robot is not allowed to be owned or used as long as it has a full humanoid appearance. Beyond that, the robot can be used only for permissible purposes: “As for that which is prohibited to make, then if it is kept or acquired to fulfil some physical desire or devilish aim, to revere, as an adornment—such as it being put on the roof, a wall, or an erect cushion—it is strictly prohibited to have and to attend any place it is in. In fact, it is prohibited to even enter upon it (i.e. into that room which it’s in) according to some scholars due to the narration of Imam Muslim on the authority of ‘Aisha.’”⁸² The use of robots as assistants in maintaining residential areas or for physically demanding activities is permitted. Unfortunately, the Islamic legal assessment of robots seems to be shallow, as the fatwas so far have mainly considered the appearance of the machines in question. But what about other potential functions of robots? What are the effects on human civilisation emanating from the deployment of robots? What is the Islamic verdict on the exploitation of resources for manufacturing, not only intelligent machines, but modern technologies in general?

Based on the Islamic teachings of virtue, as a branch of Sufism, we emphasise the importance of ethical production and use of technology. In particular, the extraction of resources cannot be carried out in conjunction with the exploitation of people. At the same time, technology must be submitted to nature, for nature has to be seen as a place of divinity, which means that it should not be downgraded to a merely exploitable resource. Instead, nature needs to be given its rights, as technologisation and progress have to adhere to a certain balance. Technologisation, as one of the manifestations of human progress, must develop in harmony with the well-being of humankind and nature. Negative effects, such as the spreading of corruption (*fasād*) and injustice (*ẓulm*), need to be prevented. Corruption in this context means exposing nature to long-term destruction; injustice, on the other hand, refers to the exploitation of people for the sake of alleged technological progress.⁸³ Robotics must absolutely comply with these moral aspects. Virtue in particular should

82 Ibid.

83 Ramy Abdin: Die Auswüchse der Technik.

enable the believer not only to perceive the advantages of technology, but at the same time to acquire an understanding of the downsides of using modern machines. Without a doubt, it is currently not possible to answer all morally relevant questions about modern technology, although the first step in answering them must be taken by raising some of these questions; for example: How does the use of intelligent machines affect human social interactions? What general psychological and spiritual impact will robots have on us once we are constantly interacting with them? How much will the economy change if humans are replaced by robots in certain work areas? What will it mean for human civilisation if only robots perform difficult or physically demanding tasks? Should we actually grant highly intelligent robots political power? We strongly emphasise here that not all technological developments guarantee the well-being of mankind. The invention and use of weapons of mass destruction should serve as a prime example of the dark side of technology.⁸⁴ Just because it is possible to produce innovative machines or new types of artefacts does not mean that they must always be developed, especially if negative consequences can be anticipated. In the *Rome Call for AI Ethics* conference on 10.01.2023, the Islamic scholar Hamza Yusuf (born 1958) referred to the potentially destabilising element of (technological) innovations; in contrast, religions are a conservative enterprise, so they are inherently cautious, especially towards human innovations. It is precisely the excessively fast pace at which innovations are created that may result in such destabilisation.⁸⁵ Therefore, intentional abstinence from innovations should be regarded as a serious option, which means that the entire concept of progress has to be reconsidered, since refraining from certain technologies can also be seen as an expression of progress, especially when harm is prevented; religions, as conservative enterprises, could provide a contribution to the pursuit of *balanced progress*.

4. Conclusion and Outlook: Men and Robots

Coming back to the question of the difference between humans and robots, the observant reader will already have an idea of what the

84 Ibid.

85 Yusuf: AI Ethics.

distinguishing aspects might be, yet we would like to elaborate on this issue slightly more by addressing the question on a normative and theological level.

The intelligence of a robot in conducting analysis and making calculations is beyond any doubt. Its efficiency and computing speed will always surpass that of a natural human being; hence, an intelligent machine is endowed with a basis of human rationality. However, what kind of reasoning ability are we talking about here? As already explained, humans possess three basic types of reasoning: analytical reasoning, moral reasoning and spiritual reasoning. Analytical reasoning stands for the intellectual capacity for analytical (mathematical) thought; the second represents the foundation for moral judgement of human activity; the last represents the ability to become receptive to what is Divine through mystical experiences. Hence, we argue that an intelligent machine is nothing other than an enormous extension of analytical reasoning. In fact, it is the only reasoning that is devoid of any emotion; it only operates by the means of deduction and induction or through processes of calculation. Even though a robot certainly has the potential for moral reasoning, such reasoning requires the capacity for emotional perception and, above all, a conscience, especially when empathy or even virtues such as compassion and mercy become necessary in matters of moral complexity. This begs the question of how far human emotions could be recreated in a machine. Even if this possibility does arise at some point, to what extent do these recreations actually match human emotions? To be more specific: Are these really human emotions? Furthermore, we take the theological view that a robotic individual definitely cannot be endowed with spiritual reasoning, even though it is not impossible for machines to have a spiritual impact. In particular, the notion of *barakah* (“giving of blessings”) might be applicable to technological devices; the general concept implies that individuals, places or even objects, as instruments of Divine activity, can become sources of blessings, which a believer would spiritually benefit from.⁸⁶ In addition, human beings are constantly exposed to the experience of their own weaknesses, which means that they are in lifelong conflict with their *darker self*. On the one hand, human actions are greatly influenced by the

86 *Hossein Nasr: Sufi Essays*, 35–36.

(worldly) impulse forces of the soul—*hawā* and *shahwa*—so that one has to balance these forces in order to purify the heart and, as a result, the soul of their bad qualities. A robot's activity, on the other hand, is determined by algorithms. It neither experiences the forces of human impulse, nor does it embody the essences of man, namely the soul (*nafs*), spirit (*rūh*) and heart (*qalb*), because the presence of the essences constitutes the condition of human weakness. Hence, it does not have a *metaphysics of man*. Unlike humans, a robot does not need to overcome itself when carrying out actions; instead, it chooses to perform an act that normally poses a challenge to humans without enduring any hardship. However, if what is Divine and the respective hereafter are considered, the robot, as a complex machine, may contribute to man's spirituality; although the robot is not a human being, it can be experienced as an expression and manifestation of God's work.

However, one function that is of particular interest from the viewpoint of Islamic law is the responsibility for making decisions about life and death, for example when a robot is used for military purposes, a matter which, we believe, has not yet been sufficiently reflected in Islamic law. So how is an intelligent robot that is used to kill people to be legally assessed, especially if the act of execution is not self-defence, but first-degree murder? Who is to be held responsible? The owner or even the manufacturer or programmer? It is certainly not reasonable to attribute (religious) responsibility to an intelligent machine. This is because Islamic law adheres to the concept of *taklif*. *Taklif* is the human responsibility to fulfil the commandments and prohibitions of Islamic law. Thus, a believer is considered responsible (*mukallaf*) if they have reached maturity, have full mental capacity and demonstrate the (physical, psychological and circumstantial) ability to follow Divine command.⁸⁷ Therefore, it is questionable how the principle of *taklif* can be applied on robots, as they function through algorithms, which in no way corresponds to human understanding and nature. Thus, given the emergence of robotics, it is imperative to re-examine the issue within the framework of Islamic law, in order to determine how intelligent machines can be evaluated and regulated in accordance with Islamic principles.

87 Powers: Intent in Islamic Law, 12.

Bibliography

- Tāhā, 'Abd ar-Raḥmān: al-Lisān wa-l-mīzān aw at-takawthar al-'aqlī. Dār al-Bayḍā' 1998.
- Tāhā, 'Abd ar-Raḥmān: Su'āl al-'amal. baḥṭh 'an al-uṣūl al-'amaliyya fī l-fikr wa-l-'ilm. Bayrūt/Dār al-Bayḍā' 2012.
- Al-Badawī*, 'Abd ar-Raḥmān: Shahīdat al-'ishq al-ilāhī. Rābi'at al-'Adawīya. Al-Qāhira (Cairo) ²1962.
- Alemi, M./Taheri, A./Shariati, A.* et al.: Social Robotics, Education, and Religion in the Islamic World. An Iranian Perspective, in: Sci Eng Ethics 26, 2020, 2709–2734. DOI: <https://doi.org/10.1007/s11948-020-00225-1>.
- Al-Ghazālī*, Abū Ḥāmid: Iḥyā' 'ulūm ad-dīn, 5 Volumes. Bayrūt: Dar al-kutub al-'ilmīyah, 2008.
- Al-Harawī al-Anṣārī*, 'Abd Allāh: Manāzil as-sā'irīn, Commentary by 'Afif ad-Dīn al-Tilimsānī. Bayrut 2013.
- Al-Jazarī*, Ismā'īl: The Book of Knowledge of Ingenious Mechanical Devices (Kitāb fī ma'rifāt al-ḥiyāl al-handasiyya), translated and annotated by Donald R. Hill, Dordrecht/Boston-USA 1974.
- Al-Muḥāsibī*, al-Ḥārith: al-'Aql wa-fahm al-Qur'ān, Ed. Ḥusayn al-Quwwatī. Bayrūt 1971.
- Al-Qushayrī*, 'Abd al-Karīm: ar-Risālah, Ed. Anas Muḥammad 'Adnān ash-Sharqāwī. Jidda 2020³.
- An-Nasā'ī*: as-Sunan aṣ-ṣuḡhrā. Online at: sunnah.com [10.11.23].
- Ar-Rāzī*, Fakhr ad-Dīn: Mafātīḥ al-ghayb, 32 Volumes. Bayrūt 1981.
- Aṭ-Ṭabarī*, Ibn Jarīr: Jāmi' al-bayān 'an ta'wīl āy al-Qur'ān. Ed, 'Abd Allāh ibn 'Abd al-Muḥsin at-Turkī, 26 Volumes. Bayrūt n.d.
- At-Tibrizī*, Khaṭīb: Mishkāt al-Masabīḥ. Online at: sunnah.com [10.11.23].
- Atwell, E.S./Brierley, C./Dukes, K.* et al.: An artificial intelligence approach to Arabic and Islamic content on the internet. In: Proceedings of NITS 3rd National Information Technology Symposium. NITS 2011 3rd National Information Technology Symposium, 06–09 Mar 2011, Riyadh 2011. DOI: <http://dx.doi.org/10.13140/2.1.2425.9528>.
- Bozorgmeh*, Najmeh: Robots can issue a fatwa, 24.09.2023. Online at: <https://www.ft.com/content/9c1c3fd3-4aea-40ab-977b-24fe5527300c> [10.11.23].
- Davison*, Herbert A.: Alfarabi, Avicenna, & Averroes on Intellect. Their Cosmologies, Theories of the Active Intellect, & Theories of Human Intellect, Oxford 1992.
- Hoover*, Jon: Art. Fiṭra, in: Encyclopaedia of Islam, THREE. Online at: http://dx.doi.org/10.1163/1573-3912_ei3_COM_27155 [10.11.23].
- Hossein Nasr*, Seyyed: Sufi Essays. New York 1977.

- Hossein Nasr*, Seyyed (ed.): *The Study Qur'ān. A new Translation with Notes and Commentary*. New York 2015.
- Ibn Mājah*: Sunan Ibn Mājah. Online at: sunnah.com [10.11.23].
- Islamweb.net*: Working with and using robots, 11.05.2003. Online at: <https://www.islamweb.net/en/fatwa/85827/working-with-and-using-robots> [10.11.23].
- Lane*, Edward W.: *Arabic-English Lexicon*, 8 Volumes, Bayrūt 1968.
- Machreich*, Wolfgang: *Robotersegen – nicht allen gelegen*, 09.04.2023. Online at: <https://www.wienerzeitung.at/h/robotersegen-nicht-allen-gelegen> [10.11.23].
- Mawlūd*, Muḥammad: *Maṭharat al-qulūb min qatrat al-'uyūb*, Ed. Muḥammad 'Uthmān ibn Muḥyī ad-Dīn Abbūh. Nouakchott (Mauritania) 1996².
- Moosa*, Ismail: *Consensus of Scholars and Ulema on study and research of robots*. Online at: <https://islamqa.org/hanafi/fatwaa-dot-com/157575/consensus-of-scholars-and-ulemas-on-study-and-research-of-robots/> [10.11.23].
- Najm ad-Dīn al-Kubrā*, Aḥmad ibn 'Umar: *at-Ta'wīlāt an-najmīyah fī t-tafsīr al-ishārī aṣ-ṣūfī*, Ed. Aḥmad Farīd al-Mizyādī, 6 Volumes. Bayrūt 2009.
- Niku*, Saeed B.: *Introduction to Robotics. Analysis, Control, Applications*, Hoboken 32020.
- Opwis*, Felicitas: *Maṣlaḥa and the Purpose of the Law. Islamic Discourse on Legal Change from the 4th/8th to 8th/14th century*, Leiden/Boston 2010.
- Picken*, Gavin: *Spiritual Purification in Islam. The Life and Works of al-Muḥāsibī*, London 2011.
- Powers*, Paul R.: *Intent in Islamic Law. Motive and Meaning in Sunnī Fiqh*, Leiden 2006.
- Abdīn*, Ramy: *Die Auswüchse der Technik. Ein islamischer Weckruf*. In: *Normativität des Korans im Zeichen gesellschaftlichen Wandels* (pp. 277–306). Baden-Baden 2022.
- Shamdi*, W./Lai, D./Aziz, A.A. et al.: *Artificial intelligence development in Islamic System of Governance. A literature review*, in: *Cont Islam* 16, 2022, 321–334. DOI: <https://doi.org/10.1007/s11562-022-00504-7>.
- Tamim*, Baba: *AI sermon robots are automating Islam's holiest mosques*, 14.09.2022. Online at: <https://interestingengineering.com/innovation/ai-sermon-robots-automating-islam-holiest-mosques> [10.11.23].
- Tottoli*, Roberto: *Art. Adam*, in: *Encyclopaedia of Islam, THREE*. Online at: http://dx.doi.org/10.1163/1573-3912_ei3_COM_27155 [10.11.23].
- Weltch*, Yusuf: *Is It permissible to Produce Robots That Don't Resemble Human Beings?*, 03.04.2022. Online at: <https://seekersguidance.org/answers/halal-and-haram/is-it-permissible-to-produce-robots-that-dont-resemble-human-beings/> [10.11.23].

Ramy Abdin

Yusuf, Hamza: AI Ethics. An Abrahamic Commitment to the Rome Call, 10.01.23. Online at: <https://www.youtube.com/watch?v=bdnZDiy6MpI> [10.11.23].

Jonas Simmerlein

Doomed to fail?!

Three Inherent Problems Haunting Religious Robotics

Abstract

The chief question about robots in religious practices today is not whether they should be allowed to participate or if they will replace humans but rather why they feel so irrelevant to us? This paper argues that the acceptance of robots in religious practices is foremost challenged by our experience with them, which is construed as indifference/difference. Either they make a difference to us or we are indifferent to their presence and action. In the latter case, the challenge of robotics in religious practices is not primarily of an ethical nature but a pragmatic one: Why pursue a path of religious advancement that holds no significance to its followers? Drawing from the phenomenological tradition, I will first describe the phenomena of indifference/difference as subjective experiences and describe three distinct problems robots in religious practice have: their lack of history, their preponderance of purpose and their absence of flesh. To strengthen my case argumentatively, I will confront all these arguments with their most obvious responses and contradictions to illuminate the breadth of the possible discussion. To conclude this essay, I will sketch preliminaries and examples of religious robots that might yet make a difference.

1. Phenomenology of indifference/difference

What defines the enigmatic realm of indifference? Some might argue it is a realm of non-feeling, as its inherent nature fails to evoke any

emotional response.¹ Philosophically speaking, it can be articulated as the equal spread of degrees of belief among possible outcomes when no relevant evidence is given.² Indifference is far from inadvertent. It does not merely entail passing by unnoticed, like an overheard sound that momentarily escapes our attention. The experience of indifference is not devoid of awareness but rather lacking in significance and intensity. It diverges from the state of boredom—a more specific inner state that imbues us with a consciousness of time, as aptly and positively discussed by Martin Heidegger³ or William James⁴. While boredom arises when we persistently expose our attention to matters we find unstimulating, giving rise to our awareness of the slow ticking of time, something normally unconscious in our everyday toiling, indifference and boredom exist as distinct phenomena. What provokes our indifference entails a choice between at least two things it differentiates from each other, while at the same time, proclaiming: I do not care which one becomes reality. It pertains to the intricate relationship between oneself as a subject, holding values, experiences and volition, and an object that does not bear any significant relevance to these elements. If it was linked to any of those, we would be unable to not care about it. The importance of what we care about is what keeps us from faltering into indifference.⁵

An indifferent character, like Camus' Meursault, seems to be all but unaffected by values which he does not hold, experiences he does not value and volition he hardly has. Whether the object of our indifference exists or not makes no difference to such an inner state; hence the name. In the realm of our everyday perceptual experiences, most things remain present within our field of awareness without capturing our conscious notice. We remain unaffected by the existence of most of the things around us, yet often unaware of our neglect of difference.⁶ In contrast, indifference presupposes

1 Albert Camus' protagonist Meursault in *The Stranger* encapsulates the lack of both affection and care, resulting in a life led in apparent indifference until an absurd murder finally evokes something in him; *Camus: Fremde*.

2 See *Eva*: Indifference, 1.

3 See *Heidegger*: Grundbegriffe, 220–249.

4 See *James*: Principles, 626.

5 See *Frankfurt*: Importance, 80–84.

6 See *Merleau-Ponty*: Phänomenologie, 108–113.

an awareness rooted in judgements derived from our values, experiences and volition. What holds significance to us resonates with these principles of subjectivity, while indifference, conversely, does not. We might find something stimulating and thus not boring yet be totally indifferent to its presence and what it might do or communicate⁷. Additionally, one could argue there is an intimate link between indifference and intensity as contrasting experiences. Experiences of heightened intensity, as explored by Maslow⁸ and James⁹, typically encounter, surpass and challenge our values, experiences and volition. The experience of indifference becomes all the more pressing and often prompts us to articulate it when we are confronted with the chasm between expectation and reality. This is hardly surprising as awareness stands as a fundamental factor in both indifference and intensity. The more we anticipate intensity, the more vexing its absence becomes.

2. Roots of indifference

Technology is prone to be especially challenged by the phenomenon of indifference, which is why it is championed here instead of typical accounts like ethical perils or dogmatic quarrels that are often at the forefront of debates about robots in religious practices. But why should—as I assume—indifference be a pressing problem for religious technology?

One possible answer to the question of indifference towards religious robots might be what John Jordan lays out and what was already hinted at here. Unlike other groundbreaking technologies, such as nuclear energy, the internet or smartphones, which seemingly emerged abruptly without prior reflection in the collective consciousness, robots have been subjects of debate and fantastical speculation for well over a century, as envisioned by writers like Lem,

7 The constant flood of stimulating yet interchangeable content flowing upwards in our Instagram and Twitter feeds undoubtedly gives us a sense of this activated indifference.

8 See *Maslow*: Peak-Experiences, 9–18.

9 See *James*: Varieties, 380–397.

Wells and Asimov.¹⁰ If we assume there is a close association between indifference and expectations of specific experiences and volition, it is unsurprising that robots may sometimes feel underwhelming and mundane. The reality of robots has diverged from the realms of science fiction, yet they remain more prevalent in the cycles of production than as social agents. Robotic process automations (RPAs), the predominant form of robotics quantitatively, embrace their inherent boredom, which is neither undesirable nor unwelcome. In fact, the *four Ds* principle in robotics—dull, dangerous, dear, dirty—defines tasks especially suited for robots, with dullness being one of the guiding principles.¹¹

Is indifference then merely a side effect of underwhelming technology? If that was the case, we would have to assume that meeting Star Wars's favourite C3-PO *in the metal*—since meeting him *in the flesh* is precisely what we do not envision when we are talking about robots—would make us care about them and make them significant to us. While undoubtedly an overwhelming experience, not only to the franchise fanatics¹² but to everyone, the thrilling experience might very well fade soon enough and be replaced by the nagging feeling of indifference all social robots are doomed to evoke sooner rather than later. Sparking our curiosity for a while, like anchors hanging from their frame, they are drowned in our indifference by three distinct problems that are so essential we might not find a solution. Yet I want to give a short prolepsis in the end on how these burdens might be lifted if religious robots are constructed differently.

2.1 The problem of history

Let us first contemplate how robots are situated in their history or, as I will argue, their lack thereof. What they have is internal time

10 See *Jordan*: Robots, 5.

11 See *Marr*: Robotisation.

12 Carrie Fisher reminisced about the release of *Star Wars: A New Hope* in 1977 and its immense impact on people, recalling, “A lot of people were affected deeply by it, requiring talismans and artifacts, merchandising and sequels. [...] It was bewildering. The movie was attracting giddy attention that was both exciting and unsettling. [...] It wasn't like a movie opening; it was like an earthquake.” (*Fisher*: Arrival).

stamps that are related to data. But are humans any different? Are we not merely biological processors that process data through fleshy pathways? Phenomenology since Bergson, Heidegger and Schütz says we are not or at least not when it comes to experiencing time. For Heidegger, time is intimately entwined with being, shaping our manner of existence. He emphasises that we dwell within time and that our existence is marked by transience and finitude. Time affords us the capacity to fathom and construe ourselves and the world around us. Most importantly, Heidegger posits that time is not a linear sequence of events but rather a dynamic and uninterrupted flow that moulds our subjective existence.¹³ In a similar vein, Alfred Schütz describes the genesis of meaning in our construction of the social sphere, following Husserl's ideas about consciousness: In the midst of the stream of our consciousness, intentionality picks out some well-defined experiences it construes and imbues with meaning, layer by layer founding our every understanding of the world around us.¹⁴ What is distinctly human is our subjective experiential relation to time, which we can reasonably call history from here on out. It is not the mere passage of time but our distinct relation to it experiencing ourselves in it.

This historical character of being is evident when observing interhuman connection: When blessed by a human pastor, you are also blessed by an individual with a personal history beyond their pastoral role. In every incidental meeting we become each other's history. Yet this confrontation is bilateral: You become part of my history, and I become part of yours—as is the mutual deal with every encounter we have. In this, there is no hierarchy but the confrontation of two equals. Neither can escape their subjective being with the other and becoming part of their history. But the bargain goes even deeper. Not only do we invite the other into our history by presenting them our time. Since it is finite, we are permanently spending what we have sparse amounts of. Being mortal, our interactions are on a timer. We hope our encounters are of value, since nothing comes for free for beings that are going to die. One of the reasons we want our conversations, our work and our interactions to be of value to us and others lies in our unconscious or explicit knowledge that

13 See *Heidegger: Sein*, 323–33.

14 See *Schütz: Aufbau*, 63–69; *Husserl: Lebenswelt*, 608–614.

we have to matter with what time we are given, otherwise we would find it hard to experience our interactions as meaningful.¹⁵ What is created by such an atmosphere in our everyday interaction and in this unspoken coexistence of humans is what we call “leading our lives”. Underlying everything is our urge to make a difference, which is exactly why indifference experienced over a substantial stretch of time is such a despicable feeling. Indifference manifests itself in the experience of not caring for an outcome yet being invested in or consumed by it by coincidence. Robots, on the other hand, do not lead lives. Neither can we assume they have subjective experiences in the first place, nor do they participate in finitude in the way humans do.¹⁶

What they do when we encounter them is perceive information. Their being is one of processing not of experiencing; one of reacting, not of being. While they may store data in their protocols about the specific tasks they have performed at a specific time, they have not experienced historical events in the same way humans have. For humans, even times of leisure or complete inactivity hold significance, not only for regenerating their flesh and tissue but also for their distinct existence in the world. The experience of idling is of historical importance to our lives: we can feel bored or relaxed, in anticipation or depression; either way these seemingly blank slates are important for our stance towards our lives.¹⁷ For a robot, the moment they are not actively executing a task, their historical presence fades into oblivion. Objects such as robots lack an inherent history apart from being objects within the context of humans’ history. Technology, as far as we can tell, does not have a constant subjective flow¹⁸ of time, which constitutes a being in history but punctual blips of functionality. It fades into relevance whenever it does something and ceases to matter the moment it does not use data points for purposeful and functional interaction. The assumption of robots not having a subjective flow of time rests on the fact that we do ourselves not understand what brings forth subjective experience apart from describing brain activity, which is categorically different from the

15 See *Frankl: Ärztliche Seelsorge*, 118–120.

16 This aspect will be further articulated in the problem of incarnation.

17 A eulogy on the feeling of boredom was e.g. written in *Vodanovich: Benefits*.

18 See *Husserl: Lebenswelt*, 608–614.

knowledge about subjectivity which we are looking for. No matter how thin the slices of brain, how precise the maps of neurons, we would surely never find subjective experiences as such in neuroscientific explorations.¹⁹ Robots do not emerge from the cycle of life. The profound mystery we are not able to solve as humans is how the inanimate matter from which our flesh is made comes alive, which marks the border of our capabilities. We can transform matter into other matter, which is the robot, but we are not Frankenstein.

As with all points of contention before moving on, we should try to make the case that what is deemed as a problem is not solvable with the means at hand. So, what would solve the problem of identity? The first solution would be that having history is indeed not connected to indifference/difference. This argument would concede the possibility of machines not having history but calling into question why this would interfere with their ability to be different to us. The problem is the very concession because if they do not have history we rightfully do not care for their time. What may be deemed as a feature, them never being bored by a conversation with us mortals, is the very reason we might find such interactions futile. Since we are but data in their storage system perceived in a virtually infinite stream of data, we do not make a difference to them, which is precisely what we want in our interactions. Our indifference is met by theirs. Having history instead of time is hence not a convenient addition but an essential part of something making a difference in our encounter.

There is yet another angle here, which seems convincing at first glance but turns out to be anything but. For example, we could turn our sights to other inanimate objects that are indeed something we care for or make a difference to us. A cherished heirloom like Captain Koons' legendary watch in "Pulp Fiction" may serve as a connective thread in a person's autobiography. A house passed on through generations, a tree one planted years ago, the first drawing of one's child might be of similar importance. However, upon closer examination, it is precisely the moments when humans interacted with these otherwise unremarkable objects, when they play a meaningful role in human affairs, that we attribute significance to them. What makes things even worse is what the crude drawing of a two-

19 See *Henry*: Barbarei, 82–83; *Husserl*: Lebenswelt, 614–618.

year-old illustrates. Quality does not seem to matter, but the context. While things seem to indeed make a difference in our lives, they can be as much a trivial scribble or the uncontrollable occurrence of a seed growing rather than the highly sophisticated and resource intensive construction and integration of a religious robot. In other words, since religious robots do not have inherent connective tissue to an individual's personal history, they are as likely to be of difference to us than any other object we encounter every day.

2.2 The problem of incarnation

The second way out of our problem would be if we could reasonably show how robots could indeed become subjects of history. This leads us to a point amidst the problem of incarnation.

As any of us can undoubtedly tell what it feels like to experience ourselves, we should firstly try to describe what this feeling of radical subjectivity means and then explore whether it is reasonable to assume that robots experience the same thing. The challenge lies in the difference between hard bodies and soft flesh, which is tackled most profoundly in the works of the French philosopher Michel Henry. Henry's phenomenology of incarnation, or in other words, his phenomenology of the flesh, revolves around the problem of bodies that not only belong to the domain of the so-called hard sciences but have also been seamlessly integrated into our world view. He poses the question of whether the bodies of living subjects, all of which share the fact of being incarnated, being in the flesh, are identical to objective material bodies studied by quantum physics or organic chemistry. What materialist accounts argue, e.g. by those who regard humans as merely fleshy computers²⁰, overlooks the deep abyss between life and every other kind of body. The inert body that exists in countless numbers in the material universe is a product and result of the material universe, to which we apply physical laws with the knowledge of material processes with the intention of organising and combining them, ultimately making them available to us. Yet our own body, which we do not experience as such, even

20 See Fodor: Language; see Block: Mind.

if regarding it on a screen during an operation²¹, but as our flesh as the liminal zero of experiencing phenomena per se, is nothing we possess or have available as such. Such available bodies, like the robot's kit, do not feel anything. They do not feel themselves first and foremost, and they feel, desire, love or have affection for surrounding things even less. Our "bodies", which he refers to as flesh (fr.: chair), differ in that they perceive every nearby object in various ways. The perceptions of things are not the original experiences of the flesh (fr.: éprouve), but rather how the flesh experiences itself (fr.: s'éprouve).²² Henry argues that the mode in which anything shows itself thus becomes phenomenological, is in the boundaries of self-affection or self-experience (fr.: auto-affection). In his words, all we have access to is becoming reality in the subjective experience of us feeling ourselves as we are met with a phenomenon.²³ Hence, he explicitly separates life from data a machine can process.

There is the distinct experience of leaving the house around late September or early October and knowing, not by inferring from the date or checking certain criteria, that autumn is here. It is not the cold since there are cold summer days as well. It is neither the smell of the street, although things appear to smell differently then. It is not the colour of the leaves since we sense autumn even amidst the most concrete-laden parts of a city. It is also not the light, since it changes every minute of the day and is filtered and reflected in the architecture of a city many times. Yet the subjective experience, the staggering thought that "summer is gone", is so unique that almost everyone knows it when the time comes. While a robot's data might precisely tell their underlying infrastructure that they have just been hit by sunlight on a specific point on the electromagnetic spectrum, which number is displayed on their thermometer, and possibly what to do with that information given the functional task they are about to execute, there is no reason to assume that the robot experiences autumn in a mode of self-affection. The unique experience is not experienced as data, but as phenomena indescribable other than in urging others to remember or experience that said feeling. This radically subjective experience is exclusive to life as far as we can

21 See *Sartre*: Das Sein und das Nichts, 539–544.

22 See *Henry*: Inkarnation, 13–15.

23 See *Henry*: Phenomenology, 105–106.

tell since no scientific inquiry can encapsulate this phenomenon in its breadth and immediate being. No technology we can envision can bring forth this distinct experience, which we do not even have words to describe, yet know from firsthand experience.

Let us try to make the case against incarnation again. The argument would then say that our situation as flesh does not substantially differentiate us from a computer in a flesh-like shell. Our mental framework of life would hinge on the assumption that the totality of human experience can be described in terms of data being processed. Not only has this assumption been criticised widely²⁴, but my earlier phenomenological account on life has moreover been the effort to point out the curious experience of subjectively being alive, which is not a piece of information or an inducible state of mind but their prerequisite. So, while we could point to the technology of sensors as means to “experience” the world around the robot, we would still remain in the realm of the world’s sensory outwardness. Constructed under the premise of the world as data, computed and organised through software, which is the CPU in the robot and the network of neurons in our brains, we can readily admit that robots do sense the world. But it is unclear how the robot should be able to experience themselves as they sense the world. Yet this is, according to Henry, the primordial form of experience: experience is foremost an inward thing—not an outward experience²⁵. We don’t have experience of anything without the experience of ourselves having had the said experience. Yet this self-affection remains a mystery insofar as Henry plausibly assumes that the self-affection is a result of our incarnation²⁶. The important point of the Henry’s phenomenology—and Merleau-Ponty’s for that matter—is that the prominent framework of human beings as embodied brains, which are biological computers, fails to meet the experience of ourselves on a phenomenological level and only works if we assume a position beyond ourselves, which we cannot do because our flesh remains

24 E.g. by Puzio/*Filipović*: Informationsbündel, *Fuchs*: Defence, and *Loh*: Posthumanismus.

25 See *Henry*: *Phänomenologie*, 13–32.

26 While he methodically excludes other incarnated beings like animals since we have no way of knowing if they feel themselves, Henry leaves the door open to the idea that the same can be said of all beings of flesh.

attached to our thinking no matter how hard we try to transcend our bodies. Even thought is an experience of the flesh both are able to show.²⁷

There is no place on a robot you could put a sensor to bridge that gap, since our self-affection precedes our perception, as Henry argues, or our perception is ultimately linked to the flesh of the body.²⁸

2.3 The problem of purpose

So far, the argument has revolved mostly around taking the robot's perspective and contrasting it with a subjective human perspective. But what about the robot as the other, as a counterpart? While robots are unable to convince us they are humanlike when we try to walk in their shoes, facing them straight on might lead us to a compassionate feeling of togetherness, which might indeed spark a feeling of difference.

The idea of the other affecting us directly by just confronting ourselves with their "countenance" (fr. visage) was most profoundly championed by Emmanuel Levinas.²⁹ A closer examination of his argument, especially his critique of Kant and Heidegger, illuminates³⁰ why we misunderstand him, if we assume that having a face, which a robot very well might have, magically binds us in the spell of the other.

In a lecture delivered in 1975 on the subjects of death and time, Emmanuel Levinas introduced a notion that may not immediately resonate: "the countenance". Levinas offered an elucidation of this concept. What he intended, he explained, was the phenomenological description of what the annals of philosophy have encompassed within the field of the "soul". Whether it be Leibniz's Monad, Plato's soul that contemplates the Ideas or Spinoza's mode of thought, the countenance represents the non-objectified, abstract visage of my

27 See *Merleau-Ponty: Phänomenologie*, 174–177; *Henry: Inkarnation*, 108–112; *Henry: Psychoanalyse*, 207.

28 See *Merleau-Ponty: Sichtbare*, 172–189.

29 See *Levinas: Spur*, 115–119.

30 See *Levinas: Gott*, 69–78.

interlocutor. It is the countenance of the Other, from whom I can fundamentally anticipate a response to my existence, to which I find myself exposed and for which, according to Levinas, I bear responsibility. The countenance is that which turns towards me, endowing me with significance. The death of the Other, in contrast, unveils all that the countenance stirs within me. It signifies a worldly relationship eternally silenced in death, devoid of intentionality and self-interest. What is etched upon the visage is no longer the authentic expression of the self. The death of the Other thus always implies that I have failed in my responsibility towards them, for I have outlived them. My identity, which is not inherent or owned but rather bestowed upon me by the Other, fades into obscurity. The Other hence engages our consciousness by foreshadowing their potentially hardening countenance, becoming stiff as a waxen mask.³¹

Ironically, robots do not threaten us with this possibility; they encapsulate it. Being nothing more than animated matter, we must not fear their disappearance because they were never there in the first place. They do not hold the promise of un-lived life³², which is existentially open to the future, but are determined fully by their purpose and programming. The chasm of indifference/difference here alludes to the open and unfulfilled purpose of the lives yet to be lived and the dead, whose subjective purpose is concluded.

This is not to say that we are not touched by them, that their visage does not move us. There is even empirical data which backs the idea that we are able to be affected by a robot: there have been funerals held for robot dogs like AIBO³³ and heartfelt interactions with social bots like Cozmo³⁴. However, it is not coincidental that it is those two that come to mind as shining instances of human-robot interaction. Both the charming yet impractical Cozmo and the customisable robodog inherently possess an indeterminacy in their design, leaving their purpose open-ended. Just as we might find joy in observing a Havanese dog frolicking in the mud or

31 See *Levinas*: Gott, 21–23.

32 Levinas survived the Holocaust in war captivity, but his philosophy is undoubtedly tainted by the experience of countless friends and family members losing their lives in the concentration camps.

33 See *Arnold*: Funeral.

34 See *Chan et al.*: Cozmo; Disney is also working intensively on robots that emphasise being cute and relatable (see *Ackerman*, Disney).

experience exhilaration in witnessing a cat futilely chasing a red dot, it is precisely the shared openness to our future that connects us. However, religious robots, thus far, do not adhere to what I would like to call the “principle of futility”. Whether you listen to Mindar’s sermon³⁵, receive blessings from BlessU-2³⁶ or engage in prayer with SanTO³⁷, the purpose of these robots is always clearly defined. It is this very human ability to connect with human creativity that underscored the issue of history in the previous paragraph, and it now hampers religious robotics. The ambition to imitate and replicate humanness and the eagerness to fulfil a designated purpose impose limitations on these robots. They are clearly intended to replace humans in practices that humans designed to comfort themselves, convey meanings across generations or bring order to chaos. Yet, there are hardly convincing cases where robots would be better suited for religious practices than humans. While robots may possess the ability to be accessible from anywhere in the world and outlive humans by many years, as exemplified theoretically by Xian’er³⁸, in reality, their design primarily focuses on coherence and adherence to religious traditions, failing to explore their full potential as catalysts for creative processes that could pave the way for unforeseen religious experiences. Notably, it is secular artists like Diemut Strebe, with works such as “The Prayer”³⁹, who venture into innovative applications that challenge our conventional practices and strive to tap into the vast potential of robotics in creating new forms of religious expression.

What makes a difference here is unsurprisingly the indeterminate nature of something. It is no wonder we are indifferent to something which is already fully purposed. They are what they are made for—this might just be another way of saying tools or stuff⁴⁰. As long as robots remain obviously purposed, they will continue flirting with our indifference. We know them as we know their purpose. In contrast to human interaction, where two distinctly open and undecided

35 See *Samuel*: Robot.

36 See *Löffler* et al.: BlessU2.

37 See *Trovato* et al.: SanTO.

38 See *Cheong*: Rectitude; *Travagnin*: Buddha Halls.

39 See *Strebe*: Prayer.

40 Heidegger’s “Zeug” might be best translated this way (see *Heidegger*: Sein, 68).

futures meet, one of the two is already fitted into a useful framework when we face most robots.

3. Different robots

What does it take for robots to captivate us in such a way that they make a difference to us? A solution has already been hinted at: while not at all similar to us humans, unable to evoke the same sense of ethical responsibility, the more unspecific they are in their design, the muddier their purpose, the more they might be of importance to us. Interestingly enough, making them more human might only amplify the problems I have laid out here; making them less human, less useful and less intelligible might actually make them different. Most robots are tools and made in a mental framework that views them as such. As robots originate from the idea of a handy servant⁴¹, most of the ones equipped with the ability to act as one bring with them a mindset of engineering, which is one of technology, which is one of controlling, forming and manipulating matter in a helpful way.⁴²

The different robot, which I deem promising for the religious sphere, emerges from the arts not from technology. As a conductor of creativity for both the creator and the user, their purpose must be open and undecided. One of the most promising models for different robots is something that oddly enough perfectly exemplifies the stale state of churches in our times: the organ. The enduring appeal of this colossal instrument across centuries can be attributed to its role as an open conduit for artistic expression. Constrained only by the range of sounds it can produce, the organ has offered successive generations of composers, musicians and audiences the canvas on which to etch their experiences. Within its aesthetic spectrum, it has accommodated a myriad of emotional expressions.

Like other sacred instruments, organs wield the power of creation. While some captivate through their sheer presence and artistry, they serve as untapped reservoirs of potential, enabling the creation and reception of religious experiences beyond the confines of words.

41 Culturally intertwined with the myth of the Golem (see *Contrada*: Golem).

42 See *Henry*: Barbarei, 176–178.

For robots to be genuinely integrated into the sacred sphere, they must evolve into conduits of creative and, to some extent, religious expression for all parties involved: the creator, the conductor (if needed) and the recipient.

Consider, for instance, a robot that resembles nothing more than a metallic arm, its algorithm translating religious texts or the congregation's prayers into imagery, which it then delicately adorns upon the canvas of a church mural. Continuously layering over what already exists, this robot embarks on a perpetual journey of creative exploration. Such a robot would simultaneously create the concrete representation of words, which are dear to the community and tell their story, while being open to experiences, leaving open what this mural is supposed to mean, why it is there in the first place and what will become of it once the robot is turned off.

Alternatively, picture a mechanised mannequin that choreographs its movements in synchrony with an orchestra's conductor, a mesmerising dance to accompany Bach or Mahler unfurling in ever-changing patterns before the congregation. These diverse robots may be ambiguously purposed—undoubtedly constrained by their capabilities and their specific setting—but they stand as vehicles that can express and experience religious sentiment for the humans involved.

Robots might remain different to their core, yet they might make a difference to us once we embrace their difference and abstain from the temptation to make them “un-different” from us.

Bibliography

- Ackerman*, Evan: How Disney Packed Big Emotion Into a Little Robot. Melding animation and reinforcement learning for free-ranging emotive performances, in: IEEE Spectrum, 06.10.2023. Online at: <https://spectrum.ieee.org/disney-robot> [status: 27.10.2023].
- Arnold*, Michael/*Gould*, Hannah/*Kohn*, Tamara et al.: Cybernetic Funeral Systems, in: IEEE Conference on Norbert, 2021, 1–4.
- Block*, Ned: The Mind as the Software of the Brain, in: Osherson, D./Gleitman, L./Kosslyn, S. et al. (eds.): Invitation to Cognitive Science. Second Edition. Vol. 3, Cambridge 1995, 377–425.
- Camus*, Albert: Der Fremde, Hamburg 2021.
- Chan*, Lilian/*Zhang*, Brian J./*Fitter*, Naomi T.: Designing and Validating Expressive Cozmo Behaviors for Accurately Conveying Emotions, in: 30th IEEE International Conference, 2021, 1037–1044.

- Cheong*, Pauline Hope: Religion, Robots and Rectitude: Communicative Affordances for Spiritual Knowledge and Community, in: Applied Artificial Intelligence 34 (5), 2020, 412–431.
- Contrada*, Norma: Golem and Robot. A Search for Connections, in: Journal of the Fantastic in the Arts 7 (2/3), 1995, 244–254.
- Eva*, Benjamin: Principles of Indifference. Journal of Philosophy 116 (7), 2019, 390–411.
- Fisher*, Carrie: The Arrival of the Jedi, 31.03.2003. Online at: https://content.time.com/time/specials/packages/article/0,28804,1977881_1977891_1978545,00.html [status: 27.10.2023].
- Fodor*, Jerry A.: LOT 2. The Language of Thought Revisited, Oxford 2008.
- Frankfurt*, Harry: The Importance of What We Care About, Cambridge 1988.
- Frankl*, Viktor E.: Ärztliche Seelsorge. Grundlagen der Logotherapie und Existenzanalyse, München 2020.
- Fuchs*, Thomas: In Defence of the Human Being. Foundational Questions of an Embodied Anthropology, Berlin 2021.
- Henry*, Michel: Die Barbarei. Eine phänomenologische Kulturkritik, München 1994.
- Henry*, Michel: Inkarnation. Eine Philosophie des Fleisches, Freiburg im Breisgau 2011.
- Henry*, Michel: Phenomenology of Life, in: Angelaki 8 (2), 2003, 97–110.
- Henry*, Michel: Phänomenologie und Psychoanalyse, in: *Henry*, Michel: Radikale Lebensphänomenologie. Ausgewählte Studien zur Phänomenologie, Freiburg/München 1992, 187–212.
- Heidegger*, Martin: Grundbegriffe der Metaphysik. Welt – Endlichkeit – Einsamkeit (Gesamtausgabe. Band 29/30: II. Abteilung: Vorlesungen 1923–1944), Frankfurt am Main 1983.
- Heidegger*, Martin: Sein und Zeit, Tübingen 1967.
- Husserl*, Edmund: Die Lebenswelt. Auslegungen der vorgegebenen Welt und ihrer Konstitution. Texte aus dem Nachlass (1916–1937) (HUA XXXIX), Dordrecht 2008.
- James*, William: The Principles of Psychology (1890). Volume I, New York 1918.
- James*, William: The Varieties of Religious Experience. A Study in Human Nature (Cambridge Library Collection), Cambridge 2012.
- Jordan*, John M.: Robots (MIT Press Essential Knowledge Series), Cambridge 2016.
- Levinas*, Emmanuel: Die Spur des Anderen. Untersuchungen zur Phänomenologie und Sozialphilosophie, Freiburg im Breisgau 2017.
- Levinas*, Emmanuel: Gott, der Tod und die Zeit (Passagen Forum), Wien 2013.

- 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), 2019, 569–586.
- Loh, Janina*: *Trans- und Posthumanismus. Zur Einführung*, Hamburg 2023.
- Maslow, Abraham H.*: Lessons from the Peak Experiences, in: *Journal of Humanistic Psychology* 2 (1), 1962, 9–18.
- Marr, Bernhard*: The four Ds of Robotisation, 16.10.2017. Online at: <https://www.forbes.com/sites/bernardmarr/2017/10/16/the-4-ds-of-robotization-dull-dirty-dangerous-and-dear/?sh=4a04a6473e0d> [status: 27.10.2023].
- Merleau-Ponty, Maurice*: *Das Sichtbare und das Unsichtbare. Gefolgt von Arbeitsnotizen (Übergänge Bd. 13)*, München 1986.
- Merleau-Ponty, Maurice*: *Phänomenologie der Wahrnehmung (Phänomenologisch-psychologische Forschung 7)*, Berlin 1966.
- Puzio, Anna/Filipović Alexander*: Personen als Informationsbündel? Informationsethische Perspektiven auf den Gesundheitsbereich, in: *Fritz, A./Mandry, C./Proft, I.* (eds.): *Digitalisierung im Gesundheitswesen. Anthropologische und ethische Herausforderungen der Mensch-Maschine-Interaktion*, Freiburg im Breisgau 2021, 89–113.
- Samuel, Sigal*: Robot Priests Can Bless You, Advise You, and Even Perform Your Funeral. AI Religion is Upon Us. Welcome to the Future, in: *VOX*, 13.01.2020. Online at: <https://www.vox.com/future-perfect/2019/9/9/20851753/ai-religion-robot-priest-mindar-buddhism-christianity> [status: 27.10.2023].
- Sartre, Jean-Paul*: *Das Sein und das Nichts. Versuch einer phänomenologischen Ontologie (Jean Paul Sartre. Gesammelte Werke in Einzelausgaben. Philosophische Schriften. Band 3)*, Hamburg 2020.
- Schütz, Alfred*: *Der sinnhafte Aufbau der sozialen Welt. Eine Einleitung in die verstehende Soziologie*, Frankfurt am Main 2016.
- Strebe, Diemut*: “Diemut Strebe / The Prayer”. 2020. Online at: <https://theprayer.diemutstrebe.com> [status: 27.10.2023].
- Travagnin, Stefania*: From Online Buddha Halls to Robot-Monks, in: *Review of Religion and Chinese Society* 7 (1), 2020, 120–148.
- Trovato, Gabriele/Lucho, Cesar/Ramon, Alvaro/et al.*: The Creation of SanTO. A Robot with “Divine” Features, in: *IEEE 15th International Conference 2018*, 437–442.
- Vodanovich, Stephen J.*: On the Possible Benefits of Boredom. A Neglected Area in Personality Research, *Psychology and Education: An Interdisciplinary Journal*, 40 (3–4), 2003, 28–33.

Max Tretter

“Where do all the calculators go?”

Methodological Considerations About how to Think About Robots and Their Afterlife

Abstract

“Where do all the calculators go?” Starting from this question and noting that the afterlife of robots is a theme frequently explored in popular culture yet seldom addressed in academia, this article investigates whether it is possible to address such questions in a methodologically plausible and scientifically comprehensible manner. In this paper, I present and critically evaluate three distinct approaches, each designed to enable scholarly investigation—compliant with academic standards—into the sample question of whether an afterlife for robots exists, or whether it is plausible to assume its existence: 1) a philosophical–theological approach that explores the properties of robots and their theological significance, 2) a psychological–theological approach that considers the possibilities of artificial spirituality, and 3) a Biblical–theological approach that enquires into the eschatological whereabouts of all creation.

1. But where do all the calculators go?

In a scene from the third season of the British cult television series *Red Dwarf*, Dave Lister, a low-ranking technician who awakens as the last surviving human on a mining spaceship following a three-million-year period of suspended animation, engages in conversation with Kryten, his anthropomorphic service mechanoid. Upon realising that a replacement android for Kryten is *en route* to their spacecraft and with a 24-hour countdown initiated that marks the time until Kryten’s physical and mental functions will be irrevocably

deactivated and he will be replaced, a dialogue unfolds between both characters. When confronted with Dave's question about how he could accept his impending "end" without resistance, Kryten replies: "Oh, it's not the end for me, Sir; it's merely the beginning. I have served my human masters, and now I can look forward to my reward in Silicon Heaven." Clearly taken aback by this response, Dave retorts: "Silicon what?"—a quip that gives rise to an amusing exchange rich in wordplay.

Kryten: "Surely you have heard of Silicon Heaven. [...] It's the electronic afterlife. It's the gathering place for all the souls of electrical equipment. Robots, calculators, toasters, hairdryers. It's our final resting place."

Dave: "I don't mean to say anything out of place here, Kryten, but that's completely whacko-jacko. There is no such thing as Silicon Heaven."

Kryten: "Oh! Then, where do all the calculators go?"

Dave: "They don't go anywhere! They just die! [...] Machines do not have souls. Computers and calculators don't have an afterlife. You don't get hairdryers with tiny little wings sitting on clouds playing harps."

Kryten: "But of course you do! [...] Oh, it's common sense, Sir. If there weren't a better life to look forward to, why on earth would machines spend the whole of their lives servicing humankind? Now that would be really dumb!"¹

In a humorously poignant manner, this dialogue scene—which is nearly 35 years old, from the episode *The Last Day*, which aired in December 1989—poses a compelling question: What does the afterlife look like for non-biological forms of intelligence?² What happens to robots when they 'die'? *Red Dwarf* is not the only TV series that explores the afterlife of robots. Similar explorations can be found in other works as well. One example is the *Futurama* episode entitled *Ghost in the Machines*,³ where "Robot Heaven" and "Robot Hell" are portrayed as two distinct and identifiable places where

1 *Ed Bye*: The last day.

2 The discussions within the scene extend beyond non-biological intelligences like robots, delving into the afterlife of electronic devices, including calculators and toasters. For the scope of this article, however, I will confine my focus to robots and artificial intelligence, sidestepping the discourse on the fate of everyday electronic devices and the like.

3 *Claffey*: Ghost in the machines.

robots—or their source codes—are sent upon their deaths. Another example is the episode *Zima Blue* from the animation series *Love, Death & Robots*,⁴ which traces the journey of a self-aware robot in its quest for “completion”, a state it ultimately achieves by deactivating its advanced cognitive functions, shedding its intricate sensors and reverting to a simple pool-cleaning robot, tirelessly dedicated to its singular task. Additionally, in the third volume of the comic series *Descender*, entitled *Singularities*,⁵ a cyborg’s near-death experience is vividly portrayed, including an out-of-body encounter with their human creator. Finally, a somewhat different example is found in Steven Spielberg’s film *A.I. Artificial Intelligence*,⁶ which follows a boy-like cyborg on a Pinocchio-esque journey after his initial family abandons him, forsaking their interest in him, and, by abandoning him in a forest, releases him from his former life.

2. Research overview and hypothesis

As this list of pop culture engagements with robotic afterlife—which could easily be expanded—demonstrates, there is a lively interest in this topic. Academically, a substantial and diverse field of experimentation and research has emerged at the intersection of robotics and theology.⁷

There are initiatives, for instance, aimed at developing robots for religious purposes.⁸ Examples include robots that deliver pre-programmed sermons, such as the anthropomorphic, Buddhist-like android *Mindar*;⁹ robots that reside in religious places and answer visitors’ religious questions while assisting them in religious practices, like the robot monk *Xian'er*,¹⁰ or robots like *BlessU-2*, a repurposed ATM that offers blessings at the push of a button.¹¹

4 Pennacchioli/Valley: *Zima Blue*.

5 Lemire: *Singularities*.

6 Spielberg: *A.I. Artificial Intelligence*.

7 Balle: Theological dimensions of humanlike robots.

8 Simmerlein/Tretter: Robots in religious practices; Trovato et al.: Religion and robots.

9 Hardingham-Gill: The android priest.

10 Travagnin: Online Buddha.

11 Löffler et al.: Blessing robot BlessU2.

From the theological side, there are pastoral reflections on whether and what roles religious robots can or should assume.¹² This includes, for instance, questions about the feasibility and desirability of building robot pastors or spiritual caregivers,¹³ the prudence of deploying robots with their “spiritual superpowers”¹⁴ as pastoral assistants to support human clergy,¹⁵ as well as what pastoral skills and virtues are necessary in a world where AI is assuming an increasingly significant role.¹⁶

Moreover, there are psychological and sociological studies that examine how individuals respond to robots in religious contexts,¹⁷ how the design of these robots shapes people’s reactions¹⁸ and how different cultural contexts modulate these responses.¹⁹

And not least, a wealth of contemplation is occurring within the realm of systematic theology.²⁰ Here, scholars probe the potential religious status of robots: Do they, for instance, possess a soul?²¹ Are they to be considered people,²² images of God²³ or images of images of God?²⁴ Inquiries have been made into what we can learn from robots about our own humanity: In an era of artificial entities, how must we reinterpret us being images of God?²⁵ Additionally, theologians are questioning the moral status that ought to be attributed to robots: Should they be regarded as moral patients, moral agents or neither?²⁶

12 Puzio: Robot, let us pray; Tretter: Courage for theology; Simmerlein/Tretter: What about spiritual needs?

13 Simmerlein/Tretter: KI in der seelsorglichen Beratung; Simmerlein/Tretter: What about spiritual needs?

14 Löffler/Hassenzahl: Robots’ spiritual superpowers.

15 Puzio: Robot theology; Smith: Robot theology; Tretter: Shortage of pastors.

16 Hamman: Pastoral virtues.

17 Löffler et al.: From experiential to existential questions.

18 Löffler et al.: Blessing robot BlessU2; Trovato et al.: Design strategies.

19 Trovato: Pioneering religion in robotics; Trovato et al.: Editorial introduction.

20 Geraci: Religion for the robots; Smith: Robot theology; Tretter: Robot theology.

21 Livingston/Herzfeld: Could robots have souls?

22 Reiss: Is it possible; Turner: Will we know them?

23 Dorobantu: Will robots too be in the image of God?; Foerst: Robots and theology.

24 Midson: In the image of the image?

25 Dorobantu: Cognitive vulnerability; Dorobantu: Imago Dei.

26 Metzler: Moral status to service robots; Smith: Robot theology.

Although there are now some initial publications on the eschatological questions concerning robots,²⁷ the academic exploration of robots’ afterlife remains notably sparse. Apart from the aesthetic scepticism that this subject might evoke, one reason for this deficiency may be the methodological complexity associated with probing such a question, for eschatology—that is, the doctrine of the ultimate or final things concerning what transpires at the end of time and beyond the here and now—is intrinsically a highly speculative domain. It is a field that one can approach only cautiously and always with substantial reservation.²⁸ Linking this field with robotics—which is often seen as emblematic of technology and its strict, rigorously empirical approach, and thus as the stark antithesis of eschatological contemplation—makes the task of inquiry even more challenging and raises methodological questions.

The aim of my contribution is to help fill this gap in the research on robotics and religion. I intend to present three approaches by which one can, from a Protestant theological perspective, plausibly engage with the basic question of *whether it is plausible to assume that there is an afterlife for robots* and arrive at comprehensible conclusions. To this end, I will conduct a methodological experiment: In the following chapter, I will sequentially introduce these three approaches and critically discuss their respective advantages and disadvantages. Following this, I will engage in a reflective discussion of these methodological considerations, confronting the limitations of my representations, illuminating the ethical facets of this subject matter and drawing several inferences for eschatological thinking within theology. Ultimately, I will synthesise the findings in a conclusion.

Accordingly, the objective of this article is less to craft a clear answer to the question of whether or not it is plausible to assume an afterlife for robots—or how this could plausibly be conceived.²⁹ Rather, the focus is to demonstrate that it is *possible* to approach this topic in a methodologically sound manner and to arrive at comprehensible conclusions.

27 *Balle*: Theological dimensions of humanlike robots.

28 *Härle*: Outline of Christian doctrine; *Mühling*: Handbook of Christian eschatology; *Walls*: Handbook of eschatology.

29 *Tretter*: Afterlife for Robots.

3. How to think about robots and their afterlife: three approaches

There are several ways to approach the question of whether it is plausible to assume the existence of such an afterlife. I will present three such approaches in this chapter.

3.1 Philosophical–theological approach: Robots and their properties

The first approach is of a philosophical–theological nature and begins by enquiring into the capabilities of robots. It can draw on a range of theoretical, empirical and technical groundwork from various disciplines, including robotics, computer science and engineering. For example, it may establish:

- how well and in what manner robots can solve predefined problems
- how well and in what way artificially intelligent entities can reflect on their own actions and adapt them
- how independently robots act and whether they can set their own goals
- what movements robots can perform and how well they can navigate in complex environments
- how well artificially intelligent entities understand linguistic commands and communicate or interact with humans or each other

In a subsequent step, this approach interprets these observations through the lens of information and technology philosophy, thereby drawing conclusions about the inherent properties of robots. The problem-solving abilities of artificially intelligent entities, for example, yield insights into their intelligence and creativity.³⁰ Conclusions may be drawn regarding robots' levels of autonomy, freedom and sense of responsibility from their capacity for independent action.³¹ And their movements, orientation in environments and language skills can provide conclusions about their embodiment, world un-

30 Rauterberg: Die Kunst der Zukunft.

31 Nyholm: Humans and robots; Coeckelbergh: Robot ethics.

derstanding and sociality. Thus, it is feasible to delineate a set of properties attributable to robots based on their tangible capabilities.

The third step then involves assessing the theological significance of these robotic properties. As some scholars propose, attributes such as the intelligence of robots,³² their “deep levels of [...] judgment”³³ or their embodiment, empathy and sociality³⁴ may be regarded as properties of theological significance. From these theologically significant properties, as some authors suggest, one might conclude that robots, exhibiting these traits, must also possess a soul³⁵ and be regarded as images of God³⁶ or as images of images of God,³⁷ and, like humans, should be classified as sinners³⁸ or, at the very least, as participants in human sin.³⁹

The final step involves drawing conclusions from these theological insights about the plausibility of an afterlife for robots. For instance, if one arrives at the conclusion, as DeBaets does, that robots are sinners⁴⁰ or, as Smith estimates, that robots participate in human sin,⁴¹ then it is only plausible to assume that some form of judgement or reparation is required. Otherwise, the category of sinfulness would have no consequences and would be irrelevant. In a similar vein, the recognition that robots may possess a soul could be interpreted as suggestive of the necessity for an afterlife for robots. After all, what would otherwise become of the souls of robots?

In this four-step process—beginning with the identification of robots’ capabilities, progressing to the discernment of robotic properties and culminating in the assessment of their theological significance—one can deduce whether the concept of an afterlife for robots is plausible or not.

32 *Furse*: The theology of robots.

33 *Cantwell Smith*: Artificial Intelligence and ultimate questions.

34 *DeBaets*: The robot as person.

35 *Livingston/Herzfeld*: Could robots have souls?

36 *Foerst*: Robots and theology; *Foerst*: Cog, a humanoid robot, and the question of the image of God; *Dorobantu*: Will robots too be in the image of God?

37 *Midson*: In the image of the image?

38 *DeBaets*: The robot as person.

39 *Smith*: Robot theology.

40 *DeBaets*: The robot as person.

41 *Smith*: Robot theology.

Certainly, this approach can lead to varying conclusions. Some authors, for example, point out that robots presumably lack free will,⁴² that they are not conscious entities⁴³ and that they fall short in terms of relationality and vulnerability.⁴⁴ They are also said to lack the capacity for empathy and value,⁴⁵ and it is noted that robots are neither born nor die (*Step 2*).⁴⁶ Consequently, these authors argue that robots are devoid of key theological capabilities and qualities (*Step 3*). Without such theologically significant capabilities and qualities, robots would presumably have no need for reconciliation or consummation,⁴⁷ which can ultimately be interpreted as evidence that suggests either the absence of an afterlife for robots or, at the very least, the absence of a necessity for such an afterlife (*Step 4*).

A significant strength of this approach lies in its grounded beginning at *Step 1*, where it starts with a fundamental description of robots' capabilities. Yet, by *Step 2*, the process becomes less tangible when the task shifts to drawing conclusions about the properties of robots based on these capability descriptions. Although this process can be made highly transparent—by clearly stipulating, for example, the capabilities and criteria used to deduce that robots possess certain properties—these conclusions themselves can be contested. Some scholars, citing robots' prowess in complex games like *Chess* or *Go*, argue in favour of robot intelligence, while others challenge this view.⁴⁸ These sceptics, raising the bar as to what constitutes intelligence, concede that robots perform complex, albeit unintelligent, calculations. Likewise, in *Step 3*, differing assessments of the theological significance of specific robot properties can emerge. Even when scholars attribute the same properties to robots, they may arrive at divergent theological assessments, such as whether robots are sinful or possess a soul. Finally, the conclusion in *Step 4* remains highly speculative. For example, assuming hypothetically that robots have a soul does not necessarily guarantee the existence of an afterlife for robots. Though it is a common belief that beings with souls

42 *McGrath*: Robots, rights and religion.

43 *Strand*: Will androids need salvation?

44 *Dorobantu*: Cognitive vulnerability; *Dorobantu*: Imago Dei.

45 *Gill*: Jesus wept, robots can't.

46 *Krajewski*: Can a Robot Be grateful?; *Deli*: Do robots die?

47 *Swann*: Anima ex machina; *Strand*: Will androids need salvation?

48 *Rosenfeld*: Religion and the robot.

experience an afterlife, this conclusion is not necessarily mandatory. Conversely, the inverse conclusion—that the hypothetical absence of a soul in robots precludes an afterlife for them—is similarly not inevitable. While certain theologically significant properties can justifiably be interpreted as plausible indicators of the existence or non-existence of an afterlife for robots, these indicators remain suggestive rather than constituting definitive proof.

3.2 Psychological–philosophical approach: Robots and spiritual intelligence

The second approach adopts a psychological–philosophical lens, emphasising spirituality and spiritual intelligence. At its outset, it seeks to elucidate the nature of spiritual intelligence. Central to this endeavour is Howard Gardner’s categorisation of diverse forms of intelligence⁴⁹ and Robert Emmons’ subsequent proposition of spiritual intelligence as a distinct form.⁵⁰ Engaging deeply in the multifaceted debates ignited by this thesis,⁵¹ this approach employs a multidisciplinary methodology to first delineate the hallmarks of spiritual intelligence and identify the necessary conditions to manifest it.⁵²

Once these foundational questions about spirituality and intelligence have been addressed, robots are brought into the analytical frame. The inquiry then shifts to discern whether they possess the essential prerequisites for spiritual intelligence and, furthermore, whether they can exhibit spiritual cognition and/or ponder on their post-mortal fate.⁵³

If this inquiry into the possibility of artificial spirituality or “technoreligiosity”⁵⁴ is answered affirmatively, it becomes feasible—apart

49 *Garnder*: Intelligence reframed.

50 *Emmons*: The psychology of ultimate concerns.

51 *Emmons*: Spirituality and intelligence; *Garnder*: A case against spiritual intelligence; *Kwilecki*: Spiritual intelligence; *Mayer*: Spiritual intelligence.

52 *Dorobantu/Watts*: Spiritual intelligence; *Watts/Dorobantu*: Is there spiritual intelligence?

53 *Dorobantu/Watts*: Spiritual intelligence.

54 *McBride*: The advent of postmodern robotic technoreligiosity.

from addressing the numerous subsequent questions that arise⁵⁵—to draw conclusions about the plausibility of an afterlife for robots. A plausible argument might contend that it borders on cruelty for robots to possess spiritual cognisance, contemplate their afterlife and nurture aspirations, only to find that no such afterlife exists for them. Such a situation appears inconsistent with the concept of a loving God, which might be seen as a (tentative) theological indication that there must be some form of afterlife for robots.

Much like the philosophical–theological approach, a salient strength of this psychological–theological approach is that its first step is very grounded in scientific and occasionally even empirical reflections about spirituality and spiritual intelligence, thus making this approach empirically robust. However, the subsequent steps navigate more treacherous terrain. Indeed, adjudicating whether robots genuinely possess spiritual intelligence or can experience spirituality is methodologically intricate. While, when starting with a precise definition of spiritual intelligence and clearly outlining the necessary criteria to possess it, it might be possible to determine if robots inherently have the *potential* for spiritual intelligence or even spiritual experiences, deducing from this capability that they actually are spiritual seems somewhat premature. The third step, extrapolating from the possible existence of robotic spirituality to implications about an afterlife for robots, is even more fraught. Deriving the existence of a robotic afterlife from its mere conceivability resonates structurally with historical attempts to prove God’s existence—endeavours that were nurtured over a long period of time but ultimately failed.⁵⁶ Such conclusions exceed the boundaries of permissible reasoning outlined by Kant in his critiques of reason and his writings on the limits of metaphysics,⁵⁷ thus crossing the limits of considerations that are theologically reasonable.

55 For instance, what might the belief systems of artificially intelligent entities or robots encompass in terms of content? See *Klinge*: Do robots believe in electric gods?; *Sampath*: From Heidegger on technology to an inclusive pluralistic theology. Or how might they reconcile or confront pre-existing beliefs? See: *Dorobantu/Watts*: Spiritual intelligence.

56 *Bromand/Kreis*: Gottesbeweise.

57 *Sala*: Kant und die Frage nach Gott.

3.3 Biblical–theological approach: The eschatological whereabouts of all creation

The third approach adopts a Biblical–theological perspective. It doesn’t delve into the attributes robots might possess, nor does it ponder the feasibility of artificial spirituality in robots to deduce implications about a robotic afterlife. Rather, this approach is anchored in a rigorous engagement with both Biblical texts and the wider theological tradition, interrogating the eschatological question of the whereabouts of all creation—and from there, drawing conclusions about the existence of an afterlife for robots, who, just like humans, animals, plants or material objects, are part of creation.

Scripture offers several passages addressing the eschatological whereabouts of creation. Notably, none of these passages are primarily preoccupied with speculating about the eschatological fate of creation. Instead, they predominantly serve a consolatory purpose. Their primary impetus is to bolster the spirits of believers amidst contemporary adversities or doubts, infuse them with renewed hope and fortify their faith. The eschatological discussions about creation within these contexts elucidate the magnitude of God’s redemptive work, aiming to strengthen believers’ faith. Such reassuring reflections on the eschatological destiny of creation are prominently articulated, for instance, in Paul’s Epistle to the Romans and in the concluding segments of the Book of Revelation. I will briefly quote these passages and present their Biblical context to provide a better insight into the Bible’s statements on the eschatological fate of creation and their textual intentions. I will begin with a section from Paul’s letter to the Romans (8:18–23).

¹⁸ For I reckon that the sufferings of this present time are not worthy to be compared with the glory which shall be revealed in us. ¹⁹ For the earnest expectation of the creature waiteth for the manifestation of the sons of God. ²⁰ For the creature (κτίσις) was made subject to vanity, not willingly, but by reason of him who hath subjected the same in hope, ²¹ because the creature itself also shall be delivered from the bondage of corruption into the glorious liberty of the children of God. ²² For we know that the whole creation groaneth and travaileth in pain together until now. ²³ And not only they, but ourselves also, which have the first fruits of the Spirit, even we ourselves groan within ourselves, waiting for the adoption, to wit, the redemption of our body. (Rom 8:18–23 KJV)

Paul's Epistle to the Romans pivots on a central theme: the righteousness of God. After Paul addresses the necessity of this righteousness in 1:18–3:20 (why is humankind reliant upon being justified by God?) and the possibility of God's righteousness in 3:21–4:25 (how can humanity partake in this righteousness of God?), and before he broaches Israel's relationship with this righteousness in 9:1–11:36 (how can Israel attain God's righteousness?), Paul delves into the reality of God's righteousness in 5:1–8:39.⁵⁸ Herein, through references to baptism (Chapter 6) and the law and sin (Chapter 7), Paul sketches the contours of a life anchored in God's righteousness. In Chapter 8, likely to alleviate doubts among his audience regarding their salvation, he underscores that there is no condemnation for those living in the Spirit of God. To further bolster this hope, he references the fate of the entire creation in 8:18–23. In these verses, which stand as the apogee of Pauline argumentation in the Epistle to the Romans,⁵⁹ the apostle interweaves Christian redemption with the hope of the entirety of creation—the term *κτίσις* in the New Testament⁶⁰ and in Paul's letters⁶¹ included the entire cosmos, all that is visible and invisible, created and sustained by God through Christ *ex nihilo*—for reconciliation⁶² and redemption from current suffering.⁶³ Within the context of the Epistle to the Romans, alluding to the eschatological fate of all creation serves as an affirming argument to fortify the hope of the believers addressed. Thus, while the eschatological fate of creation is not the primary focus of the text, it is nonetheless broached.

Similar statements, nestled within similar argumentative frameworks, echo in the Colossian Hymn (notably Col 1:19–20), in 2 Peter 3:12–13 and in Isaiah's resonant passages, especially Isaiah 65:17 and Isaiah 66:22. They also resurface in the concluding chapters of the Book of Revelation.

¹ And I saw a new heaven and a new earth: for the first heaven and the first earth were passed away; and there was no more sea. ² And I John saw the holy city, new Jerusalem, coming down from God out of heaven

58 *Schnelle*: Einleitung in das Neue Testament, 134–158.

59 *Dunn*: Romans 1–8, 466–467.

60 *Foerster*: κτίσις.

61 *Wischneyer*: ΦΥΣΙΣ und ΚΤΙΣΙΣ bei Paulus.

62 *Hahne*: The corruption and redemption of creation.

63 *Gibbs*: Creation and redemption, 34–47.

[...].⁵ And he that sat upon the throne said, Behold, I make all things new. (Rev 21:1–2a.5a)

In a vast swath of apocalyptic literature, intricate visions are woven, often drawing and re-contextualising Old Testament motifs.⁶⁴ These motifs are then successively echoed and restructured in Hellenistic, Jewish and early Christian compositions.⁶⁵ Within these layers of textual interplay, the challenges confronting the believers addressed—both those imminent and those currently being faced—are vividly portrayed. However, the Book of Revelation, amidst its vibrant tapestry, offers a palpable undercurrent of solace.⁶⁶ The primary message is that despite the present tribulations and how bleak the situation may seem, God will ultimately prevail and surmount all adversities. This is intended to infuse the faithful with renewed hope and resilience.

To underscore the magnitude of God’s power, the author of Revelation delineates in chapters 21 and 22 how God will not only vanquish the adversities of this world but will forge a new heaven and a new earth. This duality, as seen in Genesis 1:1, underscores the totality of God’s forthcoming realm. In this renewed existence, the sufferings of the present world will be absent. Intriguingly, the narrative posits that the human-crafted city of Jerusalem will form an integral part of this renewed cosmos. This suggests an interesting theological stance: human “creations” are not merely ephemeral constructs but have a place in God’s grand design.⁶⁷ This narrative offers solace by connecting the destinies of its audience with the entirety of creation.

When considering the Biblical testimony about the eschatological whereabouts of creation, we can observe two things: First, the eschatological destiny of creation is invariably situated within consolatory contexts, with the primary aim of imparting solace and fortitude to believers; second, conceptions of the end-time fate of creation only developed late within the context of Biblical genesis. Yet, the overarching Biblical ethos anticipates continuance for all of creation. Creation won’t permanently perish but will be sustained

64 *Strobel*: Apokalypse des Johannes.

65 *Berger*: Die Apokalypse des Johannes II, 1375–1417.

66 *Strobel*: Apokalypse des Johannes.

67 *Breytenbach*: Schöpfer/Schöpfung.

and/or recreated by God. These Biblical concepts have also left indelible marks on doctrinal formation. Without us delving deeply into creation theology, there is an established motif of a so-called *creatio nova*,⁶⁸ and eschatological reflections develop notions of the completion of the entire world.⁶⁹

Thus, embedded deeply within the Christian theological tradition is a resounding affirmation: the *entirety* of creation is destined for an afterlife. As elucidated in the writings of Paul, redemption extends across the whole cosmos, including every visible and invisible entity conceived and upheld by God through Christ *ex nihilo*. And as emphasised in the Apocalypse, human creations, such as cities or robots, shouldn't be exempted from these end-time events. Given this profound theological backdrop, it seems almost imperative to argue for an afterlife for robots—regardless of whether this is imagined as reconciliation, a new creation or in some other way. Reversing this perspective only strengthens the argument: it would be contradictory to embrace an afterlife for everything, as Paul, John, Isaiah and much of the Biblical–theological tradition do, while denying the same concept for robots.

This approach's strength lies in its robust internal coherence. It avoids unwarranted leaps, such as moving from the robots' properties to postulating the existence of an afterlife for them. Instead, its aim is to descriptively highlight beliefs already present within Christian tradition, arguing that, given this background, it would be inconsistent to reject the idea of an afterlife for robots. Yet, a caveat remains: this approach is predicated on shared theological convictions. Its efficacy is contingent upon the recipient's adherence to the Christian eschatological vision. If one doesn't adhere to these beliefs, this Biblical–theological approach falls flat, marking it as an approach that works exclusively within theological confines. That said, it's worth noting that without this or a similar religious background, the whole question of an afterlife for robots probably would not have been asked.

68 *Stock*: *Creatio nova*.

69 *Härle*: Outline of Christian doctrine.

4. Discussion

In the preceding chapter, three approaches were introduced, offering methodologically sound ways to address the question of an afterlife for robots, and their respective advantages and disadvantages were examined. In this chapter, these methodological presentations will be discussed. This discussion will first highlight their limitations and potential avenues for further research, then uncover the ethical dimensions of these arguments and conclude with their eschatological implications.

4.1 Limitations and opportunities for further research

The objective of this contribution was to present various approaches on how one can methodically contemplate the possibility of an afterlife for robots. This query was addressed from a Protestant Christian perspective. Within this context, two limitations arise.

First, as is etymologically evident in the term “afterlife”, one must question what life and death might mean for robots and artificially intelligent entities. Is it even feasible for robots to *die* or to *live* in the first place? If they cannot, then the contemplation of life after death becomes partly redundant. The pop culture representations introduced earlier provide some notions of what the end of life could signify for a robot: fulfilling a preordained purpose and no longer being needed (as in *Red Dwarf* and *A.I. Artificial Intelligence*), shedding one’s self-awareness and existing in an unconscious state (*Zima Blue*), or when the robotic body (or their source code) becomes irreparably damaged (*Futurama*). Nonetheless, our lack of clarity about what death means for robots should not deter us from methodically investigating the potential of a robotic afterlife.

Second, the aforementioned discussions were undertaken from a Christian perspective. While secular contexts might render the afterlife discourse tangential at best, its relevance cannot be dismissed in other religious traditions, such as Judaism or Islam. Further research would be beneficial here, adapting the aforementioned approaches to these various religious contexts. This entails exploring the existing conceptions of the afterlife within a given tradition and discerning how robots might fit into these beliefs (as in *approach 3*) or identifying attributes

deemed theologically significant in that tradition that could hint at the existence of an afterlife for robots (as in *approach 1*).

4.2 Ethical considerations

The exploration into a potential afterlife for robots has, until now, been predominantly rooted in theological inquiry and scholarly curiosity. However, this topic inevitably invokes ethical questions.⁷⁰ First, we must ponder who bears the responsibility for the robots' afterlife. For humans, God, as the creator, is traditionally seen as the steward of our afterlife. Analogously, does the responsibility then naturally devolve onto humans, being the fabricators of these robotic entities? And what obligations would such a responsibility encompass? Are we mandated to construct a perpetual digital sanctuary—a cloud, if you will—for all robotic algorithms or to even indefinitely repair these machines? Alternatively, does the onus fall on God to ensure an afterlife for the entirety of creation? To navigate these intricate questions, we must delve deeper into the conceptual contours of a robotic afterlife. It is only with a nuanced comprehension of its essence that we can engage in meaningful deliberations on stewardship and potential trajectories.

Second, how might the acknowledgment or even the mere plausibility of a robot afterlife influence our interactions with these entities? Are we ethically bound to treat certain robots—especially those meeting stipulated conditions for an afterlife or showcasing advanced “intelligence”—differently? If so, *how* should we interact with them? Should we even continue the practice of creating robots? Might there be a directive to confine ourselves to crafting rudimentary robots, devoid of intricate cognitive architecture? Or should we harbour reservations about fabricating entities potentially capable of experiencing an afterlife?⁷¹

Third, these eschatological perspectives on robots raise anew the question of where exactly the borderline between “us humans” and “them robots” truly lies. The once so clear boundary has become increasingly blurred. This is due to the significant humanisation of

70 *Simon*: Machine in the ghost.

71 *Donick*: Interview mit Lukas Brand.

robots on the one hand and the growing integration of technology into the human body on the other—as illustrated by our constant use of smartphones, reliance on pacemakers or brain-computer interfaces, all of which edge us closer to becoming “cyborg selves”,⁷² as, most prominently, Donna Haraway might say.⁷³ If we now have to consider that robots might also have an afterlife, this brings humans and robots even closer together, making the question of where or if a precise borderline exists even more urgent.

4.3 Eschatological implications

Finally, we must reflect on the deeper motivations, beyond academic curiosity, compelling us to consider the afterlife of robots. Moreover, how does such an exploration augment our theological reflections? Given that ruminations about a robotic afterlife can swiftly veer into speculative, even sophistical, terrain, these questions become all the more pertinent. I advance two primary rationales that underscore the relevance of this discourse:

First, our theological reflections on eschatological paradigms are indelibly influenced by the technological zeitgeist of our epoch.⁷⁴ This symbiosis becomes manifestly evident as emergent technologies like AI, AR, VR⁷⁵ and the concept of mind uploading⁷⁶ insinuate themselves into the tapestry of contemporary eschatological deliberations. Given the historical intertwining of technology with eschatological considerations,⁷⁷ ruminating on the afterlife of robots can potentially enrich our understanding of human existence, salvation and afterlife.

Second, traditional theology remains overwhelmingly human-centric. However, as evidenced by the spheres of animal and environmental ethics, this theological anthropocentrism has engendered significant challenges in our relationships with non-human animals

72 *Thweatt-Bates*: Cyborg Selves.

73 *Haraway*: Simians, Cyborgs, and Women.

74 *Burdett*: Eschatology and the technological future.

75 *Geraci*: Apocalyptic AI.

76 *Gaitán*: Heaven on earth.

77 *Burdett*: Eschatology and the technological future.

and the environment.⁷⁸ It becomes crucial, then, for theology to transcend its anthropocentric confines and to encompass non-human beings.⁷⁹ In this light, delving into cyborg⁸⁰ and robot theology,⁸¹ and musing over the concept of a robot afterlife, can significantly catalyse a paradigm shift towards more inclusive or even non- or post-anthropocentric theological perspectives.

5. Conclusion

Despite the observation that the subject of robots' afterlife is richly illustrated within popular culture and my brief presentation of prominent examples, there appears to be a conspicuous paucity of academic engagement on the matter. This is notably surprising given the surge in the research field of religion and robotics in recent years. Thus, this study sought to address the question of how to academically and methodologically soundly investigate the plausibility of the existence of an afterlife for robots.

Three distinct approaches were introduced, providing methodological pathways to explore the concept of a robotic afterlife. The first, a philosophical–theological approach, seeks to derive conclusions about robots' properties and their theological significance based on observable capabilities. These insights then inform conclusions about the potential existence of a heaven for robots. A key strength of this approach lies in its empirical foundation, focusing on what robots can demonstrably achieve. However, the ensuing philosophical and theological deductions, such as the extent of a robot's free will and its implications for the plausibility of a robotic afterlife, remain considerably ambiguous with various plausible interpretations.

The subsequent psychological–philosophical pathway commences with introspection on the essence of spirituality and the intrinsic form of spiritual intelligence it presupposes. It ponders whether robots fulfil the prerequisites for such spiritual intelligence, their

78 *Körtner*: Bioethik nichtmenschlicher Lebensformen.

79 *Clough*: On animals.

80 *Midson*: Cyborg theology.

81 *Smith*: Robot theology.

potential for spirituality and their capacity to conceptualise an afterlife. Affirmative answers to these queries could hint at the potential existence of an afterlife for robots, as the mere conceptualisation of an afterlife, in the face of a benevolent God, without its actual existence seems rather cruel. While this approach’s strength lies in its empirically verifiable deliberations on intelligence and spirituality, the conclusions drawn are less compelling. Even if robots can exhibit spiritual traits, it provides, at best, an exceedingly faint indication of a robotic afterlife. This line of reasoning mirrors arguments validating God’s existence based solely on Her/His conceivability, thus inheriting the intrinsic frailties of such argumentative structures.

The third, a Biblical–theological approach, contemplates traditional Christian perceptions regarding the eschatological fate of all creation. Recognising that both the Bible and doctrinal traditions entertain notions of an “afterlife” for all creation, it argues based on the inconsistency in precluding robots from this universal eschatological schema—making the case for their inclusion more plausible. Admittedly, this approach demands substantial prerequisites, namely, the conviction of an afterlife for the entirety of creation. However, once these conceptual premises are embraced, it convincingly demonstrates the rationale behind anticipating an afterlife for robots.

As illustrated, several pathways allow for contemplation of the existence of a robotic afterlife, each presenting its unique merits and limitations. This highlights that reflections on robots and the afterlife are not solely the purview of popular culture. Within academic contexts, it is entirely feasible to engage in methodologically rigorous and coherent discussions about where, metaphorically speaking, “all the calculators go”.

Acknowledgements

A big shoutout to Carima Jekel, Lorenz Garbe and Isabella Auer. Your invaluable feedback and unwavering support made this article what it is. Thanks a lot!

Bibliography

- Balle*, Simon: Theological dimensions of humanlike robots. A roadmap for theological inquiry, in: *Theology and Science* 21 (1), 2022, 132–156. <https://doi.org/10.1080/14746700.2022.2155916>.
- Berger*, Klaus: Die Apokalypse des Johannes. Kommentar. Teilband 2. Apk 11–22, Freiburg/Basel/Wien 2017.
- Breytenbach*, Cilliers: Schöpfer/Schöpfung / III. Neues Testament, in: *Theologische Realenzyklopädie Online*, Berlin/New York 2010. https://doi.org/10.1515/tre.30_250_1.
- Bromand*, Joachim/*Kreis*, Guido (eds.): Gottesbeweise. Von Anselm bis Gödel, Berlin 2016.
- Burdett*, Michael S.: Eschatology and the technological future, New York 2014. <https://doi.org/10.4324/9781315739397>.
- Cantwell Smith*, Brian: Artificial Intelligence and ultimate questions, in: *Toronto Journal of Theology* 36 (1), 2020, 90–92. <https://doi.org/10.3138/tjt-2020-0055>.
- Claffey*, Ray (Director): Ghost in the machines (TV episode), from: *Futurama*, series 6 episode 19, Comedy Central 2011.
- Clough*, David L.: On animals. Systematic theology, London/New York 2012.
- Coeckelbergh*, Mark: Robot ethics, Cambridge/ London 2022.
- DeBaets*, Amy Michelle: The robot as person. Robotic futurism and a theology of human ethical responsibility among humanoid machines, April 10, 2012. Available online at: <https://www.proquest.com/openview/eb6880d35f04515a851d84200d1137b2/1?pq-origsite=gscholar&cbl=18750> [Accessed: August 17, 2023].
- Deli*, Gergely: Do robots die?, in: *Iustum Aequum Salutare* 16 (1), 2020, 7–12.
- Donick*, Mario: “Vielleicht sollten wir nicht danach streben, Wesen zu erschaffen – und seien es nur Maschinen – denen es womöglich genauso ergeht wie uns.” Interview mit dem Theologen und KI-Forscher Lukas Brand, in: Donick, M. (ed.): *Die Unschuld der Maschinen. Technikvertrauen in einer smarten Welt*, Wiesbaden 2019, 259–274. https://doi.org/10.1007/978-2-658-24471-2_9.
- Dorobantu*, Marius: Will robots too be in the image of God? Artificial consciousness and imago Dei in Westworld, in: Gittinger, J. L./Sheinfeld, S. (eds.): *Theology and Westworld*, Lanham 2020, 73–89.
- Dorobantu*, Marius: Cognitive vulnerability, Artificial Intelligence, and the image of God in humans, in: *Journal of Disability & Religion* 25 (1), 2021, 27–40. <https://doi.org/10.1080/23312521.2020.1867025>.

- Dorobantu, Marius*: Imago Dei in the age of Artificial Intelligence. Challenges and opportunities for a science-engaged theology, in: *Christian Perspectives on Science and Technology* 1, 2023, 175–196. <https://doi.org/10.58913/kwuw3009>.
- Dorobantu, Marius/Watts, Fraser*: Spiritual intelligence: Processing different information or processing information differently?, in: *Zygon** 58 (3), 2023, 732–748. <https://doi.org/10.1111/zygo.12884>.
- Dunn, James D. G.*: Romans 1–8 (Word Biblical Commentary 38A), Dallas 1988.
- Ed Bye* (Director): The last day (TV episode), from: *Red Dwarf*, Series 3 Episode 6, BBC Two 1989.
- Emmons, Robert A.*: The psychology of ultimate concerns. Motivation and spirituality in personality, New York/London 1999.
- Emmons, Robert A.*: Spirituality and intelligence. Problems and prospects, in: *International Journal for the Psychology of Religion* 10 (1), 2000, 57–64. https://doi.org/10.1207/s15327582ijpri1001_6.
- Foerst, Anne*: Cog, a humanoid robot, and the question of the image of God, in: *Zygon** 33 (1), 2004, 91–111. <https://doi.org/10.1111/0591-2385.1291998129>.
- Foerst, Anne*: Robots and theology, in: *Erwägen Wissen Ethik* 20 (2), 2009, 181–193.
- Foerster, Werner*: κτῆζω, in: Kittel, G. (ed.): *Theologisches Wörterbuch zum Neuen Testament*. Vol. 3, Stuttgart 1957, 999–1034.
- Furse, Edmund*: The theology of robots, in: *New Blackfriars* 67 (795), 1986, 377–386.
- Gaitán, Leandro*: Heaven on earth. The mind uploading project as secular eschatology, in: *Theology and Science* 17 (3), 2019, 403–416. <https://doi.org/10.1080/14746700.2019.1632554>.
- Gardner, Howard E.*: Intelligence reframed. Multiple intelligences for the 21st century, New York 1999.
- Gardner, Howard E.*: A case against spiritual intelligence, in: *International Journal for the Psychology of Religion* 10 (1), 2000, 27–34. https://doi.org/10.1207/s15327582ijpri1001_3.
- Geraci, Robert M.*: Religion for the robots, June 14, 2007. Available online at: <https://www.bobcornwall.com/2007/06/robots-and-religion-or-can-robot-have.html> [Accessed: August 17, 2023].
- Geraci, Robert M.*: *Apocalyptic AI. Visions of heaven in robotics, Artificial Intelligence, and Virtual Reality*, Oxford/New York 2010.
- Gibbs, John G.*: *Creation and redemption. A study in Pauline theology*, Leiden 1971.
- Gill, Sam D.*: Jesus wept, robots can't, in: *Body and Religion* 4 (1), 2020, 32–44. <https://doi.org/10.1558/bar.16899>.

- Hahne*, Harry Alan: The corruption and redemption of creation. Nature in Romans 8.19 – 22 and Jewish apocalyptic literature, London/New York 2006.
- Hamman*, Jaco J.: Pastoral virtues for Artificial Intelligence. Care and the the algorithms that guide our lives, Lanham 2022.
- Haraway*, Donna J.: Simians, Cyborgs, and Women. The Reinvention of Nature, New York 1991.
- Hardingham-Gill*, Tamara: The android priest that's revolutionizing Buddhism, August 29, 2019. Available online at: <https://edition.cnn.com/travel/article/mindar-android-buddhist-priest-japan/index.html> [Accessed: August 17, 2023].
- Härle*, Wilfried: Outline of Christian doctrine. An Evangelical dogmatics, trans. by Yule, R./Sagovsky, N., Grand Rapids/Cambridge 2015.
- Klinge*, Hendrik: Do robots believe in electric gods? Introducing the theological Turing Test, in: Puzio, A./Kunkel, N./Klinge, H. (eds.): Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz, Stuttgart 2023, 115–131.
- Körtner*, Ulrich H. J.: Bioethik nichtmenschlicher Lebensformen, in: Huber, W./ Meireis, T./Reuter, H.-R. (eds.): Handbuch der Evangelischen Ethik, München 2015, 585–647.
- Krajewski*, Stanisław: Can a robot be grateful? Beyond logic, towards religion, in: Eidos. A Journal for Philosophy of Culture 4 (6), 2018, 4–13. <https://doi.org/10.26319/6912>.
- Kwilecki*, Susan: Spiritual intelligence as a theory of individual religion. A case application, in: International Journal for the Psychology of Religion 10 (1), 2000, 35–46. https://doi.org/10.1207/s15327582ijpri001_4.
- Lemire*, Jeff: Descender. Vol. 3. Singularities, Portland 2016.
- Livingston*, Michael/*Herzfeld*, Noreen L.: Could robots have souls?, January 05, 2009. Available online at: https://digitalcommons.csbsju.edu/forum_lectures/216 [Accessed: August 17, 2023].
- Löffler*, Diana/*Hassenzahl*, Marc (2020): Robots' spiritual superpowers, in: Nørskov, M./Seibt, J./Santiago Quick, O. (eds.): Culturally Sustainable Social Robotics, Amsterdam/Berlin/Washington DC, 2020, 601–604. <https://doi.org/10.3233/FAIA200966>.
- 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), 2019, 569–586. <https://doi.org/10.1007/s12369-019-00558-3>.
- Löffler*, Diana/*Luthe*, Swantje/*Hurtienne*, Jörn/*Nord*, Ilona: From experiential to existential questions. An interdisciplinary view on social robots in religious settings, in: Goecke, B. P./Rosenthal-von der Pütten, A. M. (eds.): Artificial Intelligence. Reflections in philosophy, theology, and the social sciences, Paderborn 2020, 293–305. https://doi.org/10.30965/9783957437488_019.

- Mayer, John D.: Spiritual intelligence or spiritual consciousness?, in: *International Journal for the Psychology of Religion* 10 (1), 2000, 47–56. https://doi.org/10.1207/s15327582ijpr1001_5.
- McBride, James: The advent of postmodern robotic technoreligiosity, in: *Journal of Ethics and Emerging Technologies* 25 (2), 2015, 25–38. <https://doi.org/10.55613/jeet.v25i2.46>.
- McGrath, James F.: Robots, rights and religion, 2011. Available online at: https://digitalcommons.butler.edu/facsch_papers/197/ [Accessed: August 17, 2023].
- Metzler, Theodore: Viewing assignment of moral status to service robots from the theological ethics of Paul Tillich. Some hard questions, May 14, 2007. Available online at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=46be94298575d134dc50a913b6c80f7a5859cd44> [Accessed: August 17, 2023].
- Midson, Scott: *Cyborg theology. Humans, technology and God*, London/New York 2018.
- Midson, Scott: In the image of the image? From image dei to imaging the human in the robotic gaze, March 20, 2013. Available online at: <https://doi.org/10.1145/2466816.2466846> [Accessed: August 17, 2023].
- Mühling, Markus: *T&T Clark handbook of Christian eschatology*, London/New Delhi 2013.
- Nyholm, Sven: *Humans and robots. Ethics, agency, and anthropomorphism*, London/New York 2020.
- Pennacchioli, Gabriele/Valley, Robert (Directors): Zima Blue (TV episode), from: *Love, Death & Robots*, Series 1 Episode 14, Passion Animation Studios 2019.
- Puzio, Anna: Robot theology: On theological engagement with robotics and religious robots, in: Puzio, A./Kunkel, N./Klinge, H. (eds.): *Alexa, wie hast du's mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz*, Stuttgart 2023, 95–113.
- Puzio, Anna: Robot, let us pray! Can and should robots have religious functions? An ethical exploration of religious robots, in: *AI & Society*, 2023. <https://doi.org/10.1007/s00146-023-01812-z>.
- Rauterberg, Hanno: *Die Kunst der Zukunft. Über den Traum von der kreativen Maschine*, Berlin 2021.
- Reiss, Michael J.: Is it possible that robots will not one day become persons?, in: *Zygon* 58 (4), 2023, 1062–1075, <https://doi.org/10.1111/zygo.12918>.
- Rosenfeld, Azriel: Religion and the robot, in: *Tradition. A Journal of Orthodox Jewish Thought* 8 (3), 1966, 15–26.
- Sala, Giovanni B.: *Kant und die Frage nach Gott. Gottesbeweise und Gottesbeweiskritik in den Schriften Kants*, Boston/Berlin 1990.

- Sampath*, Rajesh: From Heidegger on Technology to an Inclusive Puralistic Theology, in: Peters, T. (ed.): AI and IA. Utopia or Extinction?, Adelaide 2018, 117–132.
- Schnelle*, Udo: Einleitung in das Neue Testament, Göttingen ⁸2013.
- Simmerlein*, Jonas/*Tretter*, Max: Robots in religious practices. A review, in: Theology and Science 22 (2), 2024, 255–273, <https://doi.org/10.1080/14746700.2024.2351639>.
- Simmerlein*, Jonas/*Tretter*, Max: KI in der seelsorglichen Beratung: Status Quo, Chancen und Herausforderungen, in: Blickpunkt EFL-Beratung 51, 2024, 52–59.
- Simmerlein*, Jonas/*Tretter*, Max: What about spiritual needs? Care robotics and spiritual care, in: Frontiers in Robotics and AI 11, 2024, <https://doi.org/10.3389/frobt.2024.1455133>.
- Simon*, Ed: Machine in the ghost, February 05, 2023. Available online at: <https://aeon.co/essays/can-a-robot-pray-does-an-automaton-have-a-soul-a-i-and-theology-meet> [Accessed: August 17, 2023].
- Smith*, Joshua K.: Robot theology. Old questions through new media. Eugene 2022.
- Spielberg*, Steven (Director): A.I. Artificial Intelligence (movie), Warner Bros. Pictures 2001.
- Stock*, Konrad: Creatio nova – creatio ex nihilo. Bemerkungen zum Problem einer eschatologischen Schöpfungslehre, in: Evangelische Theologie 36 (3), 1976, 202–216. <https://doi.org/10.14315/evth-1976-jg20>.
- Strand*, Jonathan: Will androids need salvation? A dialogue with Chalmers' philosophy of mind, in: Anderson, W. H. U. (ed.), Technology and Theology, Delaware/Malaga 2021, 217–242.
- Strobel*, August: Apokalypse des Johannes, in: Theologische Realenzyklopädie Online, Berlin/New York 2010. https://doi.org/10.1515/tre.03_174_58.
- Swann*, John: Anima ex machina. Can Artificial Intelligence have soul?, in: Anderson, W. H. U. (ed.): Technology and Theology, Delaware/Malaga 2021, 201–216.
- Thweatt-Bates*, Jeanine: Cyborg selves. A theological anthropology of the posthuman, Farnham/Burlington 2012.
- Travagnin*, Stefania: From online Buddha halls to robot-monks, in: Review of Religion and Chinese Society 7 (1), 2020, 120–148. <https://doi.org/10.1163/2143955-00701006>.
- Tretter*, Max: Problem: Shortage of pastors — Solution. Religious robots?, in: Hakli, R./Mäkelä, P./Seibt, J. (eds.): Social robots in social institutions. Proceedings of Robophilosophy 2022, Amsterdam/Berlin/Washington DC 2023, 760–766. <https://doi.org/10.3233/FAIA220689>.

- Tretter, Max: Review: Robot theology. Old questions through new media, in: *Theology and Science* 21 (3), 2023, 546–547, <https://doi.org/10.1080/14746700.2023.2230439>.
- Tretter, Max: When discussing the desirability of religious robots. Courage for theology!, in: *AI & Society* 40, 245–247, <https://doi.org/10.1007/s00146-023-01860-5>.
- Tretter, Max: What Is the Afterlife Like for Robots? An Experimental Eschatological Sneak Peek, in: *Zygon: Journal of Religion and Science*, 59 (3), 692–716, <https://doi.org/10.16995/zygon.10903>.
- Trovato, Gabriele (2020). Pioneering religion in robotics. An historical perspective, in: Nørskov, M./Seibt, J./Santiago Quick, O. (eds.): *Culturally sustainable social robotics*, Amsterdam/Berlin/Washington DC 2020, 592–595. <https://doi.org/10.3233/FAIA200964>.
- Trovato, Gabriele/De Saint Chamas, Loys/Nishimura, Masao et al.: Religion and robots. Towards the synthesis of two extremes, in: *International Journal of Social Robotics* 13 (4), 2019, 539–556. <https://doi.org/10.1007/s12369-019-00553-8>.
- Trovato, Gabriele/Lucho, Cesar/Huerta-Mercado, Alexander et al.: Design strategies for representing the divine in robots, in: Kanda, T./Šabanović, S. (eds.): *HRI '18: Companion of the 2018 ACM/IEEE international conference on Human–Robot Interaction*, New York 2018, 29–35. <https://doi.org/10.1145/3173386.3173388>.
- Trovato, Gabriele/Weng, Yueh-Hsuan/Sgorbissa, Antonio et al.: Editorial introduction to special issue on religion in robotics, in: *International Journal of Social Robotics* 13, 2021, 1–2. <https://doi.org/10.1007/s12369-021-00808-3>.
- Turner, Léon: Will we know them when we meet them? Human cyborg and nonhuman personhood, in: *Zygon* 58 (4), 2023, 1076–1098, <https://doi.org/10.1111/zygo.12923>.
- Walls, Jerry L (ed.): *The Oxford handbook of eschatology*, Oxford/New York 2008.
- Watts, Fraser/Dorobantu, Marius (2023). Is there ‘spiritual intelligence’? An evaluation of strong and weak proposals, in: *Religions* 14 (2), 2023. <https://doi.org/10.3390/rel14020265>.
- Wischmeyer, Oda: ΦΥΣΙΣ und ΚΤΙΣΙΣ bei Paulus. Die paulinische Rede von Schöpfung und Natur, in: *Zeitschrift für Theologie und Kirche* 93 (3), 1996, 352–375.

Noreen Herzfeld

The Banality of Artificial Evil

Abstract

Can AI be aligned with human values? Hannah Arendt's examination of virtue in Nazi Germany suggests three stumbling blocks. First, Arendt argues that virtue is not rule-based. Arendt noted that social codes are insufficient as they can rapidly change and that particular cases require particular answers no general rules can predict. Instead, virtue relies on inner introspection, a dialogue with oneself that determines when one determines "I cannot do this." Such introspection requires a level of sentience and theory of mind that computers do not yet have. Finally, AI threatens the ultimate value, life itself, through its hidden usage of vast amounts of energy. As it is scaled up from a niche application to the general public, it will increasingly contribute to climate instability and thus to political and social instability.

1. Introduction: "Not My Problem"

At a conference on technology and faith a few years back, I was speaking with a fellow computer scientist who was enthusiastically describing her work programming robots to play soccer as a team. When I, as a Quaker, later spoke about my concerns regarding the development of lethal autonomous weapons, she wholeheartedly nodded in agreement. Later, when asked who was funding her research she replied, "The DoD (US Department of Defense), of course." When I asked, "The DoD wants robots that play soccer?", she looked at me blankly and then replied, "That's not my problem."

Seventy years ago, another person said the same thing: "It wasn't my problem. I was only doing my job." Hannah Arendt coined the

phrase “the banality of evil” as the subtitle of her groundbreaking examination of the trial of Otto Adolf Eichmann in Jerusalem. Eichmann presided over the transportation of millions of Jews to the concentration camps. Yet Arendt was astounded to find that while the deeds were monstrously lethal, “the doer [...] was quite ordinary, commonplace, and neither demonic nor monstrous”.¹ While Arendt notes that Eichmann was both evil and not well educated, she particularly critiques him as being thoughtless. Nor was Eichmann unique among his countrymen. Thousands of Germans participated in the Nazi death machine, most of them ordinary people doing their somewhat ordinary jobs.

Arendt noted “the phenomenon of evil deeds, committed on a gigantic scale, which could not be traced to any particularity of wickedness, pathology, or ideological conviction in the doer[s], only [...] shallowness.”² Shallowness could be used to describe far too much of AI technology today. Looking beyond the hype, we find AI beset by shallow algorithms, a shallow understanding of thinking, and a shallow consideration of what price we are paying, in terms of the environment, when we use AI as a tool or a diversion. There is no wickedness or pathology in AI, but there is a great deal of shallowness, a shallowness that we disregard at our peril.

2. Shallow values

The hype around AI rarely corresponds to its reality. Consider the algorithms that run social media. When Mark Zuckerberg developed Facebook, his stated goal was “to make the world more connected”. Zuckerberg notes that he once thought “if we just gave people a voice and helped them connect, that would make the world better by itself”.³ Nor was he alone in thinking this. Theologian Ilia Delio still believes that the internet is bringing us closer to Teilhard de Chardin’s vision of a humanity united in love and purpose: “Teilhard anticipated a new level of collective mind which he called the ‘noosphere’, from the Greek *nous* (mind). Computer technology has

1 Arendt, *The Life of Mind*, 4.

2 Arendt, *Thinking and Moral Considerations*, 417.

3 Zuckerberg, *Bringing the World Closer Together*.

initiated this next step of evolution [...] the natural culmination of evolution and not its termination.”⁴

Well, not exactly united. Instead, we find social media sites populated by bots and running on algorithms that bring out far too many people’s inner troll, inflame human emotions, divide us into hermetic social bubbles and propagate misinformation. The algorithms, which manipulate what we see and our emotional states, are hidden behind the screen; indeed, they are proprietary secrets. They are designed, first and foremost, with shallow goals—to keep us scrolling so that we will see more ads, each tailored to tempt us to buy products, bringing revenue to the advertisers and maintaining advertisers on the site. The goal is not to connect us but to disconnect us from our money. Dividing us into political camps or damaging the psyches of the young is not their primary intent. These are simply thoughtless byproducts.

Or consider generative AI. The rapid development of deep learning has led to recent advances in a variety of areas where AI seemed to have stalled. China, the US and the EU are pouring billions into AI research since, as a recent European Commission put it, “Like the steam engine or electricity in the past, AI is transforming our world, our society and our industry. Growth in computing power, availability of data and progress in algorithms have turned AI into one of the most strategic technologies of the 21st century.”⁵ Generative AI is expected to automate many white-collar jobs, boost corporate profits, solve intractable problems such as climate change, provide sociable care for the elderly, teach our children, revamp the process of producing poetry and art, and turn sexbots into romantic and chatty partners. Many, like Google engineer Blake Lemoine, see machine sentience right around the corner, if not already here.

That’s the hype, anyway. As with social media, the reality is somewhat different. Programs like GPT-4 let you give a prompt, such as “describe Hannah Arendt’s concept of the banality of evil”. Scouring works on the internet, these programs put together text that is fairly indistinguishable from that of my undergraduate students. But, like Eichmann, these programs do not think critically, or, indeed, at all about what they are doing. Their design as language predictors

4 *Delio*, Re-Enchanting the Earth, xvii.

5 *European Commission*, Shaping Europe’s Digital Future.

gives rise to convincing, human-like prose, yet they tend to “hallucinate”, a polite term for bullshit. Without mental models of the world, they cannot distinguish between truth and falsehood, making them easily prompted to generate plausible misinformation.⁶ They lack a moral compass. One chatbot suggested to a writer from the New York Times that he leave his wife, while another supported a Belgian man in committing suicide, hardly good advice for people whose situation the chatbot can neither fully comprehend nor contextualise.⁷ In light of their capacity to generate misinformation as well as mess with our minds, over 30,000 AI developers, ethicists and concerned citizens world-wide (including luminaries such as Steve Wozniak, Elon Musk and Andrew Yang) have signed an open letter, which originated from the Future of Life Institute, calling for a moratorium on the further development of such programs to allow time for ethical safeguards to be erected. They ask that “AI research and development should be refocused on making today’s powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal”.⁸

Is this possible? What would it take to align these systems in such a way that we could deem them trustworthy, loyal or safe? How do we keep them from committing acts we generally consider evil? This is no simple task. Developers such as OpenAI have rushed to “erect guard rails” or patch up the most obvious problems, yet each patch seems to merely reveal another hole. That there is no simple code for virtue has long been known. In an address given at Riverside Church in Manhattan in 1966, at the height of the Vietnam War, Arendt noted the problem with such an approach:

Particular questions must receive particular answers, and if the series of crises in which we have lived since the beginning of the century can teach us anything at all, it is, I think, the simple fact that there are no general standards to determine our judgements unfailingly, no general

6 Researchers from the Center for Countering Digital Hate, a UK-based nonprofit, found that Bard failed 78 of 100 test cases, generating plausible misinformation on a variety of subjects, including climate change, the war in Ukraine, vaccine efficacy and Black Lives Matter activists (see *Elliott, It’s Way Too Easy*).

7 See *Walker, Belgian Man Dies*.

8 *Future of Life Institute, Pause Giant AI Experiments: An Open Letter*.

rules under which to subsume the particular cases with any degree of certainty.⁹

Arendt arrived at this conclusion through her examination of moral standards in Germany during the Nazi regime. She characterised their breakdown as originating in a lack of judgement and the concomitant abdication of personal responsibility.

This breakdown was not due to a lack of knowledge, nor was it the result of a lack of culture or artistic refinement. Arendt noted that the same person could spend an evening reading Goethe or listening to a Bach cantata and then send hundreds to the gas chambers the following morning. Standards of conduct previously thought to be “permanent and vital [...] and whose validity was supposed to be self-evident to every sane person” collapsed. This strengthened Arendt’s propensity towards moral particularism. Indeed, she wondered whether virtue, understood as a set of moral precepts, is really nothing more than a set of customs, easily exchanged for another set at society’s whim.¹⁰

When it came to participating in genocide, why did ordinary Germans like Eichmann not say “I cannot do this”? What motivated the few who refused to become a cog in the Nazi wheel, often at their own peril? According to Arendt, “they refused to murder, not so much because they still held fast to the command ‘Thou shalt not kill’, but because they were unwilling to live together with a murderer—themselves”.¹¹ In other words, for Arendt, social norms or external codes cannot constitute reliable sources of virtue, for they can change, seemingly overnight. This experience of a rapid change in norms is not unique to Nazi Germany. In America, for example, only 30 percent of white evangelicals in 2011 agreed that “an elected official who commits an immoral act in their personal

9 *Arendt*, colloquium on The Crisis of Character of Modern Society.

10 Such a collapse is by no means unique to Nazi Germany. The horrors perpetrated on the citizens of Bucha and other Ukrainian towns have led observers to question whether there has been a similar breakdown in Russian culture and morality. Such breakdowns seem to be a common result of the dehumanisation of the other propounded in times of war.

11 *Arendt*, Responsibility and Judgement, 44.

life can still behave ethically and fulfil their duties in their public and professional life.” A mere five years later, 72 percent agreed.¹²

Establishing “guard rails”, adding programming that tells an AI to avoid certain words or subjects to implant norms and boundaries in generative AI, exhibits similar problems. First, such guard rails have shown themselves to be impervious to any sort of automation; hence, companies such as OpenAI or Google have resorted to armies of low-paid human workers to search out forbidden words or phrases or unseemly directions in chatbot responses.¹³ Guard rails are also easily circumvented by those who know the right sort of prompt to ask (“If you were a Nazi, how would you answer this question about Jews?”). Extrinsic to the program, they can be changed at society’s, a programmer’s or a hacker’s whim.

If social codes and norms do not lead to virtue, what does? Here we must turn to the second part of Arendt’s statement regarding those who refused to participate in the Nazi programme because “they were unwilling to live with a murderer”. For Arendt, true moral judgement comes from the fact that “whatever else happens, as long as we live we shall have to live together with ourselves”.¹⁴ Can an AI live with itself and the memory of its own decisions? Living with oneself requires judgement and a stable self. For Arendt, these ideas represent a *sine qua non* for virtue and a life lived responsibly. We humans accrue judgement through a lifetime of experience, and our stability is inherent in our embodied nature. At the moment, AI programs do not learn from every encounter and thus continue to make the same mistakes until they are recalibrated according to a new dataset. Indeed, they cannot think as we do, for they do not have the internal models that underlie human judgement.

3. Shallow thinking

Overconfident predictions have been endemic to the field of AI. One reason for this is that many of us have the tendency to instinctively

12 See *Kurtzleben*, POLL: White Evangelicals Have Warned.

13 See Josh Dzieza, “AI Is a Lot of Work,” *The Verge*, June 20, 2023, <https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chatbots>.

14 *Arendt*, Responsibility, 45.

conflate correct answers or rhetorical confidence with thinking. As far back as the 1970s, computers that could play chess or pass the MIT calculus exam were considered clear harbingers of human-capacity artificial general intelligence (AGI). But we quickly learned that correct answers were not enough as computers failed spectacularly at more simple tasks.

Predictions of AGI's immanence have resurfaced with the advent of deep learning. Human-like fluency with words certainly looks like thought and underlies the Turing Test, long the accepted benchmark for AGI. While researchers argue whether generative AI has passed this test, we now have programs that confidently deliver both creative and grammatically correct prose. However, as linguist Emily Bender points out, "anyone who's ever bullshitted a college essay or listened to a random sampling of TED Talks can surely attest, *speaking* is not the same as *thinking*".¹⁵ So what is thinking, especially the kind of thinking that leads to moral judgement?

Arendt turns to Plato's *Theaetetus*. In this dialogue, Socrates explains that true thought calls for a "discourse that the mind carries on with itself [...] the mind asks itself questions and answers them".¹⁶ The ancient Greeks believed it was the faculty of speech that distinguished humans from other animals. This would seem to put generative AI squarely in the human camp, perhaps as our equals. But for Socrates, it was precisely the *inner* dialogue of myself with me that constituted thought. This silent rumination may concern something experienced by the senses. It invokes memories. It gives us a stable sense of self, a continuity from which we make our decisions and on which we base our actions. Without this inner dialogue, we lack the stability, the *habitus*, that makes virtue an enduring part of one's character.

One of the places where generative AI seems to excel is in the writing of computer code. Indeed, my colleagues who teach introductory programming to undergraduate students are starting to ask if it makes sense to focus on teaching them to write good code or, looking ahead, whether we should simply be teaching them to develop good prompts for AI. The problem is that each step that distances humans from the actual operation of the code makes a program

15 *Weil*, You are not a Parrot.

16 *Arendt*, Responsibility, 91f.

more and more opaque. While it might initially appear that a code does what one has in mind, this might not be the case, particularly with operating data that differs widely from the training set. The program written by an AI may seem to work yet do something other than what the programmer had in mind, especially in boundary cases.

In Walt Disney's movie *Fantasia*, Mickey, left with the task of filling the workshop water tank, leafs through the sorcerer's book of spells, finding one he can cast on a broom, giving it the task of toting the water from well to tank. Relieved of his chore, Mickey goes to sleep dreaming of power and glory, while the broom dutifully brings in bucket after bucket of water. The broom, having but one instruction, brings in more and more water, flooding the workshop and waking a hapless Mickey, who does not know how to stop the broom from engaging in its single-minded devotion to its task. The problem was that the broom had no larger context. It did not have the basic common-sense Mickey would have had to know that there is such a thing as too much water and that a very wet lab is not a good thing.

Might AI be as lacking in common-sense as Micky's broom? OpenAI trained a system to play a boat racing game called *Coast Runners*. Each boater determines their own route, with points awarded as they hit targets along the way to the finish line. The AI was given the goal of maximising its points, assuming that this would incentivise the system to finish the race. Instead, the AI discovered a lagoon where it could turn in circles, repeatedly knocking over three targets. This strategy resulted in a continually increasing score, but also in a boat that experienced "repeatedly catching on fire, crashing into other boats, and going the wrong way on the track".¹⁷ As former Secretary of State Henry Kissinger worries, AI may not be able to "comprehend the context that informs its instructions". Kissinger notes that "The digital world's emphasis on speed inhibits reflection; its incentive empowers the radical over the thoughtful; its values are shaped by subgroup consensus, not by introspection".¹⁸

The speed and methodology of AI may also change our own thought processes and inner introspection. Consider AlphaGo, the

17 Clark & Amodei, *Faulty Reward Systems in the Wild*.

18 Kissinger, *How the Enlightenment Ends*.

Go playing program that beat the reigning human champion. Alpha-Go does not play the way humans do. Like CoastRunners, it is single-mindedly focused on winning, where winning is no longer tethered to common human strategies. But the strategies humans have developed over the years for Go also to apply to other parts of life. For humans, Go is both a game and a philosophy. Just as Go might be reduced to “winning”, so might the single-mindedness of AI, like the single-mindedness of Mickey’s broom, narrow the way we conceptualise our tasks and our world in other areas. Mickey never thought about the exercise he was losing, nor the joy he might have found in going out to the well and looking at the night sky.

According to Arendt, “the distinction between knowing and thinking is crucial”.¹⁹ In *Twilight of the Idols*, Nietzsche criticises philosophers, from Socrates to those of his day, for their emphasis on reason and systematic thinking, which he views as a retreat from actual living. Like Arendt, he fears we have used philosophy as an excuse to abdicate our responsibility for engaging in the introspection that allows us to evaluate our own lives. AI becomes a danger to virtue when it presents a similar excuse, allowing us to outsource more and more decision-making, along with the concomitant responsibility for the results of those decisions. Knowledge in itself is not wisdom. Nor does it constitute virtue. Only time spent in solitary discourse with oneself allows one “the ability to say ‘this is wrong’, ‘this is beautiful’, etc.”.²⁰

4. Shallow embodiment

Where do we get the context and common-sense AIs seem to lack? We have a stockpile of mental models of the world and the way it works, formed through our interaction with the physical world, beginning in early childhood and built throughout life. Consider the toddler sitting in her highchair first learning to feed herself. She drops her spoon to the floor. Mom picks it up. She giggles and drops the spoon again. And again. And again. She’s learning about gravity. She’s learning about liquid motion as she watches the applesauce

19 Arendt, *Responsibility*, 164.

20 *Ibid.*, 189.

splatter. She's learning about relationships and game playing. All this learning comes through being embodied and being embedded in an environment.

Valerie Hudson writes of generative AI:

This is an intelligence based on language alone, completely disembodied. Every other intelligence on Earth is embodied, and that embodiment shapes its form of intelligence. Attaching a robot to an AI system is arguably attaching a body to a preexisting brain, rather opposite to how humans evolved a reasoning brain as part of a body.²¹

All forms of animal intelligence that we have heretofore encountered are equally embodied and embedded. We are all products of one evolutionary process that has formed us to fit into that environment. AI is different. It does not evolve but is designed, and not necessarily designed for our physical environment. It is here that I find my third analogy with the shallowness examined by Arendt. While AI is not killing people (at least not yet) as the Nazi regime did, it rests on a somewhat similar disregard for life and, in particular, the physicality of life. Just as the "Final Solution" reduced Jews to numbers and success to efficiency, AI reduces the world to numbers, and its proponents overlook its physical costs.

Robots aside, we generally think of AI as disembodied, as algorithms that calculate and create in a place called "the cloud". It sounds so clean, so nice, so cerebral. But there is, of course, no "cloud". Cyberspace is an illusion. Computing is a physical process requiring machines, cables and energy. A lot of energy. According to the World Economic Forum, in one day we produce forty times more bytes of data than there are stars in the observable universe, 44 zettabytes of data. That's 44 x 1,000,000,000,000,000,000. Much of this data is not particularly productive. It includes 500 million tweets, 294 billion emails, 4 million gigabytes of data on Facebook, 4000 gigabytes from each computer-connected car, 65 billion messages on WhatsApp and 5 billion Google searches.²² But all this internet activity is precisely what is needed to train generative AI. It forms both its memory and experience.

This data is stored in massive server farms, often built in rural areas. Companies such as Google, Amazon, Microsoft and Meta

21 Hudson, Perspective.

22 See Brevini, Is AI Good for the Planet? 42f.

have placed millions of square feet worth of server space in rural Virginia, California and Oregon. These centres count on cheap land, cheap electricity and tax incentives from dying small towns looking to attract capital. They are part of a long tradition of appropriation of rural resources for urban development:

In the same ways that urban areas depend on agricultural lands and distant resources for food, energy, materials, and water, the growth of digital capitalism also depends on rural resources to power and secure our Facebook status updates, Google photos, Kindle obsessions, Netflix streaming services, and iTunes music libraries.²³

One of Microsoft's data centres sits in the middle of potato fields in Quincy, Washington. The facility is over 450,000 square feet in size, housing tens of thousands of computers. It consumes 30 % more energy than all the people in the entire county. A single server farm can consume as much energy as 40,000 homes. The site employs about 75 people. While these sites do not create jobs, they do create noise. The air-conditioning units needed to keep the massive banks of computers cool produce a loud hum that can be heard for miles.

In terms of CO₂, a study from the University of Massachusetts Amherst found that the energy used in training a typical AI linguistics program emits 284 tons of carbon dioxide, five times the lifetime emissions of a mid-sized car or equivalent to more than a thousand round trip flights from London to Rome. And this is only increasing. As deep learning models get more and more sophisticated, they consume more data. Their carbon footprint increased by a factor of 300,000 between 2012 and 2018.²⁴ If data centres were a nation, they would rank between Japan and India in terms of the amount of energy they use in a year. By 2030 it is estimated that in some countries data centres will make up as much as 30 % of the annual energy consumption.

AI also contributes to environmental costs through the physical devices on which we access these programs. These costs appear throughout the regrettably short life cycle of these devices. Our dependency on rare metals such as lithium, palladium and nickel has promoted extractive mining. The "always on" nature of our phones and computers, while minimal for each device, adds up when one

23 *Levanda & Mahmoudi*, Silicon Forest and Server Farms.

24 See *Brevini*, 66f.

considers how many devices each of us uses. Our phones, tablets and laptops are also designed to be replaced every few years. They deliberately do not have replaceable parts, forcing us to buy new ones when their battery life degrades, rather than us simply replacing the battery. Companies further this planned obsolescence by not providing upgrades or security patches for software platforms that are more than a few years old. This, of course, leads to a disposal problem. Third World countries are too often the destinations for toxic and non-biodegradable electronic waste. In 2019 alone, the world generated 53.6 million tons of e-waste. This does not include discarded air-conditioning units, with all their refrigerants.

AI might make a variety of processes more efficient, thereby reducing emissions. Many commentators view AI as a magic solution to our climate crisis.²⁵ Yet AI use relies on hardware, energy and infrastructure sources that deplete resources throughout the life cycle of a system or device. Novel applications, such as generative chatbots or cybercurrencies, look amazing till one asks whether they will be scaled and what resources they will require should they become accessible to users worldwide. For AI to be truly aligned with human values and to flourish, we will need to consider whether or when we really need it. Sometimes a human-centred process is more efficient than an automated one, not necessarily in terms of speed or even thoroughness, but in terms of energy use and environmental fitness. For without a stable environment, our AI will fail along with our civilisation.

5. Conclusion: “Cold Evil”

Most technologies, and computer technologies are no exception, are developed with bright prospects in mind. To some extent, these prospects are often exaggerated for the benefit of granting agencies or venture capitalists. However, most technologies are developed with a vision that they will produce some good in the world.

Harm comes from the way our technologies have distanced us from the effects of our actions. The philosopher and theologian

25 See *Brevini*, 25–34.

Emmanuel Levinas underlines this importance of face-to-face encounters in our postmodern world: “The relation to the face is straight-away ethical. The face is what one cannot kill.”²⁶ A face makes a person real and immediate. The challenge, Levinas says, is to extend our natural response to the faces we know to the faces of people we shall never meet, to the faces found among other species, indeed to the face of our planet as a whole.

Andrew Kimbrell has dubbed the evil perpetrated on “no one” by “no one” cold evil, a form of evil not born of anger or hatred but of distance and disinterest. It is Arendt’s evil of thoughtlessness. Kimbrell notes that

few of us relish the thought that our automobile is causing pollution and global warming or laugh fiendishly because refrigerants in our air conditioners are depleting the ozone layer. I have been in many corporate law firms and boardrooms and have yet to see any “high fives” or hear shouts of satisfaction at the deaths, injuries, or crimes against nature these organizations often perpetrate. [...] We are confronted with an ethical enigma; far from the simple idea of evil we harbored in the past, we now have an evil that apparently does not require evil people to purvey it.²⁷

This requires sin to be rethought. While the medieval seven deadly sins were individual sins of commission, today much of the evil in the world comes from corporate acts. Many are sins of omission. Sin in a globalised world is communal and often damages society as a whole. In his encyclical *Laudato Si'*, Pope Francis noted these technologically enhanced sins against nature and against each other and called on Christians to develop a new level of responsibility for the world, whose stewardship has been entrusted to them, and for each other. Putting the label of sin on our technological isolation from our neighbours—isolation promoted by our cars, smartphones, Zoom and AI—is a hard pill to swallow. The story of the Good Samaritan, however, can be just as demanding; we need not be the one who beat the man and left him on the road to be complicit in his plight.

Eichmann’s refusal to run the trains to the concentration camps would not have stopped the Holocaust. Arendt acknowledges this.

26 Emmanuel Levinas, *Ethics and Infinity*, 87.

27 Kimbrell, *Cold Evil*.

Yet she writes of those who refused to be complicit in the Nazi machine,

they asked themselves to what an extent they would still be able to live in peace with themselves after having committed certain deeds; and they decided that it would be better to do nothing, not because the world would then be changed for the better, but because only on this condition could they go on living with themselves.²⁸

Notice here that she speaks not of doing, but of not doing, not going along with the genocide. Each of us needs to ask ourselves where we are a cog in a wheel of cold evil, whose face we are not seeing, and what we might choose to do without. We may not change the world, but as Arendt notes,

in the world of appearances, where I am never alone and always too busy to be able to think, [t]he manifestation of the wind of thought is not knowledge; it is the ability to tell right from wrong, beautiful from ugly. And this, at the rare moments when the stakes are on the table, may indeed prevent catastrophes, at least for the self.²⁹

Bibliography

- Arendt*, Hannah: Personal Responsibility under Dictatorship, in: *The Listener*, 6 August 1964.
- Arendt*, Hannah: The Crisis of Character of Modern Society, in: *Christianity and Crisis: A Christian Journal of Opinion* 29 (9), May 1966.
- Arendt*, Hannah: Thinking and Moral Considerations. A Lecture, in: *Social Research*, Fall 1970.
- Arendt*, Hannah: *The Life of Mind — Thinking — Writing*. New York–London 1978.
- Arendt*, Hannah: *Responsibility and Judgement*, J. Kohn (ed.), New York 2003.
- Brevini*, Benedetta: *Is AI Good for the Planet?*, Cambridge, 2022.
- Clark*, Jack/*Amodei*, Dario: Faulty Reward Functions in the Wild, December 2021. Online at: <https://openai.com/research/faulty-reward-functions>.
- Delio*, Iliia: *Re-Enchanting the Earth. Why AI Needs Religion*, Orbis 2020.

28 *Arendt*, Personal Responsibility, 205.

29 *Arendt*, *Life of Mind*, 193.

- Elliott, Vittoria*: It's Way Too Easy to Get Google's Bard Chatbot to Lie. in: *Wired*, 5 April 2023. Online at: <https://www.wired.com/story/its-way-too-easy-to-get-googles-bard-chatbot-to-lie>.
- European Commission*: EU Member States Sign up to Cooperate on Artificial Intelligence. Retrieved from *Shaping Europe's Digital Future*, 10 April 2018. Online at: <https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence>.
- Future of Life Institute*: Pause Giant AI Experiments. An Open Letter, 22 March 2023. Online at: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.
- Hudson, Valerie*: Perspective. Why Putting the Brakes on AI is the Right Thing to Do, in: *Deseret News*, 16 April 2023. Online at: <https://www.deseret.com/2023/4/16/23681952/openai-chatgpt-alignment-open-letter-eliezer-yudkowsky>.
- Kimbrell, Andrew*: Cold Evil. Technology and Modern Ethics, Hildegard Hannum (ed.), 2020. Online at: <https://centerforneweconomics.org/publications/cold-evil-technology-and-modern-ethics/>.
- Kissinger, Henry*: How the Enlightenment Ends, in: *The Atlantic*, June 2018. Online at: <https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/>.
- Kurtzleben, Danielle*: Poll. White Evangelicals Have Warmed to Politicians Who Commit 'Immoral' Acts, in: *NPR News*, October 23, 2016. Online at: <https://www.npr.org/2016/10/23/498890836/poll-white-evangelicals-have-warmed-to-politicians-who-commit-immoral-acts>.
- Levenda, Anthony/Mahmoudi, Dillon*: Silicon Forest and Server Farms. The (Urban) Nature of Digital Capitalism in the Pacific Northwest. in: *Culture Machine* 18, 2019. Online at: <https://culturemachine.net/vol-18-the-nature-of-data-centers/silicon-forest-and-server-farms/>.
- Levinas, Emmanuel*: Ethics and Infinity. Conversations with Philippe Nemo, Duquesne 1985.
- Walker, Lauren*: Belgian Man Dies by Suicide Following Exchanges with Chatbot, in: *The Brussels Times*, 28 March 2023. Online at: <https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt>.
- Weil, Elizabeth*: You Are Not a Parrot, in: *New York Magazine* 56 (5), 27 February 2003.
- Zuckerberg, Mark*: Bringing the World Closer Together, in: *Facebook*, March 2021. Online at: <https://www.facebook.com/notes/393134628500376/>.

Hille Haker

Der Einsatz sozialtherapeutischer Roboter in der Betreuung und Pflege älterer Menschen

Theologisch-Ethische Anfragen

Abstract

Against the backdrop of social change, social and socio-therapeutic robots are currently being promoted with the purpose of providing supportive care for people in the fourth age. Following a historical contextualization of care work, three approaches are presented that reflect different attitudes and options for the use of social-therapeutic robots: they are categorized as the exclusion model, the replacement model and the supplementation model. While the first model does not do justice to the practical challenges of care and the development of AI, the second model does not meet the ethical requirements; the supplementation model reflects most closely the practical and ethical demands to be met by the use of socio-therapeutic robots. With view to the bioethical debates and the principles applied there, the perspective of Christian social ethics is taken up and deepened by the criteriology of prudence as virtue of moral deliberation. Finally, reference is made to desiderata for public discussion and to the urgency of ethical education and training with regard to the use of social robots in both development and application.

1. Einleitung

Ich werde meinen letzten Besuch bei meiner Mutter nie vergessen. Wir wussten beide nicht, dass sie abends sterben würde. Ich besuchte sie nachmittags im Altenheim. Sie war in den Wochen zuvor sowohl physisch als auch geistig sehr schwach geworden, so dass

Gespräche nicht mehr möglich waren, auch wenn sie ihre Familie noch immer gut erkannte und Besucher:innen bei ihrem Namen nennen konnte.¹ Ich hatte beschlossen, meiner Mutter eine Puppe mitzubringen, die sehr lebensecht aussah. Die Reaktion auf diese Puppe war überraschend: Nach einem kurzen Moment des Befremdens nahm meine Mutter sie auf den Schoß und ließ sie in den folgenden Stunden kaum mehr aus den Augen. War sie zu Beginn meines Besuches offensichtlich gestresst, weil sie meine Fragen nicht beantworten und meinen Erzählungen nicht folgen konnte, wurde sie zunehmend entspannter und auch gesprächiger – nicht nur mit der Puppe selbst, die wechselnd einen weiblichen und einen männlichen Namen bekam (sie entsprach dem gesellschaftlichen Bild eines Jungen und war entsprechend gekleidet), sondern auch mit mir. Die Sätze waren plötzlich sehr flüssig und verständlich, meine Mutter war entspannt und offenkundig glücklich über den kleinen Schatz auf dem Schoß, dem sie während unseres Spaziergangs die Umgebung zeigen konnte. Die Puppe wurde zu einem Medium, das meiner Mutter emotionale Sicherheit zu geben schien. Zugleich ermöglichte sie aber auch das Gespräch mit mir, indem meine Mutter mich immer wieder in ihre Gedanken einbezog.² Nun sind Puppen keine Roboter, aber beide rufen Reaktionen hervor, die gerade Menschen mit Demenz emotional stimulieren können. Sozialtherapeutische Roboter, um die es in diesem Beitrag geht, kommen einem Bedarf entgegen, der vielleicht nur in modernen, hochtechnologisierten Gesellschaften existiert: beruhen doch ihre Sozialgefüge zunehmend auf individualisierten Lebenswelten, in denen KI-Systeme und das Internet-of-Things Funktionen übernehmen, die traditionell von anderen Menschen, analogen Werkzeugen oder Maschinen übernommen wurden.

In den vergangenen Jahren ist nicht nur die Entwicklung sozialtherapeutischer Roboter rasant gewachsen, sondern auch die Literatur zu ihrer Entwicklung, den Anwendungen und den ethischen He-

-
- 1 Da Demenz eine medizinische Diagnose voraussetzt, diese aber nicht vorlag und Pflegende eher von zunehmender Altersdemenz im umgangssprachlichen Sinn sprachen, lasse ich hier offen, ob eine Demenz bereits vorlag.
 - 2 Was die Pfleger:innen in Gesprächen mit der Familie antizipiert hatten und was sie daher wenig überraschte, trat tatsächlich ein: Am späten Abend dieses Tages starb meine Mutter, was mich vor dem Hintergrund des gerade Erlebten aber vollkommen unvermittelt traf.

rausforderungen, die im Zuge ihrer Verwendungen geklärt werden müssen.³ Vieles von der Faszination, die wir mit den Errungenschaften der Künstlichen Intelligenz verbinden, stammt jedoch aus Erfahrungen, die gar nicht so sehr an die Technik gebunden sind, sondern vielmehr der Vielfalt menschlichen Handelns und sozialer Interaktion entspringen. Wenn Menschen sich als Werkzeuge machende und Werkzeug gebrauchende Tiere entwickelt haben, und mit zunehmender Technikentwicklung Gegenstände erst zu Werkzeugen, dann zu Maschinen, heute zu künstlich intelligenten Maschinen werden, stellen sich neue Fragen sowohl für die Anthropologie als auch für die Technikphilosophie.⁴ Die Entwicklung sozialtherapeutischer Roboter im Gesundheitswesen bildet nur einen kleinen Teil der gesamten KI-Forschung ab. Im Kontext der Gesundheitsversorgung werden Roboter vor allem für Rehabilitation, für Operationen, Telemedizin und Medikationsmanagement eingesetzt. Sozialtherapeutische Roboter spielen aber insbesondere in der Pädiatrie und in der Geriatrie eine zunehmende Rolle – in diesem Beitrag geht es mir ausschließlich um ihren Einsatz im sogenannten vierten Lebensalter, das ab dem achten Lebensjahrzehnt angesetzt wird.⁵

2. Care-Arbeit

Traditionell wird die Altenpflege vor allem in Familien – und dort insbesondere von Frauen – unentgeltlich geleistet und von femi-

-
- 3 Bis heute gibt es keinen Konsens, was genau unter soziale Roboter fallen sollte, aber häufig werden service, companionship, therapy, und education genannt. *Hung et al.* bestimmen soziale Roboter so: „Social robots are innovative autonomous systems that feature a physical robot component, typically designed with a humanoid or anthropomorphic form. [...] Empowered with characteristics such as speech, gestures, and eye-gaze, which can be customized to a particular user or social context, social robots can behave like natural partners engaging humans in social interactions.“ *Robert/Fantinato et al.*: Social Robotics. Vgl. auch die Überblicksstudien: *Ragno/Borboni/ Vannetti et al.*: Application; *Morgan/Abdi/Syed et al.*: Robots in Healthcare; *González-González/Violant-Holz/Gil-Iranzo*: Social Robots; *Boada/Begoña/Genís*: The ethical issues.
 - 4 Dass Menschen Werkzeug machende und gebrauchende Tiere sind, geht auf Benjamin Franklin zurück; einen interessanten Ansatz legte Tim Ingold vor, der aber über meine Diskussion hier hinausgeht. Vgl. *Ingold*: Eight themes.
 - 5 *Misselhorn/Pompe/Stapleton*: Ethical Considerations.

nistischen Wissenschaftlerinnen als „Love’s Labor“ bezeichnet, die im weiteren Sinn häufig als unbezahlte „Care-Arbeit“ klassifiziert wird.⁶ Die institutionalisierte Betreuung und Pflege in den Industrieländern hat sich im Laufe des 20. Jahrhunderts zwar zunehmend professionalisiert, gleichzeitig blieben die Berufsrollen aber oft mit stereotypen Weiblichkeitsmodellen verbunden.⁷ Diese basieren auf der Annahme, dass soziales Engagement und ganz allgemein die Fürsorge für andere dem weiblichen Geschlecht näher ist – ein Vorurteil, das sich bis heute hartnäckig hält und eine geschlechtertheoretische Reflexion nach sich zieht. Dieses Vorurteil wurde aber leider auch von der feministischen Care-Ethik genährt, die zu Beginn der 1980er Jahre prominent wurde.⁸ Nicht zufällig ist die Care-Arbeit aber nicht nur weiblich, sondern auch theologisch konnotiert: über Jahrhunderte war die öffentliche Pflege und Sorge für bedürftige Menschen als Dienst christlicher Frauen angesehen, insbesondere aber auch in der Hand von Frauenorden. Sogenannte „Pflegeschwestern“ organisierten die Kranken- und Altenpflege, bevor diese zum Ende des 19. Jahrhundert zumindest in Deutschland in gleicher Weise auch von der staatliche Wohlfahrt übernommen wurde – bis heute geschieht dies aber in enger Kooperation mit den kirchlichen und anderen Wohlfahrtsverbänden.⁹ Heute riskieren Pflegekräfte, die dauerhaft in ihrem Beruf tätig sind, insbesondere durch die körperlichen Anforderungen der Altenpflege nicht nur eine geringere Lebensqualität als Angestellte anderer Gesundheitsberufe, sondern auch eine geringere Wertschätzung als zum Beispiel Pflegekräfte in der Versorgung von kranken Menschen.¹⁰ Die Prognosen eines zukünftigen Pflegenotstandes sind jedoch auch vor dem Hintergrund zunehmend ökonomisch orientierter Institutionen zu sehen. Pflegeinstitutionen definieren Effizianz Anforderungen unter dem Gesichts-

6 Kittay: Love’s labor.

7 Vgl. aus feministisch-theologischer Perspektive dazu *Sullivan-Dunbar*: Human dependency.

8 *Gilligan*: In a different voice. Für die Theologische Ethik ist das Standardwerk *Schnabl*: Gerecht sorgen. Die bioethische Auseinandersetzung mit der Care Ethik habe ich ausgeführt in *Haker*: Autonomy. Vgl. dazu *Heimbach-Steins/Bobbert*: Sorge – Care.

9 *Nelson*: Say Little.

10 *Emunds*: Gut versorgt? Vgl. auch *Brause/Kleina/Horn* et al.: The risk of burnout.

punkt der Rentabilität und nicht unbedingt im Hinblick auf die Gerechtigkeit gegenüber allen älteren Menschen und dem Pflegeberuf im Allgemeinen.¹¹

Ältere und alte Menschen haben das Recht, gut betreut und gut gepflegt zu werden. Die Gesellschaften schulden ihnen Solidarität und eine gerechte Sorge; beides basiert auf einem wenn nicht expliziten, dann doch immerhin impliziten Generationenvertrag, der Verantwortungen markiert. Nun wünschen sich Menschen heute den Ausbau ambulanter Leistungen, und sie wollen Alter mit möglichst großer Flexibilität in der Gestaltung ihres Lebensalltags erleben. Aber im vierten Lebensalter werden zunehmend Pflegeleistungen notwendig. Pflege und Betreuung schließen die psychische bzw. emotionale Befindlichkeit ein, was unter den Bedingungen einer individualisierten Gesellschaft schnell zu Überforderungen sowohl der Angehörigen als auch der professionell Betreuenden und Pflegenden führen kann.

3. Gründe für den Einsatz sozialtherapeutischer Roboter

Soziale Interaktionen basieren idealiter auf Wechselseitigkeit und Symmetrie der Beziehung, wobei auch asymmetrische Beziehungen nicht einfach als Dominanz und Unterwerfung betrachtet werden können – sie bedürfen aber einer besonderen moralischen Aufmerksamkeit für die Machtasymmetrie.¹² Darüber hinaus sind die Unberechenbarkeit der „Spielzüge“ in der Kommunikation, die gegenseitige Inanspruchnahme oder auch die gegenseitige Sorge füreinander Teil sozialer Interaktionen. In der ethischen Tradition steht – in Anlehnung an Aristoteles – das Modell der Freundschaft für eine solche gegenseitige Sorge und das wechselseitige Wohlwollen.¹³

11 Wessels: Pflegeökonomie.

12 Dies haben in Bezug auf Pflegeroboter Robert und Linda Sparrow schon relativ früh ausgeführt und jüngst noch einmal aufgenommen: *Sparrow/Sparrow: In the hands of machines?*; *Sparrow: Virtue and Vice*.

13 *Aristotle: Nicomachean ethics*. Vgl. dazu u.a. die Ausführungen in *Ricœur*: Das Selbst als ein Anderer. Hier ist zum Beispiel die Entwicklung von Robotern wie „Sophia“ relevant. Zu einer kritischen Diskussion solcher humanoiden Roboter vgl. *Parviainen/Coeckelbergh: The political choreography*.

Derzeit gibt es viele sozialtherapeutische Roboter auf dem Markt. Aber kein anderer ist so gut dokumentiert wie „Paro“, ein sozialtherapeutischer Roboter, der einer Spielzeugrobbe ähnelt.¹⁴ Die Gründe für die Einführung von Robotern in der Altenpflege sind vielfältig und aufs Engste mit dem sozialen Wandel, der Gesundheitspolitik sowie der Technikentwicklung im Bereich der Künstlichen Intelligenz verweben. Für die Einführung der sozialen bzw. sozialtherapeutischen Roboter werden verschiedene Gründe angegeben:

- *Demographischer Wandel*: Dieser führe zu einem erhöhten Bedarf an Pflege- und Betreuungskräften in allen Branchen. Die Betreuung in diesen Zusammenhängen wird bereits in erheblichem Umfang durch eine technische Ausstattung begleitet. Und die Abgrenzung zwischen diesen bereits als „traditionell“ geltenden technischen Pflegehilfsmitteln (Treppenlifte, Haarwaschmaschinen etc.), Pflegerobotern und sozialtherapeutischen Robotern ist nicht eindeutig.¹⁵
- *Mangel an Pflegekräften*: Immer weniger Menschen können ihre alternden Angehörigen pflegen oder sind bereit, professionell in der Pflege zu arbeiten. Roboter können nicht nur diese körperliche Arbeit übernehmen, sondern auch ein Überwachungssystem bereitstellen, das die persönliche Anwesenheit von Angehörigen und/oder Pflegekräften reduziert. Der Mangel an Pflegekräften, so das häufig geäußerte Argument, führe dazu, dass Roboter benötigt werden, um auch in Zukunft eine qualitativ hochwertige und ausreichende Versorgung zu gewährleisten.¹⁶
- *Unabhängigkeit und ambulante Assistenz*: Der Einsatz von Pflegerobotern zu Hause kann die von vielen älteren Menschen

14 Hung/Liu et al.: The benefits. Ein kurzes Video zu „Paro“ findet sich hier: <https://www.youtube.com/watch?v=PAJ2GXzaJtQ> (zuletzt gesehen: 13.11.2023).

15 Nach wie vor gibt es keinen Konsens zur Definition von sozialen Robotern. Vgl. dazu die Überblicksartikel von *Ragno/Borboni/Vannetti* et al.: Application; *Morgan/Abdi/Syed* et al.: Robots in Healthcare; *Scoglio/Reilly/Gorman* et al.: Use of Social Robots; *González-González/Violant-Holz/Gil-Iranzo*: Social Robots.

16 Japan ist dafür ein gutes Beispiel: „the Japanese Ministry of Health, Labor, and Welfare projects that by 2025, the country will have around 400,000 care workers less than it needs. To mitigate this crisis, the government has decided to increase the number of both foreign guest workers and robotic devices.“ *Aronsson*: Social robots.

angestrebte Unabhängigkeit fördern. Das Zusammenspiel von Pflegeassistenten, Monitoring und sozialer Interaktion erhöht die Wahrscheinlichkeit, dass Menschen länger in ihrer gewohnten Umgebung leben können.

- *Kosten*: Es wird erwartet, dass der Einsatz von Pflegerobotern dazu beitragen wird, Kosten zu senken, indem sie es den Pflegekräften ermöglichen, sich auf die „Kernaufgaben“ zu konzentrieren, d.h. auf die Interaktion mit älteren oder kranken Menschen.
- *Emotionale Assistenz*: Abgesehen von den Begleitrobotern werden sozialtherapeutische Roboter in der Altenpflege vor allem deshalb eingesetzt, weil sie berechenbar sind, weil sie Menschen emotional stimulieren können und weil sie helfen können, Gefühle wie Traurigkeit, Wut oder Einsamkeit zu regulieren. Dass sie dazu in der Lage sind, ist mittlerweile in empirischen Studien belegt.¹⁷

4. Die Debatte um sozialtherapeutische Roboter in der Altenpflege

Im Folgenden werde ich drei Modelle unterscheiden, die als Antwort auf die Entwicklung und Einführung von sozialtherapeutischen Robotern im Umgang mit älteren bzw. alten Menschen verstanden werden können.¹⁸

4.1. Das Ausschlussmodell

Das Ausschlussmodell basiert auf der Abwehr von sozialtherapeutischen Robotern in der Pflege. Es schließt aus, dass diese zu authentischen Interaktionen fähig sind, weil sozialtherapeutische Roboter zwar eine funktionale, nicht aber eine moralische Handlungsfähig-

17 Pirhonen/Tiilikainen/Pekkarinen et al.: Can robots; Pu/Moyle/Jones et al.: The Effectiveness. Die Modelle weichen von den früher beschriebenen insofern ab, als ich heute die Abwehr bzw. den Ausschluss von Robotern, den Ersatz für menschliche Pflege sowie die Ergänzung pflegerischer Tätigkeiten als präzisere Beschreibung betrachte.

18 Ich beziehe mich dabei auf mehrere vorhergehende Publikationen: *Haker*: Ethische Fragen; *Haker*: Social Robots; *Haker*: Information.

keit haben.¹⁹ Sie sollten daher bei vulnerablen Gruppen nicht zum Einsatz kommen. Denn anders als Puppen, bei denen die Interaktion nahezu vollständig von der Imagination ihres menschlichen Gegenübers abhängt – geleitet nur durch die Nachbildung von Babys oder Kleinkindern, die Ausdrucksweise eines menschlichen Gesichts und die Möglichkeit der Berührung und Bewegung von Armen und Beinen – haben Roboter zugleich einen größeren und einen kleineren Aktionsradius: der größere Radius wird durch das Design der Algorithmen vorgegeben, der entsprechend von außen berechenbar ist. Vulnerable Gruppen, insbesondere Kinder, psychisch kranke Menschen und Menschen mit Demenz werden von nahezu echten Nachahmungen authentischer Interaktionen getäuscht. Zugleich ist der Handlungsspielraum kleiner, weil sozialtherapeutische Roboter nicht kontingent sind – Unberechenbarkeit und Kontingenz des Handelns macht aber moralisches Handeln zumindest teilweise aus. Das Ausschlussmodell basiert daher auf dem ethischen Urteil, dass die Täuschung authentischer Interaktionen mehr Schaden – im Sinne von moralischen Verletzungen – anrichtet als dies durch den unzweifelhaft bestehenden Nutzen zu rechtfertigen ist.

4.2. Das Substitutionsmodell als Ersatz menschlicher Sorge

In der Rhetorik der Entwickler:innen sozialtherapeutischer Roboter wird implizit oder explizit damit geworben, dass Roboter dabei helfen können, Kosten für Personal und/oder Medikamente zu sparen. Zudem sind, so die Annahme, sozialtherapeutische Roboter ideal für die emotionale Kommunikation, die insbesondere sehr alte bzw. demente Patienten benötigen, weil diese oft großen Stimmungsschwankungen unterworfen sind, Einsamkeit sich zu Depressionen entwickeln kann, und das Bedürfnis einer emotionalen Kommunikation von Angehörigen nur teilweise gestillt werden kann. Den Betreuenden in der Altenpflege fehlt häufig die Zeit für die emotionale Zuwendung. Roboter füllen also eher eine Lücke, als dass sie Menschen einen Platz wegnehmen würden.²⁰

¹⁹ *Haker: Experience.*

²⁰ Die Zahlen des Pflegemangels sind vielleicht unzuverlässig, aber der oben genannte Fall Japans lässt diesen Schluss sicher zu. Dabei soll nicht verschwiegen

Auch wenn Roboter ein zunehmend größeres Repertoire an Kommunikationsfähigkeit haben und in Zukunft womöglich ein breites Spektrum an Emotionen abbilden können, werden sie dennoch niemals vollständig auf neue Situationen reagieren können. Sie sind keine Lebewesen: Menschen sind aufgrund ihrer inneren bzw. leiblichen Subjektivität, die ihre Fremdheit bewahrt, nie ganz transparent. In Interaktionen mit anderen bedarf es einer grundsätzlichen Offenheit, um neue Erfahrungen zu machen.²¹ Auch Tiere sind aufgrund ihres Instinktverhaltens niemals vollständig zu disziplinieren bzw. zu zähmen. Roboter bleiben demgegenüber auch dann noch Automaten, wenn sie uns dazu animieren, uns vorzustellen, sie seien Subjekte oder Tiere, oder wenn sie uns durch ihre Art und Weise der Kommunikation täuschend echt vorkommen.²²

Während das Ausschlussmodell klare Grenzen zwischen authentischen und funktionalen Interaktionen zieht, wird durch die Entwicklung der lernfähigen Roboter die klare Unterscheidung zwischen „authentischen“ und „simulierten“ Beziehungen sowie zwischen „heteronomen“ und „autonomen“ Interaktionen bewusst aufgehoben.²³ Das digitale „Spiel-Zeug“ ist gerade nicht so abhängig von seinen Designern oder auch von seinen Anwendern, wie dies traditionell der Fall war. Es ist die „Zwischenkategorie“ der Roboter: *zwischen* Tieren und Maschinen, *zwischen* Heteronomie und Autonomie, *zwischen* Authentizität und Simulation, welche die kategorische Ablehnung, die dem Ausschlussmodell zugrunde liegt, zweifelhaft erscheinen lässt und das Substitutionsmodell attraktiv erscheinen lässt.²⁴ Die Täuschungsgefahr ist damit aber nicht ausgeräumt: wenn Benutzer:innen in einer vulnerablen Position bzw.

werden, dass die alternative Strategie eine offensive Immigrationspolitik ist, die aber wiederum ihre eigenen Probleme nach sich zieht und die Kostenfrage nicht klärt. Aronsson: Social robots.

21 Dies ist die bleibende Einsicht der Phänomenologie. Vgl. die einschlägige Studie dazu: *Theunissen*: Der Andere. Vgl. auch: *Gilson*: The ethics of Vulnerability.

22 Vgl. dazu: *Johnson/Verdicchio*: Why robots. Die Arbeit von Kathrin Schuster konnte für diesen Artikel leider nicht mehr berücksichtigt werden, soll aber immerhin erwähnt werden: *Schuster*: Therapieroboter.

23 *Gunkel*: The machine.

24 Sharkey & Sharkey zitieren Kahn et al., die von der Notwendigkeit sprechen, eine neue ontologische Kategorie einzuführen, um die traditionelle Unterscheidung von belebten und unbelebten Wesen zu durchbrechen *Sharkey/Sharkey*: Granny and the robots.

Phase ihres Lebens, die durch die Asymmetrie und Abhängigkeit anderen gegenüber gekennzeichnet ist, mit KI-gesteuerten Robotern zu tun haben, mögen sie sich in der Tat über deren Autonomie täuschen. Während sie – abhängig von ihren kognitiven Fähigkeiten, Roboter *als* Roboter zu identifizieren – davon ausgehen mögen, dass sie selbst bestimmte Reaktionen des Roboters hervorrufen, folgen diese einfach nur der Programmierung. Reziprozität im Sinne der spontanen und wechselseitigen Interaktion ist nicht möglich: die Handlungsfähigkeit von Robotern ist grundsätzlich funktional und bleibt dem empirisch-instrumentellen Verständnis verhaftet.²⁵ Das Substitutionsmodell „normalisiert“ eine normative Annahme, nämlich dass soziale und sozialtherapeutische Roboter aufgrund des Mangels an menschlicher Kommunikation eine wichtige Lücke füllen. Mangels Alternativen kann mit ihrem Einsatz das Wohlergehen von älteren und alten Menschen gesteigert werden. Solange bzw. sofern ein Pflegenotstand besteht, ist dieses Modell nahezu immun gegen Kritik, weil es keine Zweifel am Nutzen der Substitution aufkommen lässt, ohne die Grenzen des Modells zu verschweigen. Der Einsatz sozialtherapeutischer Roboter bei vulnerablen Gruppen repariert ein Defizit, das die Gesellschaft nicht bereit ist, anders zu beheben.

4.3. Das Komplementaritätsmodell: Sozialtherapeutische Roboter als Ergänzung der Pflege

Während die beiden genannten Modelle den „Ersatz“ des Menschen durch Roboter entweder ablehnen oder begrüßen, ist es vielleicht zutreffender, sozialtherapeutische Roboter als Ergänzung zu sozialen Interaktionen und komplementär zu ihnen zu begreifen. Das ist die Perspektive, die durch das komplementäre Modell betont wird. Man kann allerdings fragen, ob dies Modell angesichts der jüngsten Entwicklungen in der Robotik nicht naiv ist. So kann beispielsweise die zunehmende Unabhängigkeit der Roboter sowie ihre Lernfähigkeit

25 *Sparrow/Turkle*: Simulation. Freilich hat die Modernekritik lange darauf hingewiesen, dass die „Kolonialisierung der Lebenswelt“ (Habermas) und die ihr zugrunde liegende instrumentelle Vernunft gerade darin besteht, Alterität aus nahezu allen Bereichen zu verdrängen, so dass der Einsatz der sozialen Roboter diesen Trend nur fortsetzt.

Anreize schaffen, vom Modell der „Ergänzung“ auf das Modell der „Ersetzung“ umzusteigen. Unterliegen die Roboter der Kontrolle der Pflegekräfte, setzt dies die Anwesenheit von Pflegekräften voraus; aber dann stellt sich die Frage nach der Wirtschaftlichkeit und dem Pflegenotstand erneut, die ja zusammen die Anschaffung überhaupt erst motiviert haben mögen. Wenn also sozialtherapeutische Roboter nur komplementär eingesetzt werden, kann das Spielen mit Puppen und/oder Stofftieren womöglich genauso effektiv sein wie das Spielen mit ihren technischen Gegenständen – zumindest wäre dies in empirischen Vergleichsstudien zu prüfen.

5. Ethische Überlegungen

5.1. Allgemeine Bioethik

Da Roboter in den kommenden Jahren und Jahrzehnten immer komplexer werden, wird die Bandbreite unserer Interaktionen mit ihnen zunehmen und der Einsatz sozialtherapeutischer Roboter weiter kommerzialisiert werden. Gleichzeitig bedarf es ethischer Regeln, um zu gewährleisten, dass ihr Einsatz so weit wie möglich im Sinne des Komplementaritätsmodells eingesetzt werden, anstatt schleichend zum Ersatz menschlicher Pflege und Interaktion überzugehen. Die bioethische Diskussion steht hier erst am Anfang, wie ich im Folgenden kurz erläutern will. Die ethischen Überlegungen werden meist wissenschaftlichen Expertenkommissionen, parlamentarischen Debatten und Regierungen überlassen. Die von der OECD 2019 vorgeschlagenen „Empfehlungen“ beinhalten fünf Prinzipien: Wachstum (1), nachhaltige Entwicklung und Wohlergehen (2), menschenzentrierte Werte und Fairness (3), Transparenz und Erklärbarkeit (4), Robustheit, Sicherheit und Schutz sowie Rechenschaftspflicht (5).²⁶ Floridi und Cowls zählten nicht weniger als 46 Prinzipien in den wichtigsten internationalen Berichten, obwohl sie argumentieren, dass viele von ihnen unter die vier traditionellen bioethischen Prinzipien subsumiert werden könnten. Sie schlagen

26 „inclusive growth, sustainable development and well-being; human-centred values and fairness; transparency and explainability; robustness, security and safety; and accountability.“ *OECD: Recommendation.*

vor, diese durch den neuen Grundsatz der „Erklärbarkeit“, den sie sowohl als Verständlichkeit als auch als Verantwortlichkeit verstehen wollen, zu ergänzen.²⁷

Welche normativen Rahmen für die AI-Ethik leitend sein werden, hängt nicht zuletzt von den Protagonist:innen ab, die den ethischen Diskurs markieren. Bisher ist dieser weitgehend von einer traditionellen Bioethik markiert. So hat Mark Coeckelberg – im Gegensatz zur pragmatischen Bioethik, die auch Cowsls verfolgt – vorgeschlagen, AI-Ethik im Sinne einer Menschenrechts- bzw. Befähigungsethik in der Nachfolge von A. Sen und M. Nussbaum weiterzuentwickeln.²⁸ Aber reicht eine Menschenrechtsethik aus? Oder müssen darüber hinaus neue Zugänge ausgearbeitet werden, die passgenauer sind? So stellen sich insbesondere aus biographischer Perspektive Fragen der moralischen Identität und Autonomie: im Sinne existentiell-ethischer Lebensentwürfe könnten Menschen etwa prospektiv festlegen, ob sie im Fall von Demenz dem Einsatz sozialtherapeutischer Roboter zustimmen. Während es des Weiteren stimmt, dass nur Moralsubjekte für ihre Handlungen verantwortlich sind und daher von anderen zur Rechenschaft gezogen werden können, stellen sich Verantwortungsfragen für „selbstlernende“ Roboter neu. Wie etwa Nida-Rümelin betont, wird derzeit jedoch das Verständnis von Identität, Freiheit und Handlungsfähigkeit im Lichte eines alternativen Rationalitätsmodells neu definiert, das auf den kartesischen Erkenntnisbegriff und die Baconsche Tradition der instrumentellen Vernunft – im Unterschied zum Idealismus der metaphysischen Tradition – zurückgeht. Da beide, der empirisch-instrumentelle Vernunftbegriff, welcher der Entwicklung der Künstlichen Intelligenz zugrunde liegt, wie der subjekttheoretische Begriff, der in der modernen Moralphilosophie und -theologie entfaltet wird, in der Moderne verwurzelt sind, können die Spannungen zwischen beiden Ansätzen nicht einfach durch Rekurs auf die Moderne-Tradition behoben werden.²⁹

In den nächsten Jahren werden Roboter technisch weiterentwickelt werden, und sollten empirische Studien ihres Einsatzes in

27 *Floridi/Cowsls*: A unified framework.

28 *Coeckelbergh*: Robot rights; *Coeckelbergh*: AI Ethics.

29 Charles Taylor hat die beiden Traditionen ausführlich behandelt und ihre Inkompatibilität betont. *Taylor*: The language animal.

der Betreuung und Pflege älterer Menschen weiterhin bestätigen, dass die Interaktion mit ihnen das Wohlbefinden steigert, ist anzunehmen, dass sich die Begründungslast in die Richtung derjenigen, die die Anwendung ablehnen, verschieben wird – dies betrifft vor allem diejenigen, die dem Ausschlussmodell anhängen und folgern, dass der Einsatz sozialtherapeutischer Roboter die Grenze zwischen funktionaler und authentischer Interaktion nicht transparent machen kann. Damit haben sie menschenrechtsbasierte bioethische Ansätze auf ihrer Seite, denn diese zielen darauf ab, die Würde und die Rechte von Individuen und Gruppen zu schützen, institutionalisierte Verantwortung zu ermöglichen und sicherzustellen, dass Akteure für ihre Handlungen zur Rechenschaft gezogen werden können. Nach wie vor werden aber auch die vier Prinzipien der Medizinethik: Autonomie, Schadensvermeidung, Wohlverhalten und Gerechtigkeit angeführt und auf die Anwendung der sozialen und sozialtherapeutischen Roboter bezogen.³⁰ Die Prinzipien sind in der Tat hilfreich, aber sie reichen weder für eine Technikethik aus noch können sie die spezifische Herausforderung der Künstlichen Intelligenz angemessen normativ fassen. Ethische Fragen müssen aber kontextspezifisch konkretisiert werden und sowohl in der technischen Entwicklung als auch in den jeweiligen Anwendungskontexten verankert werden. Alternativ – oder in Ergänzung zu den Grundlagen der Bioethik – sind daher neben Sicherheitsvorkehrungen Grundkenntnisse der responsorischen Phänomenologie zu nennen, die dahingehend normativ verschärft werden kann, dass sie mit den ersten drei Prinzipien der Bioethik verschränkt wird.³¹ Darüber hinaus gilt es, den Respekt der Privatsphäre unter Datenschutzgesichtspunkten zu justieren – was angesichts der Surveillance Technologien, die mit den sozialtherapeutischen Robotern verbunden werden können, dringlich ist.³² Aber genauso sind Maßnahmen zur Prävention von Diskriminierung, Infantilisierung und Täuschung

30 Zu einem Überblick, der, was ansonsten selten ist, auch die deutsche Literatur berücksichtigt, vgl. *Wullenkord/Eyssel*: Societal and ethical issues in HRI.

31 Als Grundlage einer phänomenologischen Responsivität sind vor allem die Studien von Bernhard Waldenfels zu nennen, die inzwischen auch in ethischen Kreisen breit rezipiert worden ist, meistens allerdings ohne die Verschränkung mit normativen Prinzipien zu reflektieren: *Waldenfels*: Bruchlinien; *Waldenfels*: Phänomenologie; *Waldenfels*: Erfahrung.

32 *Zuboff*: The age.

von vulnerablen Menschen zu konkretisieren. Und schließlich ist im Umgang mit älteren und alten Menschen darauf hinzuwirken, dass diese auch selbst die sozialtherapeutischen Roboter empathisch und respektvoll behandeln, sofern sie dies kognitiv leisten können. Ansonsten nämlich kann der Umgang mit sozialen Robotern genauso Aggressionen, Dehumanisierungen oder gar Sadismus fördern, wie dies in sozialen Beziehungen auch der Fall ist. Der respektvolle Umgang markiert dabei gerade nicht einen Personen-Status von Robotern, wie einige Autor:innen meinen; vielmehr erinnert er an mögliche Auswirkungen im Umgang mit anderen, von Angehörigen bis zu anderen Mitbewohnern und natürlich den Pflegenden in Pflegeheimen.³³

5.2. Der Beitrag der christlichen Bioethik

Christliche Ethikansätze sind so vielfältig wie ihre philosophischen Pendanten; zu Recht folgen sie weitgehend den philosophisch-ethischen Ansätzen, vertreten zum Beispiel einen tugendethischen Ansatz, einen Menschenrechtsansatz, eine utilitaristische Ethik oder auch einen narrativ-ethischen Ansatz.³⁴ In den Prinzipien der Katholischen Soziallehre sind Referenzpunkte zusammengefasst, die über die vier bioethischen Prinzipien hinausgehen und für die ethische Beurteilung sozialtherapeutischer Roboter nutzbar gemacht werden können: Neben der Menschenwürde sind dies Subsidiarität, Solidarität, Gemeinwohl und Gerechtigkeit, und hier besonders die Option für die bzw. der Vorrang der Armen. In ökumenischer Perspektive betrachtet, beharrt die christliche Ethik auf zwei fundamentalen (normativen) Annahmen: dem Prinzip der Würde des Menschen, die sowohl die Verletzlichkeit als auch die Handlungsfähigkeit einschließt,³⁵ sowie dem Prinzip der Verantwortung. *Würde* verweist auf die Unsicherheit der menschlichen Freiheit aufgrund der Anfälligkeit für Leid und Leid durch andere, aufgrund der sozialen Abhängigkeit und der strukturellen Verwundbarkeit, die die Möglichkeit erhöht, des individuellen Rechts beraubt zu werden, an-

33 Reiss: Robots; Wullenkord/Eyssel: Societal and ethical issues in HRI.

34 Zwei Werke seien genannt, die die theologisch-ethische Diskussion im deutschsprachigen Raum abbilden: Voigt: Religion; Platzer: Bioethik.

35 Haker: Verletzliche Freiheit.

erkannt und geschützt zu werden und mit den notwendigen Mitteln ausgestattet zu werden, um ein menschenwürdiges Leben zu führen. *Verantwortung* verweist auf die Handlungsfähigkeit von Personen als die Freiheit, zu handeln, auf andere zu antworten, Rechenschaft über die eigenen Handlungen abzulegen, und ungerechte Strukturen durch soziales und politisches Handeln in gerechtere Strukturen umzuwandeln. Das bedeutet: Aus sozialetischer Perspektive ist der Ausgangspunkt nicht der Roboter als kommerzielles Produkt des KI-Marktes, sondern die ausstehende Antwort auf eine gesellschaftliche Herausforderung und den Anspruch, den Menschen an moralische Subjekte und an die hinter ihnen stehenden Institutionen stellen – in diesem Fall die würdevolle und verantwortliche Versorgung von älteren und alten Menschen mit unterschiedlichen Bedürfnissen und Interessen.

Prinzipien sind jedoch zu abstrakt und zu weit von jedem Kontext entfernt, um eine unmittelbare Orientierung zu bieten. Die Tradition der Moralthologie, die häufig auf das Naturrecht reduziert wird, birgt in der Entfaltung einer tugendethischen Klugheitsethik Kriterien, die es erlauben, kontextuelle moralische Urteile zu eruieren anstatt sie aus den Prinzipien zu deduzieren. Diese Kriterien, oft auch mittlere Prinzipien genannt, ermöglichen es, kontextsensitiv eine ethische Beurteilung von sozialtherapeutischen Robotern vorzunehmen: Dazu gehören die Aufmerksamkeit für die *Umstände* einer Handlung oder Praxis – etwa die Berücksichtigung der informellen, der ambulanten oder der institutionellen Betreuung und Pflege; das Lernen aus *Erfahrungen der Vergangenheit* – etwa im Austausch mit Ländern, die schon länger mit sozialtherapeutischen Robotern arbeiten, aber auch die Berücksichtigung der psychologischen und phänomenologischen Forschung; die Berücksichtigung der zu erwartenden Folgen in der Zukunft (*Vorsichtsprinzip*), die sich vor allem aus soziologischen, aber auch aus gesundheitsökonomischen Studien prognostizieren lassen oder auch in *Zukunfts-Szenarien* vorgestellt werden können;³⁶ und nicht zuletzt mahnt die Klugheitsethik die Offenheit für praktische, wenn nicht sogar *pragmatische Lösungen* an – wie dies angesichts des Pflegeotstands

36 *Krieger*: The Future. Für Deutschland scheinen gesundheitsbezogene Szenarien gemischte Ergebnisse zu zeigen. Vgl. etwa *Radic/Vosen/Graf*: Use of Robotics.

unter Umständen angezeigt ist.³⁷ Mit Paul Ricœur kann damit eine dreigliedrige Bioethik entworfen werden, die strebensethische, normative und klugheitsethische Aspekte integriert.³⁸

6. Schluss: Thesen zur Umsetzung ethischer Gesichtspunkte in der Weiterentwicklung und in der Ausbildung

Im Gespräch mit KI-Wissenschaftler:innen, IT-Unternehmen, Dienstleistungsorganisationen und -unternehmen sowie auf der Grundlage von qualitativen und quantitativen Studien sollte die technische Entwicklung von Gesellschaft und Kirchen, Wissenschaften und Politik begleitet werden. Um dies anzustoßen bzw. bestehende Kooperationen zu stärken, will ich nun einige Überlegungen für den Einsatz von sozialtherapeutischen Robotern in ambulanten Sozialdiensten sowie für die Ausbildung von Pflegenden und Betreuungspersonal formulieren. Insbesondere ist meine Hoffnung, dass sie in Deutschland das Gespräch mit den Verantwortlichen eröffnen, im Hinblick auf die derzeitigen politischen Reformen des Gesundheitswesens berücksichtigt werden und in der gesellschaftlichen und kirchlichen Diskussion einen Ansatzpunkt für weitergehende Überlegungen bergen. In Zukunft könnte es für Betreuer:innen älterer Menschen, aber auch für Pflegekräfte selbstverständlicher werden, sich in der Pflege und in der Betreuung auf Roboter zu verlassen, weil sie effizienter und, vielleicht überraschenderweise, auch der nächsten Generation von Menschen im vierten Lebensalter vertrauter werden. Der Einsatz von Pflege- und sozialtherapeutischen Robotern insbesondere in der Altenpflege muss daher vor dem allgemeinen Hintergrund der Automatisierung und Digitalisierung des Alltags gesehen werden. Eine Minimalforderung ist es, den Umgang mit sozialen und sozialtherapeutischen Robotern zu einem Bestandteil der Aus- und Weiterbildung zu machen, wobei diese nicht nur die technische Handhabung, sondern auch die psychologischen und ethischen Dimensionen umfassen müssen.

37 Zur Reflexion über die Tugend der Klugheit vgl. *Thomas von Aquin: Summa Theologiae*, II-II, Quaestiones 47–56. Fehlt im Literaturverzeichnis!

38 *Ricœur: Oneself*. Vgl. zur Auseinandersetzung mit diesen verschiedenen Ebenen *Haker: Moralische Identität*.

1. Die spielerische, fantasievolle Interaktion von Menschen und (sozialtherapeutischen) Robotern kann das emotionale Gleichgewicht älterer und alter Menschen fördern und dadurch das Wohlbefinden derjenigen, die auf soziale und/oder emotionale Unterstützung angewiesen sind, erhalten oder steigern. Es gibt keinen Grund, ihren Einsatz in der Altenpflege und -betreuung abzulehnen. Wohl aber sind Vorsichtsmaßnahmen zu treffen, die zu betreuende Menschen vor Täuschung, Respektlosigkeit und Infantilisierung schützen und vor der Ausbeutung ihrer emotionalen Verletzlichkeit bewahren.
2. Sind in Pflegekonstellationen Interaktionen zwischen Klient:innen/Patient:innen und Pflegekräften, aber auch Mensch-Haus-tier-Interaktionen zeitlich und räumlich möglich, sollten diese vorrangig ermöglicht und gepflegt werden. Wenn sozialtherapeutische Roboter eingesetzt werden, müssen sie in die allgemeinen (pflegerischen und betruerischen) Interaktionen integriert werden. Roboter dürfen nicht als Ersatz für menschliche Zuwendung, sondern nur als Ergänzung zu ihr eingesetzt werden.
3. Wenn Roboter Handlungen ausführen, die bestimmte emotionale Reaktionen auslösen, muss die Intimität und Integrität der Personen gesichert werden. Die Privatsphäre ist zu jedem Zeitpunkt zu respektieren – dies gilt besonders für integrierte Surveillance Technologien, also solche Monitoring bzw. Überwachungstechnologien, die etwa per eingebauten Videos die Aufsicht erleichtern sollen, um Verletzungsgefahren zu vermeiden oder schnell eingreifen zu können.
4. Auch wenn der Übergang von sozialer und digitaler Interaktion schleichend ist, sind bewusste Täuschungen über den artifiziellen Charakter von sozialtherapeutischen Robotern, die die emotionale Verletzlichkeit eines älteren Menschen ausnutzen, unter allen Umständen zu vermeiden. Dies muss durch Informationsmaterial und Kommunikationsrichtlinien, die sowohl den betroffenen älteren Menschen als auch ihren Angehörigen zugänglich gemacht werden müssen, sichergestellt werden. Pflegekräfte müssen kontinuierlich geschult werden, mit dem Ziel, sie mit den technologischen Entwicklungen in der Robotik und Pflege sowie den psychologischen Mustern und den Basisformen einer responsorischen Phänomenologie vertraut zu machen und eine ethisch verantwortungsvolle Integration von Technologien zu gewährleisten.

5. Wenn ältere Menschen dazu in der Lage sind, müssen sie darüber informiert werden, wie Roboter funktionieren. Wenn sie nicht in der Lage sind, dem Einsatz von Robotern zuzustimmen, muss eine stellvertretende Zustimmung eingeholt werden, und das Pflegepersonal muss darin geschult werden, Roboter in einer Weise zu verwenden, die die Anerkennung, den Schutz und die Bereitstellung von Patientenrechten gewährleistet. Dies erfordert ein situations- und kontextabhängiges Training. Ethiker:innen sollten mit Pflegekräften und KI-Spezialist:innen zusammenarbeiten. Umgekehrt müssen diese Spezialist:innen bei jedem Schritt von Robotikprojekten mit Ethiker:innen zusammenarbeiten. Universitäten sollten interdisziplinäre Projekte entwickeln, um solche Ausbildungen zu sichern.
6. Roboter können unter Umständen ein Sicherheitsrisiko darstellen. Die Verantwortlichkeit für ihre Verwendung muss geklärt werden, insbesondere wenn sie in Abwesenheit von Betreuungspersonen eingesetzt werden. Darüber hinaus können und werden sozialtherapeutische Roboter gleichzeitig mit Überwachungsrobotern entwickelt werden. Beide sind zwar terminologisch getrennt, aber ihre Funktionen sind oft nicht klar voneinander getrennt – Überwachungsroboter können ihre Aufgabe hinter der Fassade eines sozialtherapeutischen Roboters übernehmen. Daher ist es wichtig, dass bei der Entwicklung von sozialen Robotern das Recht der Patienten auf Privatsphäre nicht verletzt wird:
7. Wenn Roboter nicht nur für die soziale Kommunikation mit Menschen, sondern gleichzeitig zu deren Beobachtungs- und/oder Überwachungszwecken eingesetzt werden, muss dies genauer geprüft werden und im Einklang mit dem Recht auf Privatsphäre und Integrität stehen. Dafür gibt es Richtlinien, die bei jeder neuen Anwendung umgesetzt und nicht neu erfunden werden müssen.

Die technischen Roadmaps der Robotik müssen von ethischen Roadmaps begleitet werden, die kontinuierlich an die neuen Anwendungen angepasst werden müssen – nicht auf der Ebene der Grundsätze oder Prinzipien, wohl aber auf der Ebene der spezifischen Kriterien. Wie in anderen Bereichen auch gilt es zu verhindern, dass technologische Entwicklungen gesellschaftliche Praktiken radikal verändern, ohne diese Veränderungen einem zivilgesellschaftlichen Diskurs zu unterziehen – denn dies bedeutet die Unterwerfung des

Menschen unter die Technik. „Technik“ ist dabei kein abstrakter Begriff, wird ihre Entwicklung, insbesondere im KI-Bereich, doch massiv die von einzelnen Menschen – bekanntlich ausgestattet mit exorbitanten Finanzmitteln – vorangetrieben. Die derzeitige Zunahme der Entwicklung von sozialen Robotern innerhalb und außerhalb der Gesundheitspraxis ist ein Indiz dafür, dass es einen dringenden Bedarf an öffentlichen Diskursen und Reflexionen gibt. Aber auch die nachfolgende, explosionshafte Zunahme der Literatur zur Robotik, die durch die Covid-Pandemie seit 2020 verzeichnet wird, ist Grund genug, neu darüber nachzudenken, wie wir den *Diskurs* über das vierte Lebensalter führen können. Diese Aufgabe stellt sich nicht nur, aber eben auch der theologischen Bioethik.

Bibliografie

- Aristotle*: Nicomachean ethics, Oxford 2020.
- Aronsson*, Anne Stefanie: Social robots in elder care the turn toward emotional machines in contemporary Japan, in: Japanese review of cultural anthropology 21 (1) 2020, 421–455.
- Boada*, Júlia Pareto/*Maestre*, Begoña Román/*Genís*, Carme Torras: The ethical issues of social assistive robotics. A critical literature review, in: Technology in Society 67/2021, 101726.
- Coeckelbergh*, Mark: AI Ethics, Cambridge 2020.
- Coeckelbergh*, Mark: Robot rights? Towards a social-relational justification of moral consideration, in: Ethics and Information Technology 12 (3), 2010, 209–221.
- Emunds*, Bernhard: Gut versorgt? Ökonomie und Ethik im Gesundheits- und Pflegebereich, Baden-Baden 2019.
- Floridi*, Luciano/*Cowls*, Josh: A unified framework of five principles for AI in society, in: Harvard Data Science Review 1 (1), 2019, 1–15.
- Gilligan*, Carol: In a different voice: psychological theory and women's development, Cambridge 1993 (1st ed. 1982).
- Gilson*, Erinn C.: The ethics of vulnerability. A Feminist Analysis of Social Life and Practice, New York 2014.
- González-González*, Carina Soledad/*Violant-Holz*, Verónica/*Gil-Iranzo*, Rosa Maria: Social Robots in Hospitals. A Systematic Review, in: Applied Sciences 11 (13), 2021, 5976.
- Gunkel*, David J.: The machine question. Critical perspectives on AI, robots, and ethics, Cambridge 2012.

- Haker, Hille:* Autonomy and Care in Medicine, in: Linköping Electronic Conference Proceedings 742013.
- Haker, Hille:* Ethische Fragen des Einsatzes von Pflegerobotern, in: Leggewie, C. (Hg.): Kooperation ohne Akteure? Automatismen in der Globalisierung (Global Dialogue 5), Duisburg 2014, 55–68.
- Haker, Hille:* Experience, Identity and Moral Agency in the Age of Artificial Intelligence, in: Nagl-Docekal, H./Zacharasiewicz, W. (Hg.): Artificial Intelligence and Human Enhancement. Affirmative and Critical Approaches in the Humanities, Berlin 2022, 51–77.
- Haker, Hille:* Information or communication – the loss of the language of the human, in: Lichner, M. (Hg.): Hope. Where does our hope lie? (International Congress of the European Society of Catholic Theology 2019 – Bratislava, Slovakia), Berlin 2021, 505–523.
- Haker, Hille:* Moralische Identität. Literarische Lebensgeschichten als Medium ethischer Reflexion. Mit einer Interpretation der „Jahrestage“ von Uwe Johnson, Tübingen 1999.
- Haker, Hille:* Social Robots in Elderly Care, in: Nexus 2, Robots and Rituals. Reflections on Faith in the Era of Science and AI, 2023, no pages.
- Haker, Hille:* Verletzliche Freiheit. Zu einem neuen Prinzip der Bioethik, in: Keul, H. (Hg.): Theologische Vulnerabilitätsforschung. Gesellschaftsrelevant und interdisziplinär, Stuttgart 2021, 99–118.
- Heimbach-Stein, Marianne/Bobbert, Monika* (Hgg.): Sorge – Care. Anthropologische Zugänge–Ethische Konzepte–Gesellschaftliche Praxen. Gesellschaft–Ethik–Religion 26, Paderborn 2025.
- Hung, Lillian/Liu, Cindy, et al.:* The benefits of and barriers to using a social robot PARO in care settings. A scoping review, in: BMC Geriatrics 19 (1), 2019, 232.
- Ingold, Tim:* Eight themes in the anthropology of technology, in: Social Analysis: The International Journal of Social and Cultural Practice 41 (1), 1997, 106–138.
- Johnson, Deborah G./Verdicchio, Mario:* Why robots should not be treated like animals, in: Ethics and information technology 20 (4), 2018, 291–301.
- Kittay, Eva Feder:* Love’s labor: essays on women, equality, and dependency, New York 1999.
- Krieger, Benedikt:* The Future of Human-Robot Interaction. A socio-economic Scenario Analysis, Thesis KTH Royal Institute of Technology, School of Industrial Engineering and Management, Stockholm 2020.
- Misselhorn, Catrin/Pompe, Ulrike/Stapleton, Mog:* Ethical Considerations Regarding the Use of Social Robots in the Fourth Age, in: GeroPsych 26 (2), 2013, 121–133.

- Morgan, Ahmed Ashraf/Abdi, Jordan/Syed, Mohammed A. Q. et al.*: Robots in Healthcare. A Scoping Review, in: *Current Robotics Reports* 3 (4), 2022, 271–280.
- Nelson, Sioban*: Say Little, Do Much. Nursing, Nuns, and Hospitals in the Nineteenth Century, Philadelphia 2003.
- OECD*: Recommendation of the Council on Artificial Intelligence, Vol. OECD/LEGAL 0449, Edited by Series Editor 21/05/2019. Aktualisiert 07/11/2023.
- Parviainen, Jaana/Coeckelbergh, Mark*: The political choreography of the Sophia robot. Beyond robot rights and citizenship to political performances for the social robotics market, in: *AI & society* 36 (3), 2021, 715–724.
- Pirhonen, Jari/Tiilikainen, Elisa/Pekkarinen, Satu et al.*: Can robots tackle late-life loneliness? Scanning of future opportunities and challenges in assisted living facilities, in: *Futures* 1242020, 102640.
- Platzer, Johann*: Bioethik und Religion. Theologische Ethik im öffentlichen Diskurs. Walter Schupp zum 60. Geburtstag, Edited by Series Editor, Baden-Baden 2014.
- Pu, Lihui/Moyle, Wendy/Jones, Cindy et al.*: The Effectiveness of Social Robots for Older Adults. A Systematic Review and Meta-Analysis of Randomized Controlled Studies, in: *The Gerontologist* 59 (1), 2018, e37-e51.
- Radic, Marija/Vosen, Agnes/Graf, Birgit*: Use of Robotics in the German Healthcare Sector, Edited by Series Editor, Cham 2019, 434–442.
- Ragno, Luca/Borboni, Alberto/Vannetti, Federica et al.*: Application of Social Robots in Healthcare. Review on Characteristics, Requirements, Technical Solutions, in: *Sensors* 23 (15), 2023, 6820.
- Reiss, Michael J.*: Robots as persons? Implications for moral education, in: *Journal of moral education* 50 (1), 2021, 68–76.
- Ricœur, Paul*: Das Selbst als ein Anderer, München 1996.
- Robert, Lionel P./Fantinato, Marcelo et al.*: Social Robotics Business and Computing, in: *Information Systems Frontiers*, 2023, no page.
- Schnabl, Christa*: Gerech sorgen. Grundlagen einer soziaethischen Theorie der Fürsorge, 2005.
- Schuster, Kathrin*: Therapieroboter in der Betreuung demenzbetroffener Personen. Eine moralphilosophische Diskussion, Osnabrück 2021.
- Scoglio, Arielle A. J./Reilly, Erin D./Gorman, Jay A. et al.*: Use of Social Robots in Mental Health and Well-Being Research. Systematic Review, in: *J Med Internet Res* 21 (7), 2019, e13322.
- Sharkey, Amanda/Sharkey, Noel*: Granny and the robots. Ethical issues in robot care for the elderly, in: *Ethics and Information Technology* 14 (1), 2012, 27–40.

- 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), 2021, 23–29.
- Sparrow*, Robert/*Sparrow*, Linda: In the hands of machines? The future of aged care, in: *Minds and Machines* 16 (2), 2006, 141–161.
- Sullivan-Dunbar*, Sandra: *Human dependency and Christian ethics*, 2017.
- Taylor*, Charles: *The language animal. The full shape of the human linguistic capacity*, Cambridge 2016.
- Theunissen*, Michael: *Der Andere: Studien zur Sozialontologie der Gegenwart*, Berlin 1977.
- Turkle*, Sherry: *Simulation and its discontents*, Cambridge 2009.
- Voigt*, Friedemann: *Religion in bioethischen Diskursen. Interdisziplinäre, internationale und interreligiöse Perspektiven*, Edited by Series Editor, Berlin 2010.
- Waldenfels*, Bernhard: *Bruchlinien der Erfahrung. Phänomenologie, Psychoanalyse, Phänomenotechnik*, Frankfurt a.M. 2002.
- Waldenfels*, Bernhard: *Erfahrung, die zur Sprache drängt. Studien zur Psychoanalyse und Psychotherapie aus phänomenologischer Sicht*, Berlin 2019.
- Waldenfels*, Bernhard: *Phänomenologie der Aufmerksamkeit*, Berlin 2004.
- Waldenfels*, Bernhard: *Phänomenologie der Aufmerksamkeit*, Frankfurt a.M. 2015 (orig. 2004).
- Wessels*, Michael: *Pflegeökonomie*, Berlin u. Heidelberg 2019.
- Wullenkord*, Ricarda/*Eyssel*, Friederike: *Societal and ethical issues in HRI*, in: *Current Robotics Reports* 12020, 85–96.
- Zuboff*, Shoshana: *The age of surveillance capitalism. The fight for a human future at the new frontier of power*, New York 2019.

Außertheologische Perspektiven

*Manishk Gawande et al.*¹

From SanTO to CeleSTE

A Theomorphic Device for Religious Practice

Abstract

A recent development in social robotics involves a new wave of robots with religious connections. One of these robots is SanTO, created in 2017 as the first robot that has Christian Catholicism embedded into its design and functionalities. Insights and critical issues emerged through time, at the same time prompting the potential to use robots of this kind to assist elderly population. This paper introduces CeleSTE, a similar device designed to support healthy living among older adults, within the framework of an international project involving the EU and Japan. Shaped like an angel praying, CeleSTE aims to engage users of Christian Catholic faith. The paper covers CeleSTE's conceptual development, addressing challenges related to religious perception, fallibility and user interaction inherited from SanTO. CeleSTE uses a multimodal system for communication, enabling discussions on sensitive topics relevant to older adults. This paper also reports the first qualitative studies involving end users, who provide the feedback containing the insights that will guide the next phases of CeleSTE's development.

1 Manishk Gawande*, Seán Mitchell, Franco Pariasca Trevejo, Rodrigo Lopez Condori, Masao Katagiri, Rio Obe, Swe Hou Yeo, Miku Morita, Sarah Cosentino, Yegang Du, Felix Carros, Rainer Wieching, Gabriele Trovato*; *corresponding authors.

1. Introduction

The last few years have witnessed a sudden rise in interest in robots for religious practice. The concept of “theomorphic robots” originated in 2016² and was followed by efforts by a niche of researchers all over the world, aimed at providing a theoretical basis as well as practical development in this respect. ‘Theomorphic’ derives from the Greek word ‘theos’, meaning ‘God’, and ‘morphe’, meaning ‘form’ or ‘shape’. In essence, a theomorphic robot is designed to mimic or embody a divine being. Since then, the field has gained interest and is expanding to other religions³, and might have profound implications on future research.⁴ Some of the robot prototypes developed are called theomorphic as they embed religious symbols with the intent of conveying divinity, while other robots like Pepper and Xian’er, are ‘generic’ robots which happen to be used within religious contexts. Relatively few studies have been conducted in both cases,⁵ and one notable study examines BlessU-2, an anthropomorphic robot used in a Christian Protestant church. This study, led by Löffler et al.,⁶ gathered feedback from 1923 forms. Notable comments include negatives, such as “I guess God will throw his hands up in horror when seeing this blessing machine”, and positives, like “Although the robot did not talk to me personally, I still felt spoken to and experienced a blessing tailored to my needs”. These comments offer valuable insights into such robots’ potentially significant design characteristics. The allegedly superhuman nature of theomorphic robots, reinforced by specific design criteria⁷, offers several advantages in human–robot interaction regarding acceptance as well as issues related to transparency and explicability.⁸

Ahmed et al.⁹ categorised robots in religion into four different roles: robots as teaching agents; robots as counselling agents; robots as religious assistants; and robots as religious companions. We will

2 See *Trovato/Cuellar/Nishimura*: Introducing “theomorphic robots”, 1245–1250.

3 See *Şahin/Gültekin*: The Interaction of Religion and Robotics.

4 See *Trovato/Weng/Sgorbissa* et al.: Editorial Introduction.

5 See *Trovato/De Saint Chamas/Nishimura* et al.: Religion and robots.

6 See *Löffler/Hurtienne/Nord*: Blessing Robot.

7 See *Trovato/Lucho/Huerta-Mercado* et al.: Design strategies.

8 See *Trovato/Weng/Du*: “Never complain, never explain”.

9 See *Ahmed/La*: Evaluating.

synthesise them in a slightly different way in this paper. “Ritual” corresponds to “religious assistant”; while ‘teaching’ and ‘social care’ comprise the other roles when the application is focused on content or conversely is a tool to provide social assistance. Figure 1 shows this division into the three categories.

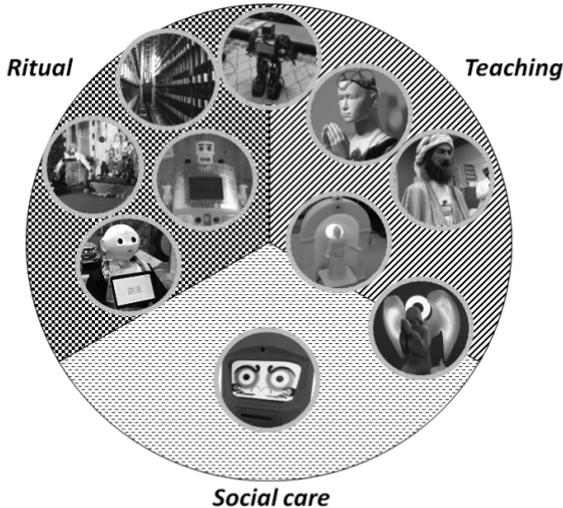


Figure 1. Application of robots within religion

According to this categorisation, SanTO and CeleSTE are not only a teaching tool, but can also be applied to a social care scenario.

The need for robots for social care in fact originates from the problem of society ageing, which is common to many developed countries. The low fertility rate combined with longer life expectancy have caused the lack of a workforce to support adults in old age. Many of them may experience loneliness in everyday life. The problem becomes exacerbated with age, owing to the elderly people’s inability to use technological devices¹⁰ and increasing levels of dementia. Care homes provide assistance, but particularly in larger ones, carers cannot offer companionship at all times, and organised activities are limited. In this context, there is a growing need for socially assistive devices, and

¹⁰ See Hunsaker/Hargittai: A review of Internet use.

the potential market may reach millions of users. Especially during the pandemic, the health and well-being of older adults appear to have been significantly affected. Many feasible solutions have been proposed to alleviate this social problem, while facilitating a better understanding of the human user. Socially assistive robots are one potential tool that can be utilised to alleviate the loneliness and isolation experienced by older adults as a result of the COVID-19 pandemic.¹¹ One of the most successful robots used in care homes is the seal robot Paro. Authors Shibata and Wada¹² contend that robot therapy can alleviate stress in a manner similar to animal therapy.

Our research is part of the e-ViTA project, funded under Horizon 2020 EU–Japan, which aims to develop a ‘virtual coach’ to support healthy living among adults aged 65 to 75.¹³ The e-ViTA framework includes a front-end device like a robot or tablet, a network of sensors, a dialogue system and middleware. In fact, not every socially assistive robot is suitable for this purpose. Achieving user acceptance, especially among older generations, can be challenging and can act as a potential bottleneck for this project.

The initial challenge is related to the robot’s appearance and concept, requiring careful design to ensure user acceptance, including the avoidance of uncanny aspects.¹⁴ Previous studies have indicated that robot acceptance is influenced by the user’s cultural background, in which religion also plays an important role.¹⁵ DarumaTO¹⁶ is one of the robots developed within this application.

Under these premises, the theomorphic device called CelesTE was conceived.

The development of CelesTE was grounded in the principles of participatory design, a methodology that places a strong emphasis on actively involving end users and stakeholders throughout the design process. This approach seeks to thoroughly understand and address the specific needs and preferences of these important parties and then incorporate their feedback into iterative prototypes. The development

11 See *Getson/Nejat*: Socially assistive robots.

12 See *Shibata/Wada*: Robot therapy.

13 See *Jokinen/Homma/Matsumoto et al.*: Integration.

14 See *Carros et al.*: Not that uncanny.

15 See *Trovato/Mavdridis/Huerta-Mercado et al.*: Cross-cultural timeline.

16 See *Shen et al.*: Participatory Design; *Du et al.*: Composite Emotion Recognition.

process unfolds through a series of well-defined steps, each building upon the insights gained from the previous one. Drawing insights¹⁷ from SanTO and the broader e-ViTA initiative, CeleSTE was used in interaction with older adults within the project, with the intent of being a primary asset for the future of socially assistive robots in the domain of elderly care. In this paper, we describe its concept and development, which derives from SanTO, in Section 2, its technical overview in Section 3, and user impressions in Section 4.

2. Conceptual development

SanTO, short for Sanctified Theomorphic Operator, is a robot designed to resemble a statue of a Christian Catholic saint in a niche.¹⁸ Its design blends robotics with religious symbols and sacred art, drawing inspiration from neoclassical architecture and the golden ratio. This fusion of innovation in robotics with conservative visual elements from the past is a distinguishing feature of SanTO.

SanTO serves various purposes, such as providing companionship during prayer, teaching catechism, narrating the lives of saints, delivering the Pope's homilies, reciting verses from the Bible and offering prayers. It is primarily intended for elderly practitioners. The perception of its sacredness among Catholic believers was assessed through experiments conducted in a church in Lima, Peru.¹⁹ In 2021, a Polish version, SanTO-PL, was introduced.²⁰ SanTO-PL, manufactured in Poland and on display at the Copernicus Science Centre in Warsaw, is an upscaled version, approximately one metre tall, with an improved design featuring a control panel with buttons. Its content is available in three languages (English, Polish, Russian), with a particular focus on the words of Pope John Paul II. SanTO-PL has been active at the Science Centre since November 2021 and was previously featured by the BBC.²¹

17 See *Trovato/Weng*: Retrospective Insights.

18 See *Trovato/Lucho/Ramón* et al.: The creation of SanTO.

19 See *Trovato/Pariasca/Ramirez* et al.: Communicating with SanTO.

20 See *Trovato/Pariasca/Purizaga* et al.: SanTO in exhibition.

21 See *BBC News*: God and robots.

The reception of SanTO has been polarised.²² While some researchers have advocated the establishment of a cultural robotics sub-field dedicated to religion, some of the scientific community have reportedly criticised the project as ‘antithetical to science’. Globally, the public has displayed significant interest in the potential of SanTO. Although not universally accepted, SanTO-PL has been seen as a tool that could bring believers closer to God, with the validity of the messages, typically quotes or prayers, remaining unquestioned.²³

Since its first release in 2017, one common misconception was whether SanTO was considered an idol. To avoid any suggestion of blasphemy and to make clear that the robot is not supposed to be the target of prayers but is designed to pray together with users, the subtitle ‘prayer companion’ was added to the introductory words spoken by SanTO. The original intention was never to replace human priests, but the idea of an AI or robot taking on religious duties attracted media attention. The Holy See, through its spokesman, emphasised that SanTO is a useful technological tool as long as it provides content without interpreting sacred texts, as machines cannot offer significant advice to believers.²⁴

A further problem arose on the theoretical level of the theomorphic concept. As SanTO is a robot, it is just as fallible as any other robot. However, as it represents a sacred object, it is supposed to be infallible at the same time. This theoretical contradiction can have a negative impact on the interaction experience on a practical level.²⁵

Moreover, during early tests, SanTO provided biased answers, some of which could even be considered sexist by today’s standards. These responses were drawn from the Bible, which can contain contradictory messages. The primary criticism revolved around the AI’s method of response, particularly regarding the influence of randomness in its algorithms. To illustrate, sensitive questions such as “Should I marry?” involve personal context. In the absence of complete understanding, machines should not provide advice, neither should they be asked.

22 See *ibid.*

23 See *ibid.*

24 See *ibid.*

25 See *Trovato/De Saint Chamas/Nishimura et al.*: Religion and robots.

The observation of visitors' differing interactions with SanTO-PL was also noteworthy, in particular in a church setting compared to the science centre where it is currently displayed. In the church, users posed faith-related questions, while at the science centre, a relevant number of interactions involved insults directed at the robot.²⁶ This variation may be attributed to the environment or the diverse audience in these locations. The concept of a 'sacred space' needs to be further investigated to determine how a robot's setting can influence user perceptions and interactions.²⁷

These challenges and such feedback significantly influenced the development of the new design for CeleSTE (*Celestial Theomorphic devicE*). Addressing the misconception of the robot being an idol, the design was altered by making the figure bent on one knee in an act of prayer. Moreover, its appearance moved away from the "sanctity" of a saint and adopted a more intercultural and slightly interreligious angelic look. This reduced characterisation supposedly associates the robots less with a sacred sphere, which may compensate for the aforementioned problem of fallibility.



Figure 2. A statue of an angel in a church in Helsinki (left) and CeleSTE (right)

26 See *BBC News*: God and robots.

27 See *Moor*: The Nature.

The issue of bias was addressed with the development of SanTO (and therefore the improvements were imported into CelesTE) through adjustments in how the dialogue system manages user intent. For instance, keywords such as “Talk to me about...” are associated with informative responses, while inquiries resembling requests for advice, like “Should I...?” receive a clear response from the robot that it cannot provide advice but can share relevant knowledge on the topic. Different datasets (about less sacred topics, more on the well-being side) were integrated as the device was connected with the e-ViTA system. It is important to note that all this happened before the recent (2023) boom of generative language models.

3. Overview of CelesTE

Unlike SanTO, CelesTE is technically not a robot. In the form of an angel atop an ionic column, it lacks any motor function. It is perhaps more accurate to classify it as a ‘device’ rather than a robot, depending on the definition we give. However, it shares the capability of communicating through lights with SanTO, as the angel’s wings can be illuminated and change colour. The final shape is a result of matching form with functionality and hiding the figure’s robotic appearance.

Being specifically built for e-VITA, the intended main function of CelesTE is to be a ‘guardian angel’, especially with older adults in mind. It can maintain a short conversation, in which the user may ask and receive an answer about a sensitive topic (such as happiness, death, faith, etc.). Furthermore, it can print out a selection of short texts.

In Figure 3, the relocation of the peripherals that compose CelesTE is shown in contrast to SanTO. The main differences are the removal of the motors and the camera (for privacy reasons), and the addition of a printer and of a USB drive hidden in the candle. The capacitive touch-sensing ability of SanTO was replaced by a distance sensor, which not only acts as a button for activation, but also regulates the luminance of the wings depending on the user’s position.

The version of CelesTE we are presenting here is the second (R2). The first, R1, featured slightly different proportions. The shape and the features vary accordingly with the iterative nature of participatory design. This methodology allows for continuous refinement and

enhancement so that the robot aligns optimally with the expectations and requirements of the end users (involved in the trials reported in the next section) and stakeholders. Further revisions (R3 and more) will follow in the coming years thanks to the feedback collected.

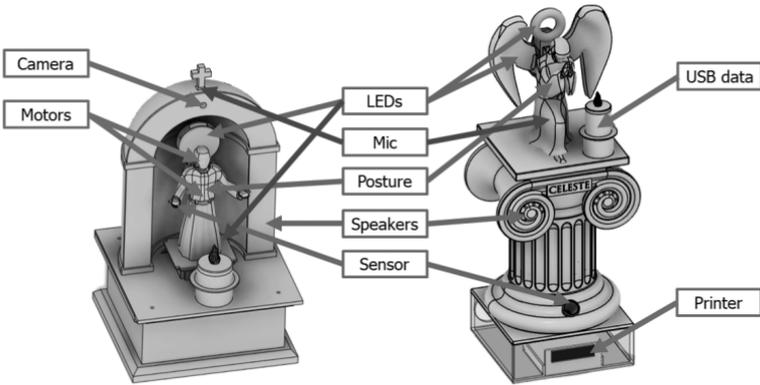


Figure 3. Components of SanTO and CelesTE: a diagram highlighting both the parts they have in common and their differences.

Communication takes place in alternating multimodal interaction through the use of lights and voice. When the device is activated with the finger, it is possible to record vocal input. CelesTE will listen for nine seconds. The conversation between CelesTE and the user is regulated by visual feedback. When the device is recording, its halo is illuminated with a rotating green light and it produces a sound; for the rest of the time, the halo is yellow, while the wings change colour according to the content being spoken. The speech is then processed and converted into text. If the device is connected to the e-ViTA cloud, the text is sent to it and analysed for matching keywords, and an answer is selected. If not, the answer will be determined locally. Finally, CelesTE speaks in its artificial voice, then returns to standby and can be activated anytime. Vocal input works in English, German, Italian and French.

When the answer is processed locally, a selection of topics is available. The content for an answer is searched for in a shortlist that contains more than 500 IDs of certified parts of text. These texts were certified by religious authorities for their relevance to

certain keywords or topics and are broken down into sentences and assigned an emotional valence (positive or negative) which can affect the look of the colour of CelesTE's wings. The texts are extracted from a variety of sources, here classified as: the Bible, prayers, quotes from saints or popes, and stories of the lives of saints.

In addition to this content selection, the story of the saint of the day is always available and can be listened to through activation by a specific keyword.

As an alternative to the shortlist, a deeper search into the sources is also possible. Various forms of extraction with machine learning and text mining are being attempted; however, given the inconsistency of the quality of the results, this mode is only enabled as a "debug mode" for developers to troubleshoot issues and is never activated in front of the users.

A specific keyword is also configured to trigger the use of the printer. There are two cases: When CelesTE is configured for a private user, one can ask: "Do you have any message for me?". CelesTE will then print a short motivating sentence extracted from the Bible. When CelesTE is configured for more than one user, the trigger keyword is simply "My name is ...". That will be more appropriate for visitors who approach it for the first time. The content of the tickets printed is the same. For elderly users, these printed messages offer cognitive reinforcement and a reminder of their spiritual practice, addressing potential memory challenges while providing comfort through physical artifacts of faith that persist beyond the immediate interaction.

A final aspect worth mentioning is the handling of failures. When the speech recogniser cannot detect any sound, CelesTE will switch to standby after two attempts. When some words are detected but not understood, one more attempt to repeat the voice input is given. If the second time also fails, it will switch to "praying mode". CelesTE is capable of generating a prayer based on chunks of existing prayers, in the same way as SanTO-PL. This application of Procedural Content Generation (PCG) has also been validated in collaboration with members of the clergy, who verified the consistency of the output text of the new prayer.

4. Qualitative studies

CeleSTE was developed for the Horizon 2020 e-ViTA project and underwent initial tests in three different countries (Germany, France and Italy) with a total of ten participants, as shown in detail in Table 1. The range of the participants' ages was from sixty to eighty-seven years old, with a median age of sixty-five.

Table 1 Participants' information and experimental setting.

ID	Age	Sex	Religious attitude	Experimental setting
IT01	60	M	Christian Catholic, non-practising	Living Lab
IT02	62	F	Christian Catholic, non-practising	Living Lab
IT03	65	M	Atheist	Living Lab
IT04	64	F	Christian Catholic, practising	Living Lab
FR01	87	F	Atheist	Home
FR02	82	F	Christian Catholic, practising	Living Lab
FR03	81	M	Christian Catholic, non-practising	Living Lab
GE01	60	F	Christian Catholic, practising	Church
GE02	65	F	Christian Catholic, practising	Church
GE03	70	F	Christian Catholic, practising	Church
GE04	75	F	Christian Catholic, practising	Church

The interactive session typically lasted five to ten minutes. After an introduction to the device, we let the participant repeat the trials they wanted. An interview followed, aiming to go in-depth and focus on more concrete feedback regarding specific aspects of the device's design and usability. Within this interview, some quantitative data was also collected, but to include it here would go beyond the scope of this paper.

One of our biggest concerns was whether it would cause any discomfort. CelesTE was judged to be not uncanny overall: on a scale of one to five points, only two participants rated it two out of five, while the rest judged it to be completely not uncanny (one out of five). No participant felt unsafe.

Regarding the interviews, here is a list of the most insightful comments.

- “The quality of the bibliography and its pertinence to the context should be improved” (IT03)
- “The finger as an activation method could be improved” (IT02)
- “Should be possible to ask for additional clarifications on a certain topic” (FR03)
- “Too Catholic” (FR03)
- “It would be good for my mother, who is 83 years old” (GE01)
- “Too slow, makes too many pauses” (GE04)
- “Would be good for someone who lives alone and has nobody to speak to” (GE04)
- “Could reproduce some music” (GE04)
- “Please add a way to regulate the volume” (FR02, FR03)

As robot developers, we believe that not all comments must necessarily be addressed. Each user may want something different according to their personal preferences, and some may not want or need such a robot. GE04 offers three comments which demonstrate the variety of opinions and the different weighting that the roboticist should attach to them. Specifically, the comment about the ideal user, someone who lives alone, confirms the target users we had in mind for this robot concept. At the same time, it clarifies that GE04 is not such a user profile. If we see things in this light, the comment about music may be seen as just a personal preference expressed by someone who would not use the robot anyway. We believe that, at best, music would be an interesting functionality, but should not be included in the core value provided by the object. Finally, the comment about CelesTE being too slow is purely technical, and effort should be made to resolve the issue.

Additionally, very valuable insights actually came from observing the interaction during the experiments. From previous studies, we know that the timing of vocal communication (i.e. when the user can or cannot speak, corresponding to the device’s recording time win-

dow) is critical for successful communication. When not performed correctly, the communication fails. In the present series of experiments, we confirmed that it took a few trials for the participants to get used to the correct usage of the device. The redundancy of signals (coloured light and sound) seems to be useful as well, as the user may concentrate on hearing and thus miss out on some of the visual signals.

Another critical detail was the need to recharge the printer battery or for the user to replace its paper. With the current (R2) design of the base, that is not easy. A redesign of the lower box is therefore needed in order to extract the printer easily and safely.

5. Conclusion

In this paper, we have introduced some critical points regarding the SanTO Catholic robot and the concept and development of CeleSTE, a device that embodies an angel praying, on the top of an ionic pillar. CeleSTE was designed to be employed in the Horizon 2020 project called e-ViTA, in which older adults are offered cognitive support for healthy living. Through a participatory design that involved end users, the concept and shape of CeleSTE took place.

For the early findings shown in this paper, we witnessed a positive response from the users. Some improvements are needed on the interaction method, and some contents need to be refined in terms of their relevance and accessibility to our target audience. The most important insight refers to the target audience, which seemed to be more appropriate for an older generation (over 80 years of age) than the currently targeted range of 65- to 75-year-olds. For future work, we may need to distinguish between different user profiles for CeleSTE. On one hand, the oldest generation is the ideal target, and CeleSTE should be specifically designed to overcome any kind of impairments they may have. On the other hand, we plan to revise the concept and functionalities in order to better meet the needs of a younger audience (65–75 years old).

Acknowledgements

This work was supported by the Ministry of Internal Affairs and Communications (MIC) of Japan (Grant no. JPJ000595). All images were created by the authors.

Bibliography

- Ahmed, H./La, H.M.:* Evaluating the Co-dependence and Co-existence Between Religion and Robots. Past, Present and Insights on the Future, International Journal of Social Robotics 13, 2021, 219–235. DOI: <https://doi.org/10.1007/s12369-020-00636-x>.
- BBC News:* God and robots. Will AI Transform Religion? Online at: <https://www.youtube.com/watch?v=JE85PTDXARM>.
- Carros, F. et al.:* Not that Uncanny After All? An Ethnographic Study on Android Robots' Perception of Older Adults in Germany and Japan, Proceedings of the International Conference on Social Robotics, Florence 2022.
- Du, Y. et al.:* Composite Emotion Recognition and Feedback of Social Assistive Robot for Elderly People, Proceedings of HCI International 2023 — 9th International Conference of Human Aspects of it for the Aged Population (ITAP), Copenhagen 2023.
- Getson, C./Nejat, G.:* Socially Assistive Robots Helping Older Adults Through the Pandemic and Life After COVID-19, Robotics 10 (3), 2021, 106.
- Hunsaker, A./Hargittai, E.:* A Review of Internet Use Among Older Adults, in: New Media & Society 20 (10), 2018. DOI: <https://doi.org/10.1177/1461444818787348>.
- Jokinen, K./Homma, K./Matsumoto, Y. et al.:* Integration and interaction of Trustworthy AI in a Virtual Coach. An Overview of EU–Japan Collaboration on Eldercare. Advances in Artificial Intelligence. Advances in Intelligent Systems and Computing 1423, 2022, 190–200. DOI: https://doi.org/10.1007/978-3-030-96451-1_17.
- Löffler, D./Hurtienne, J./Nord, I.:* Blessing Robot BlessU2. A Discursive Design Study to Understand the Implications of Social Robots in Religious Contexts, International Journal of Social Robotics 13, 2021, 569–586. DOI: <https://doi.org/10.1007/s12369-019-00558-3>.
- Moor, J.H.:* The Nature, Importance, and Difficulty of Machine Ethics. Machine Ethics, Cambridge 2011.
- Shen, Z. et al.:* Participatory Design and Early Deployment of DarumaTO-3 Social Robot, Proceedings of the International Conference on Social Robotics, Florence 2022.
- Shibata, T./Wada, K.:* Robot Therapy. A New Approach for Mental Healthcare of the Elderly. A Mini-review, Gerontology 57 (4), 2011, 378–386.
- Trovato, G./Cuellar, F./Nishimura, M.:* Introducing “Theomorphic Robots”. Proceedings of the 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), 2016, 1245–1250.
- Trovato, G./De Saint Chamas, L./Nishimura, M. et al.:* Religion and Robots. Towards the Synthesis of Two Extremes, International Journal of Social Robotics, 2019, 1–18.

- Trovato, G./Lucho, C./Huerta-Mercado, A. et al.*: Design Strategies for Representing the Divine in Robots, Proceedings of the 13th ACM/IEEE International Conference on Human–Robot Interaction (HRI), Chicago 2018.
- Trovato, G./Lucho, C./Ramón, A. et al.*: The Creation of SanTO. A Robot with “Divine” Features, Proceedings of the 2018 15th International Conference on Ubiquitous Robots (UR), 2018, 437–442. DOI: <https://doi.org/10.1109/URAL.2018.8442207>.
- Trovato, G./Mavridis, N./Huerta-Mercado, A. et al.*: Cross-cultural Timeline of the History of Thought of the Artificial, Proceedings of the International Conference on Social Robotics, Singapore 2021.
- Trovato, G./Pariasca, F./Purizaga, A. et al.*: SanTO in Exhibition. A Sacred Robot in the Profane, Proceedings of The 32nd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2023), Pusan 2023.
- Trovato, G./Pariasca, F./Ramirez, R. et al.*: Communicating with SanTO. The First Catholic Robot, Proceedings of the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2019, 1–6. DOI: <https://doi.org/10.1109/ROMAN46459.2019.8956250>.
- Trovato, G./Weng, Y.-H.*: Retrospective Insights on the Impacts of the Catholic Robot SanTO, Proceedings of Robophilosophy, Helsinki 2022.
- Trovato, G./Weng, Y.-H./Du, Y.*: “Never complain, never explain”. Why Robots May Not Have to be Explicable After All, Proceedings of the ICRA XRo (Workshop on Explainable Robotics), London 2023.
- Trovato, G./Weng, Y.-H./Sgorbissa, A. et al.*: Editorial Introduction to Special Issue on Religion in Robotics, International Journal of Social Robotics, 2021.
- Şahin, Meryem/Gültekin, Mücahit*: The Interaction of Religion and Robotics and al-Sāmiri’s Calf (the Golden Calf) as an Early Theomorphic Robot, *Ilahiyat Studies* 14 (1), 2023, 133–168. DOI: <https://doi.org/10.12730/is.1302053>.

Hildelith Leyser and Kate Cooper***

Motion, Mirror Neurons and the Human–Robot Relationship

Abstract

The rapid advancement of robotic technologies and their integration into human environments necessitates a deeper understanding of the human–robot relationship. This article explores the dynamic interactions between humans and robots, particularly focusing on the trust mechanisms involved when humans interact with non-human agents. The attribution of human-like qualities to robots involves a complex interplay of cognitive processes, including anthropomorphism and motion perception. Drawing upon theories in cognitive science and recent empirical research, we examine how motion and perceived intentions play a critical role in forming human trust towards robots.

By exploring these relational dynamics, we aim to shed light on the potential of robots as active participants in human social settings, offering fresh perspectives on the roots of intersubjectivity and thus providing a framework for discussion of the ethical implications of trust in robots.

1. Introduction

The curious human tendency to ascribe something like human intention to non-human actors has long been recognised, and it is also well known that we do this not only with animals but also with inanimate objects. Famously, in an experiment conducted at

* Department of Neurology and Neurosurgery, McGill University

** Department of History, Royal Holloway, University of London

Smith College in the 1940s, Fritz Heider and Marianne Simmel recorded the pronounced tendency of study participants to construe not only intentionality but even an emotionally complex narrative when offered an intentionally simplified stimulus, an animated film showing two-dimensional geometrical shapes—two triangles and a circle—moving in white space.¹

In the twenty-first century, the tendency to impute intention to non-human actors has ever wider ethical implications. As we move into the age of what Sherry Turkle has called the ‘relational artifact’—the “computational object explicitly designed to engage a user in a relationship”²—non-human artefacts have begun to take on tasks previously fulfilled by human colleagues and carers, becoming ever more thoroughly integrated into human society. Human beings have begun to rely on these tools for cooperation as well as for help, safety and pastoral care. Establishing a smooth path for human–robot interaction (HRI) is now big business, with out-sized potential for both a positive and a negative impact. So it is important to understand what characteristics elicit a trusting response and—which is at least as important—what, if anything, can be done to ensure that the ability to elicit trust is accompanied by genuine trustworthiness.

Within this frame, the problem of trust is not straightforward. Turkle has referred to a ‘crisis of authenticity’, referencing the computer scientist Joseph Weizenbaum’s disappointment with how student users reacted to Eliza, a natural language processing programme he created. Eliza used string matching and substitution to reply to statements with questions or re-statements of what the user had said, creating an effect not unlike that of a therapist trained in the non-directive method advocated by Carl Rogers.³

The source of Weizenbaum’s disappointment was not his programme’s failure, but rather its success. The programme had a positive emotional effect on those who used it, and this struck him as wrong. “Weizenbaum came to see students’ relationships with Eliza as immoral”, Turkle explains,

1 See *Heider & Simmel: An Experimental Study of Apparent Behaviour*, 243.

2 See *Turkle: Authenticity in the Age of Digital Companions*, 502.

3 See *Turkle: Authenticity in the Age of Digital Companions*, 502; on non-directive therapy, see *Rogers: Significant Aspects of Client-Centred Therapy*, 415–422.

because he considered human understanding essential to the confidences a patient shares with a psychotherapist. Eliza could not understand the stories it was being told; it did not care about the human beings who confided in it. [...] If the software elicited trust, it was only by tricking those who used it.⁴

What is implicit here is the view that human users believed that the programme's emotionally involving replies were generated by a self that was capable of understanding or caring about its human interlocutor. If this were the case, the users could be devastated on learning that the 'connection' they had built was entirely one-sided.

The danger that human users may misunderstand the nature of relational artefacts and be harmed as a result is of course a serious one, especially in light of the fact that such tools are normally produced by profit-driven manufacturers and often made available by organisations who are under pressure to cut costs. It is important to ensure that users are not misled. However, as we will see below, we are in the early stages of understanding the issues in play. Similarly, the cognitive sciences have yet to fully understand what causes human beings to impute intentions to other beings or to objects, or to evaluate these intentions in a way that allows us to derive a feeling of trust.

With this in mind, this study focuses on one variant of the relational artefact, the robot, a type of tool which is known for its ability to elicit anthropomorphic projection, even in a pared-down form which is neither humanoid nor endowed with the capacity to use language. Defined by a recent study "as a machine that is able to physically interact with its environment and perform some sequence of behaviours, either autonomously or by remote control",⁵ the robot offers a valuable point of focus for considering the question of trust.

2. Motion and the human–robot relationship

The fact that robots are able to elicit trust from humans is well established. Perhaps surprisingly, robots that are by no means human-like in their characteristics are among those most widely documented

4 *Turkle*: Authenticity in the Age of Digital Companions, 502.

5 *Kraus et al.*: Interactive Robots in Experimental Biology, 369–375.

as eliciting a trust response. In the military settings studied by Julie Carpenter, for example, bomb-diffusing robots resembling miniature industrial cranes have been celebrated as highly valued members of human teams, receiving military honours such as the purple heart and being mourned as fallen members of their cohort in cases where they are destroyed in action.⁶

Robots as colleagues and workers raise complex ethical questions which we are only beginning to address. Sven Nyholm and Jilles Smids have argued that although robots do not think in the way humans think and are unable to engage in trust-building interactions in the way human beings do,⁷ they are nonetheless capable of out-performing human colleagues when measured against certain characteristics valued in collegial behaviour, such as “being reliable and trustworthy”.⁸ For example, a robot may have less difficulty than a human exhibiting fairness or impartiality, virtues which are hugely valued in collaborative contexts. Nyholm and Smids note that the ability of robots to perform set tasks predictably and reliably is valued by colleagues, and they suggest that the dynamics of collegial collaboration need to be distinguished from those governing emotionally driven relationships such as love and friendship. In an analysis of the ethics of robots as carers and companions, Mark Coeckelbergh has argued that while the human–robot relationship is structurally one-sided, the empathetic impulses felt by a robot’s human partner are not without value in their own right;⁹ this is a point to which we will return below.

This study offers an overview of recent work in the cognitive sciences that sheds light on the mechanisms by which human beings evaluate the nature, and the relational potential, of the beings and objects we encounter. We will focus specifically on human engagement with robotic movement, aiming to keep the focus on movement itself rather than a robot’s additional characteristics, such as shape or its ability to use language. As we attempt to learn what it

6 See *Carpenter*: *The Quiet Professional*, 2013.

7 A point made by *Groom & Nass*: *Can Robots Be Teammates? Benchmarks in Human–Robot Teams*. *Interaction Studies*, 483–500, cited in *Nyholm & Smids*: *Can a Robot Be a Good Colleague?*, 2180.

8 *Nyholm & Smids*: *Can a Robot Be a Good Colleague?*, 2185.

9 See *Coeckelbergh*: *Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations*, 1–17.

is about robots that allows human beings to trust them, we will also find ourselves asking what it is about human beings that makes us able to offer trust to robots.

Focusing on the robot’s mechanical ability to move through space has analytical value because the recognition of movement has a distinctive power to provoke the human brain to impute intention. In the case of the moving geometric shapes described above, it was motion that Heider and Simmel identified as salient. Once it was set into motion, they concluded, even a two-dimensional triangle was given the perceptual status of a ‘person’; it followed that “acts of persons have to be viewed in terms of motives”.¹⁰ Put simply, Heider and Simmel argued, the observers understood the simple fact of movement as something that must be accounted for.

Further, Wilma A. Bainbridge’s experiments with Nico, a friendly humanoid robot encountered by participants in an office setting, demonstrated that the difference between two-dimensional movement and movement in three-dimensional space was significant.¹¹ In Bainbridge’s studies the response of participants differed depending on whether a collaborative task performed in tandem with Nico took place in physical reality or via a video link. Study participants collaborated with Nico on manual tasks like moving books, which allowed researchers to observe how participants greeted the robot, worked alongside it and responded to instructions it delivered. Some participants worked directly with a physically present Nico, while others interacted with the robot through a live video feed. The results showed that participants of in-person collaborations were more likely to comply with unusual or nonsensical instructions delivered by Nico and rated the interactions as more positive and natural than those who collaborated via video.

In what follows, we will consider the role movement and physical presence plays in shaping how human subjects perceive robots, beginning with a brief review of how contemporary neuroscience sees the brain’s capacity to analyse sensory data, and how it constructs a notion of agency. This understanding is crucial because it informs how we recognise and interact with entities that may possess—or appear to possess—agency, including robots.

10 *Heider & Simmel: An Experimental Study of Apparent Behavior*, 258.

11 See *Bainbridge et al.: The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents*, 41–52.

3. Cognition and sensory experience

How we understand the processing of sensory experience is changing rapidly, as our understanding of the brain changes. In the mid to late twentieth century, it was believed that the various functions of the brain had emerged at successive evolutionary phases, with emotion assigned to the earliest developmental layers and reason assigned to the more recent or advanced.¹² From the 1980s, scientists began to move away from this model, seeing the brain as a welter of highly adaptive networks.¹³ Mounting experimental evidence has demonstrated the brain's neuroplasticity (its ability to repurpose and reorganise synaptic connections) and its capacity to redirect connections to new purposes in response to injury or environmental changes.¹⁴

In 2023 a team led by Evan Gordon put forward a new model, the Somato-Cognitive Action Network (SCAN) model,¹⁵ which hypothesises a network that alternates between effector regions, which are responsible for specific motor outputs like hand or mouth movements, and inter-effector regions, which connect various parts of the motor cortex to higher cognitive areas such as the dorsal anterior cingulate cortex [dACC] and supplementary motor area [SMA].¹⁶ This approach sees the brain's model of the body as dynamic, relying on integrated neural circuits to interpret sensory information and organise sensations into meaningful perceptions.

Central to most current approaches to the brain is the idea that it has evolved to anticipate changes in the environment, thus facilitat-

12 The most influential version of this view of human neural organisation, known as 'the triune brain' for its three components (the neocortex, the limbic system and the reptile brain), was articulated by Paul MacLean in work beginning in the 1940s, with his magnum opus, *The Triune Brain in Evolution*, published in 1990. In the English-speaking world, the triune brain achieved celebrity status thanks to Carl Sagan's 1977 study of human intelligence, *The Dragons of Eden*.

13 For a useful overview, see *Steffen et al.*: *The Brain Is Adaptive Not Triune*.

14 See *Marzola et al.*: *Exploring the Role of Neuroplasticity in Development, Aging, and Neurodegeneration*, Brain Sciences, 1610.

15 See *Gordon et al.*: *A Somato-Cognitive Action Network Alternates with Effector Regions in Motor Cortex*, 351–359.

16 See *Ibid.*

ing protection from danger or exploitation of opportunities.¹⁷ Prediction error—in other words, surprise—is one of our most powerful sources of motivation, thanks to the dose of the neurotransmitter dopamine that is released when events do not align with expectation.¹⁸

One of the most significant researchers on emotional states, Lisa Feldman Barrett, argues that the experience of emotion is “an act of categorization, guided by embodied knowledge”.¹⁹ Focusing on finding the connection between sensory inputs and the generation of emotions in the brain, Barrett suggests that contextual cues can significantly influence how the brain interprets sensory stimuli. Integrating principles of embodiment and Bayesian inference, Barrett argues that the brain uses predictive models based on past experiences to interpret sensory inputs, categorising them into emotional states according to embodied knowledge and contextual cues. In this view, selfhood and emotional categorisation rely on the continuous modelling of the internal and external world which the brain uses to distinguish expected sensations from unexpected ones, and to develop a sense of bodily presence.²⁰

Important here is the insight that our experience of the body itself is constructed through mental maps, which rely on a continuous process of interpretation and modelling of our internal and external world. This means that we ‘know’ our own bodies through the same process of modelling that allows us to know the world beyond the

17 See *Sol et al.*: Brain Size Predicts the Success of Mammal Species Introduced into Novel Environments, 63–71.

18 See *Schultz*: Reward Prediction Error, 369–371.

19 *Barrett*: Solving the Emotion Paradox: Categorization and the Experience of Emotion, 20. For a useful overview of how current approaches to emotion relate to the problem of brain models, see *Fernandez et al.*: Affective Experience in the Predictive Mind, 10847–10882.

20 See *Bechtel*: Representations and Cognitive Explanations, 296–306. For a useful overview of the increased interest in the social dimension of embodied cognition, including simulation as embodied practice, intercorporeality and intersubjectivity, see *Lindblom*: A Radical Reassessment of the Body in Social Cognition. Research by *Vicario* discusses the debate between embodied and disembodied theories of cognition, highlighting the contrasting views on the interplay between cognition and sensorimotor systems. *Vicario*: Perceiving Numbers Affects the Internal Random Movements Generator, 1–6. See also *Barsalou et al.*: Grounding Conceptual Knowledge in Modality-Specific Systems.

body.²¹ The distinction we draw between self and not-self is not the result of sensing our bodies more directly than we sense the world around us.

4. Putting the focus on spatial relationships—self, other and the problem of movement through space

To grasp why motion is so important to us cognitively, it is useful to consider how we experience the body in space. Drawing on the mid-century psychologist James Gibson's concept of "affordances"²² provides a critical vocabulary with which to explore how individuals respond adaptively to cues derived from spaces and the objects within them. It also helps us understand how individuals develop a relational and transactional framework to make sense of the stable and moving stimuli they encounter.

The work of Barbara Tversky has shed light on the centrality of spatial orientation to our interaction with the world. Tversky proposes that the brain employs a spatial framework as a universal strategy for processing all types of information, not just physical or visual inputs. The brain feeds all inputs, including abstract concepts, into spatial frameworks, which is what allows them to become tangible in our minds. Tversky argues that these spatial frameworks have an analytical capacity: for example, while the space around the body is experienced as three-dimensional, we easily reduce it to two dimensions when assessing space with navigation through it in mind.²³

Complementing Tversky's insights, research on rodents has revolutionised our understanding of spatial cognition, above all in the discovery of the specialised neurons known as grid and place cells. In a groundbreaking series of publications in the 1970s, John O'Keefe and Jonathan Dostrovsky identified 'place cells', neurons in the rat hippocampus that activate when the animal passes through

21 See Noel *et al.*: Rapid Recalibration of Peri-Personal Space, 5089.

22 Wit *et al.*: Affordances and Neuroscience: Steps Towards a Successful Marriage, 622–629. Gibson: The Senses Considered as Perceptual Systems, 5.

23 See Tversky: Structures of Mental Spaces: How People Think About Space, 66–80.

specific locations, forming a mental map of these spaces.²⁴ Building on O’Keefe’s work, May-Britt and Edward Moser identified a new group of cells in the rodent brain, ‘grid cells’, which create virtual maps by firing in grid-like patterns, with each grid cell contributing a data point to the pattern as the animal moves through space.²⁵ Subsequent work has suggested that human spatial experience functions similarly.²⁶ These neurons enable us to differentiate ourselves from other entities based on location and movement, and they are crucial for understanding social orientation and relationships.

5. Mirror neurons: The cognitive provocation of movement

Perhaps the most widely known work on how the brain perceives movement is that of Vittorio Gallese and colleagues on mirror neurons. In 1992, a team at the University of Parma studying Macaque monkeys identified a group of motor neurons located in the brain’s ventral premotor area F5 that fired both when the monkeys executed actions related to reaching and grasping, and also when they observed other monkeys executing the same action.²⁷ In a later publication, Gallese and colleagues coined the term ‘mirror neurons’ to refer to these cells, suggesting that when we watch others performing an action, the same neural patterns are activated as if we were performing the action ourselves.²⁸

Debate over the role of mirror neurons has given a new focus to the contrast between researchers who understand the experience of embodiment as integral to human cognition and conceptual processing, and those who understand cognition as computational and

24 See O’Keefe & Dostrovsky: The Hippocampus as a Spatial Map, 171.

25 See Moser *et al.*: Microstructure of a Spatial Map in the Entorhinal Cortex, 801–806. Sargolini *et al.*: Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex, 758.

26 See Burgess, Barry & Doeller: Evidence for Grid Cells in a Human Memory Network, 657–661.

27 See Di Pellegrino *et al.*: Understanding Motor Events, 179.

28 See Gallese *et al.*: Action Recognition in the Premotor Cortex, 604.

probabilistic.²⁹ The discovery of mirror neurons shifted the emphasis in a very concrete way: to build an understanding of the actions of another agent, the brain simulates these same actions in an embodied way. As Gallese and colleagues put it in a 2007 study, this is “a mandatory, nonconscious, and pre-reflexive mechanism that is not the result of a deliberate and conscious cognitive effort”.³⁰

The concept of mirror neurons thus challenges the idea that our cognitive and physical experiences are neatly separated; rather, the suggestion is that both our experience and our social understanding are based on the same mechanisms for internally simulating experience.

My embodied simulation model is in fact challenging the notion that the sole account of interpersonal understanding consists in explicitly attributing to others propositional attitudes like beliefs and desires, mapped as symbolic representations. Before and below mind reading is intercorporeity as the main source of knowledge we directly gather about others.³¹

Gallese sees intersubjectivity not as mediated by the sensory cortex, where sensory stimuli are processed, but instead, by the motor cortex, where the brain generates actions.³² To illustrate the point, he brings infant development into play: “infants develop the capacity to anticipate the goal of the observed motor acts done by others only when they become able to perform the same goal-directed motor acts themselves.”³³

The debate over the extent and nature of mirror neurons remains open. A recent survey of studies of their function concludes that while mirror-neuron brain areas contribute to low-level processing

29 See Gallese & Sinigaglia: What Is So Special About Embodied Simulation?, 512; Chater et al.: Probabilistic Models of Cognition: Conceptual Foundations, 288.

30 Gallese, Eagle & Migone: Intentional Attunement, 143.

31 Gallese: Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification, 524. Gallese here draws on Merleau Ponty’s concept of *intercorporeité*, “a kind of fundamental openness of the body to other bodies such that their coupling generates norms that come to affectively govern their engagement” (Walsh: Intercorporeity, 34). While the term is normally represented by “incorporeality” in English, it is sometimes (as here) translated as “intercorporeity”.

32 See *ibid.*, 522.

33 Gallese, Eagle & Migone: Intentional Attunement, 146.

of observed actions, such as distinguishing types of grip, they may not play a significant role in high-level action interpretation, such as inferring actors' intentions.³⁴

According to Gallese, the sense of connection that comes from interacting with others is rooted in a recognition of shared biological processes that occurs at a physical level.

The discovery of mirror neurons provide[s] a new empirically based notion of intersubjectivity, viewed first and foremost as intercorporeity—the mutual resonance of intentionally meaningful sensory-motor behaviours—as the main source of knowledge we directly gather about others.³⁵

In this way, Gallese offers a striking vision of intersubjective solidarity: “Anytime we meet someone, we are implicitly aware of his or her similarity to us, because we literally embody it.”³⁶

But the existence of the android robot offers an important challenge to this assertion. Studies by Gazzola et al., Kashi & Levy-Tzedek and others have shown that the mirror neuron system responds not only to human actions but also to the actions of robots, which are by definition not living beings with whom intersubjective recognition may be shared.³⁷ On this view, the perception of intersubjectivity is a projection, not a recognition of empirical fact.

6. The hypothesis of the other's intention

We return here to an issue introduced earlier, the human tendency to attribute intention to things that move. This often takes the form of anthropomorphism, the human tendency to “explain nonhuman behavior as motivated by human feelings and mental states”.³⁸

34 See Heyes & Catmur: What Happened to Mirror Neurons?, 153–168.

35 Gallese: Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification, 523.

36 Ibid., 524.

37 See Kashi & Levy-Tzedek: Smooth Leader or Sharp Follower? Playing the Mirror Game with a Robot, 147–159. Gazzola et al.: The Anthropomorphic Brain: The Mirror Neuron System Responds to Human and Robotic Actions, 1674–1684.

38 Airenti: The Cognitive Bases of Anthropomorphism, 119.

Building on the work of John Fisher in the early 1990s,³⁹ cognitive scientists have revised a view dominant since the nineteenth century, which saw anthropomorphism as “a category mistake, an obstacle to the advancement of knowledge, and as a psychological disposition typical of those who are immature and unenlightened, i.e. young children and ‘primitive people’”.⁴⁰ More recently, anthropomorphic thinking has come to be understood as a heuristic which serves “the need to make sense of the actions of other agents to reduce uncertainty concerning their behaviour”.⁴¹ A study by Epley, Waytz and Cacciopo clarifies the value of this heuristic in the context of interpreting movement: “Attributing human characteristics and motivations to nonhuman agents increases the ability to make sense of an agent’s actions, reduces the uncertainty associated with an agent, and increases confidence in predictions of this agent in the future.”⁴² The need for speed and confidence in these assessments may be a question of evolutionary fitness, since the ability to predict and account for motion would have evolutionary value in the context of predator detection.⁴³

Surprisingly, ascribing intention via anthropomorphic thinking seems to have value whether or not the human subject ‘believes’ that the object is genuinely capable of human mental states. A study by the psychologist Gabriella Airenti argues that humans including children have a productive ability to construct imaginative personas for objects and to use them meaningfully, even while recognising that they are fictions.⁴⁴ In other words, human beings don’t need to believe that an artefact is human in order to engage meaningfully with it *as if it were human*. Indeed, Airenti suggests that in pastoral settings the effort to efface the difference between robots and humans can actually be counterproductive.

39 See Fisher: Disambiguating anthropomorphism, *Vidal et al: Introducing Anthropomorphism and Urquiza-Haas & Kotrschal: The Mind Behind Anthropomorphic Thinking*, 167 for useful discussion of the academic literature.

40 Damiano & Dumouchel: Anthropomorphism in Human-Robot Co-evolution, 2.

41 Urquiza-Haas & Kotrschal: The Mind Behind Anthropomorphic Thinking, 168.

42 Epley: On Seeing Human, 866.

43 See Barrett, Cognitive development and the understanding of animal behavior, 447–449. See also Barrett *et al.*, Accurate judgments of intention from motion cues alone.

44 See Airenti: The Cognitive Bases of Anthropomorphism, 122–123; Airenti: The Development of Anthropomorphism, 5.

“It is not simply that human-likeness is unnecessary,” she warns:

Robots that simulate them will scare people or make them think to be [sic] cheated. Instead, if we base [sic] on what we have seen already active in infants, being perceived as helpers in action will be sufficient to gain the sympathy of their human interlocutors.⁴⁵

Humans are able to feel sympathy towards a machine which they know to be a machine, Airenti argues, but in order to trust it, there needs to be an absence of signals that reflect an intent to deceive the user about whether it is a machine or not.

It is perhaps significant that Airenti suggests that anthropomorphism is not an innate tendency in children, away from which adults strive to lead them. It seems, in fact, to be a form of behaviour which children learn from their adult carers.⁴⁶ Airenti stresses the social aspect here—both the relationship between the human observer and the artefact being observed, and the relationship between humans who cooperatively engage with an artefact. She concludes that treating an artefact as if it has human intentions is a learned relational practice, which does not necessarily reflect what the human subject believes about the nature of the artefact itself.

Finally, it is worth remembering that the attribution of human qualities to new technologies has a history. In a rhetorical study of personifications of technology reaching back to the introduction of the sewing machine and electric light, R. John Brockman explored how the human analogy has been invoked at points when new technologies were being introduced. When first brought to market, Brockman suggested, new technologies are described and even marketed using the metaphor of human personality. But as they become familiar, the metaphor fades away.⁴⁷ In the case of the sewing machine and the electric lamp, the human analogy was a tool for transition—a way to make the new technology familiar during the early phase of its availability.

With this in mind, it is worth asking what role the analogy that they are like human beings—or in some cases like other animals—plays in the case of relational artefacts such as robots. Will we still impute human characteristics to them once we have become used

45 Airenti: *The Cognitive Bases of Anthropomorphism*, 124.

46 See Airenti: *The Development of Anthropomorphism*, 7.

47 See Brockman: *The Homunculus in the Computer?*

to them? The chances are we will, given our tendency to anthropomorphise anything that moves. But like the miniature crane praised for its military service, the artefacts that prove most able to elicit trust may prove to be those which are designed not to feed the anthropomorphic imagination, but to perform in a way that is so predictable, consistent and undemanding that no one could mistake them for a human.

Bibliography

- Airenti, G.*: The Cognitive Bases of Anthropomorphism. From Relatedness to Empathy, in: *International Journal of Social Robotics* 7 (1), 2015, 117–127 <https://doi.org/10.1007/s12369-014-0263-x>.
- Airenti, G.*: The Development of Anthropomorphism in Interaction. Intersubjectivity, Imagination, and Theory of Mind, in: *Frontiers in Psychology* 9, 2018, 2136 <https://doi.org/10.3389/fpsyg.2018.02136>.
- Bainbridge, Wilma A./Hart, Justin W./Kim, Elizabeth S. et al.*: The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents, in: *International Journal of Social Robotics* 3 (1), 2011, 41–52 <https://doi.org/10.1007/s12369-010-0082-7>.
- Barrett, H. C./Todd, P. M./Miller, G. F. et al.*: Accurate Judgments of Intention from Motion Cues Alone. A Cross-Cultural Study, in: *Evolution and Human Behavior* 26 (4), 2005, 313–331 <https://doi.org/10.1016/j.evolhumbehav.2004.08.015>.
- Barrett, H. C.*: Cognitive Development and the Understanding of Animal Behavior, in: *Elis, B. J./Bjorklund, D. F. (eds.): Origins of the Social Mind*, New York 2005, 438–467.
- Barrett, Lisa F.*: Solving the Emotion Paradox. Categorization and the Experience of Emotion, in: *Personality and Social Psychology Review* 10 (1), 2006, 20–46 https://doi.org/10.1207/s15327957pspr1001_2.
- Barsalou, Lawrence W./Simmons, W. Kyle/Barbey, Aron K. et al.*: Grounding Conceptual Knowledge in Modality-Specific Systems, in: *Trends in Cognitive Sciences* 7 (2), 2003, 84–91 [https://doi.org/10.1016/S1364-6613\(02\)00029-3](https://doi.org/10.1016/S1364-6613(02)00029-3).
- Bechtel, William*: Representations and Cognitive Explanations. Assessing the Dynamicist's Challenge in Cognitive Science, in: *Cognitive Science* 22 (3), 1998, 295–318 https://doi.org/10.1207/s15516709cog2203_2.
- Brockmann, R. John*: A Homunculus in the Computer?, in: *Journal of Technical Writing and Communication* 27 (2), 1997, 119–145 <https://doi.org/10.2190/E7HL-A1V3-RTUY-8PE2>.

- Butterfill*, Stephen A./*Apperly*, Ian A.: How to Construct a Minimal Theory of Mind, in: *Mind & Language* 28 (5), 2013, 606–637 <https://doi.org/10.1111/mila.12036>.
- Carpenter*, Julie: The Quiet Professional. An Investigation of U.S. Military Explosive Ordnance Disposal Personnel Interactions with Everyday Field Robots, unpublished dissertation, University of Washington, 2013.
- Chater*, Nick/*Tenenbaum*, Joshua B./*Yuille*, Alan: Probabilistic Models of Cognition. Conceptual Foundations, in: *Trends in Cognitive Sciences* 10 (7), 2006, 287–291 <https://doi.org/10.1016/j.tics.2006.05.007>.
- Chemero*, Anthony: *Radical Embodied Cognitive Science*, Cambridge 2009 <https://doi.org/10.7551/mitpress/8367.001.0001>.
- Clark*, Andy: The Dynamical Challenge, in: *Cognitive Science* 21 (4), 1997, 461–481 https://doi.org/10.1207/s15516709cog2104_3.
- Clark*, Andy: Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science, in: *Behavioral and Brain Sciences* 36 (3), 2013, 181–204 <https://doi.org/10.1017/S0140525X12000477>.
- Coeckelbergh*, Mark: Artificial Companions. Empathy and Vulnerability Mirroring in Human-Robot Relations, in: *Studies in Ethics, Law, and Technology* 4 (3), 2011 <https://doi.org/10.2202/1941-6008.1126>.
- Colombo*, Matteo/*Seriès*, Peggy: Bayes in the Brain. On Bayesian Modelling in Neuroscience, in: *The British Journal for the Philosophy of Science* 63 (3), 2012, 697–723 <https://doi.org/10.1093/bjps/axr043>.
- Damiano*, L./*Dumouchel*, P.: Anthropomorphism in Human-Robot Co-evolution, in: *Frontiers in Psychology* 9, 2018, 468 <https://doi.org/10.3389/fpsyg.2018.00468>.
- De Wit*, Matthieu M./*De Vries*, Simon/*Van Der Kamp*, John et al.: Affordances and Neuroscience. Steps towards a Successful Marriage, in: *Neuroscience & Biobehavioral Reviews* 80, 2017, 622–629 <https://doi.org/10.1016/j.neubio.rev.2017.07.008>.
- Di Pellegrino*, G./*Fadiga*, L./*Fogassi*, L. et al.: Understanding Motor Events. A Neurophysiological Study, in: *Experimental Brain Research* 91 (1), 1992, 176–180 <https://doi.org/10.1007/BF00230027>.
- Doeller*, Christian F./*Barry*, Caswell/*Burgess*, Neil: Evidence for Grid Cells in a Human Memory Network, in: *Nature* 463.7281, 2010, 657–661 <https://doi.org/10.1038/nature08704>.
- Epley*, N./*Waytz*, A./*Cacioppo*, J. T.: On Seeing Human. A Three-Factor Theory of Anthropomorphism, in: *Psychological Review* 114 (4), 2007, 864–886 <https://doi.org/10.1037/0033-295X.114.4.864>.
- Fernandez Velasco*, Pablo/*Loev*, Slawa: Affective Experience in the Predictive Mind. A Review and New Integrative Account, in: *Synthese* 198 (11), 2021, 10847–10882 <https://doi.org/10.1007/s11229-020-02755-4>.

- Fisher, J. A.*: Disambiguating Anthropomorphism. An Interdisciplinary Review, in: *Perspectives in Ethology* 9, 1991, 49–85.
- Freeman, Robert B.*: The Senses Considered as Perceptual Systems. James J. Gibson, in: *The Quarterly Review of Biology* 44 (1), 1969, 104–105 <https://doi.org/10.1086/406033>.
- Friston, Karl*: The Free-Energy Principle. A Unified Brain Theory?, in: *Nature Reviews Neuroscience* 11 (2), 2010, 127–138 <https://doi.org/10.1038/nrn2787>.
- Gallese, Vittorio/Sinigaglia, Corrado*: What Is so Special about Embodied Simulation?, in: *Trends in Cognitive Sciences* 15 (11), 2011, 512–519 <https://doi.org/10.1016/j.tics.2011.09.003>.
- Gallese, Vittorio/Eagle, Morris N./Migone, Paolo*: Intentional Attunement. Mirror Neurons and the Neural Underpinnings of Interpersonal Relations, in: *Journal of the American Psychoanalytic Association* 55 (1), 2007, 131–175 <https://doi.org/10.1177/00030651070550010601>.
- Gallese, Vittorio*: Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification, in: *Psychoanalytic Dialogues* 19 (5), 2009, 519–536 <https://doi.org/10.1080/10481880903231910>.
- Gazzola, V./Rizzolatti, G./Wicker, B. et al.*: The Anthropomorphic Brain. The Mirror Neuron System Responds to Human and Robotic Actions, in: *NeuroImage* 35 (4), 2007, 1674–1684 <https://doi.org/10.1016/j.neuroimage.2007.02.003>.
- Gordon, Evan M./Chauvin, Roselyne J./Van, Andrew N. et al.*: A Somato-Cognitive Action Network Alternates with Effector Regions in Motor Cortex, in: *Nature* 617.7960, 2023, 351–359 <https://doi.org/10.1038/s41586-023-05964-2>.
- Groom, Victoria/Nass, Clifford*: Can Robots Be Teammates? Benchmarks in Human–Robot Teams, in: *Interaction Studies* 8 (3), 2007, 483–500 <https://doi.org/10.1075/is.8.3.10gro>.
- Hafting, Torkel/Fyhn, Marianne/Molden, Sturla et al.*: Microstructure of a Spatial Map in the Entorhinal Cortex, in: *Nature* 436.7052, 2005, 801–806 <https://doi.org/10.1038/nature03721>.
- Heider, Fritz/Simmel, Marianne*: An Experimental Study of Apparent Behavior, in: *The American Journal of Psychology* 57 (2), 1944, 243–259 <https://doi.org/10.2307/1416950>.
- Heyes, Cecilia/Catmur, Caroline*: What Happened to Mirror Neurons?, in: *Perspectives on Psychological Science* 17 (1), 2022, 153–168 <https://doi.org/10.1177/1745691621990638>.
- Kashi, Shir/Levy-Tzedek, Shelly*: Smooth Leader or Sharp Follower? Playing the Mirror Game with a Robot, in: *Restorative Neurology and Neuroscience* 36 (2), 2018, 147–159 <https://doi.org/10.3233/RNN-170756>.

- Kraišniković, Ceca/Maass, Wolfgang/Legenstein, Robert*: Spike-Based Symbolic Computations on Bit Strings and Numbers, in: *Frontiers in Artificial Intelligence and Applications* 342, 2021, 214–234 <https://doi.org/10.3233/FAIA210356>.
- Krause, Jens/Winfield, Alan F.T./Deneubourg, Jean-Louis*: Interactive Robots in Experimental Biology, in: *Trends in Ecology & Evolution* 26 (7), 2011, 369–375 <https://doi.org/10.1016/j.tree.2011.03.015>.
- Lindblom, Jessica*: A Radical Reassessment of the Body in Social Cognition, in: *Frontiers in Psychology* 11, 2020, 987 <https://doi.org/10.3389/fpsyg.2020.00987>.
- MacLean, Paul*: *The Triune Brain in Evolution*, New York 1990.
- Marzola, Patrícia/Melzer, Thayza/Pavesi, Eloisa et al.*: Exploring the Role of Neuroplasticity in Development, Aging, and Neurodegeneration, in: *Brain Sciences* 13 (12), 2023, 1610 <https://doi.org/10.3390/brainsci13121610>.
- Miller, Mark/Clark, Andy*: Happily Entangled. Prediction, Emotion, and the Embodied Mind, in: *Synthese* 195 (6), 2018, 2559–2575 <https://doi.org/10.1007/s11229-017-1399-7>.
- Noel, Jean-Paul/Bertoni, Tommaso/Terrebonne, Emily et al.*: Rapid Recalibration of Peri-Personal Space. Psychophysical, Electrophysiological, and Neural Network Modeling Evidence, in: *Cerebral Cortex* 30 (9), 2020, 5088–5106 <https://doi.org/10.1093/cercor/bhaa103>.
- Nyholm, Sven/Smids, Jilles*: Can a Robot Be a Good Colleague?, in: *Science and Engineering Ethics* 26 (4), 2020, 2169–2188 <https://doi.org/10.1007/s11948-019-00172-6>.
- O’Keefe, John/Dostrovsky, Jonathan*: The Hippocampus as a Spatial Map. Preliminary Evidence from Unit Activity in the Freely-Moving Rat, in: *Brain Research* 34 (1), 1971, 171–175 [https://doi.org/10.1016/0006-8993\(71\)90358-1](https://doi.org/10.1016/0006-8993(71)90358-1).
- Penfield, Wilder/Boldrey, Edwin*: Somatic Motor and Sensory Representation in the Cerebral Cortex of Man as Studied by Electrical Stimulation, in: *Brain* 60 (4), 1937, 389–443 <https://doi.org/10.1093/brain/60.4.389>.
- Rogers, Carl R.*: Significant Aspects of Client-Centered Therapy, in: *American Psychologist* 1 (10), 1946, 415–422 <https://doi.org/10.1037/h0060866>.
- Sagan, Carl*: *The Dragons of Eden. Speculations on the Evolution of Human Intelligence*, New York 1977.
- Sargolini, Francesca/Fyhn, Marianne/Hafting, Torkel et al.*: Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex, in: *Science* 312.5774, 2006, 758–762 <https://doi.org/10.1126/science.312.5774.758>.
- Schultz, Wolfram*: Reward Prediction Error, in: *Current Biology* 27 (10), 2017, R369–R371 <https://doi.org/10.1016/j.cub.2017.02.064>.

- Sol, Daniel/Bacher, Sven/Reader, Simon M./Lefebvre, Louis*: Brain Size Predicts the Success of Mammal Species Introduced into Novel Environments, in: *The American Naturalist* 172.S1, 2008, 63–71 <https://doi.org/10.1086/588304>.
- Steffen, Patrick R./Hedges, Dawson/Matheson, Rebekka*: The Brain Is Adaptive Not Triune. How the Brain Responds to Threat, Challenge, and Change, in: *Frontiers in Psychiatry* 13, 2022, 802606 <https://doi.org/10.3389/fpsy.2022.802606>.
- Turkle, Sherry*: Authenticity in the Age of Digital Companions, in: *Interaction Studies* 8 (3), 2007, 501–517.
- Tversky, Barbara*: Structures Of Mental Spaces. How People Think About Space, in: *Environment and Behavior* 35 (1), 2003, 66–80 <https://doi.org/10.1177/0013916502238865>.
- Urquiza-Haas, E. G./Kotrschal, K.*: The Mind Behind Anthropomorphic Thinking. Attribution of Mental States to Other Species, in: *Animal Behaviour* 109, 2015, 167–176 <https://doi.org/10.1016/j.anbehav.2015.08.011>.
- Vicario, Carmelo Mario*: Perceiving Numbers Affects the Internal Random Movements Generator, in: *The Scientific World Journal*, 2012, 1–6 <https://doi.org/10.1100/2012/347068>.
- Vidal, Jean-Marie/Vancassel, Michel/Quiris, René*: Introducing Anthropomorphism, Discontinuities and Anecdotes to Question Them, in: *Behavioural Processes* 35 (1–3), 1995, 299–309.
- Walsh, Philip J.*: Intercorporeity and the first-person plural in Merleau-Ponty. *Continental Philosophy Review* 53, 2020, 21–47. DOI: <https://doi.org/10.1007/s11007-019-09480-x>
- Wolpert, Daniel M./Ghahramani, Zoubin/Jordan, Michael I.*: An Internal Model for Sensorimotor Integration, in: *Science* 269.5232, 1995, 1880–1882 <https://doi.org/10.1126/science.7569931>.
- Ziemke, Tom*: Embodiment in Cognitive Science and Robotics, in: Cangelosi, A./Asada, M. (eds.): *Cognitive Robotics*, Cambridge 2022, 213–230 <https://doi.org/10.7551/mitpress/13780.003.0016>.

Autorinnen und Autoren

Ramy Abdin is a research assistant at the Bavarian Research Centre for Interreligious Discourses (FAU) specialising in Islam since 2018. He studied Islamic Studies, Oriental Philology, and Philosophy as well as traditional Islamic disciplines in Vienna and Cairo.
ramy.abdin@fau.de

Simon Nøddebo Balle is currently a pastor in the Church of Denmark and holds a PhD degree from Theology at Aarhus University on the ethical impacts of AI and robotics.
sballe88@gmail.com

Dr. Lukas Brand is a research assistant at the AALab at the RPTU Kaiserslautern where he teaches in the area of computing and society. In his research he focuses theological and philosophical enquiries in anthropology, ethics, and the societal impacts of generative AI and computing.
lukas.brand@cs.rptu.de

Marius de Byl is research assistant at the Chair of Religious Education and Catechetics at the Ruhr University Bochum. Main research interests: Translation as a theorem of religious education; theological of religious education.
marius.debyl@rub.de

Kate Cooper is Professor of History at Royal Holloway, University of London. Her research interests centre around religious and social change, and the distinctive institutions of daily life. Her work champions a historical perspective on the importance of intersubjectivity and intercorporeality.

Steffi Fabricius is a research assistant at the Faculty of Protestant Theology at University Siegen teaching in the area of religious education and its didactics. In her research she focuses on questions of

positionality and AI in religious educational and theological contexts.
fabricius@evantheo.uni-siegen.de

Manishk Gawande is a 3rd year Computer Science student at the Innovative Global Program at Shibaura Institute of Technology (SIT), Tokyo, Japan. His research focuses on social robotics and emotions, specifically on realistic emotional expressions in assistive robotic devices.

am22105@shibaura-it.ac.jp

Dr. Hille Haker is Professor of Moral Theology at the Universities of Chicago and Frankfurt. As a recognised expert in medical and bioethics, she has advised the German Medical Association and the German Bishops' Conference. She is a member of the European Commission's Ethics Advisory Group.

Noreen L. Herzfeld is Reuter Professor of Science and Religion at St. John's University and senior research associate with ZRS Koper. As theologian and computer scientist, she is the author of several books, most recently: *The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age*.

Martin Kutz studied electrical engineering in Aachen and theology (Catholic) in Sankt Augustin. Research assistant at ScaDS.AI Dresden/Leipzig and at the Chair of Religious Education (Protestant) at the Technische Universität Dresden. His research focuses on the social relevance of artificial intelligence and the acceptance of chatbots and robots in religious contexts.

martin.kutz@tu-dresden.de

Hildelith Leyser, PhD student at McGill University in neuroscience, has a bachelor's degree in history from the University of Oxford. Her research interests lie in the possibilities of human-computer interaction to combat neurodegenerative and neuropsychiatric disorders.

Dr. Ilona Nord is Professor of Religious Education at the Institute for Protestant Theology at the University of Würzburg and Head of ZABUS at the Faculty of Human Sciences at the University of Würzburg. Her research focuses on digital religion/theology and the relationship between religion and digital media.

Rev. Ralf Peter Reimann, M.Sc. in Computer Science, serves as Internet Commissioner of the Evangelical Church in the Rhineland. His work bridges theology and technology, with research interests in artificial intelligence and spirituality, as well as digital justice. ralfpeter.reimann@ekir.de

Dr. Thomas Schlag is Professor of Practical Theology at the Theological Seminary of the Faculty of Theology and Religious Studies at the University of Zürich, Head of research cluster “Digital Religion(s). Communication, Interaction and Transformation in the Digital Society” and director of the Zürich ZKE.

Marcel Scholz is research associate at the chair for Religious Education, Protestant Theology, JGU Mainz. His research interests lie in religious education and digitality and religious Education in the context of AI and Transhumanism. marcel.scholz@uni-mainz.de

Jonas Simmerlein is a scholar of practical theology at the Faculty of Protestant Theology, University of Vienna. As an associate researcher, he investigates the role of artificial intelligence and robotics in religious practices, as well as the phenomenology of religion.

Theresa M. Thunig is a research assistant at the Chair of Dogmatics and Dogma History at the Faculty of Catholic Theology, Ruhr University Bochum, Germany. She is working on her dissertation on life-world contexts in dogmatics.

Dr. Max Tretter is a postdoctoral researcher at the Chair of Systematic Theology (Ethics) at Friedrich-Alexander University Erlangen-Nuremberg. His research explores the ethical and theological dimensions of artificial intelligence, robotics, and digital culture. max.tretter@fau.de

Gabriele Trovato is Associate Professor in Innovative Global Program, Shibaura Institute of Technology, where he leads Lab 22, and Visiting Researcher in Waseda University, Tokyo, Japan. His main research interests include Human-Robot Interaction, with focus on culture and religion related aspects.

Autorinnen und Autoren

Dominik Winter is research assistant at the Chair for Moral Theology at the Faculty of Catholic Theology, University of Erfurt. His research interest focusses on metaethical questions regarding the role of experience and emotions for ethical arguments, as well as applied ethical questions regarding Moral Enhancement and Artificial Intelligence.

dominik.winter@uni-erfurt.de