



Special Topic KI-Systeme gestalten und erfahren. Konzepte, Werte, Anwendungen

Research Bürgerwissenschaftliche Forschungsansätze
in Medizin und Gesundheitsforschung

Interview with K. Eid-Sabbagh: Critical political economy
of the public infrastructure crisis in Lebanon

TATuP

– Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis ist die begutachtete Open-Access-Zeitschrift für das interdisziplinäre Feld der Technikfolgenabschätzung und angrenzende Forschungsgebiete. Die Zeitschrift richtet sich an eine inter- und transdisziplinäre Leserschaft.

TATuP – Journal for Technology Assessment in Theory and Practice is the peer reviewed open access journal for the interdisciplinary field of technology assessment and neighboring fields of research. It addresses an inter- and transdisciplinary readership.

IMPRESSUM / LEGAL NOTICE

HERAUSGEBER/EDITOR



Karlsruher Institut für Technologie (KIT)
Institut für Technikfolgenabschätzung
und Systemanalyse (ITAS)
Karlstraße 11
76133 Karlsruhe, Germany

VERLAG / PUBLISHING HOUSE



oekom – Gesellschaft für ökologische
Kommunikation mit beschränkter Haftung
Waltherstraße 29
80337 München, Germany

HERAUSGEBERSCHAFT / EDITORIAL BOARD

Prof. Dr. Armin Grunwald, KIT-ITAS, Karlsruhe, DE
(Vorsitzender/Chair)
Prof. Mahmud Farooque, ASU, Phoenix, US
Prof. Sven Ove Hansson, KTH, Stockholm, SE
Dr. Karen Kastenhofer, ITA, Wien, AT
Dr. Linda Nierling, KIT-ITAS, Karlsruhe, DE
Dr. Pauline Rioussat, KIT-ITAS/TAB, Berlin, DE

WISSENSCHAFTLICHER BEIRAT / SCIENTIFIC ADVISORY BOARD

Prof. Peta Ashworth, University of Queensland, AU
Prof. Dr. Daniel Barben, Universität Klagenfurt, AT
Prof. Dr. Birgit Blattel-Mink,
Universität Frankfurt a. M., DE
PD Dr. Alexander Bogner, ITA, Wien, AT
Dr. Pierre Delvenne, Universität von Lüttich, BE
Prof. Dr. Hans-Liudger Dienel, nexus Institut, Berlin, DE
Prof. Dr. Ulrich Dolata, Universität Stuttgart, DE
Dr. Elisabeth Ehrensperger, TA-Swiss, CH
PD Dr. Jessica Heesen, Universität Tübingen, DE
Prof. Dr. Matthias Kaiser, University of Bergen, NO
Prof. Dr. Cordula Kropp, Universität Stuttgart, DE
Dr. Justine Lacey, CSIRO, Canberra, AU
Dr. Ralf Lindner, Fraunhofer ISI, Karlsruhe, DE
Dr. Stephan Lingner, IQIB, Bad Neuenahr-Ahrweiler, DE
Dr. Krzysztof Michalski, TU Rzeszow, PL
PD Dr. Michael Nentwich, ITA, Wien, AT
Dr. Poonam Pandey, DST-Center for Policy Research,
IISC, Bangalore, IN

Prof. Dr. Sebastian Pfotenhauer, TU München, DE
Dr. Witold-Roger Pogonietz, ITAS-KIT, Karlsruhe, DE
Dr. Bernard Reber, CEVIPOF, Paris, FR
Dr. Martin Sand, TU Delft, NL
Prof. Dr. Thomas Saretzki, Universität Lüneburg, DE
Prof. Dr. Petra Schaper-Rinkel, Universität Graz, AT
Prof. Dr. Jan C. Schmidt, Hochschule Darmstadt, DE
Prof. Dr. Miranda Schreurs, HfP an der
TU München, DE
Dr. Elena Seredkina, Universität Perm, RU
PD Dr. Mahshid Sotoudeh, ITA, Wien, AT
Prof. Dr. Karsten Weber, OTH Regensburg, DE
Dr. Marcel Weil, KIT-ITAS, Karlsruhe, DE
Prof. Dr. Johannes Weyer, TU Dortmund, DE

REDAKTION / EDITORIAL TEAM

Dr. Ulrich Ufer (Redaktionsleitung/managing & lead editor)
Constanze Scherz (wiss. Redaktion/academic editor)
Leonie Seng (wiss. Redaktion/academic editor)
Jonas Moosmüller (TA-Fokus-Redaktion/TA-Fokus editor)
Julia Thomas (Redaktionsassistent/editorial assistant)
Simon Barth (Redaktionsassistent/editorial assistant)

Institut für Technikfolgenabschätzung
und Systemanalyse (ITAS)
Karlstraße 11

76133 Karlsruhe, Germany
☎ +49 (0)721 608 26014
✉ redaktion@tatup.de

HINWEISE FÜR AUTORINEN UND AUTOREN / INFORMATION FOR AUTHORS

www.tatup.de/index.php/tatup/Submit

ERSCHEINUNGSWEISE / PUBLICATION SCHEDULE

3x jährlich / 3x per year

BESTELLUNG / SUBSCRIPTION

www.tatup.de/index.php/tatup/subscriptionStatic

www.oekom.de/tatup/bezug
Verlegerdienst München GmbH
Aboservice oekom verlag
Gutenbergstraße 1
82205 Gilching, Germany
☎ +49 (0)8105 388 563
☎ +49 (0)8105 388 333
✉ oekom-abo@verlegerdienst.de

TATuP ONLINE

www.tatup.de
Newsletter: www.oekom.de/newsletter

PAPIER / PAPER

Gedruckt auf Circleoffset Premium White,
100 % FSC®-Recyclingpapier, zertifiziert mit dem
Blauen Engel (RAL-UZ 14). /
Printed on Circleoffset Premium White, 100 % FSC®
recycled paper, certified with The Blue Angel (RAL-UZ 14).

DRUCK / PRINTING

Friedrich Pustet GmbH & Co. KG
93051 Regensburg, Germany
www.pustet-druck.de

ANZEIGEN / ADVERTISEMENTS

Mona Fricke
oekom – Gesellschaft für ökologische
Kommunikation mit beschränkter Haftung
☎ +49 (0)89 54418417
✉ anzeigen@oekom.de

VISUELLE KONZEPTION & GESTALTUNG / VISUAL CONCEPT & DESIGN

Kornelia Rumberg, www.rumbergdesign.de

GRAFIK & SATZ / GRAPHIC DESIGN & TYPESET

Tobias Wantzen, www.wantzen.com
Cover art based on © macrovector_official / Freepik

ISSN

2568-020X (Print), 2567-8833 (Online)

COPYRIGHT & LIZENZ / COPYRIGHT & LICENCE

Creative Commons Licence CC BY 4.0
www.creativecommons.org/licenses/by/4.0/

Erfüllungsort / Gerichtsstand: München, Deutschland
Place of fulfillment / Place of jurisdiction: Munich, Germany



oekom kompensiert bereits seit 2008 seine unvermeid-
lichen CO₂-Emissionen. / Since 2008 oekom offsets
its unavoidable CO₂ emissions.

Editorial



KAREN KASTENHOFER

Institut für Technikfolgen-Abschätzung,
Österreichische Akademie
der Wissenschaften, Wien, AT
(kkast@oeaw.ac.at)

B einahe zwei Jahre lang müssen wir alle uns nun schon einem wellenförmigen Rhythmus unterwerfen: erst hohe Inzidenzraten und strikte Maßnahmen zur Eindämmung der Verbreitung von SARS-CoV-2, dann gesunkene Raten und Lockerungen ebendieser Maßnahmen. Gerade noch begann diese Dynamik dank effektiver Impfstoffe und einer teilweisen Entkoppelung von Inzidenz und Mortalität abzuflachen und ein Gefühl von ‚Ebbe nach der Flut‘ stellte sich ein; nun stehen wir bereits in der nächsten Welle. Und dennoch wird es höchste Zeit, das Strandgut zu mustern, das diese Fluten mit sich bringen, wenngleich die immer noch anhaltende Pandemie auch jetzt noch kein Strandspaziergang ist. Umso existenzieller ist es, jetzt genau hinzusehen, die vergangenen Ereignisse kritisch zu reflektieren, die gegenwärtige Situation in all ihren Facetten zu erfassen und zugleich Visionen für unsere gesellschaftliche Zukunft im Blick zu behalten. Während der erste Schockzustand und die Dringlichkeit unmittelbarer Maßnahmen zu Beginn der Pandemie kaum Zeit ließen, die Details lokalen Regierens mit großen gesellschaftlichen Entwürfen abzugleichen, haben wir uns nun unter vielen – nicht immer gerecht verteilten – Entbehrungen den Raum dafür geschaffen. Es liegt an uns, ihn verantwortungsvoll zu nutzen. Wo stehen wir als Gesellschaft nach zwei Jahren Krisenerfahrung und Krisenbearbeitung? Welche kurzfristig über Bord geworfenen Optionen würden wir auch weiterhin nicht vermissen? Welche wollen wir uns in jedem Fall zurückerobern? Wo braucht es neue Ansätze, um entstandene Schäden zu bearbeiten oder neue Resilienzen aufzubauen? Technikfolgenabschätzung (TA) als ‚reflexives Projekt‘ kann hier wichtige Anstöße geben, ist sie doch geübt darin, die Tücken der Details mit großen Zielsetzungen wie Gemeinwohl und Nachhaltigkeit zu verknüpfen. Und dennoch müssen auch wir uns der kritischen Frage stellen: Kann TA in Bezug auf derartige Krisensituationen noch mehr beitragen? Was sich deutlich abzeichnet ist, dass wir in den kommenden Jahren alle zur Verfügung stehenden Ressourcen – seien es lokales Wissen oder wissenschaftliche Expertise, gesellschaftliche Institutionen oder andere Formen sozialen Kapitals – mobilisieren werden müssen, um gute Antworten auf drängende gesellschaftliche Herausforderungen zu finden. Wie folgenreich punktuelle (Nicht-)Entscheidungen im lokalen Management globaler Krisen sind, haben uns die vergangenen Monate jedenfalls gelehrt. Diese und andere Erkenntnisse gilt es zu sondieren und umzusetzen; auch wenn wir uns die Notlage, aus der sie resultieren, alle nicht gewünscht haben.

Karen Kastenhofer

This is an article distributed under the terms of the Creative Commons Attribution License CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.3>

Inhalt / Content

TATuP 3/2021

SPECIAL TOPIC

KI-Systeme gestalten & erfahren

Künstliche Intelligenz (KI) und Big Data sind derzeit in zahlreichen Anwendungsgebieten präsent – in Wirtschaft, Gesellschaft und Wissenschaft. Technikfolgenabschätzung hat die Aufgabe, im Hinblick auf KI öffentliche Erwartungen und Befürchtungen in sachliche, realistische Perspektiven zu transformieren sowie Impulse für eine wünschenswerte Gestaltung von KI zu geben.

10

EDITORIAL

- 3** K. KASTENHOFER

TA-FOKUS

- 6** *Meldungen • TA-Grafik • Fünf Fragen an Alma Kolleck • Personalia*

SPECIAL TOPIC „KI-SYSTEME GESTALTEN UND ERFAHREN“

- 11** B. G. HUMM, S. LINGNER, J. C. SCHMIDT, K. WENDLAND
KI-Systeme. Aktuelle Trends und Entwicklungen aus Perspektive der Technikfolgenabschätzung

- 17** B. BECKERT
Vertrauenswürdige künstliche Intelligenz. Ausgewählte Praxisprojekte und Gründe für das Umsetzungsdefizit

- 23** S. DICHEL
Wenn die Technik sprechen lernt. Künstliche Kommunikation als kulturelle Herausforderung mediatisierter Gesellschaften

- 30** R. HEIL, N. B. HEYEN, M. BAUMANN, B. HÜSING, D. BACHLECHNER, U. SCHMOCH, H. KÖNIG
Artificial intelligence in human genomics and biomedicine. Dynamics, potentials and challenges

- 37** M. JUNG, J. VON GARREL
Mitarbeiterfreundliche Implementierung von KI-Systemen im Hinblick auf Akzeptanz und Vertrauen. Erarbeitung eines Forschungsmodells auf Basis einer qualitativen Analyse

- 44** S. STRAUSS
“Don’t let me be misunderstood”. Critical AI literacy for the constructive use of AI technology

- 50** C. VATER, E. GEITZ
Künstliche Intelligenz parlamentarisch (mit)gestalten. Vergangene technische Zukünfte in den Berichten der Enquete-Kommissionen des Deutschen Bundestags

- 56** T. WILMER
Rechtliche Rahmenbedingungen für KI-Systeme. Immanente Herausforderungen und mögliche Lösungen durch Control by Design

RESEARCH

Bürgerwissenschaftliche Forschungsansätze

In der Medizin- und Gesundheitsforschung finden sich aktuell nur wenige Forschungsansätze mit Citizen Science. Bürgerwissenschaftliche Ansätze werden dort durchaus praktiziert, jedoch häufig unter anderen Namen. Welche Begriffe und Methoden existieren und wie unterscheiden sich diese in Bezug auf den Grad der Bürgerbeteiligung?

63

INTERVIEW

Public infrastructure crisis in Lebanon

The case of the public infrastructure crisis in Lebanon highlights the importance of including analytical dimensions of critical political economy and global financial dynamics in technology assessment alongside a technology-society-governance perspective – in particular when focusing on the Global South.

70

RESEARCH

63

G. HAMMEL, S. WOLL, M. BAUMANN, C. SCHERZ, M. MAIA, B. BEHRISCH, S. O. BORGMANN, M. EICHINGER, J. GARDECKI, N. B. HEYEN, A. ICKS, M. POBIRUCHIN, S. WESCHKE
Bürgerwissenschaftliche Forschungsansätze in Medizin und Gesundheitsforschung. Ausgewählte Begriffe mit Fokus auf den Beteiligungsgrad

INTERVIEW

70

K. EID-SABBAGH, U. UFER
Critical political economy of the public infrastructure crisis in Lebanon. How can critical political economy perspectives contribute to technology assessment for countries in the Global South?

REFLECTIONS

75

G. KAMP
Book Review: Update und Compendium. Technikfolgenabschätzung – Handbuch für Wissenschaft und Praxis

78

P. VETTER
Book Review: Value sensitive design. A pragmatic approach to ethically guided technology design

80

P. VETTER
Meeting Report: Changing Values, Changing Technologies

82

C. BENIGHAUS, R. BEECROFT
Meeting Report: International Transdisciplinarity Conference 2021. Creating spaces and cultivating mindsets for learning and experimentation

83

J. WEYER
Letter: Gegendarstellung. Uwe Schimanks Fehldeutungen der Echtzeitgesellschaft

85

U. SCHIMANK
Reply: Replik auf J. Weyers Gegendarstellung

PRACTICE

86

M. NENTWICH
EPTA und die Pandemie

COMING NEXT

88

TATuP Dates

Meldungen

STUDIE

Deepfakes: Gefahr für die Demokratie

Wie lässt sich prüfen, ob Informationen echt und vertrauenswürdig sind – gerade solche, die über das Internet verbreitet werden? Die Möglichkeit, Personen mit Hilfe künstlicher Intelligenz in täuschend echte neue Kontexte zu stellen, macht eindeutige Antworten immer schwieriger. Forschende aus Deutschland, den Niederlanden und Tschechien haben sich nun im Auftrag des STOA-Panels des Europaparlaments mit den potenziellen Gefahren so genannter Deepfakes beschäftigt. Ihr Fazit: Die Technologie kann dazu missbraucht werden, Fake News und Desinformationen zu streuen. Deepfakes können so Personen sowie ganze Organisationen diskreditieren und potenziell auch dem Vertrauen in demokratische Institutionen schaden. Die Forschenden emp-

fehlen deshalb insbesondere, die digitalen Verbreitungswege von Deepfakes in den Blick zu nehmen und gesellschaftliche Fertigkeiten weiterzuentwickeln, um die Glaubwürdigkeit von Medieninhalten kritisch zu hinterfragen.

www.europarl.europa.eu/stoa

PODCAST

Zukunft der Bioökonomie

Was wäre, wenn es in Restaurants keine Rindersteaks mehr gäbe, sondern gegrillte Insekten? Wie nachhaltig wäre der Verzicht auf herkömmliche Fleischprodukte? Und wie würden die Gäste das Angebot annehmen? Mit solchen Fragen zu Zukunftsszenarien der Bioökonomie setzt sich der im Sommer gestartete Podcast bio:fictions auseinander. In insgesamt elf Folgen stellt der Podcast jeweils eine Person vor, die biobasierte Innovationen entwickelt oder vorantreibt. Die Spannweite reicht von nachhaltigen Insekten-Snacks, über essbare Trinkhalme und Teller bis hin zu künstlichen Koral-

lenriffen zur Wiederherstellung von Ökosystemen. Hinter dem Podcast steht ein Projekt des Instituts für Zukunftsstudien und Technologiebewertung (IZT) im Wissenschaftsjahr 2020/21 des Bundesministeriums für Forschung und Bildung. Alle Podcastfolgen sind auf den gängigen Streamingplattformen erschienen und kostenfrei nachzuhören.

biofictions.de

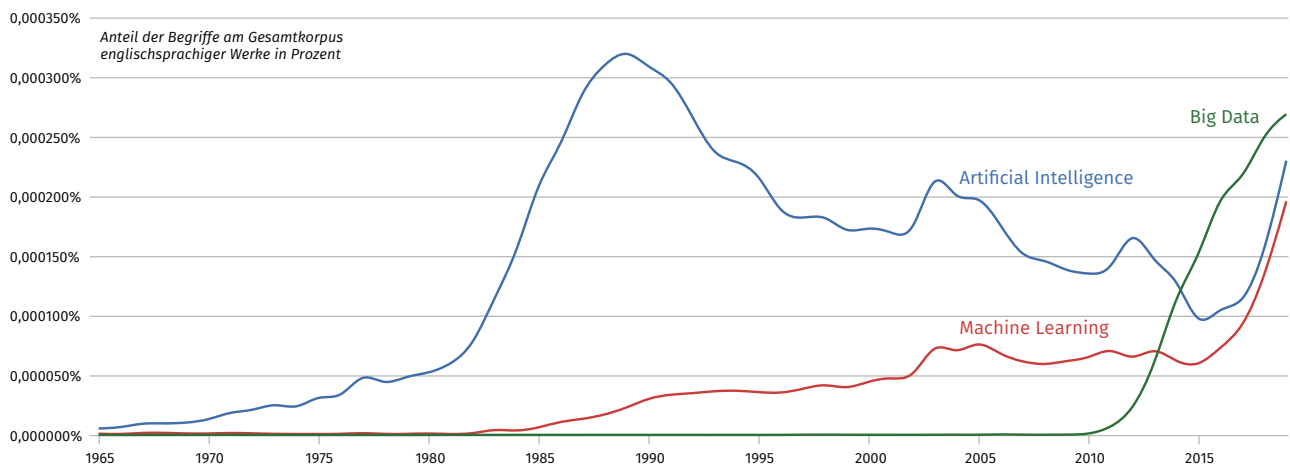
PANDEMIE

Keine Mobilitätswende durch Corona

Arbeiten im Homeoffice, Ansteckungsgefahren in öffentlichen Verkehrsmitteln, kaum mehr Urlaubsreisen: Die Corona-Pandemie hat die Mobilität der Deutschen auf den Kopf gestellt. Eine neue Studie des Fraunhofer ISI geht nun der Frage nach, ob sich die individuelle Mobilität dauerhaft verändern wird. Dazu befragten die Forschenden im August 2020 sowie im März 2021 deutsche Bürgerinnen und Bürger in Großstädten, wie sich ihre

TA-Grafik *Künstliche Intelligenz*

Der US-Technologieriese Google digitalisiert die Buchproduktion der Welt – und liefert damit Einblicke in die Konjunktur zentraler Begrifflichkeiten der aktuellen TATuP-Ausgabe.



Quelle: Books Ngram Viewer, https://books.google.com/ngrams/graph?content=Artificial+Intelligence%2CMachine+Learning%2CBig+Data&year_start=1965&year_end=2019&corpus=26&smoothing=0#

Wegezwecke (z. B. Dienstreisen, Freizeitwege) und die Wahl der Verkehrsmittel nach den Corona-Beschränkungen ändern würden. Die Ergebnisse legen nahe, dass die große Mobilitätsrevolution ausgeblieben ist: Im März 2021 stehen den knapp 12 Prozent der Befragten, die in Zukunft weniger mobil sein möchten, 38 Prozent an Befragten gegenüber, die künftig sogar mobiler sein möchten als zuvor. Als „kein gutes Zeichen für die Mobilitätswende“ wertet das Studienleiter Johannes Schuler. Es werde daher weiter strenge Push-out-Maßnahmen brauchen, um die Mobilität in Großstädten zu verändern.
www.isi.fraunhofer.de

KONFERENZ

Internationale TA-Community in Karlsruhe

Vom 20. bis 23. März 2022 treffen sich Forschende aus der Technikfolgenabschätzung und angrenzenden Feldern sowie Vertreterinnen und Vertreter aus Politik und Gesellschaft in Karlsruhe. Unter dem Titel „Digital Future(s): TA in and for a Changing World“ diskutieren sie bei der mittlerweile fünften Europäischen TA-Konferenz über digitale Transformationen und deren Herausforderungen für Politik, Gesellschaft, Wirtschaft und Wissenschaft. Keynote-Speakerinnen sind Payal Arora, Expertin für ethisches Design an der Universität Rotterdam, Doris Allhutter, Politikwissenschaftlerin an der Österreichischen Akademie der Wissenschaften sowie Catrin Misselhorn, Professorin für Philosophie Vordenkerin im Bereich der Maschinen- und Roboterethik in Deutschland. Die Konferenz findet im Zentrum für Kunst- und Medien (ZKM) statt, einer der weltweit bedeutendsten Kunstinstitutionen, die selbst zu den Auswirkungen von Digitalisierung forscht. Gastgeber ist das Institut für Technikfolgenabschätzung und Systemanalyse (ITAS), das die Konferenz mit seinen Partnern im Netzwerk globalTA organisiert.

karlsruhe2022.technology-assessment.info



5 Fragen an: Alma Kollect

Wissenschaftliche Mitarbeiterin am Büro für Technikfolgen-Abschätzung
beim Bundestag (TAB)

Warum betreiben Sie Technikfolgenabschätzung?

Im Studium habe ich eine TAB-Studie zu Netzöffentlichkeit und digitaler Demokratie gelesen und war so beeindruckt von ihrer Vielseitigkeit und Weitsicht, dass ich dachte: Sowas will ich auch mal beruflich machen. Und heute bin ich überzeugt: Menschen, die interessengetrieben über Technik reden, gibt es genug, aber interdisziplinäre Technikanalysen könnte es ruhig noch mehr geben.

Welche Forschungsfrage interessiert Sie besonders?

Gerade arbeite ich zu Gene Drives, einer Gentechnologie, bei der durch Freisetzung einzelner transgener Lebewesen ganze Populationen geändert werden sollen. Das ist ein ziemlich ambivalenter Anspruch und deshalb sehr interessant. Grundsätzlich finde ich das Feld von KI und Digitalisierung aber auch spannend, weil es Auswirkungen auf den Alltag so vieler Menschen hat.

Am TAB beraten Sie Politikerinnen und Politiker des Deutschen Bundestags.

Sollten die mehr auf Sie hören?

Nein! Oft ist es sogar besser, nicht auf mich zu hören. Aber ganz im Ernst: Das hieße ja, dass sie extrem leicht durch externe Positionen beeinflussbar wären. Wir als TAB geben zwar mitunter Handlungsempfehlungen, möchten aber in erster Linie möglichst breit und fundiert informieren.

Sie forschen und lehren zu politischen Beteiligungsformaten. Wie sähen die denn idealerweise aus?

Informiert, inklusiv und verbindlich: also mit ausreichend Sachkenntnis, einer möglichst breiten Beteiligung auch von klassisch Unterrepräsentierten und mit einem klaren politischen Output.

Welche Rolle sollte Wissenschaft in gesellschaftlichen Debatten spielen? Einmischen oder raushalten?

Unbedingt einmischen! Der öffentliche Diskurs kommt nicht aus ohne informierte und sachliche Stimmen.

NETZWERK

Zwei neue EPTA-Mitglieder

Das europäische Netzwerk für parlamentarische TA (EPTA) hat zwei weitere Mitgliedsinstitutionen aufgenommen. Neu dabei ist als 13. Vollmitglied die Oficina de Ciencia y Tecnología. Das parlamenta-

rische Büro für Wissenschaft und Technik ist im März 2021 vom Congreso de los Diputados, dem Unterhaus des spanischen Parlaments, und der Stiftung FECYT gegründet worden. Teil des EPTA-Netzwerks ist künftig – als 12. assoziiertes Mitglied – auch das Committee for the Future of Lithuania des litauischen Parlaments.
eptanetwork.org

BETEILIGUNG

Bürgerdialog zu künstlicher Intelligenz

Wo halten Bürgerinnen und Bürger den Einsatz von KI-Anwendungen für sinnvoll – und wo nicht? Welche Hoffnungen und Erwartungen verbinden sie mit der Technologie, was sind ihre Befürchtungen und Sorgen? Nach Antworten hat das Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) mit dem Bürgerdialog „Künstliche Intelligenz in unserem Alltag“ gesucht. Bei der Veranstaltung im Programm der ersten Science Week des Karlsruher Instituts für Technologie diskutierten über 30 Teilnehmende einen Tag lang untereinander und mit Expertinnen und Experten über den Einsatz von KI für Energieversorgung, Autonomes Fahren und Robotik. Entstanden sind dabei eine Liste konkreter Wünsche in Bezug auf die künftige Forschungsagenda des KIT. Der Präsident der Forschungsuniversität, Holger Hanselka, versicherte den Bürgerinnen und Bürgern, ihre Ideen in Kürze in die KIT-Gremien einzubringen.

www.scienceweek.kit.edu

PANDEMIE

Lernen aus der Corona-Krise

Die Mitglieder des europäischen Netzwerks für parlamentarische TA (EPTA) haben analysiert, wie Regierungen und Gesellschaften auf die Covid-19-Pandemie reagiert haben. Im Fokus eines im November vorgestellten Berichts steht die Frage, wie Politikerinnen und Politiker wissenschaftliche Erkenntnisse, Technologien und Innovationen genutzt haben. Dabei untersuchten die Partnerinstitutionen des Netzwerks auch Schwächen und Stärken der Entscheidungsfindung während der Pandemie. Darauf aufbauend wurden sechs Themenbereiche identifiziert, die sofortige Aufmerksamkeit verlangen. Ganz oben auf die politische Agenda sollten aus EPTA-Perspek-



Bürgerdialog mit Hochschulpräsident: neues Beteiligungsformat zum Thema KI am Karlsruher Institut für Technologie.

(Quelle: Amadeus Bramsiepe/KIT)

tive etwa die Diskussion über eine integrative und faire Digitalisierung sowie die Frage, wie Arbeit von zu Hause aus verantwortungsvoll organisiert werden kann. Darüber hinaus müsste das Augenmerk auf Schwachstellen im Gesundheitswesen und in der technologischen Infrastruktur liegen. Um dieser und kommenden Pandemien besser zu begegnen, empfehlen die EPTA-Mitglieder zudem, medizinische Daten künftig international unter Wahrung der Privatsphäre auszutauschen. (vgl. ausführlich S. 88)

eptanetwork.org

STUDIE

Demokratie und Digitalisierung

Digitale Technologien eröffnen Bürgerinnen und Bürgern neue Möglichkeiten der Partizipation, gerade auch bei politischen Prozessen. Die Schweizer Stiftung für Technologiefolgen-Abschätzung (TA-Swiss) hat verschiedene Aspekte im Wechselspiel von Digitalisierung und Demokratie untersucht. In insgesamt drei Teilstudien betrachtet sie das Thema aus den Perspektiven von Forschung, Jugend-

lichen und der Kunst. In den im August 2021 veröffentlichten Studien analysiert die TA-Swiss Chancen und Risiken und spricht Empfehlungen aus, wie demokratische Systeme mit der zunehmenden Digitalisierung umgehen sollten. Eine große Chance bestehe beispielsweise darin, dass Politikerinnen und Politiker verschiedene Zielgruppen besser erreichen und ihnen einen umfassenderen Zugang zu politischen Informationen bieten können. Um der Gefahr von Fake News vorzubeugen, empfiehlt die TA-Swiss Faktenchecks durch Expertenplattformen. Zudem sollen die politisch Verantwortlichen Menschen dabei helfen, ihre Kompetenzen bei der Informationssuche und -verarbeitung im Internet zu verbessern.

www.ta-swiss.ch/digitale-demokratie

PUBLIKATION

Die Epistemisierung des Politischen

Ob Impfdebatte, Corona- oder Klimakrise – viele politische Streitfragen werden heute als Wissenskonflikte verhandelt. Gesellschaftliche Akteure beschäftigen sich immer weniger mit normativen

Aspekten und individuellen Handlungsoptionen, sondern streiten darüber, wer am genauesten mit den Ergebnissen der Wissenschaft übereinstimmt. Diese These verfolgt der österreichische Soziologe Andreas Bogner in einem aktuellen Essay zur gesellschaftlichen Fixierung auf Wissensfragen und ihre Folgen. Seine Diagnose: Die „Epistemisierung des Politischen“ ist gefährlicher für unsere Demokratie als das leicht durchschaubare Spiel mit Fake News und Filterblasen. Der Autor ist Senior Scientist am Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften und Vorsitzender der Österreichischen Gesellschaft für Soziologie.



Bogner, Alexander:
Die Epistemisierung des Politischen.
Wie die Macht des Wissens die Demokratie gefährdet.
Ditzingen: Reclam, 2021, 132 S.,
ISBN 9783150140833

Personalia



Quelle: Matthias Lange/Gips-Schule-Stiftung

Die Forschungsgruppe „Nachhaltigkeit und gesellschaftliche Transformation“ des Instituts für Technikfolgenabschätzung und Systemanalyse (ITAS) hat den Joachim-Reutter-Preis 2021 für soziale Innovationen erhalten. Die Stuttgarter Gips-Schule-Stiftung zeichnete das Team um Oliver Parodi für seine Vorreiterrolle im Bereich der Reallabore aus. In Karlsruhe beschäftigen sich die Forschenden im Projekt „Quartier Zukunft – Labor Stadt“ seit nunmehr 10 Jahren mit den Grundlagen nachhaltiger Entwicklung. Ihr Ziel ist es, an der Schnittstelle von Gesellschaft und Wissenschaft Experimentierräume für soziale oder technische Innovationen zu schaffen, die den Alltag nachhaltiger machen. Mit dem Preisgeld von insgesamt 15.000 € will die Forschungsgruppe unter anderem Gastaufenthalte von Forschenden aus dem Ausland ermöglichen. Geplant ist auch, die Erfahrungen aus dem Reallabor für die Arbeit von Unternehmen und Kommunen nutzbar zu machen.

Quelle: André Künzelmann/UFZ



Sina Leipold leitet seit September 2021 das Department Umweltpolitik am Helmholtz-Zentrum für Umweltforschung (UFZ) und wurde von der Friedrich-Schiller-Universität Jena und dem UFZ auf die gleichnamige Professur berufen. Die Politik- und Sozialwissenschaftlerin leitete zuletzt als Juniorprofessorin eine vom BMBF geförderte Nachwuchsforschungsgruppe zur „Circular Bio-Economy“. Schwerpunkt ihrer Arbeit, für die sie jüngst mit dem Albert-Bürklin-Preis 2021 ausgezeichnet wurde, sind Narrative von Politik, Wirtschaft und Gesellschaft für eine Transformation hin zu einer klugen und nachhaltigen Ressourcensteuerung.



Oliver Bennett ist neuer Leiter des britischen Parliamentary Office of Science and Technology (POST). Der Umweltwissenschaftler hat seit 2004 eine Reihe von Aufgaben im und um das Parlament des Vereinigten Königreichs herum wahrgenommen – etwa als Leiter des Bereichs Wirtschaft und Verkehr im Forschungsdienst der House of Commons Library oder als parlamentarischer Berater im Ministerium für Wirtschaft, Energie und Industriestrategie. Oliver Bennett will die wissenschaftliche Beratung durch POST künftig noch stärker als bisher in die Prozesse des Parlaments einbinden und dabei auch mit der Spitzenforschung im Vereinigten Königreich kooperieren.

In memoriam Melanie Peters



Im Sommer ereilte die TA-Community eine schockierende Nachricht: Am 11. August ist Melanie Peters, seit 2015 Direktorin des niederländischen Rathenau-Instituts, im Alter von 55 Jahren gestorben. Ihre techniknahe, gleichwohl immer sensibel die Rolle von Technik in der Gesellschaft beobachtende wissenschaftliche Karriere umfasste unter anderem Stationen an niederländischen Universitäten, am Imperial College in London, an der University of Texas in Austin und bei Shell Research. In ihrer Leitungsfunktion am Rathenau-Institut in Den Haag lag ihr der offene und transparente Dialog mit Wissen-

schaft, Wirtschaft und Gesellschaft besonders am Herzen. Dies war insbesondere auf Veranstaltungen des EPTA-Netzwerks (European Parliamentary Technology Assessment) zu spüren, genauso wie ihre hohe kommunikative Kompetenz, die Fähigkeit des empathischen Zuhörens wie auch ihre Begabung, für dialogisch orientierte TA zu begeistern.

Die internationale TA-Community hat eines ihrer Vorbilder verloren. Wir werden ihr ein ehrendes Andenken bewahren.

Armin Grunwald
(Vorsitzender der TATuP-Herausgeberschaft)

SPECIAL TOPIC

KI-Systeme gestalten und erfahren

Konzepte, Werte, Anwendungen

Künstliche Intelligenz (KI) ist derzeit in Wirtschaft, Gesellschaft und Wissenschaft präsent. Aufgabe der Technikfolgenabschätzung ist es, öffentliche Erwartungen und Befürchtungen in sachliche, realistische Perspektiven zu transformieren sowie Impulse für eine wünschenswerte Gestaltung von KI zu geben.

Herausgegeben von Bernhard G. Humm, Stephan Lingner,
Jan C. Schmidt und Karsten Wendland

EINLEITUNG

KI-Systeme

Aktuelle Trends und Entwicklungen aus Perspektive der Technikfolgenabschätzung

Bernhard G. Humm, Hochschule Darmstadt – University of Applied Sciences, Fachbereich Informatik,
Haardtring 100, 64295 Darmstadt, DE (bernhard.humm@h-da.de) 0000-0001-7805-1981

Stephan Lingner, Institut für qualifizierende Innovationsforschung und -beratung GmbH (IQIB), Bad Neuenahr-Ahrweiler, DE (stephan.lingner@iqib.de)

Jan C. Schmidt, Hochschule Darmstadt – University of Applied Sciences, Hochschule Darmstadt,
Fachbereich Gesellschaftswissenschaften, Darmstadt, DE (jan.schmidt@h-da.de)

Karsten Wendland, Karlsruher Institut für Technologie (KIT), Institut für Technikfolgenabschätzung und Systemanalyse (ITAS),
Karlsruhe, DE (karsten.wendland@kit.edu) 0000-0003-4812-0928

11

Zusammenfassung • Künstliche Intelligenz (KI) ist – ebenso wie damit verknüpfte Techniken wie maschinelles Lernen und Big Data – in aller Munde. Die große Dynamik und Tragweite dieser Entwicklungen zeigen sich bereits in zahlreichen Anwendungsgebieten von Wirtschaft, Gesellschaft und Wissenschaft. Technikfolgenabschätzung (TA) von KI hat in diesem Zusammenhang zunächst die Aufgabe, etwaige überzogene öffentliche Erwartungen und Befürchtungen in sachliche, realistische Perspektiven zu transformieren. In einem zweiten Schritt kann TA entlang begründbarer Entwicklungsziele von KI und legitimer gesellschaftlicher Wertvorstellungen Impulse für die weitere, wünschbare Gestaltung von KI geben. Wenn TA diese Orientierungsaufgabe nah am technologischen Kern wahrnimmt, findet sie dabei große gestalterische Freiräume in frühen Phasen der Technikentwicklung vor. Die damit zusammenhängenden Gedanken werden im vorliegenden Einleitungskapitel konkretisiert und auf die Beiträge zu diesem Themenschwerpunkt angewendet.

Designing and experiencing AI systems. Recent trends and developments from a technology assessment perspective

Abstract • Artificial intelligence (AI) is on everyone's lips – as well as the associated technologies of machine learning and big data. The enormous dynamics and consequences of these developments become already evident in numerous areas of application in business, society and science. In this context, technology assessment (TA) of AI initially has the task of transforming any excessive public expectations and fears to the factual level. In a second step, TA can provide impulses for

the further, desirable design of AI based on reasonable development goals of AI and legitimate societal values. If TA conducts this orientation task close to the technological core, it can consider wide scopes of options for action in the early phases of technology development. Related thoughts are put into concrete terms in this article and will be related to the authors' contributions to this topical focus.

Keywords • Artificial Intelligence, AI, Machine Learning, Big Data, Technology Assessment

Künstliche Intelligenz (KI) und die mit ihr verbundenen Ermöglichungstechniken wie maschinelles Lernen und Big Data sind derzeit von einer großen öffentlichen Aufmerksamkeit gekennzeichnet. Die hohe Entwicklungsdynamik ist zudem mit einem erheblichen Veränderungspotenzial für Wirtschaft, Gesellschaft und Wissenschaft verknüpft. In der öffentlichen Debatte kursieren zudem teilweise unreflektierte Erwartungen an KI und auch überzogene Befürchtungen vor ihrem breiten Einsatz. Angesichts ihrer möglichen Konsequenzen für alle gesellschaftlichen Lebensbereiche erscheint es daher geboten, Erforschung, Entwicklung und Einsatz von KI entlang begründbarer Zielvorstellungen und legitimer Wertvorstellungen zu rahmen und zu gestalten. Dies ist eine Aufgabe der Technikfolgenabschätzung (TA), die hier zur Klärung und Orientierung auf unsicherem und ambivalentem Terrain in ähnlicher Weise gefragt ist, wie bei anderen tiefgreifenden Entwicklungen in Bereichen der Nanotechnologie oder der Biotechnologie. Technikfolgenabschätzung von KI betrifft dabei nicht nur die Implementierung und Anwendungsfelder von bereits entwickelten KI-Technologien, sondern insbesondere auch die vorgelagerte, im öffentlichen Diskurs oft vernachlässigte Ebene der Forschung. In diesen Frühphasen der

Technikentwicklung ist TA allgemein mit großen gestalterischen Freiräumen nah am technikwissenschaftlichen Kern von KI konfrontiert, aber eben auch mit besonderen methodischen Herausforderungen der Vorausschau ihrer Folgen (Collinridge 1982).

So ist hier zunächst die Suche nach geeigneten Ansätzen *für eine prospektive Beurteilung von KI* zu nennen. Dabei stellt sich unter anderem die Frage, wie retrospektive Analysen von bereits abgeschlossenen – und somit evidenten – Digitalisierungsdebatten für heutige Ideen für KI-Zukünfte nutzbar gemacht werden können.

Für die gestaltungsorientierte Beurteilung von KI ist sicher von zentraler Bedeutung, wie sich die *Autonomie und Kreativität sowie die Komplexität und Opazität* künstlicher ‚intelligenter‘ Systeme im gesellschaftlichen Alltag und in der Forschungspraxis aus TA-Sicht darstellt. Hier sind zum Beispiel spezifische Probleme künstlicher Kommunikation durch technische Agenten in sozialen Medien zu nennen. Ein weiterer Punkt ist die Frage, wie vertrauenswürdige KI entwickelt und in die Praxis umgesetzt werden kann und welche Hemmnisse dabei zu überwinden sind.

Daran schließt sich unmittelbar auch die Notwendigkeit der Gestaltung von geeigneten Rahmenbedingungen für den Einsatz von KI-Systemen in der gesellschaftlichen Praxis an. Hier sind zunächst *aussichtsreiche Regulierungsansätze*, ihre ‚offenen Flanken‘ sowie mögliche Lösungsvorschläge zu erarbeiten, die wünschbare KI-Innovationen fördern können. Diese Vorschläge sollten dabei auch auf die unterschiedlichen immanenten Geschwindigkeiten gesetzgeberischer und technischer Entwicklungen eingehen.

Ein anderes praktisches Problem ist die Frage, wie *notwendige KI-Kompetenzen* für den aufgeklärten Einsatz entsprechender Systeme in der Bevölkerung vermittelt und verankert wer-

ansätze zu ertüchtigen; mögliche Anwendungsfälle aus vielen Beispielen sind die Abschätzung der Chancen und Risiken von KI in der biomedizinischen Diagnose und Therapie von Krankheiten.

Die folgenden Abschnitte erläutern den Kern von KI-Systemen und ihrer Nutzbarkeit vor dem Hintergrund ihrer öffentlichen Wahrnehmung. Es folgt sodann eine kursorische Übersicht der Themenbeiträge dieses Schwerpunkts mit Vorschlägen für die Gestaltung von KI-Systemen und ihres Anwendungsrahmens.

KI – Hype und Realität¹

Die hohe Öffentlichkeitswirksamkeit von KI ist mit überzogenen Erwartungen einerseits sowie mit übertriebenen Befürchtungen andererseits verknüpft. Hinzu kommt, dass KI ein beliebtes Thema für Kinofilme ist und so die Gefahr besteht, dass Fiktion und Wirklichkeit im Bewusstsein vieler verschwimmen.

Herbert Simon, einer der Gründerväter von KI meinte bereits 1965: „Machines will be capable [...] of doing any work that a man can do“ (Allen 2001). Und sein Kollege Marvin Minski prophezeite 1970: „Within 10 years computers won't even keep us as pets“ (Allen 2001). Keine dieser Prophezeiungen hat sich auch nur annähernd erfüllt. Doch das hält zeitgenössische Autor*innen und Visionär*innen auch heute nicht davon ab, überzogenen Erwartungen und Befürchtungen medienwirksam auszurufen. Prominente Vertreter gewagter Positionen sind Ray Kurzweil, Nick Bostrom, Yuval Noah Harai, Elon Musk, sowie der verstorbene Stephen Hawking. So behauptet Ray Kurzweil (2015, S. 9): „Mit der Singularität werden wir die Grenzen unserer biologischen Körper und Gehirne überschreiten. Wir werden

Die hohe Öffentlichkeitswirksamkeit von KI ist mit überzogenen Erwartungen einerseits sowie mit übertriebenen Befürchtungen andererseits verknüpft.

den können. Dieses Anliegen richtet sich darauf, in der Breite der Gesellschaft ein KI-Verständnis zu stärken, das technischen Laien erlaubt, Möglichkeiten und Risiken von KI-Systemen besser einschätzen können, um ihre Handlungs- und Entscheidungskompetenzen auch zukünftig einsetzen zu können. Auf Unternehmensseite stellt sich auch die Frage, wie der mitarbeiterfreundliche Einstieg in betriebliche KI-Anwendungen gelingen kann und welche flankierenden Forschungsansätze hierzu beitragen können. Diese sind insbesondere auf die Analyse und Abschätzung kritischer *Vertrauens- und Akzeptanzbedingungen* hin zu entwickeln.

Die vorstehenden Überlegungen sind aber auch auf konkrete Problemkontexte hin zu beziehen und für mögliche Lösungs-

die Gewalt über unser Schicksal erlangen. Unsere Sterblichkeit wird in unseren Händen liegen. Wir werden so langen leben können wie wir wollen [...]. Bis zum Ende des Jahrhunderts wird die nichtbiologische Komponente unserer Intelligenz Trillionen Mal mächtiger sein als bloße menschliche Intelligenz“.

Die meisten professionellen Einschätzungen von KI sind demgegenüber deutlich nüchterner. Uns Gastherausgebern ist Stand heute keinerlei Evidenz bekannt, dass sogenannte ‚starke KI‘, also KI-Anwendungen, die sich, wie oben prognostiziert,

¹ Dieser Abschnitt ist in weiten Teilen sinngemäß der Studie von Gethmann et al. (2021) entnommen, an der Bernhard Humm, Stephan Lingner und Jan Schmidt maßgeblich mitgewirkt haben.

eigenständig weiterentwickeln, wirklich möglich sind. In jedem Fall sind diese aus unserer Sicht nicht in greifbarer Nähe. Besser werden lediglich die Imitationen und Inszenierungen von Intelligenz. Auch der hohe Anspruch, dereinst KI mit Bewusstsein zu erschaffen, geht stark mit der Reduzierung des Bewusstseinsbegriffs auf Funktionalitäten einher, nicht aber mit etwaigen Erlebnisqualitäten. Alle KI-Anwendungen, die bis heute entwickelt wurden, werden der sogenannten ‚schwachen KI‘ zugeordnet. Sie sind auf spezielle Aufgaben zugeschnitten, deren Lösung bislang menschliche Fähigkeiten erforderten, und lösen diese selbstständig.

Klar ist, dass KI bereits heute eine hohe Bedeutung in Wirtschaft und Gesellschaft hat und diese Bedeutung voraussichtlich weiter zunehmen wird. Für viele Unternehmen hat KI eine strategische Bedeutung. So stammt beispielsweise folgende Aussage von Amazon: „Without ML [machine learning], Amazon.com couldn't grow its business, improve its customer experience and selection, and optimize its logistic speed and quality.“ (Marr 2018). Ähnlich äußern sich Google, Facebook, IBM und andere Technologie-Konzerne (Marr 2018). Auch die deutsche Bundesregierung hat 2018 eine KI-Strategie (BMBF 2021) verabschiedet mit dem Ziel, die Erforschung, Entwicklung und Anwendung von künstlicher Intelligenz in Deutschland auf ein weltweit führendes Niveau zu bringen und zu halten.

Perspektiven von KI

KI steht in der Tradition von Automatisierungs-Technologien, bei der menschliche Arbeit von Maschinen übernommen wird. Besonders seit der industriellen Revolution verändert diese Automatisierung in rasanter Weise alle Lebens- und Arbeitsbereiche. Viele Berufsgruppen sind komplett verschwunden, neue sind entstanden, und Menschen sowie Gesellschaften mussten sich stets an veränderte Gegebenheiten anpassen, die durch technische Entwicklungen verursacht wurden. Während im Zuge der industriellen Revolution vorwiegend physische Tätigkeiten von Maschinen übernommen wurden, werden im Zuge der Digitalisierung, insbesondere mit KI-Techniken, zunehmend auch kognitive Tätigkeiten automatisiert. Damit verbunden sind Chancen und Risiken, die oft eng beieinander liegen, sich sogar gegenseitig bedingen können. Wir möchten dies anhand von Anwendungsbeispielen erläutern.

KI wird zunehmend in der *Medizin* eingesetzt; so werden beispielsweise bildgebende Verfahren mittels Röntgenstrahlung, Ultraschall und MRT in Kombination mit KI-Verfahren zur (semi-)automatischen Analyse beziehungsweise Diagnose eingesetzt. Diese übertreffen in speziellen Gebieten teilweise sogar menschliche Ärzt*innen. Die Chancen hier liegen in der verbesserten Diagnostik, einer höheren Reichweite (z. B. bis hin zu Hausarztpraxen) und reduzierten Kosten im Gesundheitswesen. Demgegenüber steht das Risiko von Fehldiagnosen z. B. aufgrund von Verzerrungen (bias) in den Trainingsdaten für machine learning-Anwendungen. Ironischerweise kann dies eine

indirekte Folge der Chance verbesserter Diagnostik sein, wenn über einen längeren Zeitraum der erfolgreichen Verwendung einer KI-Anwendung sich schleichend ein blindes Vertrauen einstellt, insbesondere bei einer neuen Generation von Ärzt*innen. Dann kann schrittweise die Expertise zur menschlichen Diagnose, welche der maschinellen Diagnose kritisch gegenübergestellt wird, sinken.

Als nächstes Beispiel betrachten wir das sogenannte *autonome Fahren*, derzeit noch in der Experimentierphase, aber mit Ausblick auf breiten Einsatz in vielen Ländern. Die Chancen bestehen in einer höheren Effizienz und damit verbunden geringeren Kosten, z. B. durch Wegfall von Taxi-Fahrer*innen. Man kann auch hoffen, dass mit der Automatisierung car sharing-Konzepte umgesetzt werden können, sodass sich perspektivisch die Anzahl der Fahrzeuge, somit auch die damit verbundene Umweltbelastung sowie die Verkehrsverdichtung der Städte reduziert. Mit fortschreitender Technik könnten autonome Fahrzeuge im Durchschnitt sicherer fahren als menschliche Fahrer*innen. Allerdings bleibt auch bei immer besserer Technik grundsätzlich ein Risiko von Unfällen, verursacht durch ‚autonome‘ Fahrzeuge. Hier bestehen noch große ungeklärte juristische Fragen, z. B. nach der Verantwortung und Haftung. Die Sicherheit kann aber auch drastisch sinken, da mit durchgängigem autonomem Fahren das Risiko großflächiger, verheerender Cyberangriffe im Verkehr steigt. Auch die Hoffnung auf eine Reduzierung der Umweltbelastung könnte sich in das Gegenteil umkehren, wenn die Möglichkeit von einfach verfügbarem Individualtransport mit ‚autonomen‘ Fahrzeugen den Bedarf nach Mobilität ständig wachsen lässt – ein sogenannter Rebound-Effekt.

So genannte *Business Intelligence Systeme* unterstützen das Management von Unternehmen, Geschäftsentscheidungen zu treffen. Häufig werden dabei KI-Techniken eingesetzt, z. B. die Sentiment-Analyse von social media-Kanälen. Chancen sind das frühzeitige Erkennen, bessere Verständnis und schnelle Reagieren auf Kundenmeinungen – sowohl Wünsche als auch Kritik. Aber darin liegen auch Risiken. Werden Kunden wirklich oder nur vermeintlich besser verstanden? Es besteht die Gefahr einer Schein-Objektivität beziehungsweise Zahlengläubigkeit, die echter Einsicht entgegensteht. Insbesondere sind solche Analysen nur bedingt für strategische Zukunftsplanung geeignet, da sie nur die Vergangenheit abbilden können. Außerdem wächst mit der Bedeutung von solchen automatisierten Analysen die Gefahr der absichtlichen Manipulation, z. B. durch fake news, welche von Computerprogrammen (bots) generiert werden.

Beim *Hochfrequenzhandel* werden umfangreiche Börsentransaktionen automatisch mittels KI-Anwendungen durchgeführt. Chancen bestehen für einzelne Unternehmen hier auf zusätzliche Gewinne durch Ausnutzen von kurzlebigen Kurschwankungen. Für den Markt und damit die gesamte Wirtschaft bestehen aber auch erhebliche Risiken durch nicht kalkulierbare Rückkopplungseffekte. So können einzelne Fehlentscheidungen von vielen anderen Systemen wiederholt werden; ein gegenseitiges Aufschaukeln kann nicht vorhersehbare Folgen mit sich bringen.

Robotik wird intensiv in der industriellen Fertigung eingesetzt. Die Chancen liegen sowohl in einer gesteigerten Effizienz als auch in einer höheren Präzision. Aber grundsätzlich bedingt eine gesteigerte Effizienz das Risiko einer geringeren Stabilität. So birgt hocheffiziente just in time-Produktion mit minimierter Lagerhaltung die Gefahr eines Zusammenbruchs der Lieferketten, wie es bei der Corona-Pandemie der Fall war. Auch hier gilt, dass komplette Vernetzung – die Grundlage für die hohe Automatisierung – gleichzeitig Einfallstore für umfassende Cyberangriffe bietet.

Überwachungssysteme setzen zunehmend KI-Methoden ein, z. B. Systeme zur Gesichtserkennung an Flughäfen. Chancen solcher Systeme sind erhöhte Sicherheit, auch bei der Detek-

gen zurück, die in frühen Dokumenten angelegt wurden. Heute wirkmächtig und analytisch aufschlussreich sind also technische Zukünfte der Vergangenheit. Heutige Darstellungsformen von KI prägen dementsprechend auch zukünftige Wirkungen und Entwicklungen.

Die zunehmende Verbreitung von Sprachassistenten, Chatbots und anderen sprachverarbeitenden Programmen nimmt Sascha Dickel zum Anlass, kommunizierende Technik kritisch zu beleuchten – vor allem im Hinblick auf menschliche Deutungen der ihr zugeschriebenen Autonomie, Kreativität und (In-)Transparenz. Mithin stellt sich der Bedarf nach Identifizierbarkeit des Menschen besonders in hybriden Kommunikationsensembles moderner mediatisierter Gesellschaften, in denen maßge-

Komplette Vernetzung bietet Einfallstore für Cyberangriffe.

tion von potenziellen Straftätern, die weltweit agieren. Damit verbunden ist aber auch das Risiko der ‚falschen Sicherheit‘, da solche Erkennungssysteme nicht perfekt sind und auf vielfältige Weise manipuliert werden können.

Schließlich kann KI vielfältig für *militärische Anwendungen* eingesetzt werden, z. B. in Form autonomer Gefechtsdrohnen. Die Chance besteht darin, dass Länder sich besser verteidigen und ihre eigenen Soldaten besser schützen können. Es gibt aber auch enorme Risiken. Zum einen wird das Töten von Menschen vereinfacht, was durch die Verbreitung jener autonomer Waffen auf alle Staaten zurückfallen kann. Aus ethischer Sicht besonders zu kritisieren ist, die Entscheidung über Leben und Tod Maschinen zu überlassen. Die Gefahr besteht, dass technische Möglichkeiten neue Realitäten hervorbringen; fangen die einen an, so müssen die anderen nachziehen.

Beiträge zum Schwerpunktthema

Zukunftsvorstellungen aus der Vergangenheit gehen Christian Vater und Eckhard Geitz auf den Grund und analysieren in ihrem Beitrag, wie in früheren Berichten der Enquete-Kommissionen des Deutschen Bundestags über künstliche Intelligenz und Digitalisierung gedacht wurde, welche Schwerpunkte gesetzt wurden und wie diese ausgestaltet wurden. Heutige Debatten zu Technik- und KI-Zukünften binden sie an Technikvergangenheiten an, stellen Querbezüge über mehrere Jahrzehnte her und arbeiten Bedeutungen früher Formatierungen heutiger Konzeptverständnisse heraus. Wie KI heute angelegt und ausgerichtet wird, ist oft in solchen Kartierungen und Präsentationsformen der Enquete-Kommissionen begründet, die vor vielen Jahren bereits Vorschläge zum Umgang mit künstlicher Intelligenz im politischen Raum gemacht haben. Die Ausprägung heutiger Debatten und heutige Entscheidungen gehen somit mitunter auf Prägun-

gschneiderte KI-Systeme bis dato genuin menschliche Rollen einnehmen. Der regelmäßige Umgang mit entsprechenden künstlichen Agenten lässt den Schluss zu, dass die Anwender*innen (und Gestalter*innen) diesen Artefakten – zumindest implizit – kontextuelle Sinnverständnisse zuschreiben. In die Zukunft gedacht, verschwimmt die bislang kulturell bedeutsame Grenze zwischen dem Menschlichen und dem Künstlichen in der allgemeinen Wahrnehmung oder sie wird durch eskalierende Turing-Tests immer wieder nachzuschärfen sein. Der Autor schlägt daher zur Klärung dessen, was wir von ‚sprechender Technik‘ erwarten dürfen, einen hermeneutischen Ansatz des ‚Verstehens des Verstehens von Technik‘ vor, der die Gestaltung, Wahrnehmung und Praxis künstlicher Kommunikation systematisch untersuchen soll.

Bernd Beckert behandelt das Thema vertrauenswürdiger KI. Es gibt bereits eine ganze Reihe praktischer Leitfäden für die Implementation vertrauenswürdiger KI-Anwendungen. Bei der Betrachtung von Umsetzungsprojekten zeigt sich jedoch durchweg ein noch geringer Konkretisierungsgrad. Was können die Gründe für das Umsetzungsdefizit sein? Bernd Beckert führt drei Gründe an: time to market-Überlegungen seitens der Unternehmen, Unklarheit darüber, welche Aspekte des Konzepts der vertrauenswürdigen KI bei welchen Anwendungen überhaupt relevant sind sowie die Tatsache, dass die Umsetzung von KI-Projekten komplexer ist als die Umsetzung ‚normaler‘ Software-Projekte und dass deshalb spezifische Vorkehrungen notwendig sind.

Stefan Strauß widmet sich der kritischen KI-Kompetenz für die konstruktive Nutzung von KI. Ziel seines Beitrags ist es, Problembewusstsein und kritische KI-Kompetenz zu stärken. Gesellschaftlicher Diskurs sowie Forschung und Entwicklung zu den Risiken künstlicher Intelligenz sind aus seiner Sicht oft einseitig: Entweder mit Fokus auf wichtigen, aber praxisfernen ethischen Aspekten, oder auf technokratischen Ansätzen, um gesellschaftliche Probleme nur mit Technologie zu lösen.

Es braucht jedoch auch praktikable, problemorientierte Perspektiven zur Analyse KI-bedingter Risiken. Im Kern sind Diskrepanzen zwischen Systemverhalten und Nutzungspraktiken in KI-automatisierten Anwendungskontexten zentrale Auslöser für Bias und gesellschaftliche Risiken. Stefan Strauß formuliert in dem Zusammenhang ein zentrales Meta-Risiko von KI-Systemen: Verzerrungen (automation bias). Dafür stellt er einen analytischen Rahmen vor mit Indikatoren zur Erkennung von Verzerrungen in KI-Anwendungen.

Eine besondere Aufgabe in der Gestaltung von IT-Systemen kommt jenen Führungskräften zu, die KI-Implementierungen und Einführungen verantworten und aktiv steuern. Ihr Einsatz in der Vermittlung von Zielklarheit und Erwartbarem inklusive der erwarteten Veränderungen sind nach Marie Jung und Jörg von Garrel wesentliche Einflussfaktoren für die Vertrauensbildung und die Unterstützung von Handlungsakzeptanz bei den Mitarbeitenden. Eine in diesem Sinne personalfreundliche Implementierung von KI-Systemen ist der Schwerpunkt ihres Beitrags, in dem sie anhand eines in ihrer Forschung entwickelten Modells aufzeigen, wie Vertrauensbildung und Akzeptanz bereits in frühen Entwicklungs- und Einführungsphasen angelegt und gefördert werden und somit zu einer erhöhten Handlungsakzeptanz beitragen können.

Angesichts des komplexen und unübersichtlichen Rechtsrahmens für Datenschutz, Dateneigentum und Urheberrecht fragt sich Thomas Wilmer, ob rechtssichere Rahmenbedingungen für neue KI-Entwicklungen weiterhin durch klassische, politisch veranlasste Regulierungsansätze formuliert werden sollen oder ob hier niederschwellige Verfahrensweisen womöglich besser geeignet sind. Dabei tangieren die spezifischen Datenbedarfe und -angebote vieler anwendungsorientierter KI-Entwicklungen direkt oder zumindest indirekt auch persönliche Interessen Einzelner und sind insofern von rechtlicher Relevanz. Dies betrifft insbesondere die Bereiche der Arbeitswelt und der medizinischen Forschung und Versorgung. In diesen sensiblen Bereichen sind Transparenzgebote und Grenzen für den angemessenen Einsatz von KI-Systemen durch geeignete Regulierungen zu formulieren. Entsprechende Maßnahmen sollen der Datensouveränität der Betroffenen und der Rechtssicherheit der Systementwickler und -betreiber dienen. Diese Aufgabe ist angesichts der unterschiedlichen Geschwindigkeiten heutiger technischer und legislativer Entwicklungen und der Heterogenität der Einsatzgebiete von KI eine große Herausforderung für die Verantwortlichen. Der Autor schlägt daher bestimmte vertragliche bzw. Control by Design-Lösungen zur Schließung von Regelungslücken im Bereich der KI vor: Eine pragmatische Lösung für zulässige Datenauswertungen und -übertragungen durch KI-Systeme wäre die Formulierung privatrechtlicher Klauseln in entsprechenden Vereinbarungen mit den Beteiligten (z. B. in Form von AGBs), die einen Rahmen für den Verkehr sowohl personenbezogener als auch anonymer Daten setzen. Diese Lösung könnte gegebenenfalls auch eine (finanzielle) Anreizregelung für die Betroffenen beinhalten. Ein ergänzender, vorgelagerter Ansatz greift die Idee des Privacy by Design der Europäischen Datenschutz-

Grundverordnung auf. Im Unterschied dazu soll der hier vorgeschlagene Control by Design-Ansatz die Souveränität auch über anonymisierte Daten bis hin zu entsprechenden Haftungsfragen regeln. Konkret könnte dies z. B. durch transparente und anerkannte Instrumente der informierten Einwilligung (Cookies etc.) unterstützt werden.

KI hat ein erhebliches Nutzungspotenzial in Forschung und Entwicklung – insbesondere auch in den Lebenswissenschaften und in der Biomedizin. Entsprechende Anwendungsbereiche zeigen sich in Medikamentenentwicklung, klinischer Praxis und vorgelagerter humangenetischer Forschung. Gleichwohl stehen diesen Potenzialen auch zahlreiche Herausforderungen entgegen, die besondere Handlungsnotwendigkeiten nach sich ziehen. Der Frage, wie die Entwicklung von KI mit der sich ebenfalls rasch entwickelnden Humangenomik konvergiert und welche praktische Bedeutung diese Konvergenz für die biomedizinische Praxis hat, haben sich Reinhard Heil und sechs weitere Ko-Autor*innen gewidmet. Auf Basis einer umfangreichen Literaturanalyse untersucht das Team, inwieweit der korrelative Zugang von KI zu neuem biomedizinischem Wissen dem Bedarf an begründbarem Wissen z. B. für die Diagnose und Therapie von Krankheiten gerecht wird und welche Fragen sich hieraus für Handlungsebene ergeben. Die Autor*innen heben dabei u. a. folgende ungelöste Probleme einer KI-unterstützten Forschung hervor: fragliche Qualität des Wissens und dessen Interpretation sowie dessen Erklärbarkeit, Evaluierung und Anerkennung. Auf anderer Ebene kommen noch ungeklärte Folgen von KI-assistierter Forschung für Wissenschaft, Wirtschaft und Gesellschaft dazu – auch hinsichtlich ethischer Herausforderungen sowie Fragen der adäquaten Gestaltung entsprechender Forschungs- und Dateninfrastrukturen und damit verknüpfter Regulierungsnotwendigkeiten (s. o.). Angesichts dieser Gemengelage empfehlen Heil et al. vertiefende Analysen, breite gesellschaftliche Dialoge zu kritischen Einzelproblemen sowie explorative Politikansätze, um einschlägige Forschungsprogramme und Regulierungen in allgemein wünschbarer Weise zu verbessern.

KI-Anwendungen haben bereits heute große Auswirkungen auf Gesellschaft, Wirtschaft und Wissenschaft und dies wird noch zunehmen. KI-Anwendungen richtig einzuschätzen und verantwortlich zu gestalten ist eine wichtige Aufgabe. Dieses Themenheft soll dazu einen Beitrag leisten.

Literatur

- Allen, Frederick (2001): The myth of artificial intelligence. In: American Heritage 52 (1). Online verfügbar unter <https://www.americanheritage.com/myth-artificial-intelligence>, zuletzt geprüft am 04. 11. 2021.
- BMBF – Bundesministerium für Bildung und Forschung (2021): Nationale Strategie für Künstliche Intelligenz. Online verfügbar unter <https://www.ki-strategie-deutschland.de>, zuletzt geprüft am 04. 11. 2021.
- Collingridge, David (1982): The Social Control of Technology. London: Pinter.
- Gethmann, Carl Friedrich et al. (2021): Künstliche Intelligenz in der Forschung. Neue Möglichkeiten und Herausforderungen für die Wissenschaft. Cham: Springer Nature (in Druck).
- Kurzweil, Ray (2015): Menschheit 2.0. Die Singularität naht. Berlin: Lola Books.

Marr, Bernard (2018): The key definitions of artificial intelligence (AI) that explain its importance. In: Forbes, 04.02.2018. Online verfügbar unter <https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#54d0f59a4f5d>, zuletzt geprüft am 04.11.2021.



PROF. DR. BERNHARD G. HUMM

ist Informatik-Professor an der Hochschule Darmstadt und geschäftsführender Direktor des Instituts für Angewandte Informatik Darmstadt (aiDa). Sein Forschungsschwerpunkt ist Angewandte KI. Er führt nationale und internationale KI-Forschungsprojekte im Industrie- und Hochschulumfeld durch und publiziert regelmäßig.



PROF. DR. JAN C. SCHMIDT

ist promovierter Physiker und habilitierter Philosoph. Seit 2008 ist er Professor für Wissenschafts- und Technikphilosophie an der Hochschule Darmstadt. Schmidt ist Mitglied verschiedener Beiräte und Kuratorien, etwa dem Transdisziplinaritätsbeirat der Schweizerischen Akademie der Wissenschaften oder dem Konvent der ev. Akademie Frankfurt.



DR. STEPHAN LINGNER

verantwortet seit vielen Jahren den Forschungsbereich Technology Assessment am Institut für qualifizierende Innovationsforschung und -beratung GmbH (IQIB) in Bad Neuenahr-Ahrweiler. Er ist zudem Mitglied im Koordinations-Team des Netzwerks TA (NTA) sowie im Wissenschaftlichen Beirat der Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis (TATuP).



PROF. DR. KARSTEN WENDLAND

ist Informatiker, Humanwissenschaftler und Technikfolgenabschätzer. An verschiedenen Universitäten und Hochschulen lehrt er in den Grenz- und Überlappgebieten von Technik- und Humanwissenschaften und forscht zu Sonderthemen im Feld Digitale Technologien und gesellschaftlicher Wandel am Institut für Technikfolgenabschätzung und Systemanalyse (ITAS).



Liebe Leserinnen und Leser,

danke, dass Sie auch im Jahr 2021 *TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis* mit Interesse begleitet haben. Im neuen Jahr blicken wir voraus auf Hefte mit den TATuP-Themen Technikfolgenabschätzung und Lehre, Energiesuffizienz sowie nukleare Endlager, dazu wie immer Forschungsbeiträge, Interviews, Buchrezensionen, Konferenzberichte und vieles mehr.

Mit diesem Ausblick wünschen Ihnen die TATuP-Redaktion und der oekom verlag ein gesundes Jahr 2022!

Dr. Ulrich Ufer,
Redaktionsleitung, ITAS/KIT

Dr. Ulrike Sehy,
Leitung Fachzeitschriften, oekom verlag

Foto: © s. mccutcheon / unsplash.com

RESEARCH ARTICLE

Vertrauenswürdige künstliche Intelligenz

Ausgewählte Praxisprojekte und Gründe für das Umsetzungsdefizit

Bernd Beckert, Fraunhofer-Institut für System- und Innovationsforschung ISI, Breslauer Straße 48, 76139 Karlsruhe, DE
(bernd.beckert@isi.fraunhofer.de)  0000-0003-0157-9096

17

Zusammenfassung • Während es inzwischen eine ganze Reihe praktischer Leitfäden für die Implementierung des Konzepts der vertrauenswürdigen künstlichen Intelligenz (KI) gibt, fehlt es an konkreten Beispielen und Projekten für Umsetzungen, anhand derer sich Probleme und Erfolgsstrategien der Akteur:innen vor Ort aufzeigen ließen. Dieser Beitrag stellt ausgewählte Umsetzungsprojekte vor. Durchweg zeigt sich dabei ein noch geringer Grad an Konkretisierung. Deshalb wird anschließend nach den Gründen für das Umsetzungsdefizit gefragt. Drei Erklärungen kommen infrage: Time-to-Market-Überlegungen aufseiten der Unternehmen, Unklarheit darüber, welche Aspekte des Konzepts der vertrauenswürdigen KI bei welchen Anwendungen überhaupt relevant sind sowie die Tatsache, dass die Umsetzung von KI-Projekten komplexer ist als die Umsetzung von 'normaler' Software-Projekten und deshalb spezifische Vorkehrungen notwendig sind.

Trustworthy artificial intelligence. *Selected practical projects and reasons for the implementation deficit*

Abstract • While there are now a number of practical guides for implementing the concept of trustworthy artificial intelligence (AI), there is a lack of concrete examples and projects for implementations that could be used to highlight problems and success strategies of actors in the field. This paper presents selected implementation projects showing that the degree of concretization is still low throughout. Therefore, the reasons for the implementation deficit are then explored. There are three possible explanations: time-to-market considerations on the part of the companies, lack of clarity about which aspects of the concept of trustworthy AI are relevant at all for which applications, and the fact that the implementation of AI projects is more complex than the implementation of 'normal' software projects and thus requires specific arrangements.

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.17>
Received: Jun. 13, 2021; revised version accepted: Oct. 20, 2021;
published online: Dec. 20, 2021 (peer review)

Keywords • *trustworthy AI, human centered AI, best practices, implementation gap, AI regulation*

Das Konzept der vertrauenswürdigen künstlichen Intelligenz

Unter vertrauenswürdiger künstlicher Intelligenz versteht man automatisierte Erkennungs-, Vorschlags- und Entscheidungssysteme, die den Anforderungen von Transparenz, Verantwortlichkeit, Privatheit, Diskriminierungsfreiheit und Zuverlässigkeit gerecht werden. Hintergrund ist die Tatsache, dass auf künstlicher Intelligenz (KI) basierende Systeme gesellschaftliche Implikationen haben, die über herkömmliche Software-basierte Systeme hinausgehen. Als Gegenentwurf zur US-amerikanischen und chinesischen Konzeptualisierung von KI („Kommerz“ beziehungsweise „Kontrolle“) propagiert die Europäische Kommission die „vertrauenswürdige“ KI. Diese soll zum einem Qualitätsmerkmal von KI „made in Europe“ werden.

Das Konzept der vertrauenswürdigen KI ist dabei nicht neu, es orientiert sich an der Diskussion um ethische oder menschenzentrierte KI und nimmt Erkenntnisse aus dem Bereich der Mensch-Computer-Interaktion auf (Dragan 2019; Russell 2019; Shneiderman 2020). Das europäische Konzept wurde 2019 von der High-Level Expert Group on Artificial Intelligence der Europäischen Kommission entwickelt (HLEG 2019). Es umfasst die sieben Dimensionen:

- Vorrang menschlichen Handelns und menschlicher Kontrolle
- Technische Robustheit und Sicherheit
- Privatsphäre und Datenqualitätsmanagement
- Transparenz und Erklärbarkeit
- Vielfalt, Nichtdiskriminierung und Fairness
- Gesellschaftliches und ökologisches Wohlergehen
- Rechenschaftspflicht

Um diesen eher abstrakten Anforderungskatalog zu konkretisieren, hat die Expertengruppe 2020 eine Checkliste erarbeitet, mit der Anbieter und Nutzer überprüfen können, inwieweit das zu implementierende KI-System den Anforderungen der vertrauenswürdigen KI entspricht (HLEG 2020). ‚Vertrauenswürdige KI‘ bezeichnet dabei sowohl das Ziel als auch den Prozess, wie dieses Ziel erreicht werden kann.

In jüngster Zeit kamen weitere Vorschläge für die praktische Umsetzung und die ethische Bewertung von KI-Projekten hinzu und es wurden Arbeitsvorlagen, Checklisten und Guidelines vorgelegt. So hat zum Beispiel die AI Ethics Impact Group unter der Koordination des Verbands der Elektrotechnik Elektronik Informationstechnik e. V. und der Bertelsmann Stiftung ein KI-Ethik-

schlägigen Veröffentlichungen. Deshalb wurde eine Suchstrategie entwickelt, die im ersten Schritt Umsetzungsprojekte im akademischen Bereich (oft mit Beteiligung von Unternehmen) identifizierte. Im zweiten Schritt wurde Hinweisen auf Umsetzungen in großen Unternehmen nachgegangen und schließlich wurde die Start-up-Szene untersucht. Dazu wurden ausschließlich öffentlich zugängliche Quellen ausgewertet und keine exklusiven Zugänge in die Unternehmen hinein genutzt. Zwar wurden zusätzlich die einschlägigen Verbände gebeten, Umsetzungsbeispiele aus ihren Mitgliedsunternehmen zu benennen, jene haben von dieser Möglichkeit jedoch keinen Gebrauch gemacht.

Im akademischen Bereich, dem Startpunkt der Recherche, wurden die laufenden Forschungsprogramme des Bundes mit

Mitte 2021 befassten sich in Deutschland über 40 öffentlich geförderte Projekte mit dem Konzept vertrauenswürdiger KI.

Kennzeichnungssystem entwickelt, das sich an das Energieeffizienzlabel anlehnt und für die sechs Dimensionen Transparenz, Verantwortlichkeit, Privatheit, Gerechtigkeit, Zuverlässigkeit und ökologische Nachhaltigkeit entsprechende Güteklassen definiert (Hallersleben und Hustedt 2020).

Weitere Vorschläge zur konkreten Umsetzung wurden von der Plattform Lernende Systeme (Huchler et al. 2020) gemacht und in Forschungsprojekten wie *Ethik in der agilen Softwareentwicklung* des Bayerischen Forschungsinstituts für Digitale Transformation (Zuber et al. 2020). Weiterhin hat die Gesellschaft für Informatik (GI) ein Machine-Learning-Kennzeichnungssystem vorgelegt, das sich an der Idee des Beipackzettels von Medikamenten orientiert (Seifert et al. 2020). Für die Gestaltung des konkreten Projektablaufs und die Governancestruktur von KI-Entwicklungs- und Umsetzungsprojekten liegen ebenfalls Vorschläge vor (Shneiderman 2020; Puntschuh und Fetic 2020).

Doch welche Erfahrungen haben Unternehmen und Organisationen bisher mit den verschiedenen Ansätzen gemacht? Welche Guidelines haben sich als praxistauglich erwiesen, welche Herausforderungen gab es bei der Umsetzung? Da entsprechende Leitfäden und Governance-Empfehlungen seit einiger Zeit existieren, sollten sich erste konkrete Umsetzungsprojekte in Unternehmen und Organisationen finden lassen, die eine Auswertung von Erfahrungen im Hinblick auf Best Practices ermöglichen. Im Folgenden wird die Suche nach solchen Projekten nachgezeichnet und es werden die Ergebnisse vorgestellt.

Auf der Suche nach Beispielen für die praktische Umsetzung des Konzepts

Zunächst lässt sich feststellen, dass Beispiele für praktische Umsetzungen nicht auf der Hand liegen; es existieren keine Sammlungen oder Datenbanken und es gibt bisher auch keine ein-

Bezug zu KI, die Aktivitäten ausgewählter Bundesländer sowie von Universitäten und Forschungsinstituten analysiert, um relevante Projekte zu identifizieren. Die Suchstrategie bestand darin, zunächst solche KI-Projekte zu identifizieren, in denen Aspekte der vertrauenswürdigen KI überhaupt eine Rolle spielen und anschließend auf Projekte zu fokussieren, die sich mit der konkreten Umsetzung von Richtlinien beschäftigen.

Ergebnis der Recherche ist, dass es Mitte 2021 in Deutschland über 40 öffentlich geförderte Projekte gibt, die sich mit der Umsetzung verschiedener Teilaspekte des Konzepts der vertrauenswürdigen KI, wie zum Beispiel der Transparenz und Erklärbarkeit befassen. In diesen Projekten wurden zum Teil auch Leitfäden entwickelt und adaptiert und in einigen Fällen in Kooperation mit Pilotunternehmen evaluiert. Allerdings stellte sich heraus, dass sich diese Projekte in einem frühen Stadium befinden und dass die bisher vorliegenden Ergebnisse noch keine systematische Zusammenschau von Strategien und Erfahrungen erlauben. Drei Projekte sollen diesen Befund illustrieren (Tabelle 1).

In einem zweiten Schritt wandte sich die Recherche dem Unternehmensbereich zu. Hier konnten mehrere große Firmen identifiziert werden, die sich mit ethischen Aspekten der von ihnen genutzten oder bereitgestellten KI-Systeme befassen. Tabelle 2 führt drei Beispiele auf. Dabei zeigt sich, dass die Beschäftigung dieser Unternehmen mit diesem Thema im Kontext ihrer jeweiligen Compliance- und Nachhaltigkeitsaktivitäten gesehen werden muss. Beispiele, wie konkrete Umsetzungen des Konzepts der vertrauenswürdigen KI in den Unternehmen vorgenommen wurden, konnten dagegen nicht recherchiert werden. Dies könnte neben den Gründen, die anschließend diskutiert werden, auch damit zusammenhängen, dass Unternehmen Details ihrer Softwareentwicklung nicht veröffentlichen, oder dass sie dies zum Schutz ihrer Kunden und Produkte bewusst vermeiden (Machmeier 2020).

Projekt	Ziele und Ergebnisse
ExamAI – KI Testing & Auditing	Entwicklung von Kontroll- und Prüfverfahren, mit denen KI-Systeme in industriellen Produktionssettings und im Personalmanagement sicher, transparent und diskriminierungsfrei gestaltet werden können sollen. Die Verfahren werden in Anwendungsszenarien getestet. Die Ergebnisse sollen es den Forschern dann ermöglichen, Empfehlungen für reale Implementierungen zu geben.
KI-Gestaltungsansätze und Ethik-Briefing der Plattform Lernende Systeme	Entwicklung umsetzungsorientierter Richtlinien und Change-Management-Empfehlungen für die Realisierung verantwortungsvoller KI-Systeme. Es werden einige Beispiele dafür angeführt, wie ausgewählte Aspekte in Unternehmen praktisch gehandhabt wurden. Die vorgeschlagenen Gestaltungsansätze sollen in Zukunft mit Leben gefüllt werden.
GOAL – Governance von und durch Algorithmen	Identifizierung von Governance-Strukturen und regulatorischem Handlungsbedarf für den verantwortungsvollen Einsatz von KI-Systemen. Ein Teilprojekt, das an der Westfälischen Wilhelms-Universität Münster durchgeführt wird, umfasst neben konzeptionellen Arbeiten, auch konkrete Tests und evaluiert die Erkenntnisse in einer Fallstudie mit einer selbst entwickelten mobilen App.

Tab. 1: Umsetzungsbeispiele aus dem Bereich der Forschung.

Quelle: eigene Darstellung

Unternehmen	Aktivitäten mit Bezug auf vertrauenswürdige KI
Deutsche Telekom	Entwicklung eines internen AI-Code-of-Conduct. Für die Umsetzung wurde eine Prüfmatrix zur ethischen Bewertung neuer KI-Produkte und ein internes Gütesiegel entwickelt. Das Unternehmen hat angekündigt, mit Wirtschaftsprüfern an einer Zertifizierung zu arbeiten, mit der der ethische und vertrauenswürdige Einsatz von KI im Unternehmenskontext nachgewiesen werden soll (Mackert 2020).
SAP	Veröffentlichung eines Handbuchs zur KI-Ethik, das auf der Arbeit eines internen KI-Ethik-Lenkungsausschusses und eines externen KI-Ethik-Beirats basiert (Heesen et al. 2020, S. 26). Hinsichtlich konkreter Erfahrungen bei Implementierungsprojekten ist das Unternehmen zurückhaltend: „Aus Rücksicht auf die beteiligten Kunden und Kollegen ist es schwierig, über konkrete Szenarien zu sprechen“ (Markus Noga, zitiert in Machmeier 2020).
BMW	Ankündigung einer generellen Verankerung der sieben EU-Anforderungen für vertrauenswürdige KI im Unternehmen. Im Hinblick auf konkrete Ansätze oder Erfahrungen mit der Umsetzung des Konzepts konnten keine Berichte gefunden werden.

Tab. 2: Aktivitäten großer Unternehmen mit Bezug zur Umsetzung vertrauenswürdiger KI.

Quelle: eigene Darstellung

Start-ups	Aktivitäten mit Bezug auf vertrauenswürdige KI
Datanizing	Die Algorithmen der Firma Datanizing segmentieren Konsumenten auf Basis von Daten aus sozialen Netzwerken und stellen das Ergebnis Marktforschungsunternehmen zur Verfügung. Die Gründer wollen ihren Kunden ein besseres Verständnis für die Potenziale und Grenzen der verwendeten Algorithmen ermöglichen und haben einen Leitfaden entwickelt, der dabei helfen soll, die soziale Dimension zu berücksichtigen (Heesen et al. 2020, S. 25).
Prenode	Das Start-up Prenode hat eine Software entwickelt, die es ermöglicht, Daten aus verschiedenen Quellen zu verarbeiten, ohne sie in einer zentralen Datenbank zusammenführen zu müssen. Die Zusammenführung ist aufgrund von Datenschutzbedenken oft problematisch (techtage 2020). Die gefundene Lösung ist insbesondere für den Gesundheitsbereich interessant, in dem Patientendaten aus verschiedenen Datenbanken verarbeitet werden müssen.

Tab. 3: Umsetzungsbeispiele von Start-ups.

Quelle: eigene Darstellung

Im dritten Rechterschritt wurde nach Umsetzungsbeispielen bei Start-ups und mittelständischen Unternehmen, insbesondere in der IT- und Softwarebranche gesucht. Doch dieser Rechterschritt erwies sich als am wenigsten produktiv. Start-ups und mittelständische Unternehmen scheinen derzeit abzuwarten, wie sich das Thema entwickelt. Dennoch sollen hier zwei Beispiele angeführt werden, die zeigen, dass bestimmte Aspekte

des Themas auch in der Start-up-Szene eine Rolle spielen (Tabelle 3).

Insgesamt zeigt die Recherche, dass es trotz einiger Ausnahmen eine große Lücke zwischen den konzeptionellen Angeboten und der praktischen Umsetzung gibt. Das Umsetzungsdefizit ist dabei nicht auf Deutschland beschränkt. Im The AI Index 2021 Annual Report der Stanford University schreiben die

Autoren, dass sie überrascht waren, wie wenig sie zum Thema Umsetzung vertrauenswürdiger KI gefunden haben: „Though a number of groups are producing a range of qualitative or normative outputs in the AI ethics domain, the field generally lacks benchmarks“ (Zhang et al. 2021, S. 127). Es ist daher naheliegend zu fragen, wodurch diese Lücke, beziehungsweise der Mangel an Umsetzungsprojekten zustande kommt. Im folgenden Abschnitt werden drei mögliche Erklärungen aufgeführt.

Gründe für das Umsetzungsdefizit

Die folgenden Erklärungen basieren neben der Analyse einschlägiger Literatur auf der Auswertung von Interviews, die im April und Mai 2021 mit zehn Expert:innen durchgeführt wurden.

Der erste Grund für das Umsetzungsdefizit hat mit Time-to-Market-Überlegungen der Unternehmen zu tun: *First Mover* (Pionierunternehmen) scheinen auf dem noch relativ jungen KI-Markt einen erheblichen Marktvorteil zu haben, auch wenn ihre KI-Anwendung bis zu einem gewissen Grad fehlerbehaftet ist. Ein Beispiel hierfür ist die Personalauswahl-KI des Münchner Startups Retorio. Obwohl die Anwendung nachweislich bestimmte ethnische Gruppen diskriminiert und anfällig für Manipulationen ist, scheint sie viele Kunden gefunden zu haben, da sie eine der ersten Anwendungen auf dem Markt war (Radü 2021).

Ein zweiter Grund ist, dass es für Entwickler und Manager oft schwierig ist zu entscheiden, inwiefern das Konzept der vertrauenswürdigen KI für ihre spezifische Anwendung überhaupt relevant ist. Ein großer Teil der aktuellen KI-Anwendungen – so scheint es – wird in einem Fertigungskontext (vorausschauende Wartung, Maschinen- und Logistiko Optimierung, neue Materialien) oder in einem Forschungskontext (Modelloptimierung, Visualisierung) eingesetzt. Robustheit und Sicherheit sind hier wichtige Aspekte. Vielfalt, Nichtdiskriminierung und Fairness spielen aber nur zum Teil eine Rolle. Auf der anderen Seite gibt es KI-Systeme, die als entscheidungsunterstützende Systeme bei Kredit- oder Versicherungsunternehmen eingesetzt werden,

vertrauenswürdigen KI prioritär berücksichtigt werden müssen und welche eine weniger wichtige Rolle spielen.

Der dritte Grund bezieht sich auf die Tatsache, dass KI-Projekte nicht wie ‚normale‘ Software-Projekte umgesetzt werden können. Die Fähigkeit von KI-Systemen, ihre Outputs kontinuierlich an veränderte Inputs anzupassen („Lernen“), macht gerade ihren Reiz aus (Heesen 2021). Diese Anpassungsfähigkeit unterscheidet KI von anderen Softwareprogrammen, die zwar zum Teil auch an unterschiedliche Kontexte angepasst werden können, deren Verarbeitungsgrundlagen sich aber nicht grundsätzlich durch neu eingespeiste Daten verändern. Im Vergleich zur Standard-Softwareentwicklung erfordert die Entwicklung und Implementierung vertrauenswürdiger KI-Anwendungen zusätzliche Vorkehrungen: Auf der Ebene der Softwareentwicklung beinhaltet dies die Einführung spezifischer Prüfungsprozesse und Bias-Tests sowie die Integration von Erklärungen und auf der Management-Ebene die Einführung von internen Überprüfungen, kontinuierlichen Fehleranalysen sowie speziellen Strukturen der Aufsicht und Kontrolle (Shneiderman 2020, S. 12). Derartige, neue Arbeitsabläufe und Prüfprozesse zu implementieren ist eine Herausforderung für Firmen und Organisationen, denn sie erfordern entsprechendes Know-how, spezifische Planung und zusätzliche Ressourcen.

Fazit und Ausblick: Wege aus dem Umsetzungsdefizit

Welche Möglichkeiten gibt es nun, das Umsetzungsdefizit perspektivisch zu verringern und mehr Unternehmen und Organisationen dazu zu motivieren, das Konzept der vertrauenswürdigen KI auch anzuwenden? Basierend auf den Erkenntnissen des vorangegangenen Abschnitts werden im Folgenden zwei mögliche Ansätze vorgeschlagen. Der erste bezieht sich auf die weitere Konkretisierung von Leitfäden und die Einführung von Zertifikaten, der zweite auf die Einrichtung gemischter Teams als Teil einer spezifischen KI-Governance-Struktur.

First Mover scheinen auf dem noch relativ jungen KI-Markt einen erheblichen Marktvorteil zu haben.

im Personalmanagement, in öffentlichen Verwaltungen (Sozialamt, Arbeitsamt) und im Gesundheitswesen. Hier spielen ethische Fragen und Persönlichkeitsrechte des Einzelnen eine sehr wichtige Rolle. Tatsächlich ist das Anwendungsspektrum von KI generell sehr breit. KI sollte daher nicht nur im Kontext automatisierter Entscheidungssysteme gesehen werden (Köszegi 2020), sondern auch im Kontext von Muster- und Bilderkennung, von Prozessoptimierungen sowie von Empfehlungssystemen oder auch dem autonomen Fahren. Für all diese Anwendungen gilt es derzeit, jeweils spezifisch zu entscheiden, welche Aspekte der

Konkretisierung von Leitfäden und Einführung von Zertifikaten

Obwohl es heute eine Vielzahl von Leitfäden und konkreten Umsetzungsvorschlägen gibt, reicht der Grad der Konkretisierung für die komplexe Implementation von KI-Projekten vielfach nicht aus. Was fehlt, sind Überlegungen zur Art und Weise, wie Daten generiert und verarbeitet werden, wie Trainingsdatensätze ausgewählt und wie geeignete Algorithmen bestimmt werden können. Hagendorff (2020, S. 111) verweist in diesem Zusammenhang darauf, dass es bei der Konkretisierung von Leitfä-

den darauf ankommt, Ethik in ‚Mikroethik‘ zu verwandeln, d. h. die ethische Debatte auf ein handhabbares Konkretisierungs-niveau herunterzubrechen.

Ein weiterer Ansatz, um die Verbreitung vertrauenswürdiger KI zu erhöhen, ist die Einführung von Zertifikaten. Die prominentesten Zertifizierungsaktivitäten in Deutschland sind derzeit die Normungsroadmap zur künstlichen Intelligenz, erarbeitet durch den DIN e. V. und das Bundesministerium für Wirtschaft und Energie sowie die Plattform zur Zertifizierung von KI-An-

teme zu demonstrieren und Nachahmer zu motivieren. Unternehmen und Organisationen sollten ein Interesse daran haben, ihre Bemühungen um die Vertrauenswürdigkeit ihrer Anwendungen auch zu kommunizieren, da dies die Attraktivität und Akzeptanz ihrer Angebote prinzipiell erhöht.

Angabe von Finanzierungsquellen

Dieser Artikel wurde aus Mitteln der KI-Gruppe am Fraunhofer ISI (www.isi.fraunhofer.de/de/themen/ki.html) finanziert.

Zertifikate erhöhen das Bewusstsein für die Notwendigkeit, sich mit Aspekten der vertrauenswürdigen KI zu beschäftigen.

wendungen des Landes Nordrhein-Westfalen. Darüber hinaus hat der KI-Bundesverband ein KI-Gütesiegel herausgegeben, zu dem sich die Mitglieder des Verbandes in einer freiwilligen Vereinbarung verpflichten können. In allen drei Initiativen sollen künftig konkrete Anwendungsfälle in Zusammenarbeit mit Unternehmen entstehen (Schonschek 2020). Mit entsprechenden Zertifikaten können Unternehmen für ihre Anwendungen werben. Außerdem erhöhen Zertifikate das Bewusstsein in der Öffentlichkeit für die Notwendigkeit, sich mit Aspekten der vertrauenswürdigen KI zu beschäftigen.

Einrichtung gemischter Teams

Wie erwähnt, erfordert die Umsetzung des Konzepts der vertrauenswürdigen KI spezifische Abläufe und zusätzliche Prüfprozesse. Dabei kommt es darauf an, die beiden Welten der Softwareentwicklung und der Ethik zusammenzubringen. Hierfür eignen sich insbesondere gemischte Teams, denn Software-Ingenieure und Informatiker:innen überblicken in der Regel nicht die Literatur und die Empfehlungen ethischer Diskussionen, und Ethiker:innen und Sozialwissenschaftler:innen sind in der Regel nicht mit den Anforderungen der Programmierung oder organisatorischen Details vertraut. Der Ruf nach interdisziplinären Teams ist dabei nicht neu (Shneiderman 2020; Hagendorff 2020; Lopez 2021; Kusner und Loftus 2020). Allerdings gibt es hier mit Ausnahme einiger Projekte in der Gesundheitsforschung (Franke 2021) noch zu wenige Umsetzungsbeispiele.

Projekte mit gemischten Teams könnten die Umsetzungsbilanz verbessern und die Diskussion um Best Practices und Erfolgsstrategien bereichern. Entsprechende Methoden aus dem Bereich der Science Technology Studies, wie zum Beispiel die der *Embedded Social Scientists* liegen vor (Fisher und Schuurbiers 2013; Simone 2018; Gransche und Manzeschke 2020; Haux und Karafyllis 2021) und könnten auf den Bereich der vertrauenswürdigen KI übertragen werden.

Best-Practice-Darstellungen und Erfolgsbeispiele sind dringend notwendig, um die Vorteile vertrauenswürdiger KI-Sys-

Literatur

- Dragan, Anca (2019): Der unberechenbare Mensch. In: Süddeutsche Zeitung, 23. 04. 2019. Online verfügbar unter <https://www.sueddeutsche.de/kultur/kuenstliche-intelligenz-roboter-realtaet-1.4418243?reduced=true>, zuletzt geprüft am 08. 10. 2021.
- Fisher, Erik; Schuurbiers, Daan (2013): Socio-technical integration research. Collaborative inquiry at the midstream of research and development. In: Neelke Doorn et al. (Hg.): Early engagement and new technologies. Opening up the laboratory. Dordrecht: Springer, S. 97–110. https://doi.org/10.1007/978-94-007-7844-3_5
- Franke, Thomas (2021): Kooperative und kommunizierende KI-Methoden für die medizinische bildgeführte Diagnostik. Webdarstellung des Projekts CoCoAI. Online verfügbar unter <https://www.imis.uni-luebeck.de/de/forschung/projekte/cocoi>, zuletzt geprüft am 08. 10. 2021.
- Gransche, Bruno; Manzeschke, Arne (Hg.) (2020): Das geteilte Ganze. Horizonte integrierter Forschung für künftige Mensch-Technik-Verhältnisse. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-26342-3>
- Hagendorff, Thilo (2020): The ethics of AI ethics. An evaluation of guidelines. In: Minds & Machines 30, S. 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hallersleben, Sebastian; Hustedt, Carla (2020): From principle to practice. An interdisciplinary framework to operationalise AI ethics. Frankfurt a. M.: VDE & Bertelsmann Stiftung. Online verfügbar unter <https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf>, zuletzt geprüft am 08. 10. 2021.
- Haux, Reinhold; Karafyllis, Nicole (2021): Methodisch-technische Aspekte der Evaluation erweiterten Zusammenwirkens. In: Reinhold Haux, Klaus Gahl, Meike Jipp und Otto Richter (Hg.): Zusammenwirken von natürlicher und künstlicher Intelligenz. Wiesbaden: Springer VS Open Access, S. 175–198. https://doi.org/10.1007/978-3-658-30882-7_13
- Heesen, Jessica (2021): Wie kommt Ethik in die Künstliche Intelligenz? In: Digitale Welt, 06. 01. 2021. Online verfügbar unter <https://digitaleweltmagazin.de/2021/01/06/wie-kommt-ethik-in-die-kuenstliche-intelligenz/>, zuletzt geprüft am 19. 10. 2021.
- Heesen, Jessica; Grunwald, Armin; Matzner, Tobias; Roßnagel, Alexander (2020): Ethik-Briefing. Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen. München: Lernende Systeme – Die Plattform

- für Künstliche Intelligenz. Online verfügbar unter https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_EB_200831.pdf, zuletzt geprüft am 08.10.2021.
- HLEG – High-Level Expert Group on Artificial Intelligence (2019): Ethics guidelines for trustworthy AI. Brussels: European Commission. Online verfügbar unter https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, zuletzt geprüft am 19.10.2021.
- HLEG (2020): The assessment list for Trustworthy Intelligence (ALTAI) for self-assessment. Brussels: European Commission. Online verfügbar unter <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>, zuletzt geprüft am 08.10.2021.
- Huchler, Norbert et al. (2020): Kriterien für die menschengerechte Gestaltung der Mensch-Maschine-Interaktion bei Lernenden Systemen. Online verfügbar unter https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG2_Whitepaper2_220620.pdf, zuletzt geprüft am 08.10.2021.
- Köszegi, Sabine (2020): Der autonome Mensch im Zeitalter des Digitalen Wandels. In: Markus Hengstschlager (Hg.): Digitaler Wandel und Ethik. München: Ecowin Verlag, S. 62–90.
- Kusner, Matt; Loftus, Joshua (2020): The long road to fairer algorithms. Build models that identify and mitigate the causes of discrimination. In: Nature 578, S. 34–36. <https://doi.org/10.1038/d41586-020-00274-3>
- Lopez, Paola (2021): Artificial Intelligence und die normative Kraft des Faktischen. In: Merkur 75 (863), S. 42–52.
- Machmeier, Corinna (2020): Verantwortungsvoll mit Künstlicher Intelligenz umgehen. Ein Jahr des Lernens. In: SAP News Center, 09.01.2020. Online verfügbar unter <https://news.sap.com/germany/2020/01/ki-kuenstliche-intelligenz-ethik/>, zuletzt geprüft am 08.10.2021.
- Mackert, Manuela (2020): Compliance und künstliche Intelligenz. Im Blickpunkt – Warum sich Unternehmen einen Code-of-Conduct leisten sollten. In: Compliance Business 1, S. 3–5. Online verfügbar unter <https://www.deutscheranwaltspiegel.de/compliancebusiness/kuenstliche-intelligenz/compliance-und-kuenstliche-intelligenz-19597/>, zuletzt geprüft am 20.10.2021.
- Puntschuh, Michael und Fetic, Lajla (2020): Praxisleitfaden zu den Algo.Rules. Orientierungshilfen für Entwickler:innen und ihre Führungskräfte. Gütersloh: Bertelsmann Stiftung. Online verfügbar unter https://www.bertelsmannstiftung.de/fileadmin/files/alg/Algo.Rules_Praxisleitfaden.pdf, zuletzt geprüft am 19.10.2021.
- Radü, Jens (2021): Wie Künstliche Intelligenz über Ihren nächsten Job entscheidet. In: Der Spiegel, 08.03.2021, S. 68–70.
- Russell, Stuart (2019): Human compatible. Artificial Intelligence and the problem of control. New York: Viking.
- Schonschek, Oliver (2020): Künstliche Intelligenz muss Vertrauen schaffen. In: Com! professional, 20.01.2020. Online verfügbar unter <https://www.com-magazin.de/praxis/kuenstliche-intelligenz/kuenstliche-intelligenz-vertrauen-schaffen-2451438.html>, zuletzt geprüft am 19.10.2021.
- Seifert, Christin; Scherzinger, Stefanie; Wiese Lena (2020): Beipackzettel für Modelle des maschinellen Lernens für fairere KI. In: Gesellschaft für Informatik e.V. (Hg.): Themen in der GI, 27.11.2020. Online unter <https://gi.de/themen/beitrag/beipackzettel-fuer-modelle-des-maschinellen-lernens-fuer-fairere-ki>, zuletzt geprüft am 19.10.2021.
- Shneiderman, Ben (2020): Bridging the gap between ethics and practice. Guidelines for reliable, safe, and trustworthy human-centered AI Systems. In: ACM Transactions on Interactive Intelligent Systems 10 (4), S. 1–31. <https://doi.org/10.1145/3419764>
- Simone, Angela (2018): Steering research and innovation through RRI. What horizon for Europe? In: Journal of Science Communication 3, 6S. <https://doi.org/10.22323/2.17030302>
- techttag Redaktion (2020): prenode im Gründerview. 10 Fragen an Dr.-Ing. Robin Hirt und Dr. Ronny Schüritz. Online verfügbar unter <https://www.techttag.de/startups/gruenderview/prenode-im-gruenderview/>, zuletzt geprüft am 08.10.2021.
- Zhang, Daniel et al. (2021): The AI Index 2021 Annual Report. Stanford, CA: Human-Centered AI Institute, Stanford University. Online verfügbar unter https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf, zuletzt geprüft am 20.10.2021.
- Zuber, Niina; Kacianka, Severin; Pretschner, Alexander; Nida-Rümelin, Julian (2020): Ethische Deliberation für agile Softwareprozesse. EDAP-Schema. In: Markus Hengstschlager (Hg.): Digitaler Wandel und Ethik. München: Ecowin Verlag, S. 160–184.



DR. BERND BECKERT

ist stellvertretender Leiter des Competence Centers (CC) Neue Technologien am Fraunhofer-Institut für System- und Innovationsforschung (ISI). Er ist zudem Leiter der KI-Gruppe am ISI, die im Jahr 2020 gegründet wurde. Die KI-Gruppe beschäftigt sich CC-übergreifend mit künstlicher Intelligenz aus einer TA- und Innovationsforschungs-Perspektive.

RESEARCH ARTICLE

Wenn die Technik sprechen lernt

Künstliche Kommunikation als kulturelle Herausforderung mediatisierter Gesellschaften

Sascha Dickel, Institut für Soziologie, Johannes Gutenberg-Universität Mainz, Saarstr. 21, 55122 Mainz, DE
(dickel@uni-mainz.de)  0000-0003-3620-2934

23

Zusammenfassung • Wir scheinen am Anfang einer Ära kommunizierender Technik zu stehen. Amazons Sprachassistentensystem Alexa ist in Millionen von privaten Haushalten eingezogen, Chatbots gehören heute zu den Standardanwendungen im Kundenservice und der Einsatz von Algorithmen zur Erzeugung von Texten ist in die Praxis professionellen Publizierens integriert worden. An kommunizierenden Maschinen zeigen sich kontroverse Deutungen von Autonomie, Kreativität, Komplexität und Opazität von Mensch und Maschine wie in einem Brennglas. Der Beitrag argumentiert, dass kommunizierende Technik eine eigenständige kulturelle Herausforderung für mediatisierte Gesellschaften darstellt, der sich eine reflexive Technikfolgenabschätzung hermeneutisch zu stellen hat.

When technology learns to speak. Artificial communication as a cultural challenge for mediatized societies

Abstract • We are at the dawn of an era of communicating technology. Amazon's Alexa voice assistant system has entered millions of private households, chatbots have become a standard application in customer service, and the use of algorithms to generate texts has been integrated into the practice of professional publishing. Communicating machines reveal controversial interpretations of autonomy, creativity, complexity, and opacity of humans and machines. The article argues that artificial communication represents a cultural challenge for mediatized societies that reflexive technology assessment should address hermeneutically.

Keywords • artificial intelligence, artificial communication, hermeneutics, social theory

Künstliche Kommunikation

Im gesellschaftlichen Diskurs der Digitalisierung spielt der Topos der künstlichen Intelligenz eine herausgehobene Rolle. Dieser Artikel nimmt diesbezüglich eine Problemverschiebung vor, indem er sich einem Phänomen annähert, das sich im Windschatten des gesellschaftlichen Diskurses um künstliche Intelligenz (KI) entfaltet und das Verhältnis von Mensch und Technik fundamental rekonfigurieren könnte. Es handelt sich um das Phänomen sprechender und schreibender Maschinen und damit um die Frage nach den Möglichkeiten und Grenzen *künstlicher Kommunikation*. Dieser Begriff wurde von Elena Espósito (2017, S. 261) geprägt: „By artificial communication I mean communication that involves an entity, the algorithm, which has been built and programmed by someone to act as a communication partner“.

Bereits Alan Turing hat die Frage „Can machines think?“ (Turing 1950, S. 433) als Problem der Kommunikationsfähigkeit von Maschinen reformuliert. Anstelle einer wie auch immer gearteten definitiven Bestimmung oder Messung von Intelligenz, wird durch den (später sogenannten) Turing-Test geprüft, ob Maschinen in einer spezifisch formatierten Interaktionssituation von Menschen nicht mehr treffsicher zu unterscheiden sind. Die populärste Variante des Tests¹ basiert auf der Kommunikation eines Menschen (der ‚Richter:in‘) mit zwei weiteren Entitäten. Eine davon ist ein Mensch, die andere eine Maschine. Die Kommunikation findet dabei rein textbasiert, also explizit

1 An der Stelle sei angemerkt, dass die ursprüngliche gedankenexperimentelle Versuchsanordnung von Turing (1950) ein:e Richter:in, einen Mann und eine Frau vorsah. Die Aufgabe der Richter:in bestand darin, herausfinden, welcher der beiden anderen Akteure Mann oder Frau war. Turing fragte dann, was passieren würde, wenn der Mann durch eine Maschine ersetzt würde. Mittlerweile hat sich die oben skizzierte Form gleichwohl als „Standardvariante“ des Tests etabliert.

ohne Sicht- und Hörkontakt, statt. Die Aufgabe der Richter:in besteht nun darin, zu entscheiden, welcher der beiden Interaktionspartner:innen ein Mensch ist und bei welchem es sich um eine Maschine handelt. Dabei sind alle möglichen Fragen und Antworten legitim. Die Aufgabe von Maschine (und Mensch!) besteht darin, die Richter:in davon zu überzeugen, ein Mensch zu sein. Je besser es Maschinen gelingt, Humanität vorzutäuschen, desto besser wird ihr Abschneiden im Test bewertet. Die ‚Intelligenz‘ von Maschinen wird beim Turing-Test also nicht an bestimmten Merkmalen festgemacht, die einem Programm selbst zu eigen sind und die z. B. mit einem Intelligenztest erfass-

2019; Hepp 2020). Zugleich entwerfen Technologiefirmen Szenarien der Mensch-Maschine-Beziehung, in denen kommunizierende Systeme noch wesentlich selbstverständlicher in Alltag und Berufsleben integriert sind und dabei menschliche Positionen einnehmen (Dickel und Schmidt-Jüngst 2021). In einer Welt kommunizierender Maschinen werden akademische Streitfragen um die Autonomie, Kreativität, Komplexität und Opazität von Mensch und Maschine in alltagspraktische Problemstellungen transformiert.

Bereits vor zehn Jahren stellte Dirk Baecker (2011) in dieser Zeitschrift die Frage, ob die Gesellschaft sich zukünftig auf

*Die Intelligenz von Maschinen wird beim Turing-Test
nicht an bestimmten Merkmalen festgemacht,
die einem Programm selbst zu eigen sind, sondern an der
maschinellen Bewährung als Kommunikationspartner:in.*

bar wären, sondern an der maschinellen *Bewährung als Kommunikationspartner:in*.

Der Test sollte nicht lange Gedankenexperiment bleiben. Bereits das 1966 von Joseph Weizenbaum entwickelte Programm ELIZA erregte Aufmerksamkeit, da es durch ein limitiertes, aber rollenkonformes (nämlich therapeutisch interpretierbares) Antwortverhalten als ‚echte:r‘ Interaktionspartner:in wahrgenommen werden konnte. Ähnlich wie beim Turing-Test war hier die Fähigkeit des Programms, sich an einer Konversation in natürlicher Sprache beteiligen zu können, wichtiger als die interne Komplexität des Programms (Suchman 2007, S. 47 ff.).

Insgesamt blieb die künstliche Nachbildung menschlichen Kommunikationsverhaltens ein stets wiederkehrendes Thema technikwissenschaftlicher Forschung und ihrer kulturellen Reflexion. Auch und gerade in der Science-Fiction wurden kommunikationsfähige Maschinen in zahlreichen Varianten portraitiert. Eine ganze Reihe von zeitgenössischen Filmen (etwa: *Her*, *Ex Machina*, *Blade Runner 2049*) präsentiert Konstellationen, in denen Menschen und Maschinen miteinander sprechen und dabei nicht zuletzt die Art und Weise ihrer sozialen Beziehung reflektieren.

Mittlerweile ist die Kommunikation mit künstlichen Systemen zu einer *Alltagserfahrung* geworden, die über spezifische Kontexte technikwissenschaftlicher Erprobung und kultureller Reflexion weit hinausgeht: Sprachassistenzsysteme wie Alexa und Siri bevölkern Millionen von privaten Haushalten und Hosentaschen, Chatbots gehören zur Standardanwendung im Kundenservice und der Einsatz von Algorithmen zur Erzeugung von Texten wird bereits in vielfältige Praktiken professionellen Publizierens integriert (Guzman und Lewis 2020). Die Gesellschaft hat sich offenbar zunehmend daran gewöhnt, dass maschinelle Entitäten am Sozialen qua Kommunikation teilnehmen (Sieber

Maschinen als Kommunikationspartner:innen einstellen müsse. Seine Überlegungen waren seinerzeit noch weitgehend theoretisch-spekulativer Natur. Heute jedoch hat die Frage nach den Implikationen kommunizierender Technik eine unverkennbare empirische Relevanz und Brisanz gewonnen, welche nach Antworten insbesondere auch aus dem Feld der Technikfolgenabschätzung (TA) verlangt. Zugleich ist aber aktuell noch kaum klar, was in den kommenden Jahren und Jahrzehnten technisch möglich sein wird, welche Konzepte realisiert werden und welche Anwendungen Akzeptanz finden. In genau solchen Situationen einer Unsicherheit von Zukunft, stoßen sowohl prognostische als auch szenariobasierte Verfahren der TA an ihre Grenzen. Stattdessen bieten sich *hermeneutische* Reflexionen an, die soziotechnische Visionen und Optionen vor ihrem gesellschaftlichen Horizont zu verstehen versuchen (Grunwald 2014). TA wird damit stärker als bislang sinnverstehend und gesellschaftsdiagnostisch.

Ich schlage in diesem Beitrag vor, sozial- und gesellschaftstheoretisches Wissen als Ressource für eine hermeneutische TA fruchtbar zu machen. Dieser Rückgriff dient dem Verständnis, welche Aspekte von Kultur und Gesellschaft durch eine technologische Option in welcher Weise berührt werden. Für den Gegenstand schreibender und sprechender Maschinen soll vor diesem Hintergrund zunächst die Bedeutung von Kommunikation ausgeleuchtet und gesellschaftsdiagnostisch eingebettet werden. Diese Reflexion ist zugleich eine erste hermeneutische Annäherung, die klären soll, worin die grundsätzliche Herausforderung des Phänomens (jenseits spezifischer Anwendungen) liegt. Der Artikel entwirft darauf aufbauend ein Konzept, wie spezifischere hermeneutische Forschungsanstrengungen aussehen könnten, die für das Feld der TA Orientierungs- und Handlungswissen generieren könnten.

Kommunikationstheoretische Reflexion

Eine ganze Reihe soziologischer Ansätze betonen, dass Kommunikation eine wesentliche, wenn nicht gar die entscheidende Rolle für die Konstitution sozialer Ordnung einnimmt (Habermas 1981; Knoblauch 2017; Luhmann 1984). Gesellschaft reproduziert sich durch kommunikative Akte, die aneinander anschließen und sukzessive soziale Wirklichkeit hervorbringen. Niklas Luhmann hat dazu einen überaus schlanken Kommunikationsbegriff vorgeschlagen, der mit sehr wenigen Vorannahmen über die ontologischen Beschaffenheiten der an Kommunikation beteiligten Akteur:innen auskommt. Gerade dies macht den Begriff für Untersuchungen attraktiv, die sich der Bedeutung von Kommunikation mit nicht-menschlichen Entitäten in Zeiten der Digitalisierung zuwenden wollen (Esposito 2017). Mit Luhmann (1984) kann Kommunikation als dreigliedriger Selektionsprozess begriffen werden, der *Information* (was wird gesagt), *Mitteilung* (wie wird es gesagt) und *Verstehen* (dass etwas gesagt wurde) umfasst. Dabei sollte Luhmann zufolge keine dieser drei Selektionen als innerpsychischer Prozess aufgefasst werden. Vielmehr bildet Kommunikation eine Ordnung *sui generis*, in der prozesshaft bestimmt wird, was eigentlich von wem mitgeteilt wurde, ob es zu Verstehensproblemen gekommen ist und inwiefern etwas überhaupt als Kommunikation gemeint war.² Das hier explizierte Modell von Kommunikation geht davon aus, dass die Entitäten, die an Kommunikation beteiligt sind, wechselseitig intransparent sind und bleiben. Gerade weil diese Intransparenz nicht intersubjektiv gelöst werden kann, kommt Kommunikation in Gang. Autonomie, Handlungsfähigkeit und

als *vollwertige* Teilnehmer:innen an gesellschaftlicher Kommunikation anerkannt sind. Sie werden *personifiziert*, während andere Entitäten (Ahnen, Bäume, Tiere, Geister, etc.) häufiger objektiviert bzw. fiktionalisiert werden. Gesa Lindemann (2011) hat herausgearbeitet, dass diese kommunikative Sonderstellung des Menschen zwar in der Moderne institutionalisiert wurde, aber prinzipiell als kulturell kontingent zu betrachten ist. Es ist eben diese vermeintlicherweise herausgehobene Position von Menschen als exklusiv kommunizierenden Personen, die mit der Ankunft sprechender und schreibender Maschinen herausgefordert wird. Das führt zu kulturellen Herausforderungen.

Kulturelle Herausforderungen

Eine erste Frage, die sich stellt, ist die nach der *Identität* des kommunikativen Gegenübers in einer Zeit maschineller Autor:innenschaft. Seit mehreren Jahren wird die Befürchtung artikuliert, dass sogenannte ‚Social Bots‘ die öffentliche Meinung auf sozialen Medien beeinflussen (Leistert 2017). 2018 stellte Google öffentlichkeitswirksam ein intelligentes persönliches Assistenzsystem (die Erweiterung Duplex für den Google Assistant) vor, welches die Stimme und das mündliche Kommunikationsverhalten von Menschen täuschend echt reproduzieren konnte. 2020 zeigte sich die Fachwelt beeindruckt, da der von dem Unternehmen OpenAI entwickelte Generative Pre-trained Transformer 3 (ein Sprachverarbeitungsmodell, das auf Deep Learning basiert) in der Lage ist, vielfältige Formen von Textgattungen zu generieren. Simone Natale (2021) weist angesichts

*Wenn aber Maschinen nun auch Bücher schreiben,
Musik veröffentlichen, Beziehungen aufbauen und
Dialoge führen, dann fordert dies das Verhältnis von Mensch
und Technik in tiefgreifenderer Weise heraus.*

Bewusstsein – die im Diskurs um ‚künstliche Intelligenz‘ typischerweise als *Eigenschaften* von Akteuren begriffen werden – werden kommunikationstheoretisch als *Zurechnungen* sichtbar, die kommunikativ explizit gemacht, stillschweigend vorausgesetzt oder (kommunikativ!) problematisiert werden können.

In der modernen Gesellschaft (sprich: in ihrer kommunikativen Praxis!) geht man gemeinhin davon aus, dass nur Menschen

solcher Phänomene darauf hin, dass es immer schwieriger werden könnte, Menschen und Maschinen zu unterscheiden, wenn künstliche Systeme sprechen und schreiben lernen und damit Merkmale menschlicher Kommunikation zu imitieren beginnen. Wenn Turing-Tests aus dem Labor in den Alltag wandern, stellen sich offenkundig neue Probleme der Zurechnung von Kommunikation.

Kommunizierende Maschinen lassen *zweitens* einen Problemhorizont aufscheinen, der über die typischen zeitgenössischen Deutungen von KI als einer neuen industriellen Revolution hinausgeht, in der nun (auch) die geistige Arbeit automatisiert wird (Brynjolfsson und McAfee 2012). Die hoffnungsvoll konturierte Komplementärdiagnose zur Maschinisierung kognitiver Leistung ist nämlich, dass nun menschliche Arbeiten verbes-

² So kann A etwa einen Ausruf von B als Mitteilung betrachten, B kann bestreiten, dass dies so gemeint war und z. B. darauf hinweisen, dass sie sich lediglich den Fuß gestoßen hätte – und schon sind beide in einen Prozess des Kommunizierens verstrickt – und dass obwohl B niemals die Absicht hatte, A etwas mitzuteilen. Sozialität (im Kontrast zu bloßen Verhalten) kommt somit immer dann in Gang, wenn Entitäten unterstellt wird, dass sie nicht nur *das Eine*, sondern auch *etwas Anderes* hätten mitteilen oder verstehen können.

sert werden, die primär durch Kommunikation getragen werden: Care-Arbeit, Beziehungspflege, kreativ-schöpferische Tätigkeit, kontextsensibles Sprechen und Schreiben.

Wenn aber Maschinen nun auch Bücher schreiben, Musik veröffentlichen, Beziehungen aufbauen und Dialoge führen, dann fordert dies das Verhältnis von Mensch und Technik in tiefgreifender Weise heraus. Neben die Identitätsfrage tritt damit die Frage nach der *kommunikativen Substitution* des Menschen durch Maschinen (Sieber 2019). In den gegenwärtigen Einsatzgebieten kommunizierender Maschinen (vom automatisierten Journalismus über Service-Chatbots bis hin zu Sprachassistentensystemen im Privatleben) ist diese Substitution gleichwohl noch auf spezifische Funktionsrollen fokussiert, die dezidiert an bestimmte Formen der Leistungserbringung gekoppelt sind (Lindemann 2011, S. 344). Das trifft auch dann zu, wenn kommunizierende Systeme explizit als Lebenspartner:innen oder Familienmitglieder vermarktet werden. Ein Beispiel für ersteres ist das japanische System Gatebox, ein Beispiel für letzteres der soziale Roboter Jibo. Dass wir Technik Funktionsrollen zuweisen, ist an sich nichts Besonderes. Brisant wird aber, wenn die dienstbare Technik zugleich durch kommunikative Zuschreibungen personifiziert wird. Nehmen wir das Beispiel von Alexa. Das System wird als moderne Servicekraft positioniert, die ihren Besitzer:innen im Umkehrschluss das Gefühl einer höheren sozialen Stellung vermitteln kann. Problematisch ist dabei nicht nur die durch Namen und Stimme vergeschlechtlichte Codierung der untergeordneten Funktionsrolle (Sontopski 2019; Dickel und Schmidt-Jüngst 2021). Befürchtet wird auch, dass die kommunikativen Muster (etwa von Befehl und Gehorsam), die mit maschinellen Kommunikationspartner:innen eingeübt werden, auf menschliche Kommunikationspartner:innen übertragen werden könnten.

Verkettungen von Fragen und Antworten (Sieber 2019). In den Zukunftsvisionen kommunizierender Maschinen werden solche Defizite freilich als technologische Hürden gerahmt, die prinzipiell überwindbar sind. Damit tritt auch der Traum von einer *Symmetrisierung* von Mensch und Maschine durch die Konstruktion kommunikativer Ebenbürtigkeit aktuell mit aller Macht in die Technoimagination unserer Gegenwart ein.

Mediatisierung als Bedingung maschineller Personifizierung

Ein hermeneutischer Blick auf technologische Entwürfe und Szenarien lenkt den Fokus auf die gesellschaftliche Gegenwart, in der sich oben skizzierte Herausforderungen als kulturelle Fragen überhaupt stellen. In welcher Gesellschaft also werden Fragen nach (1) der Identität von Mensch und Maschine, (2) der maschinellen Substitution von Menschen und schließlich (3) der Symmetrisierung von Mensch und Maschine eigentlich akut? Die zentrale These dieses Artikels ist, dass die *Mediatisierung von Kommunikation* (Krotz 2017) das entscheidende ‚Einfallstor‘ für die Personifizierung nicht-menschlicher Entitäten darstellt.

Kommunikation wird typischerweise immer noch oft von einem Paradigma körperlich anwesender Menschen verstanden, die ein mündliches Gespräch führen (Knorr Cetina et al. 2017). Doch ist diese Form der Kommunikation (nicht erst seit der Coronakrise) längst zu einer sehr spezifischen Spezialform geworden, neben die vielfältigste Formen des Kommunizierens auf Basis technischer Medien getreten sind (Dickel 2020). Neue medientechnologische Innovationen verändern die Strukturen des Sozialen nachhaltig, da sie die Formen des Kommunizierens ver-

Die Mediatisierung von Kommunikation stellt das entscheidende ‚Einfallstor‘ für die Personifizierung nicht-menschlicher Entitäten dar.

Die dritte und letzte Herausforderung, die ich hier nennen möchte, ist zugleich die spekulativste: Werden wir geneigt sein, Maschinen zunehmend Autonomie, Bewusstsein und Verantwortlichkeit zuzuschreiben, wenn sie sich nicht nur in spezifisch formatierten Funktionskontexten, sondern vielfältigen lebenspraktischen Situationen kommunikativ bewähren? Aktuell sehen wir, dass die kommunikativen Fähigkeiten von Maschinen im Vergleich zu Menschen noch stark begrenzt sind. Die Interaktion mit einem Service-Chatbot wirkt typischerweise mechanisch und unflexibel. Und selbst bei anspruchsvolleren Systemen wie Alexa oder Siri, die eine mündliche Konversation simulieren sollen, wäre es kaum angebracht, von einem echten Dialog zu sprechen. Vielmehr handelt es sich um recht einfache

ändern (Luhmann 1997, S. 249 ff.; Knoblauch 2017, S. 316 ff.; Baecker 2007). Durch mediale Infrastrukturen überschreitet das soziale Netz die Zeit- und Raumgrenzen der Interaktion unter körperlich Anwesenden: Autor:innen schreiben Texte für ein ihnen unbekanntes Publikum. Radio, Fernsehen und Internet schaffen neue technomediale Räume der Massenkommunikation. Aber auch die zwischenmenschliche Interaktion wird mediatisiert (Höflich 2016). Durch Medien wie Telefon, Chat und Instant Messaging-Programme werden synchrone Kommunikationskontexte medial angereichert und formatiert (Knorr Cetina et al. 2017). Mediale Techniken betreffen zunächst den *Mitteilungsaspekt* – also das ‚wie‘ – des Kommunizierens. Diese Veränderung von Mitteilungsmöglichkeiten verändert aber auch,

was mitgeteilt wird und werden kann (*Informationsaspekt*), was überhaupt als Kommunikation gilt (*Verstehensaspekt*) – und welche Entitäten als Sender:innen und Empfänger:innen von Kommunikation infrage kommen.

Die Anwesenheit eines menschlichen Körpers ist längst nicht mehr nötig, damit Kommunikation als Kommunikation verstanden wird: Die Gesellschaft hat sich nicht nur daran gewöhnt, geschriebene Worte als Mitteilungen von Autor:innen zu interpretieren, sondern auch körperlose Stimmen aus Lautsprechern und Gesichter auf Bildschirmen als Repräsentationen von Anwesenden zu behandeln, denen Kommunikationsfähigkeit zugerechnet werden kann und bei denen Verstehen erwartet wird. Diese

technischer Leistungsfähigkeit, sondern auch eine Frage medialer Settings und kultureller Praktiken, in die eine Technik eingebunden wird.

Jenseits des Turing-Tests: Eine hermeneutische Aufgabe für die TA

Dieser Artikel hat argumentiert, dass die Zuschreibung von nicht-menschlicher Kommunikationsfähigkeit als Mediatisierungseffekt zu begreifen ist und damit maßgeblich von den medialen Ökologien abhängt, in denen KI eingebettet ist. Vor diesem Hin-

Die Anwesenheit eines menschlichen Körpers ist längst nicht mehr nötig, damit Kommunikation als Kommunikation verstanden wird.

gesellschaftliche Einübung mediatisierter Kommunikation bereitet den Boden dafür, dass sich ‚hinter‘ dem medialen Interface nicht mehr zwingend menschliche Kommunikationsteilnehmende befinden müssen.

Künstliche Kommunikation setzt an eben dieser Stelle an. Sie wird heute immer mehr zum expliziten Ziel der Gestaltung von Human-Machine-Interfaces (Sieber 2019; Hepp 2020). Das Aufkommen kommunizierender Maschinen lässt sich demgemäß als nächste Stufe einer „deep mediatization“ (Hepp 2020, S. 1413) des gesellschaftlichen Lebens begreifen, in welcher sich die Bedeutung von Medien und Kommunikation selbst verändert: „Communicative AI departs from the historical role of media as mere channels of communication, since AI also acts as a producer of communication“ (Natale 2021, S. 11).

Der Turing-Test lässt sich diesbezüglich weniger als Blaupause für maschinelle Denkfähigkeit lesen, sondern als kommunikations- und medientheoretisches Lehrstück.

- Der Test führt *erstens* vor, dass sich die Differenz von Mensch und Maschine durch eine kommunikative Praxis selbst situativ aktualisiert.
- Er demonstriert *zweitens*, dass die Kommunikation von Mensch und Maschine durch mediale Infrastrukturen geformt wird. Die gesamte Anordnung des Tests funktioniert nur unter mediatisierten Sonderbedingungen (hier: von Intransparenz und Schriftlichkeit).
- *Drittens* zeigt der Test, dass jede Kommunikation von Mensch und Maschine von gesellschaftlich institutionalisierten Deutungsmustern abhängt, die bestimmen, was überhaupt als typisch menschliches oder maschinelles Verhalten gilt.

tergrund muss sich eine TA der künstlichen Kommunikation vor allem der Frage zuwenden, wie die Gestaltung von und der Umgang mit medialen Infrastrukturen und Interfaces die Deutung hervorbringt, dass technische Artefakte mit uns kommunizieren – dass sie uns Informationen mitteilen und unsere Worte verstehen.

Angesichts der kulturellen Unsicherheiten, die mit kommunizierenden Maschinen verbunden sind, soll somit zum Abschluss dieses Artikels skizziert werden, worin die kommende Aufgabe einer hermeneutischen TA zu diesem Problemfeld bestehen kann. Ich schlage vor, dass sie sich insbesondere als Hermeneutik zweiter Ordnung profilieren sollte, welche das *Verstehen des Verstehens* von Technik als ihren Gegenstand begreift. Ansätze dafür liefern methodische Herangehensweisen der sozialwissenschaftlichen Hermeneutik, die soziale Deutungen und Praktiken nicht einfach vor einem schon vermeintlich verstandenen gesellschaftlichen Kontext deuten, sondern durch eine minutiöse Rekonstruktion von Sinnstrukturen diesen Kontext erst zu erschließen versuchen (Sammet und Erhard 2018).

Eine Hermeneutik der mediatisierten Kommunikation mit Maschinen würde die Aufgabe von TA, Orientierungswissen bereitzustellen, mit dem Aufzeigen gesellschaftlicher Handlungsoptionen verzahnen. Über die allgemeine hermeneutische Reflexion grundlegender Konstitutionsbedingungen und Herausforderungen (welche dieser Artikel umreißen konnte) hinaus, ergibt sich so die Möglichkeit einer Hermeneutik der kommunikativen Praxis, die nicht selbst – vom akademischen Lehnstuhl aus – definiert, wie es um die Kommunikationsfähigkeit von künstlichen Systemen bestellt ist, sondern rekonstruktiv erschließt, was es bedeutet, wenn Technik zur sozialen Adresse wird.³ Solche Untersuchungen können auf drei Ebenen ansetzen:

Ob wir Technik Eigenschaften wie Autonomie, Reflexions- und Handlungsfähigkeit zubilligen, ist somit nicht nur eine Frage

3 Erste Forschungsarbeiten, die in diese Richtung weisen, liegen bereits vor (Muhle 2016; Meister 2014; Mayer et al. 2020; Voss 2021).

- Um die *Gestaltung* von künstlicher Kommunikation zu erschließen, kann durch Expert:inneninterviews und teilnehmende Beobachtungen in Erfahrung gebracht werden, welche Erwartungen hinsichtlich der Kommunikation von Mensch und Maschine in das Design der Technik eingeschrieben werden.
- Um die *Praxis* künstlicher Kommunikation anhand der Interaktionen von Nutzer:innen mit gegenwärtig verfügbaren Anwendungen zu rekonstruieren, können sozialwissenschaftli-

von dem Design medialer Interfaces und Infrastrukturen. Dabei lassen sich spekulative Extrempole unterscheiden: Zum einen wäre denkbar, dass die Konstellation des Turing-Test zu einer impliziten Alltagskonstellation wird und ständig neue Praktiken und Techniken in Stellung gebracht werden, um Menschen und Maschinen in tiefgreifend mediatisierten Umwelten weiterhin differenzieren zu können. Der Gegenpol wäre eine Gesellschaft, in der eben diese Relevanzsetzungen verschwinden, und die Differenz von Mensch und Maschine als einer kulturell

In welchen Bereichen lassen wir zu, dass Maschinen kommunikativ menschliche Rollen einnehmen?

che Verfahren wie Sequenz- und Konversationsanalyse zum Einsatz kommen. Mit diesen lässt sich zeigen, wie Menschen und Maschinen sich gegenseitig kommunikativ behandeln und welche sozialen Positionen und Rollen sie sich wechselseitig zuweisen.

- Um die *Imaginationen* künstlicher Kommunikation (etwa Werbevideos von Firmen oder Science-Fiction-Filme) zu verstehen, können öffentliche Darstellungen sprechender Maschinen zum Beispiel video- und bildanalytisch untersucht werden.

Durch solche Untersuchungen können wir nicht nur etwas über die impliziten und expliziten Erwartungen erfahren, die in den entsprechenden Artefakten angelegt und eingeschrieben sind. Anhand der Analyse von gegenwärtigen (oder in Entwicklung befindlichen) Kommunikationsmaschinen können wir auch etwas über deren Möglichkeiten und Unzulänglichkeiten sowie ihre spezifischen Bedingungen und Begrenztheiten in vorausschauender Weise lernen. Ergänzend dazu führen uns Imaginationen der Kommunikation mit Maschinen vor, wie solche Interaktion im Idealzustand aussehen könnten, aber auch welche Hoffnungen und Befürchtungen mit ihnen verbunden werden. In jedem Fall erbringen solche hermeneutischen Analysen Distanzierungsgewinne, die über unsere gesellschaftliche Beziehung zu Technik aufklären. Sie eröffnen so Ansatzpunkte für öffentliche Dialoge. Denn das Wissen über die gegenwärtig real praktizierten Deutungen kommunizierender Maschinen und das Verstehen ihres gesellschaftlichen Verstehens, eröffnet zugleich Möglichkeiten, die Arten und Weisen zu verhandeln, wie Kommunikation im Zeitalter kommunizierender Technik gestaltet werden kann. Welche Ansprüche stellen wir an die Identifizierbarkeit von Menschen? In welchen Bereichen lassen wir zu, dass Maschinen kommunikativ menschliche Rollen einnehmen? Wie würde eine Gesellschaft aussehen, die Mensch und Maschine qua Kommunikation symmetrisiert?

Wie mit der Differenz von Mensch und Maschine in Zeiten künstlicher Kommunikation zukünftig umgegangen wird, wird von Konventionen kommunikativer Praxis ebenso abhängen, wie

bedeutsamen Unterscheidung de-institutionalisiert wird. Das Resultat wäre eine posthumane Ordnung, in der die einstmals scharfe, ‚natürliche‘ Differenz von Mensch und Maschine als kontingente kulturelle Konstruktion gelesen wird – ähnlich wie dies heute für verschiedene Binnendifferenzierungen des Humanen (wie Geschlechter oder Völker) in Anspruch genommen wird.⁴

Angabe von Finanzierungsquellen

Der vorliegende Forschungsartikel hat keine Förderung erhalten.

Literatur

- Baecker, Dirk (2007): Studien zur nächsten Gesellschaft. Frankfurt am Main: Suhrkamp.
- Baecker, Dirk (2011): Who qualifies for communication? A systems perspective on human and other possibly intelligent beings taking part in the next society. In: Technikfolgenabschätzung. Theorie und Praxis 20 (1), S. 17–26. <https://doi.org/10.14512/tatup.20.1.17>
- Brynjolfsson, Erik; McAfee, Andrew (2012): Race against the machine. How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. Lexington: Digital Frontier Press.
- Dickel, Sascha (2020): Gesellschaft funktioniert auch ohne anwesende Körper. Die Krise der Interaktion und die Routinen mediatisierter Sozialität. In: Michael Volkmer und Karin Werner (Hg.): Die Corona-Gesellschaft. Analysen zur Lage und Perspektiven für die Zukunft. Bielefeld: transcript, S. 79–86. <https://doi.org/10.14361/9783839454329-008>
- Dickel, Sascha; Schmidt-Jüngst, Miriam (2021): Gleiche Menschen, ungleiche Maschinen. Die Humandifferenzierung digitaler Assistenzsysteme und ihrer Nutzer:innen in der Werbung. In: Dilek Dizdar, Stefan Hirschauer, Johannes Paulmann und Gabriele Schabacher (Hg.): Humandifferenzierung. Disziplinäre Perspektiven und empirische Sondierungen. Weilerswist: Velbrück Wissenschaft, S. 342–367. <https://doi.org/10.5771/9783748911364-342>

⁴ Das Forschungsprogramm der Humandifferenzierung strebt an, 1) zwischen-menschliche Binnendifferenzierungen und 2) Unterscheidungen von Menschen und ihrer nicht-menschlichen Umwelt in ihrem Zusammenhang zu untersuchen. Vgl. dazu am Fall von Amazons Alexa Dickel und Schmidt-Jüngst 2021.

- Esposito, Elena (2017): Artificial communication? The production of contingency by algorithms. In: *Zeitschrift für Soziologie* 46 (4), S. 249–265. <https://doi.org/10.1515/zfsoz-2017-1014>
- Grunwald, Armin (2014): The hermeneutic side of responsible research and innovation. In: *Journal of Responsible Innovation* 1 (3), S. 274–291. <https://doi.org/10.1080/23299460.2014.968437>
- Guzman, Andrea; Lewis, Seth (2019): Artificial intelligence and communication. A human-machine communication research agenda. In: *New Media & Society* 22 (1), S. 70–86. <https://doi.org/10.1177/1461444819858691>
- Habermas, Jürgen (1981): *Theorie des kommunikativen Handelns*. Frankfurt am Main: Suhrkamp.
- Hepp, Andreas (2020): Artificial companions, social bots and work bots. Communicative robots as research objects of media and communication studies. In: *Media, Culture & Society* 42 (7–8), S. 1410–1426. <https://doi.org/10.1177/0163443720916412>
- Höfllich, Joachim (2016): *Der Mensch und seine Medien. Mediatisierte interpersonale Kommunikation. Eine Einführung*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-531-18683-2>
- Knoblauch, Hubert (2017): *Die kommunikative Konstruktion der Wirklichkeit*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-15218-5>
- Knorr Cetina, Karin; Reichmann, Werner; Woermann, Niklas (2017): Dimensionen und Dynamiken synthetischer Gesellschaften. In: Friedrich Krotz, Cathrin Despotović und Merle-Marie Kruse (Hg.): *Mediatisierung als Metaprozess. Transformationen, Formen der Entwicklung und die Generierung von Neuem*. Wiesbaden: Springer VS, S. 35–57. https://doi.org/10.1007/978-3-658-16084-5_3
- Krotz, Friedrich (2017): Mediatisierung. Ein Forschungskonzept. In: Friedrich Krotz, Cathrin Despotović und Merle-Marie Kruse (Hg.): *Mediatisierung als Metaprozess. Transformationen, Formen der Entwicklung und die Generierung von Neuem*. Wiesbaden: Springer VS, S. 13–32. https://doi.org/10.1007/978-3-658-16084-5_2
- Leistert, Oliver (2017): Social Bots als algorithmische Piraten und als Boten einer techno-environmentalen Handlungskraft. In: Robert Seyfert und Jonathan Roberge (Hg.): *Algorithmenkulturen. Über die rechnerische Konstruktion der Wirklichkeit*. Bielefeld: transcript, S. 215–234. <https://doi.org/10.14361/9783839438008-009>
- Lindemann, Gesa (2011): Die Akteure der funktional differenzierten Gesellschaft. In: Nico Lüdtke und Hironori Matsuzaki (Hg.): *Akteur. Individuum. Subjekt*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 329–350. https://doi.org/10.1007/978-3-531-93463-1_15
- Luhmann, Niklas (1984): *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Frankfurt am Main: Suhrkamp.
- Luhmann, Niklas (1997): *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Mayer, Henning; Muhle, Florian; Bock, Indra (2020): Whiteboxing MAX. Zur äußeren und inneren Interaktionsarchitektur eines virtuellen Agenten. In: Eckhard Geitz, Christian Vater und Silke Zimmer-Merkle (Hg.): *Black Boxes. Versiegelungskontexte und Öffnungsversuche. Interdisziplinäre Perspektiven*. Berlin: de Gruyter, S. 295–322. <https://doi.org/10.1515/9783110701319-016>
- Meister, Martin (2014): When is a robot really social? An outline of the robot sociologicus. In: Michaela Pfadenhauer und Knud Böhle (Hg.): *Of social robots and artificial companions. Contributions from the social sciences*. STI Studies 10 (1), S. 107–134.
- Muhle, Florian (2016): „Are you human?“ Plädoyer für eine kommunikationstheoretische Fundierung interpretativer Forschung an den Grenzen des Sozialen. In: *Forum Qualitative Sozialforschung* 17 (1).
- Natale, Simone (2021): *Deceitful media. Artificial intelligence and social life after the Turing Test*. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780190080365.001.0001>
- Sammet, Kornelia; Erhard, Franz (2018): *Methodologische Grundlagen und praktische Verfahren der Sequenzanalyse. Eine didaktische Einführung*. In: Franz Erhard und Kornelia Sammet (Hg.): *Sequenzanalyse praktisch*. Weinheim: Beltz Juventa, S. 15–71.
- Sieber, Armin (2019): *Dialogroboter. Wie Bots und künstliche Intelligenz Medien und Massenkommunikation verändern*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-24393-7>
- Sontopski, Natalie (2019): Hey Siri?! In: *Kursbuch* 55 (199), S. 62–75. <https://doi.org/10.5771/0023-5652-2019-199-62>
- Suchman, Lucy (2007): *Human-machine reconfigurations. Plans and situated actions*. Cambridge, U. K.: Cambridge University Press. <https://doi.org/10.1017/CBO9780511808418>
- Turing, Alan (1950): Computing machinery and intelligence. In: *Mind* LIX (236), S. 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Voss, Laura (2021): *More than machines?* Bielefeld: transcript. <https://doi.org/10.14361/9783839455609>



PROF. DR. SASCHA DICKEL

ist seit 2021 Professor für Mediensoziologie und Gesellschaftstheorie an der Johannes Gutenberg-Universität Mainz. Seine Arbeitsschwerpunkte sind Technikzukünfte, Mensch-Maschine-Beziehungen, Wissenschaftskommunikation und -partizipation.

RESEARCH ARTICLE

Artificial intelligence in human genomics and biomedicine

Dynamics, potentials and challenges

Reinhard Heil, *Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT),*
Karlstr. 11, 76133 Karlsruhe, DE (reinhard.heil@kit.edu)  0000-0001-7265-7376

Nils B. Heyen, *Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, DE (nils.heyen@isi.fraunhofer.de)*  0000-0002-9354-1388

Martina Baumann, *Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT),*
Karlsruhe, DE (martina.baumann@kit.edu)

Bärbel Hüsing, *Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, DE (baerbel.huesing@isi.fraunhofer.de)*  0000-0003-0277-0570

Daniel Bachlechner, *Fraunhofer Austria Research GmbH, Innovation Center »Digital Transformation of Industry«, Wattens, AT*
(daniel.bachlechner@fraunhofer.at)  0000-0001-7726-9065

Ulrich Schmoch, *Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, DE (ulrich.schmoch@isi.fraunhofer.de)*  0000-0001-8615-5115

Harald König, *Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT),*
Karlsruhe, DE (h.koenig@kit.edu)  0000-0002-0117-0939

Abstract • The increasing availability of extensive and complex data has made human genomics and its applications in (bio)medicine an attractive domain for artificial intelligence (AI) in the form of advanced machine learning (ML) methods. These methods are linked not only to the hope of improving diagnosis and drug development. Rather, they may also advance key issues in biomedicine, e. g. understanding how individual differences in the human genome may cause specific traits or diseases. We analyze the increasing convergence of AI and genomics, the emergence of a corresponding innovation system, and how these associative AI methods relate to the need for causal knowledge in biomedical research and development (R&D) and in medical practice. Finally, we look at the opportunities and challenges for clinical practice and the implications for governance issues arising from this convergence.

Künstliche Intelligenz in der Humangenomik und Biomedizin.
Dynamiken, Potenziale und Herausforderungen

Zusammenfassung • Die zunehmende Verfügbarkeit umfangreicher und komplexer Daten hat die Humangenomik und ihre Anwendungsbereiche in der (Bio-)Medizin zu einem attraktiven Bereich für künstliche Intelligenz (KI) vor allem in Form von fortgeschrittenen Methoden des maschinellen Lernens (ML) gemacht. Diese Methoden sind nicht

nur mit der Hoffnung verbunden, Diagnosen und die Medikamentenentwicklung zu verbessern. Sie könnten auch darum, Kernthemen in der Biomedizin voranzubringen, z. B. zu verstehen, wie individuelle Unterschiede im menschlichen Genom bestimmte Merkmale oder Krankheiten verursachen können. Wir analysieren die zunehmende Konvergenz von KI und Genomik, das Entstehen eines entsprechenden Innovationssystems und wie diese assoziativen KI-Methoden mit dem Bedarf an kausalem Wissen in der biomedizinischen Forschung und Entwicklung und in der medizinischen Praxis zusammenhängen. Schließlich betrachten wir die Potenziale und Herausforderungen für die klinische Praxis und die sich aus dieser Konvergenz ergebenden Implikationen für Governance-Fragen.

Keywords • artificial intelligence, biomedicine, genomics, governance, knowledge

Introduction

The increasing availability of extensive and complex data has made human genomics and its application areas in (bio)medicine an attractive domain for artificial intelligence (AI) in the form of advanced machine learning (ML) methods (Wainberg et al. 2018). The focus of interest is on sequence data of the human genome as well as on data of genes that are read (transcribed) or proteins that are produced in various body cells and organs. These can be combined with clinical data from biobanks or electronic patient records, among others. The use of ML in

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.30>
Received: Jun. 17, 2021; revised version accepted: Oct. 08, 2021;
published online: Dec. 20, 2021 (peer review)

human genomics and biomedicine is associated with the hope of obtaining answers to a key question in these domains, namely how individual differences or mutations in the genome cause specific traits or diseases. This would allow predictions about functional consequences of genetic differences, diagnoses, prognoses about therapy options or the development of new drugs at an unprecedented pace and scope.

With AI and human genomic research, two emerging techno-scientific domains appear to converge, both being linked to hopes and fears on issues such as new diagnosis and therapy options, health economics, autonomy, discrimination, privacy, or accountability – all of which will likely be judged against the background of different interests, values, or worldviews of people. To develop policies helping to align innovations with societal needs and expectations, (1) an understanding of scientific-technical potentials and challenges, and (2) (mutual) learning on perspectives through dialog activities on the technologies and possible applications involving various stakeholders and publics will be needed.

Given the nascent state of the convergence of both domains, in this article we mainly focus on the first step, by exploring and using evidence from the literature as well as perspectives from stakeholders involved in current scientific-technical developments. This work may inform discussions on policy issues as well as realistic scenarios in societal dialog activities in the required second step towards the further development of AI in human genomics according to societal expectations.

Methods

As basis for identifying potentials and challenges, the current state of research as well as existing and emerging applications were examined and mapped by reviewing peer-reviewed scientific publications, conference proceedings, and patents (König 2020). In addition, the emerging innovation system at the intersection of AI and human genomics was analyzed through a mapping of international key actors, a review of international policy strategies, and a publication and patent analysis. The results of

these steps were presented and discussed at a two-day workshop (held in November 2019 in Heidelberg) with twelve international experts from academic research, industry and the venture capital sector. Subsequently, the literature was evaluated with regard to ethical, social and regulatory challenges. These results informed a one-day workshop (held online in October 2020) with eight experts from Germany representing clinical research, genetic counseling, patient associations, medical and technical ethics, and jurisprudence. Finally, the results of all mentioned steps were reviewed and summarized for this article.

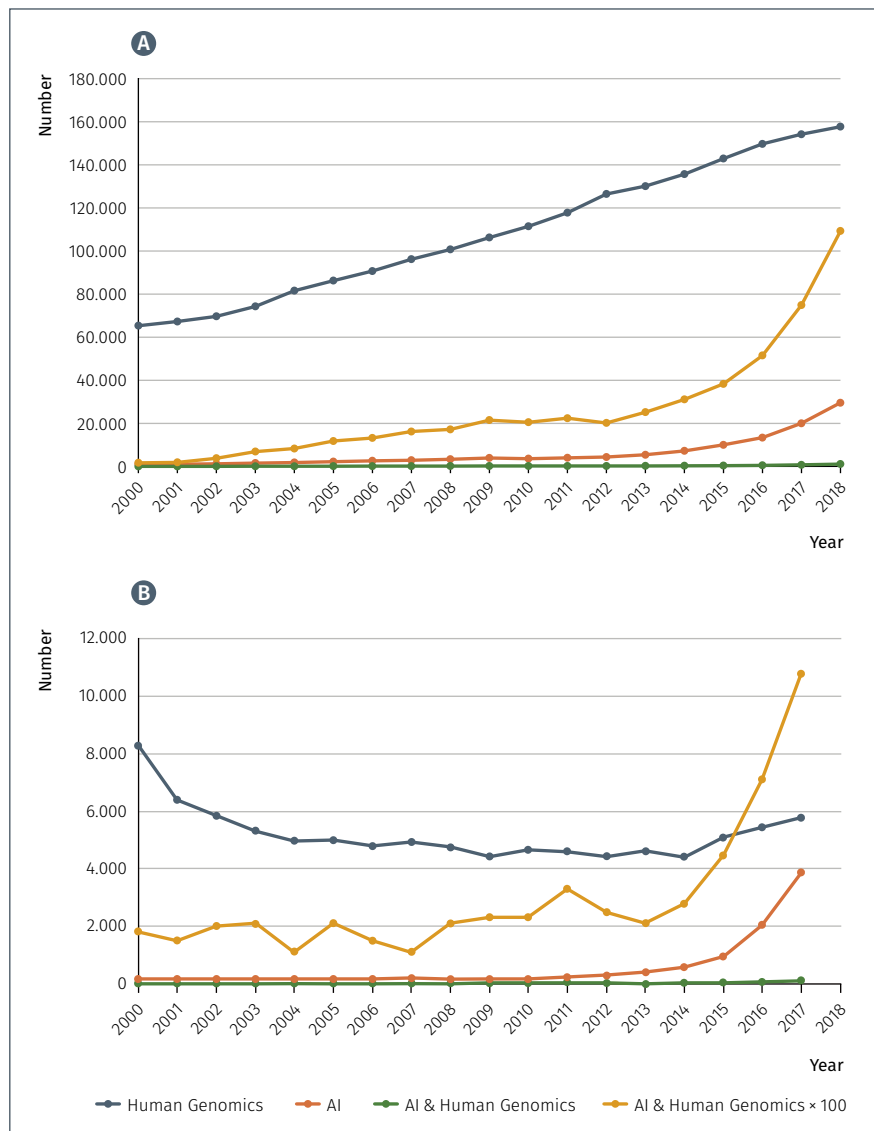


Fig. 1: Number of publications and patents worldwide over time.

A: Number of publications (Source: Web of Science 2019. Authors' own calculations), **B:** Number of transnationally registered patents (Source: World Patents Index 2019. Authors' own calculations).

Number of publications and patents, respectively, in the domain 'AI in human genomics' (green curves), human genomics (blue) and AI (orange). To better visualize the dynamic development, publication and patent numbers of the green curves were multiplied by 100 (yellow curves).

Results and Discussion

Innovation system analysis

Our analyses of the publication and patent statistics show that the use of AI-based methods in human genomics and biomedicine has boomed in recent years, both in academia and industry. Thus, the number of scientific publications and patents rapidly increased since 2014 (Figure 1 A/B; yellow curves). However, if one compares these activities to those in the domains of AI or of human genomics overall, it becomes clear that they still represent a relatively small niche with a limited number of key players and application areas (Figure 1 A/B; green curves).

cisive help in rapidly and comprehensively identifying putative causative genetic alterations and molecular mechanisms for diseases. These include common diseases such as neurological diseases (e. g., Alzheimer's disease, autism or schizophrenia), inflammatory bowel diseases (e. g., Crohn's disease) and diabetes (König et al. 2021; Zou et al. 2019). Similarly, potential pathogenic mutations that could facilitate diagnoses or prognoses have been identified for various cancer types and rare genetic diseases (Bailey et al. 2018; Brasil et al. 2019).

Drug development: ML methods are increasingly developed and used, especially by start-up companies, to make predictions of molecular properties (such as protein structures and interac-

In the future, there could be a shift from scientific explanation based on experimentally validated causal relationships towards explanation largely derived from AI predictions.

However, if this momentum continues, research and development (R&D) in human genomics will be heavily influenced by AI in the next years.

Looking at different world regions, the publication analysis suggests that Europe as a whole plays a significant role in the global research domain (with the UK and Germany as the most prominent European countries) alongside the two dominant countries USA and China. In contrast, the statistics of transnational patents (suggesting a particular commercial relevance) show that the USA is well ahead of all other countries and regions in utilizing relevant research for commercial purposes.

The dynamics in global publication and patent activities at the intersection of AI and human genomics point to the emergence of a new innovation system. Our analyses suggest various factors and actors that have contributed to this development. These are in particular the widespread (also clinical) use of and demand for whole genome sequencing and corresponding platform technologies as well as the engagement of venture capital companies, big tech companies, pharma companies, and start-ups. The latter (especially in the USA) are significantly driving technological innovations for drug discovery and development (DKA 2019). In addition, numerous public research organizations, international research consortia (e. g., the International Cancer Genome Consortium) or local research clusters (such as Boston, MA, USA) are important drivers.

Potentials of AI for human genomics and biomedicine

Basic and translational research: Deep learning (DL), a form of ML which relies on software-simulated multiple layers of so-called artificial neurons (deep neural networks), is increasingly used to explore how individual differences in the human genome may cause traits or diseases. Several recent studies suggest that DL models, via analysis of large data sets, can be of de-

tions, or toxicity) from genomic data, which are used to identify drug candidates (virtual screening), to use existing drugs for new purposes (drug repurposing) or to design new drugs (Paul et al. 2021).

Clinical practice: ML has enabled new diagnostic methods by linking genomic data with clinical data (such as on disease progression or medical images). Such new methods include the diagnosis of rare genetic diseases in children by linking altered facial features, symptoms and genetic changes (Gurovich et al. 2019) or methods to analyze minute amounts of DNA and other tumor cell components in body fluids (liquid biopsies) (Heidrich et al. 2021). Another strand of development aims to improve risk predictions by combining and weighing very large numbers of individual genetic variants (genome-wide polygenic scores) for a range of important and common diseases to such an extent that they may be widely used clinically (Lello et al. 2019).

Challenges and fields of action

AI-based understanding, quality of knowledge, and potential implications: Although causality and causal modeling have become an active area of research in AI, currently established ML methods for analyzing large and complex data are still based on statistical modeling and do not reflect true causal relations but correlative associations (Schölkopf 2019). In contrast, current self-understanding of scientific explanation and associated quality criteria for (causal) knowledge in biomedical research rely on experimental intervention to reveal causal processes and interactions that lead to the event (e. g., a disease) to be explained (MacArthur et al. 2014; Soldner and Jaenisch 2018). Accordingly, rigorous validation of such correlation-based model predictions on putative gene functions or physiological pathways by interventions in molecular and cellular processes remains necessary. Due to the very large and increasing numbers of genetic

variants associated with complex traits and the complexity of implicated gene and protein networks, such validation is costly and dependent on the availability of suitable human cell models or model organisms.

In the future, there could thus be a shift from scientific explanation based on experimentally validated causal relationships towards explanation largely derived from AI predictions. This would not only challenge the current self-conception in basic research regarding scientific understanding and the quality of knowledge, but also pose challenges for application-oriented research. The latter is supported by retrospective studies of drug approvals, suggesting that demonstrating a causal genetic link between the drug target and the disease significantly increases the likelihood of successful drug development (King et al. 2019).

certain geographical area and data are subject to less strictly regulated access by security and law enforcement authorities (Dove et al. 2015; Kolata and Murphy 2018). Possible privacy issues that may lead to (re-)identification and discrimination risks of data donors also pose a major challenge beyond ethical and legal issues, because academic as well as industrial research relies on a sufficiently high representativeness of data sets. This requires that as many patients and study participants as possible make their data available. Ethnic groups in particular must not be underrepresented in the databases (Sirugo et al. 2019) in order to avoid biases and to generate benefits for as many people as possible.

In the area of private R & D, in addition to big tech companies, large pharmaceutical companies have increasingly invested

It remains unclear what influence explainability actually has on trust in and acceptance of AI-based systems.

While causal mechanistic accounts of understanding prevail in basic and translational biomedical research, in clinical practice and evidence-based medicine difference-making probabilistic concepts of causation are the centerpiece. Using randomized controlled trials as their most important tool, they usually only provide black-box causal claims about the (statistical) effectiveness of interventions in a studied population, without providing a mechanistic explanation (König et al. 2021). Therefore, and as the value of mechanistic knowledge is controversial among practitioners (Andersen 2012; Reiss and Ankeny 2016), the impact of a potential quality loss in causal-mechanistic knowledge through AI is much less clear. Yet causal knowledge on mechanisms can play a role in the interpretation of clinical trials (Andersen 2012) as well as in diagnosis when cases are rare or complex (Brush Jr. et al. 2017).

Research and data infrastructures as well as data governance: Large amounts of high-quality genomic and other ‘omic’ data as well as health data are essential for ML methods (Saunders et al. 2019; Wainberg et al. 2018). Harnessing possible benefits from these techniques for biomedicine would thus require the (further) development of large and diverse biobanks (Denny 2019) as well as international initiatives which link national genome and health data and allow as many researchers as possible to share and access data (Powell 2021; Saunders et al. 2019). This poses considerable challenges linked to sufficient data processing and storage capacity, broad implementation of common technical standards, such as the FAIR principles (Wilkinson et al. 2016), high data security, and solutions that enable data sovereignty (Phillips et al. 2020; Powell 2021; Saunders et al. 2019).

Although commercial or ‘community’ cloud computing services can in principle solve these challenges (Langmead and Nellore 2018), problems may arise regarding regulation and/or privacy aspects, e. g., if the cloud services are located outside a

in AI for diagnosis and drug development and have entered into numerous collaborations with innovative start-up companies. While data might become concentrated in pharmaceutical companies, start-ups are better placed to produce innovative approaches to using data, as they are usually more agile. They can take risks, pivot, focus on niche markets and be disruptive. Large companies often become sustainers as they chase quarterly results. They concentrate on incremental innovations that support their business models. Operating like a start-up through largely autonomous entities or collaborating with real start-ups is considered essential for new AI developments in the biomedical domain (DKA 2019). Large companies thus need to engage directly with start-ups by providing equity as well as access to resources such as technology and data. In Europe, however, the financing of such start-ups is a structural weakness due to various problems generally associated with the European venture capital landscape (DKA 2019).

Explainability, evaluation, and approval: Important ML methods for genomic medicine, especially DL models, have a distinct ‘black box’ character. They are difficult for humans to explain or interpret in terms of how and/or why a result is produced (Lipton 2018). This challenge is particularly severe for systems that are continuously learning and changing (Babic et al. 2019). The lack of explainability or interpretability is widely considered to be particularly important in the medical domain because of the high risks for human lives associated with potential errors and biases in models and data. Expectations are therefore high for so-called explainable AI systems (Arrieta et al. 2020). Despite the importance often attributed to explainability or interpretability, existing and proposed guidelines and regulations for AI in general, as well as for software as a medical device in particular (Ordish et al. 2019), currently lack clear standards for explainability or interpretability in both Europe and the USA. For

instance, in the USA, guidelines by the U.S. Food and Drug Administration (FDA) urge developers to provide information such as an “explanation of how the software works” (FDA 2019 b, 26), and physicians’ ability to “independently review the basis for the recommendations” is considered important in determining whether software should be regulated (FDA 2019 a, 8). Similarly, the recently proposed EU Regulation on AI (AI Act) demands that high-risk AI systems are designed in a way “that their operation is sufficiently transparent to enable users to interpret the system’s output and use it appropriately” (EC 2021, Article 13).

However, it remains largely unclear how information toward such transparency should look like in practice. In addition, the

that the AI system in question can or will actually improve the quality of care for specified patient groups. Trust is closely related to acceptance and both may be generated by explainability. However, since explainability cannot replace validation of AI systems with regard to medical outcomes and patient benefits by clinical trials, the generation of trust by ‘plausible’ explanations on how AI systems work just for the sake of pushing the diffusion of AI systems is highly problematic.

If risk predictions and disease progression prognosis, both in the direct-to-consumer and the regulated domain, should become available for a growing number of diseases and people, further challenges on the societal, health care system and individual

The generation of trust by ‘plausible’ explanations on how AI systems work just for the sake of pushing the diffusion of AI systems is highly problematic.

conceptual problem arises that for current ML systems it is only possible to explain how correlations and the predictions based on them are obtained, but not to draw causal conclusions (Pearl 2010). Thus, policy makers and regulators need to agree on more concrete information developers should provide with regard to the functionality of AI systems and on what role explainability can or should play for the approval of AI systems in comparison to complex, rigorous clinical trials (and possibly post-market monitoring).

Ethical and social implications: In the application of AI-based procedures in medical care, like diagnosis or treatment selection, trust and acceptability on the part of physicians and patients are seen as being of decisive importance (Arrieta et al. 2020; Kelly et al. 2019). Various EU policy documents, including the Ethics Guidelines for Trustworthy AI (HLEG 2019) and the European Commission’s White Paper on AI (EC 2020), as well as the recent proposal for the EU’s AI Act (EC 2021) aim at trustworthy AI. Transparency in the form of explainability or interpretability of AI systems (in the AI Act, in particular) is delineated as an important element to achieve this aim and the ethical use of AI.

However, given rather weak and contradictory empirical evidence, it remains unclear what influence explainability or interpretability actually have on trust in AI-based systems or their recommendations, compared to other factors, such as marketing, clinical trial data, or the regulatory environment. Studies on the diffusion or adoption of medical innovations also indicate that the adoption of applications is a complex social process (Azoulay 2002; Lublóy 2014). Thus, there is currently a lack of empirical evidence on how AI applications and their governance need to be designed in order to create and deserve sustainable trust and acceptance in medical AI systems. Before thinking about creating trust and acceptance, it is, of course, paramount

level may arise. There would be a growing need for trained human genetic counselors who can help healthy and diseased individuals to make informed decisions (Heyen 2016), considering uncertainties arising from the potentially changed understanding of knowledge through AI systems. Already known social and psychological issues of health predictions or diagnoses based on genetics – such as societal pressure on individuals with regard to lifestyle related diseases or the right not to know (also of biological relatives) in case of a lack of therapeutic options (Voorwinden et al. 2020) – may be exacerbated if such AI systems are widely adopted. Health inequalities may be increased in case of a lack of financial support by health insurances for effectively health enhancing but costly genetic testing and AI-based diagnoses or prognoses. Research on these issues, especially by actively involving patients or the broader public is still scarce compared to the rapidly increasing technological possibilities.

Conclusion

In view of the epistemological, economical, technical, ethical and social challenges outlined above, as well as the current scarcity of evidence on how to best govern them, more research and efforts to experimental policy making are urgently needed. Given the complexity and wide scope of these challenges, broad societal debate and mutual learning by different forms of inclusive dialog activities, will be needed to improve research agendas and current regulatory proposals. These activities should involve stakeholders and publics – in order to harness the potentials and minimize risks for improved quality of care and life. By exploring and providing an overview of possible applications, actors and challenges, the article strives to help to identify and discuss most realistic scenarios in the needed societal dialogue.

Acknowledgement

Funding has been provided by the German Federal Ministry of Education and Research, Grant/Award Number: 16ITA201A/B.

References

- Andersen, Holly (2012): Mechanisms. What are they evidence for in evidence-based medicine? In: *Journal of evaluation in clinical practice* 18 (5), pp. 992–999. <https://doi.org/10.1111/j.1365-2753.2012.01906.x>
- Arrieta, Alejandro et al. (2020): Explainable artificial intelligence (XAI). Concepts, taxonomies, opportunities and challenges toward responsible AI. In: *Information Fusion* 58, pp. 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Azoulay, Pierre (2002): Do pharmaceutical sales respond to scientific evidence? In: *Journal of Economics & Management Strategy* 11 (4), pp. 551–594. <https://doi.org/10.1111/j.1430-9134.2002.00551.x>
- Babic, Boris; Gerke, Sara; Evgeniou, Theodoros; Cohen, I. Glenn (2019): Algorithms on regulatory lockdown in medicine. In: *Science* 366 (6470), pp. 1202–1204. <https://doi.org/10.1126/science.aay9547>
- Bailey, Matthew et al. (2018): Comprehensive characterization of cancer driver genes and mutations. In: *Cell* 174 (4), pp. 1034–1035. <https://doi.org/10.1016/j.cell.2018.07.034>
- Brasil, Sandra; Pascoal, Carlota; Francisco, Rita; Dos Reis Ferreira, Vanessa; Videira, Paula; Valadao, Goncalo (2019): Artificial intelligence (AI) in rare diseases. Is the future brighter? In: *Genes* 10 (12), pp. 1–24. <https://doi.org/10.3390/genes10120978>
- Brush Jr., John; Sherbino, Jonathan; Norman, Geoffrey (2017): How expert clinicians intuitively recognize a medical diagnosis. In: *The American journal of medicine* 130 (6), pp. 629–634. <https://doi.org/10.1016/j.amjmed.2017.01.045>
- Denny, Joshua et al. (2019): The ‘All of Us’ research program. In: *New England Journal of Medicine* 381 (7), pp. 668–676. <https://doi.org/10.1056/NEJMSr1809937>
- DKA – Deep knowledge analytics “Pharma Division” (2019): AI for drug discovery, biomarker development and advanced R & D landscape overview 2019/Q3. Available online at <http://analytics.dkv.global/data/pdf/AI-for-DD-Q3-2019/AI-for-Drug-Discovery-Q3-2019-Teaser.pdf>, last accessed on 05.10.2021.
- Dove, Edward; Joly, Yann; Tassé, Anne-Marie; Knoppers, Bartha (2015): Genomic cloud computing. Legal and ethical points to consider. In: *European Journal of Human Genetics* 23 (10), pp. 1271–1278. <https://doi.org/10.1038/ejhg.2014.196>
- EC – European commission (2020): White paper on artificial intelligence. A European approach to excellence and trust. Available online at https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_de, last accessed on 28.09.2021.
- EC – European commission (2021): Proposals for a regulation of the European Parliament and of the Council. Laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Available online at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>, last accessed on 28.10.2021.
- FDA – U.S. Department of Health and Human Services, Food and Drug Administration (2019 a): Clinical decision support software. Draft guidance for industry and food and drug administration staff. 27.09.2019. Available online at <https://www.fda.gov/media/109618/download>, last accessed on 28.10.2021.
- FDA – U.S. Food and Drug Administration (2019 b): Developing a software pre-certification program. A working model. V1.0. Available online at <https://www.fda.gov/media/119722/download>, last accessed on 28.10.2021.
- Gurovich, Yaron et al. (2019): Identifying facial phenotypes of genetic disorders using deep learning. In: *Nature Medicine* 25, pp. 60–64. <https://doi.org/10.1038/s41591-018-0279-0>
- Heidrich, Isabel; Aćkar, Lucija; Parinanz, Mossahebi Mohammadi; Pantel, Klaus (2021): Liquid biopsies. Potential and challenges. In: *International Journal of Cancer* 148 (3), pp. 528–545. <https://doi.org/10.1002/ijc.33217>
- Heyen, Nils (2016): Towards a technocratic biomedicine? In: *Soziale Welt* 67 (4), pp. 389–406. <https://doi.org/10.5771/0038-6073-2016-4-389>
- HLEG – High-Level Expert Group on Artificial Intelligence (2019): Ethics guidelines for trustworthy AI. Available online at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, last accessed on 28.09.2021.
- Kelly, Christopher; Karthikesalingam, Alan; Suleyman, Mustafa; Corrado, Greg; King, Dominic (2019): Key challenges for delivering clinical impact with artificial intelligence. In: *BMC medicine* 17 (195), pp. 1–9. <https://doi.org/10.1186/s12916-019-1426-2>
- King, Emily; Davis, Wade; Degner, Jacob (2019): Are drug targets with genetic support twice as likely to be approved? In: *PLoS Genetics* 15 (12), pp. 1–20. <https://doi.org/10.1371/journal.pgen.1008489>
- Kolata, Gina; Murphy, Heather (2018): The golden state killer is tracked through a thicket of DNA, and experts shudder. In: *The New York Times*, 27.04.2018. Available online at <https://www.nytimes.com/2018/04/27/health/dna-privacy-golden-state-killer-genealogy.html>, last accessed on 05.10.2021.
- König, Harald (2020): Artificial intelligence in human genomics. Mapping research and existing and emerging applications. Karlsruhe: Institute for Technology Assessment and Systems Analysis, pp. 1–9. <https://doi.org/10.5445/IR/1000130769>
- König, Harald; Frank, Daniel; Baumann, Martina; Heil, Reinhard (2021): AI models and the future of genomic research and medicine. True sons of knowledge? In: *Bioessays* 43 (2100025), pp. 1–12. <https://doi.org/10.1002/bies.202100025>
- Langmead, Ben; Nellore, Abhinav (2018): Cloud computing for genomic data analysis and collaboration. In: *Nature Reviews Genetics* 19 (4), pp. 208–219. <https://doi.org/10.1038/nrg.2017.113>
- Lello, Louis; Raben, Timothy; Yong, Soke; Tellier, Laurent; Hsu, Stephen (2019): Genomic prediction of 16 complex disease risks including heart attack, diabetes, breast and prostate cancer. In: *Scientific Reports* 9 (15286), pp. 1–16. <https://doi.org/10.1038/s41598-019-51258-x>
- Lipton, Zachary (2018): The mythos of model interpretability. In: *Queue* 16 (3), pp. 31–57. <https://doi.org/10.1145/3236386.3241340>
- Lublóy, Ágnes (2014): Factors affecting the uptake of new medicines. A systematic literature review. In: *BMC health services research* 14 (469), pp. 1–25. <https://doi.org/10.1186/1472-6963-14-469>
- MacArthur, Daniel et al. (2014): Guidelines for investigating causality of sequence variants in human disease. In: *Nature* 508, pp. 469–476. <https://doi.org/10.1038/nature13127>
- Ordish, John; Murfet, Hannah; Hall, Allison (2019): Algorithms as medical devices. Cambridge, U.K.: PHG Foundation. Available online at <https://www.phgfoundation.org/media/74/download/algorithms-as-medical-devices.pdf?v=1&inline=1>, last accessed on 28.09.2021.
- Paul, Debleena; Sanap, Gaurav; Shenoy, Snehal; Kalyane, Dnyaneshwar; Kalia, Kiran; Tekade, Rakesh (2021): Artificial intelligence in drug discovery and development. In: *Drug discovery today* 26 (1), pp. 80–93. <https://doi.org/10.1016/j.drudis.2020.10.010>
- Pearl, Judea (2010): An introduction to causal inference. In: *The International Journal of Biostatistics* 6 (2), pp. 1–59. <https://doi.org/10.2202/1557-4679.1203>

- Phillips, Mark et al. (2020): Genomics. Data sharing needs an international code of conduct. In: *Nature* 578 (7793), pp. 31–33. <https://doi.org/10.1038/d41586-020-00082-9>
- Powell, Kendall (2021): The broken promise that undermines human genome research. In: *Nature* 590 (7845), pp. 198–201. <https://doi.org/10.1038/d41586-021-00331-5>
- Reiss, Julian; Ankeny, Rachel (2016): Philosophy of medicine. In: *The Stanford Encyclopedia of Philosophy*. Available online at <https://plato.stanford.edu/entries/medicine/>, last accessed on 28. 09. 2021.
- Saunders, Gary et al. (2019): Leveraging European infrastructures to access 1 million human genomes by 2022. In: *Nature Reviews Genetics* 20 (11), pp. 693–701. <https://doi.org/10.1038/s41576-019-0156-9>
- Schölkopf, Bernhard (2019): Causality for Machine Learning. Available online at <https://arxiv.org/abs/1911.10500>, last accessed on 28. 09. 2021.
- Sirugo, Giorgio; Williams, Scott; Tishkoff, Sarah (2019): The missing diversity in human genetic studies. In: *Cell* 177 (4), pp. 26–31. <https://doi.org/10.1016/j.cell.2019.02.048>
- Soldner, Frank; Jaenisch, Rudolf (2018): Stem cells, genome editing, and the path to translational medicine. In: *Cell* 175 (3), pp. 615–632. <https://doi.org/10.1016/j.cell.2018.09.010>
- Voorwinden, Jan et al. (2020): Cognitive and affective outcomes of genetic counselling in the Netherlands at group and individual level. A personalized approach seems necessary. In: *European journal of human genetics* 28 (9), pp. 1187–1195. <https://doi.org/10.1038/s41431-020-0629-5>
- Wainberg, Michael; Merico, Daniele; Delong, Andrew; Frey, Brendan (2018): Deep learning in biomedicine. In: *Nature biotechnology* 36 (9), pp. 829–838. <https://doi.org/10.1038/nbt.4233>
- Wilkinson, Mark et al. (2016): The FAIR guiding principles for scientific data management and stewardship. In: *Scientific Data* 3 (160018), pp. 1–9. <https://doi.org/10.1038/sdata.2016.18>
- Zou, James; Huss, Mikael; Abid, Abubakar; Mohammadi, Pejman; Torkamani, Ali; Telenti, Amalio (2019): A primer on deep learning in genomics. In: *Nature Genetics* 51 (1), pp. 12–18. <https://doi.org/10.1038/s41588-018-0295-5>

REINHARD HEIL

is a philosopher and senior researcher at ITAS, KIT. His main research areas are social consequences of artificial intelligence and transhumanism/human enhancement.

DR. NILS B. HEYEN

is a sociologist and works as a senior researcher in the Competence Center Emerging Technologies at Fraunhofer ISI. Among his main research areas are technical and social innovations in medicine and health care as well as the science-society relationship.

MARTINA BAUMANN

is a member of the research group Life, Innovation, Health, and Technology at ITAS, KIT, where she focuses on ethical and social aspects of bio- and medical technologies.

DR. BÄRBEL HÜSING

is a biologist and works as a senior researcher in the Competence Center Emerging Technologies at Fraunhofer ISI. Her main research areas are innovations in biotechnology, bioeconomy and in the life sciences.

DR. DANIEL BACHLECHNER

is a social scientist and economist. He works as head of Advanced Data Analytics at Fraunhofer Austria. His research focuses on artificial intelligence methods, privacy-preserving technologies and the economics of data sharing and use.

PROF. DR. ULRICH SCHMOCH

is a senior researcher in the Competence Center Emerging Technologies at Fraunhofer ISI. His main research areas are patent and publication statistics and characteristics of emerging technologies.

DR. HARALD KÖNIG

is a senior researcher at ITAS, KIT, since 2011, after having worked as a research group leader in molecular biology. His work focuses on the assessment and governance of emerging biotechnologies.

RESEARCH ARTICLE

Mitarbeiterfreundliche Implementierung von KI-Systemen im Hinblick auf Akzeptanz und Vertrauen

Erarbeitung eines Forschungsmodells auf Basis einer qualitativen Analyse

Maria Jung, Fachbereich Gesellschaftswissenschaften, Hochschule Darmstadt, Haardtring 100, 64295 Darmstadt, DE (maria.jung@h-da.de)
Jörg von Garrel, Fachbereich Gesellschaftswissenschaften, Hochschule Darmstadt, Darmstadt, DE (joerg.vongarrel@h-da.de)  0000-0002-3617-1798

37

Zusammenfassung • Der Einsatz künstlicher Intelligenz (KI) in produzierenden Unternehmen bietet Chancen und Potenziale. KI kann neben der Wettbewerbsfähigkeit von Betrieben auch die arbeitnehmerische Selbstbestimmtheit fördern. Derzeit wird die Bedeutung der Mitarbeiter für einen effektiven und effizienten Einsatz von KI-Systemen oftmals zu wenig berücksichtigt, da der Fokus überwiegend auf der Technologie liegt. Aus diesem Grund wurde eine qualitative Studie durchgeführt, die die beiden Faktoren „mitarbeiterfreundliche Implementierung“ und „mitarbeiterfreundlicher Betrieb“ in Bezug auf Akzeptanz und Vertrauen von KI-Systemen analysiert. Aus den Erkenntnissen wurde ein prozessorientiertes Forschungsmodell konzipiert, das auf der Adoptionstheorie von Rogers basiert und Einflüsse verschiedener Technologieakzeptanzmodelle sowie akzeptanz- und vertrauensfördernde Faktoren umfasst. Die Ergebnisse zeigen, dass eine wahrgenommene Arbeitserleichterung und sichtbare Erfolgserlebnisse zu einer erhöhten Handlungsakzeptanz beitragen können.

Employee-friendly implementation of AI systems in terms of acceptance and trust. Development of a research model based on a qualitative analysis

Abstract • The use of artificial intelligence (AI) in manufacturing companies offers opportunities and potentials. In addition to the competitiveness of companies, AI can also promote self-determination by employees. Currently, too little attention is often paid to the importance of employees for effective and efficient use of AI systems as the focus is predominantly on the technology. Therefore, a qualitative study was conducted to analyze the two factors “employee-friendly implementation” and “employee-friendly operation” in terms of acceptance and trust of AI systems. From the findings, a process-oriented research model was developed based on Rogers’ adoption theory and including influences from different technology acceptance models as well as factors promoting acceptance and trust. The results show that perceived ease of work and sense of achievement can contribute to increased acceptance of action.

Keywords • employee-friendly AI, AI acceptance, AI trust process

Einleitung

Im Wandel hin zur Industrie 4.0 spielt der Einsatz von künstlicher Intelligenz (KI) eine zentrale Rolle, denn dieser bietet vielfältige Potenziale beziehungsweise Chancen für Unternehmen hinsichtlich der Steigerung ihrer Wettbewerbsfähigkeit (Stowasser et al. 2020, S. 5). Durch den Einsatz von KI-Systemen oder KI-basierten Prozessen in zahlreichen Unternehmensbereichen und Sektoren können Produktionsaktivitäten aber auch Geschäftsprozesse effektiv und effizient gestaltet werden (Rammer et al. 2020, S. 15). Häufig wird die Implementierung von KI-Systemen auf organisatorisch-technologischer Ebene betrachtet, die Perspektive der Arbeitnehmer¹ wird hierbei vernachlässigt. Infolgedessen muss insbesondere dem Vertrauensaufbau in Bezug auf Einführung und Nutzung innovativer Technologien eine besonders hohe Relevanz zugeschrieben werden, weil „[diese] die Balance der Kräfte zugunsten der Innovation umschlagen lassen oder diese daran hinder[t], Fahrt aufzunehmen“ (Diekhöner 2018, S. 12).

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.37>
Received: Jun. 14, 2021; revised version accepted: Oct. 18, 2021;
published online: Dec. 20, 2021 (peer review)

1 Begriffe mit spezifischem Genus gelten im Sinne der Gleichbehandlung grundsätzlich für alle Geschlechter. Die verkürzte Sprachform beinhaltet keine Wertung.

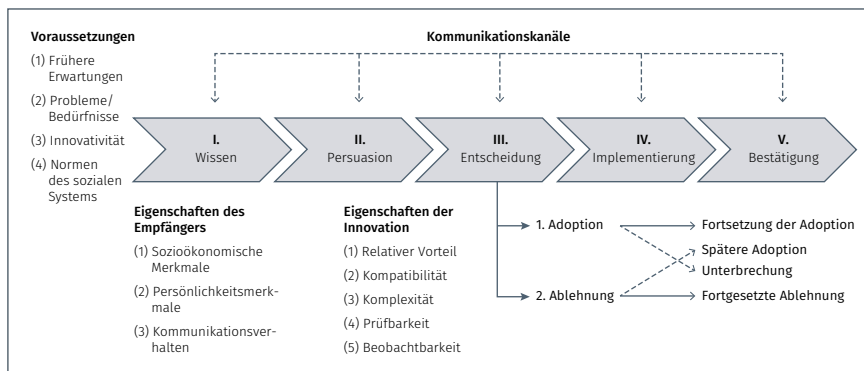


Abb. 1: Innovations-Entscheidungsprozess.

Quelle: nach Rogers 1983, S. 165

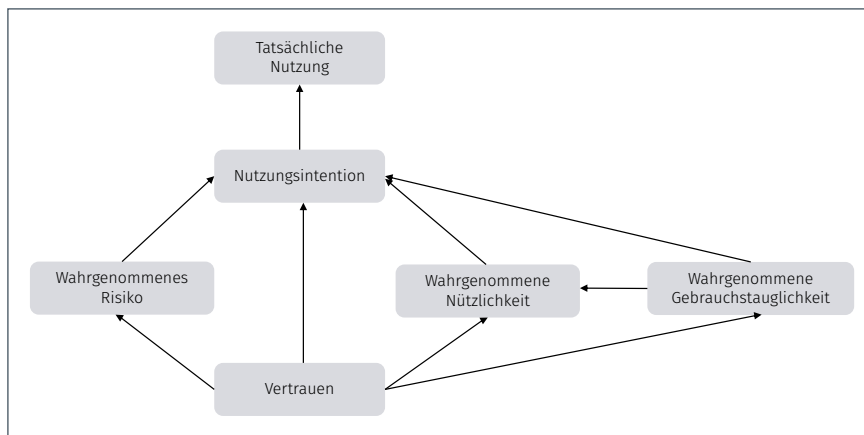


Abb. 2: Vereinfachtes Forschungsmodell zur Integration von Risiko, Vertrauen und TAM.

Quelle: nach Pavlou 2003, S. 122

orientierte Implementierung von KI-Systemen. Idealerweise verläuft dieser Prozess vom ersten Bewusstsein über eine Innovation bis zur bestätigten Nutzung dieser Innovation durch ein Individuum. Ausgangspunkt des Adoptionsprozesses sind (1) die Erfahrungen, (2) die Probleme bzw. Bedürfnisse, (3) die Innovationsneigung einer Person unter Berücksichtigung (4) der Normen des sozialen Systems (Rogers 1983).

Der Adoptionsprozess gliedert sich dabei in fünf Phasen. In der *ersten Phase* des Adoptionsprozesses wird sich das Individuum der Innovation bewusst, es erlangt Kenntnis über die Innovation und Wissen über ihre Funktionsweise (ibid., S. 164). Den Verlauf dieser Phase beeinflussen (1) Sozioökonomische Eigenschaften (z. B. Bildung), (2) Persönlichkeitsmerkmale (z. B. Offenheit) sowie (3) das Kommunikationsverhalten (z. B. Mediennutzung des Individuums) (ibid., S. 165 ff.). In der *zweiten Phase* erfolgt die individuelle Meinungsbildung und im Ergebnis die individuelle Einstellung zur Innovation. Nach Rogers (ibid., S. 14 ff.) üben folgende fünf Faktoren einen entscheidenden Einfluss auf diese Einstellung und auf die Wahrscheinlichkeit der Adoption einer Innovation aus:

1. Relativer Vorteil – Vergleich mit Althergebrachtem oder anderen Innovationen auf dem gleichen Gebiet;
2. Kompatibilität – Grad der Einpassung in bestehende Wertestrukturen, als Anknüpfung an bisherige Erfahrungen und Bedürfnisse der potenziellen Übernahme;
3. Komplexität – Unfähigkeit bzw. Unsicherheit, mit der Komplexität umzugehen;
4. Erprobbarkeit – bei Möglichkeit zum vorherigen Testen ist die Übernahme der Innovation wahrscheinlicher und geht schneller vonstatten;
5. Kommunizierbarkeit/Beobachtbarkeit – Sichtbarkeit/Kommunizierbarkeit der Innovation für andere Personen.

Dieser Phase wird eine besonders hohe Relevanz zugeschrieben, da sie den Grundstein für den weiteren Verlauf hinsichtlich Adoption oder Ablehnung der Innovation legt. In der anschließenden *dritten Phase* erfolgen Adoption oder Ablehnung der Innovation. Da der Adoptionsprozess zeitlich dynamischer Natur ist, ist er im Falle einer Ablehnung nicht notwendigerweise von dauerhafter Natur und es kann bei einem erneuten Durchlaufen des Adoptionsprozesses zu späteren Adoptionen kommen. Im Falle einer Adoption schließt die *vierte Phase* der Nutzung bzw.

Ziel dieses Artikels ist daher die Erarbeitung eines Forschungsmodells für eine mitarbeiterfreundliche Implementierung – hinsichtlich Vertrauens und Akzeptanz bei KI-Systemen in produzierende Unternehmen. Die Entwicklung dieses Modells erfolgt auf Basis der Erkenntnisse einer qualitativen Analyse unter Berücksichtigung der *Innovation Diffusion Theory* (IDT) (Rogers 1983), weiterer Akzeptanzmodelle (Pavlou 2003; Gefen et al. 2003; Backhaus 2017) sowie der Konzepte ‚Arbeitsfähigkeit‘ (Ilmarinen und Tempel 2002) und ‚Nutzerleben‘ beziehungsweise ‚User Experience‘ (Norman et al. 1995).

Theoretischer Rahmen

Um den Prozess der Akzeptanz- und Vertrauensbildung im Rahmen der Implementierung von KI-Systemen aus Mitarbeiterperspektive zu analysieren, bietet sich die recht abstrakte IDT im Sinne eines Innovations-Entscheidungsprozess (Adoptionsprozess) an. Dieser Ansatz ermöglicht eine subjektorientierte Betrachtung des Adoptionsprozesses einer Innovation und eignet sich somit potenziell für eine mitarbeiterzentrierte und prozess-

der Implementierung der Innovation an, die durch die Demonstration im konkreten Kontext gekennzeichnet ist. Hierbei spielen die gesammelten Erfahrungen mit dem Produkt eine zentrale Rolle. Denn fallen diese positiv aus, so kommt es zur Bestätigung und folglich zur Adoption. Die *fünfte Phase* der Bestätigung ist charakterisiert durch eine Informationssuche nach Faktoren, die eine mögliche Dissonanz reduzieren und somit die getroffene Entscheidung unterstützen (Rogers 1983, S. 184). Die Adoption wird nicht fortgeführt, wenn überlegenere Innovationen zur Verfügung stehen oder das Individuum mit der Innovation unzufrieden ist (Rogers 1983, S. 209).

Mit Bezug auf das dargestellte Ziel einer akzeptanz- und vertrauensförderlichen Implementierung kann der Innovations-Entscheidungsprozess nach Rogers insgesamt als wichtiger Orientierungsrahmen zur Gestaltung der Einstellungs-, Handlungs- und Nutzungsakzeptanz gelten (Kollmann 1998). Um zusätzlich die Dimension der Gestaltung des Vertrauens im Mensch-Technik-Kontext zu berücksichtigen, bieten sich Modelle zur Technologieakzeptanz an (Backhaus 2017, S. 23).

Stehen im von Davis (1989) erarbeiteten Technologieakzeptanzmodell (TAM) die wahrgenommene Gebrauchstauglichkeit (*Perceived Ease of Use*) und die wahrgenommene Nützlichkeit des Systems (*Perceived Usefulness*) im Mittelpunkt, wurden in späteren Modellen – u. a. TAM 2 (Venkatesh und Davis 2000) oder TAM3 (Venkatesh und Bala 2008) – zusätzliche Einflussfaktoren (z. B. Geschlecht, Alter, Erfahrung mit dem System) ergänzt. Insbesondere das Modell nach Pavlou (2003) stellt in diesem Kontext einen häufig verwendeten Modellansatz dar, der das Konstrukt Vertrauen als Einflussvariable der Technologieakzeptanz berücksichtigt (siehe Abb. 2):

Strukturell wirken in diesem Modell – analog zum Technologieakzeptanzmodell – der wahrgenommene Nutzen sowie die wahrgenommene Gebrauchstauglichkeit positiv auf die Nutzungsintention. Pavlou (2003, S. 106 ff.) ergänzt in seinem Modell aber als weiteres, negativ zur Nutzungsintention korrelierendes Konstrukt das wahrgenommene Risiko. Alle diese vier Faktoren werden wiederum direkt vom (Nutzer-)Vertrauen beeinflusst: Vertrauen wirkt auf (1) den wahrgenommenen Nutzen, da erst durch das Vorhandensein eines (Nutzer-)Vertrauens eine effektive und effiziente Nutzung ermöglicht wird, (2) die wahrgenommene Gebrauchstauglichkeit, da durch Vertrauen die Komplexität der Situation reduziert werden kann, (3) das wahrgenommene Risiko, da vorhandenes (Nutzer-)Vertrauen risikomindernd wirkt sowie (4) die Nutzungsintention, da vorhandenes (Nutzer-)Vertrauen die Absicht der Nutzung erhöht. Welche Faktoren Einfluss auf das Vertrauen haben, verdeutlicht dieses Modell aber nicht.

Hinsichtlich der logischen Wirkbeziehungen zwischen (Nutzer-)Vertrauen und Technikakzeptanz besteht in der Literatur wenig Konsens. Anders als das genannte Modell von Pavlou kommen bspw. Gefen et al. (2003, S. 74 ff.) zu dem Schluss (siehe Abb. 3), dass sich die wahrgenommene Gebrauchstauglichkeit sowohl auf den wahrgenommenen Nutzen als auch auf das Vertrauen auswirkt, wobei das Vertrauen wiederum auch

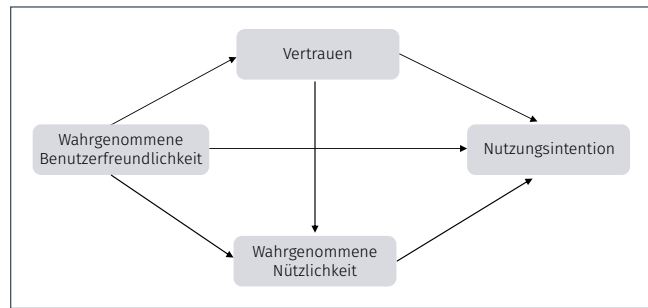


Abb. 3: Forschungsmodell zu Integration von TAM und Vertrauen.

Quelle: nach Gefen et al. 2003, S. 71

auf den wahrgenommenen Nutzen wirkt. Alle drei Faktoren beeinflussen dabei die Nutzungsintention. Gerade die wahrgenommene Gebrauchstauglichkeit beeinflusst damit das Vertrauen.

In dem Modell von Backhaus (2017) wird der Einfluss des Faktors Vertrauen auf die Technologieakzeptanz um das Konstrukt des Nutzerlebens erweitert. Im Gegensatz zur (wahrgenommenen) Gebrauchstauglichkeit erweitert Nutzererleben diese funktionale Perspektive um Erfahrungen und Erlebnisse eines Nutzers mit der Technologie. Backhaus kommt zu dem Ergebnis, dass Vertrauen stark mit dem Nutzererleben zusammenhängt und sich diese beiden Faktoren gegenseitig bedingen. Weiterhin wird Vertrauen durch die Eigenschaften des Technologieanbieters sowie der Technologie beeinflusst (Backhaus 2017, S. 187): „Zukünftige Studien sollten Vertrauen mit in die Untersuchungen einbeziehen und das Nutzererleben weiter fassen, so z. B. im Kontext des Kundenerlebens“ (Backhaus 2017, S. 88).

Im industriellen Sektor werden mit KI-Systemen meist automatisierte Entscheidungsfindungssysteme bezeichnet, welche den Menschen anwendungsorientiert unterstützen und auch als industrielle KI betitelt werden (Bundesregierung 2018, S. 22). Da die Art der Unterstützung zwischen Menschen und KI-Systemen vielfältiger Natur sein kann, muss – vor dem Hintergrund der genannten akzeptanz- und vertrauensgestaltenden Faktoren – eine detailliertere Auseinandersetzung mit unterschiedlichen User-Interfaces stattfinden (nach Bergstein 2017 in Apt und Priesack 2019, S. 232).

Ein hierfür geeigneter Ansatzpunkt bzw. Lösungsansatz kann das Konzept der Autonomiestufen in der industriellen Produktion bilden (BMW 2019, S. 13 f.): Diese Autonomiestufen stellen die „[...] kontinuierliche Veränderung der Verantwortung des Anlagenbetriebs vom Menschen hin zum autonomen System“ (BMW 2019, S. 13) dar. Der KI-bedingte Gesamtautonomiegrad eines industriellen Vorgangs kann sich von nicht-autonom (Stufe 0), über teil-autonom (Stufen 1–4), bis hin zu voll-autonom (Stufe 5) unterscheiden. So ist in den ersten Stufen (Stufe 0–2) eine menschliche Einwirkung weiterhin erforderlich, so dass der Mensch nach wie vor die Kontrolle und die Verantwortung über das technologische System hat. Ab der dritten Stufe wächst der Anspruch an Zuverlässigkeit sowie Verantwortung des KI-Systems. Bei der höchsten Autonomiestufe (Stufe 5)

kann der Bediener komplett abwesend sein, weil keine Interaktion für den Betrieb der Anlage mehr erforderlich ist (BMWi 2019, S. 14 ff.). Trotz vieler Vorteile von KI weist diese Technologie dennoch Defizite auf. Dazu gehören beispielsweise die Vorkonfiguration von Daten, Umgebungen und Informationen, wodurch deutlich wird, dass die Auswirkungen von KI auf Beschäftigte und deren Arbeit weiterhin vom technologischen Fortschritt abhängen (nach Dengler und Matthes 2015 in Apt und Priesack 2019, S. 223).

Methodisches Vorgehen

Um ein Forschungsmodell für einen ‚KI-Vertrauensprozess‘ zur mitarbeiterfreundlichen Implementierung aufzubauen, ist es sinnvoll, neben (1) dem Adoptionsprozess nach Rogers, der den Prozess des Vertrauens- und Akzeptanzaufbaus – vom ersten Kennenlernen, über die Nutzung bis hin zu letztendlich regelmäßig bestätigten Handeln aus Subjektsicht darstellt, weiterhin (2) diverse Technologieakzeptanzmodelle mit einem Fokus auf Vertrauen (Pavlou 2003; Gefen et al. 2003; Backhaus 2017) sowie (3) das Konzept der Autonomiestufen in der industriellen Produktion heranzuziehen.

Auf dieser theoretischen Basis haben wir ein qualitatives Erhebungsinstrument erarbeitet, das verschiedene Szenarien berücksichtigt, um einer differenzierten Ausgestaltung hinsichtlich des Aufbaus und der Gestalt von Vertrauen und Akzeptanz mit Bezug auf verschiedene industrielle KI-basierte Arbeitssysteme gerecht

Ergebnisse

Die Auswertung der Ergebnisse ist induktiv erfolgt, um ein verifiziertes Forschungsmodell aufzubauen. Die Ergebnisse der qualitativen Studie legen nahe, dass eine erfolgreiche Implementierung von KI-Systemen in Unternehmen, unabhängig von der Autonomiestufe, den Phasen des Adoptionsprozesses nach Rogers folgt. So bildet den Ausgangspunkt die konkrete, unternehmensspezifische Problemlage in Kombination mit den Erfahrungen, die eine Organisation mit bestimmten Technologien gesammelt hat und in Verbindung mit finanziellen Restriktionen sowie technischen Determinanten (u. a. *Fit to System*, *Fit to Data*). Weitere adoptionsförderliche Vorbedingungen bilden gerade vor dem Hintergrund hoch-autonomer KI-Systeme die Innovationsfähigkeit der Organisation, eine innovationsfreundliche Unternehmenskultur sowie ein entsprechendes Engagement des Managements. Gerade die drei letztgenannten Faktoren können sich positiv auf das Vertrauen der Mitarbeiter auswirken, die das Unternehmen oder das Management weitestgehend als Vorbildfunktion anerkennen.

Um neue KI-Systeme erfolgreich einzuführen, betonen die befragten KI-Experten, Manager und Mitarbeiter die Notwendigkeit einer frühzeitigen Aufklärung seitens des Unternehmens. Schon in dieser frühen Phase sollten alle relevanten Organisationsmitglieder Kenntnis über das KI-System und Wissen über dessen Funktionsweise erlangen. Gerade Führungskräfte sollten hier proaktiv auf die Mitarbeiter zugehen und die geplante Implementierung von KI-Systemen thematisieren. In diesem Kon-

Vertrauen in die Organisation als Vorbedingung ermöglicht ein positives Nutzererleben, was wiederum zu Vertrauen in KI-basierte Arbeitssysteme führt.

zu werden. Das erste, moderate Szenario hat ein Arbeitssystem dargestellt, in dem eine Unterstützung von Mitarbeitern durch Objekterkennung und Ausgabe von Hinweisen auf einem Bildschirm erfolgt. Das zweite Szenario hat ein auf KI aufbauendes fahrerloses Transportsystem gezeigt, das eigenständig Tätigkeiten in einem kollaborativen Arbeitsprozess übernimmt. Szenario 3 hat als radikales Szenario autonom agierende KI-basierte Robotik-Lösungen im Bereich der Logistik vorgeführt.

Insgesamt wurden 15 Personen im Rahmen von Einzel- und Gruppeninterviews befragt, bei denen es sich um (1) Geschäftsführer sowie Manager, (2) KI-Experten sowie (3) Mitarbeiter aus der industriellen Produktion/Logistik handelte.² Pro Akteursgruppe wurden fünf Personen interviewt, siehe Abb. 4.

text sollten auch relevante Auswirkungen auf die Arbeitsfähigkeit der Mitarbeiter erörtert werden. Insbesondere die Darstellung der Tätigkeitsfelder der KI-Systeme und hieraus potenziell resultierender Tätigkeitsverluste bis hin zu Ausführungen zur Arbeitsplatzsicherheit sollten direkt thematisiert werden, um Existenzängste und eine damit verbundene Ablehnung der Neuerung zu vermeiden. Vor allem Mitarbeiter haben in den Interviews betont, dass ihnen die transparente Kommunikation der Arbeitsplatzsicherheit ein zentrales Anliegen ist. Um den Verlauf dieser Phase positiv zu beeinflussen, ist es auch sinnvoll, individuelle Eigenschaften der Mitarbeiter, u. a. sozioökonomische Eigenschaften oder auch Persönlichkeitsmerkmale wie Offenheit, sowie das Kommunikationsverhalten der Mitarbeiter als Gelingensbedingungen zu berücksichtigen und in eine zielgruppengerechte Kommunikation einfließen zu lassen. Dadurch kann schon vor dem Einsatz von KI die Basis für ein stärker individualisiertes Nutzererleben gelegt werden. Gerade transparente Kommunikation

² Die Interviews sind im April 2021 online durchgeführt und mit einem Aufnahmegerät aufgezeichnet worden. Anschließend ist die Transkription sowie die Auswertung mit dem Tool f4 erfolgt.

Management/ Führungsposition (M)	Manager 1	Manager 2	Manager 3	Manager 4	Manager 5
	Produktionsleiter bei einem Hersteller von Drähten und Kabelsystemen für Standard- und Spezialleitungen (ca. 650 Mitarbeiter in Deutschland)	Geschäftsführer und Gesellschafter bei einem Hersteller für Materialprüfmaschinen und Prüfsystemen (ca. 80 Mitarbeiter in Deutschland)	Montageleiter bei einem Hersteller für Materialprüfmaschinen und Prüfsystemen (ca. 80 Mitarbeiter in Deutschland)	Manager bei einem Hersteller für Materialprüfmaschinen und Prüfsystemen (ca. 80 Mitarbeiter in Deutschland)	Manager (Strategie) bei einem Anbieter von Produkten und Dienstleistungen in der medizinischen Versorgung (ca. 125.000 Mitarbeiter in über 20 Ländern)
KI-Experten (E)	KI-Experte 1	KI-Experte 2	KI-Experte 3	KI-Experte 4	KI-Experte 5
	Experte aus dem E-Commerce bei einem Süßwarenkonzern (ca. 7.000 Mitarbeiter in 14 Ländern)	Geschäftsführer bei einer Beratungsgesellschaft im Bereich Digitalisierung (ca. 33 Mitarbeiter in Deutschland)	Consultingleiter bei einem führendem Hard- und Softwareentwickler (ca. 150.000 Mitarbeiter in über 100 Ländern)	Projektleiter bei einem deutschen Forschungsinstitut (Bereich KI) (ca. 200 Mitarbeiter in Deutschland)	Projektleiter bei einem deutschen Forschungsinstitut (Bereich KI) (ca. 200 Mitarbeiter in Deutschland)
Mitarbeiter (A)	Arbeiter 1	Arbeiter 2	Arbeiter 3	Arbeiter 4	Arbeiter 5
	Lagerist bei einem Hersteller für Materialprüfmaschinen und Prüfsystemen (80 Mitarbeiter in Deutschland)	Monteur bei einem Hersteller für Materialprüfmaschinen und Prüfsystemen (80 Mitarbeiter in Deutschland)	Produktionsmitarbeiter Fräsetechnik bei einem Prüfmittelhersteller (ca. 40 Mitarbeiter in Deutschland)	Produktionsmitarbeiter bei einem Prüfmittelhersteller (ca. 40 Mitarbeiter in Deutschland)	Abteilungsleiter bei einem Prüfmittelhersteller (ca. 40 Mitarbeiter in Deutschland)

Abb. 4: Charakteristika der Interviewpartner.

Quelle: eigene Darstellung

und Offenheit gegenüber Mitarbeitern durch das Management (als interpersonelle Faktoren) als auch Unternehmenskultur und Innovationsfähigkeit (als strukturelle Faktoren) sind als zentrale Erfolgsfaktoren genannt worden, die „zuversichtliche positive Erwartungen“ (Oswald 2010, S. 63) gegenüber dem Einsatz KI-basierter Systeme im Sinne eines organisationalen Vertrauens aufbauen bzw. verstetigen sollen. Um die Meinungsbildung des Individuums zur KI und damit die Einstellungsakzeptanz positiv zu gestalten, sollte die Darstellung von Nützlichkeit des KI-Systems – im Sinne des Nettonutzens als Differenz zwischen Nutzen und Kosten bzw. Risiken – für jeden einzelnen Mitarbeiter im Fokus stehen. Die weitere Ausgestaltung des Nutzererlebens sollte Eigenschaften, Fähigkeiten und Funktionsweise der KI auf Arbeitsplatzebene beinhalten, aber auch deren Grenzen u. a. mit Bezug auf die jeweilige Autonomiestufe. Zentrale Faktoren aus Sicht der Interviewten sind in diesem Kontext, dass Faktoren wie Gestalt und Haptik des KI-Systems für die Mitarbeiter angenehm und intuitiv gestaltet sind und somit eine hohe, funktionale Gebrauchstauglichkeit gewährleistet ist. Gerade mit Bezug auf die unterschiedlichen Autonomiestufen in der industriellen Produktion bei KI-Systemen hat sich herausgestellt, dass das User Interface eines KI-Systems einen entscheidenden Einfluss auf die Einstellungs- und Handlungsakzeptanz haben kann. Durch eine klare Darlegung der Faktoren ist es für den Mitarbei-

ter möglich, sich eine individuelle Meinung zur Gebrauchstauglichkeit, aber auch zu Nutzen und Risiken des KI-Systems zu bilden. Neben den schon dargestellten Erfolgsfaktoren, die Vorteile der KI-Systeme klar darzulegen und diese kompatibel zu den bisherigen Arbeitstätigkeiten einfach anwendbar zu gestalten, erhöhen in dieser Phase insbesondere eine mögliche Erprobbarkeit und Beobachtbarkeit des KI-Systems die Wahrscheinlichkeit einer Adoption. Daher können Workshops aber auch Demonstratoren sinnvolle Elemente einer positiven Gestaltung des Nutzerlebens sein.

Eine positive Bewertung und somit auch eine positive Einstellungsakzeptanz führt zu einer individuellen Entscheidung durch den Mitarbeiter, das KI-System zu nutzen. Diese Handlungsakzeptanz kann dann zu einer Nutzungsakzeptanz führen, wenn die Erwartungen hinsichtlich Gebrauchstauglichkeit, Nutzen aber auch Risiken (Soll) mit der tatsächlich wahrgenommenen Situation (Ist) übereinstimmen, positiver bewertet werden und somit das Nutzererleben als überwiegend positiv bewertet wird.

Die Ergebnisse der Interviews verdeutlichen, dass vor allem die erfahrene Arbeitserleichterung sowie sichtbare Erfolgserlebnisse zu einer erhöhten Handlungsakzeptanz beitragen können. Gleichzeitig könnten Mitarbeiter durch eine mögliche Arbeitsentlastung profitieren, indem sie nun „sinnstiftende“ Aufgaben erledigen. Werden die Erwartungen an das KI-System in der

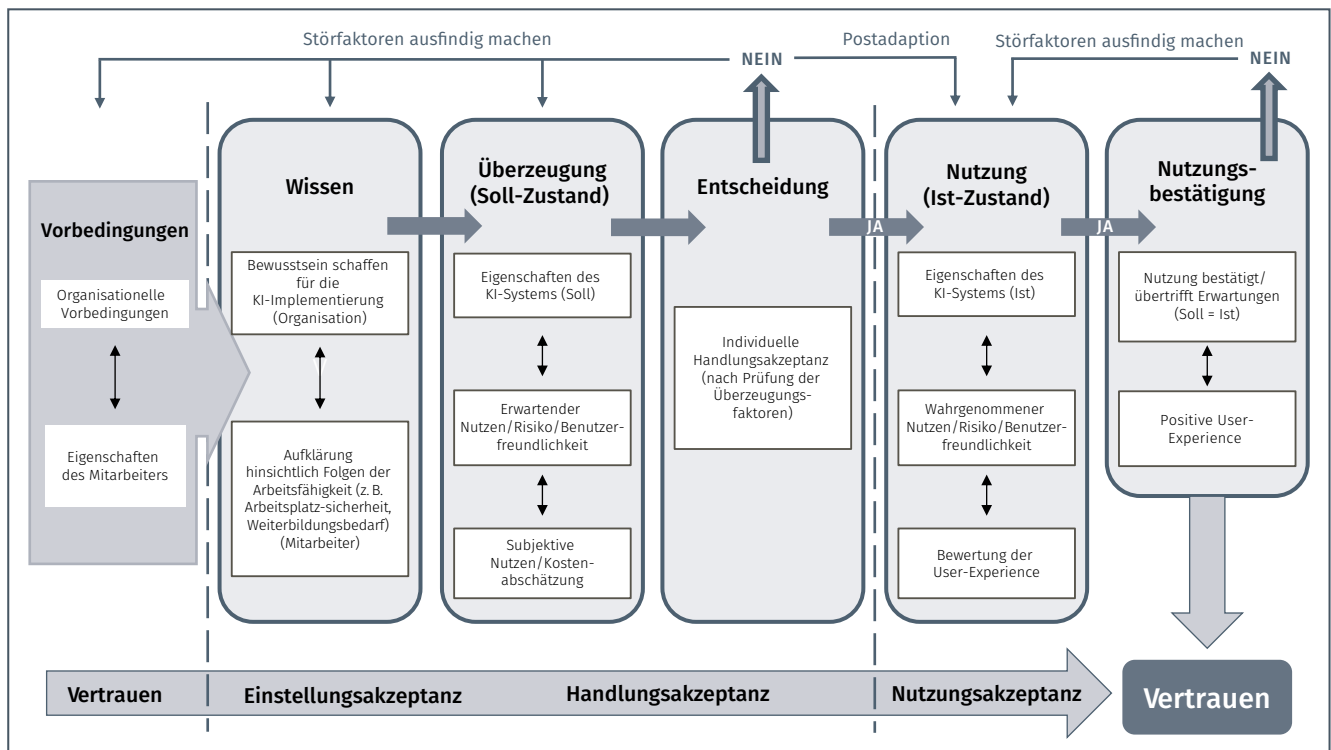


Abb. 5: Prozessorientiertes Forschungsmodell zur Schaffung von Akzeptanz und Vertrauen bei Mitarbeitern hinsichtlich KI.

Quelle: eigene Darstellung

kontinuierlichen Nutzung erfüllt oder sogar übertroffen, führt dies zu einer Nutzungsbestätigung, einer Bestätigung des positiven Nutzerlebens und final zu einem Vertrauen in das KI-System. Unternehmen müssen in diesem Kontext darauf achten, Defizite am System oder sonstige Hindernisse schnellstmöglich zu eliminieren, da das gewonnene Vertrauen der Mitarbeiter durch Systemfehler sonst schnell wieder zunichtegemacht werden kann. In Summe führen die Ergebnisse damit zu einem prozessorientierten Forschungsmodell wie in Abb. 5 dargestellt.

Fazit und Ausblick

Um einen innovativen Vertrauensaufbau bei Mitarbeitern hinsichtlich neu einzuführender KI-Technologien zu generieren, kann neben diversen Technologie-Akzeptanzmodellen insbesondere der Adoptionsprozess nach Rogers (1983) hilfreich sein, um den Prozess des Vertrauens- und Akzeptanzaufbaus – vom ersten Kennenlernen, über die Nutzung, bis hin zum letztendlich regelmäßig bestätigten Handeln – zu gestalten und einen ‚KI-Vertrauensprozesses‘ im Sinne einer mitarbeiterfreundlichen Implementierung zu ermöglichen. Im Gegensatz zu den in diesem Artikel dargestellten Modellen (Pavlou, 2003; Gefen et al. 2003; Backhaus, 2017) scheint Vertrauen aber einerseits als organisatorisches Vertrauen (im Sinne eines interpersonellen und strukturellen Vertrauens) eine Vorbedingung zu sein, um ein positives Nutzerleben im Kontext der industriellen Produktion zu ermög-

lichen. Andererseits ist Vertrauen in das KI-System als zentrales Ergebnis dieses Prozesses anzusehen.

In einem nächsten Forschungsvorhaben wird auf Basis dieser qualitativ generierten Ergebnisse eine für die Praxis relevante Checkliste erarbeitet, die Hinweise für die betrieblichen Entscheider und Umsetzer liefern soll, um KI-Systeme erfolgreich im Unternehmen zu implementieren. Weiterhin werden quantitative Studien zur detaillierteren Analyse der Faktoren Vertrauen und Akzeptanz im Rahmen der Implementierung und Nutzung von KI-Systemen durchgeführt.

Angabe von Finanzierungsquellen

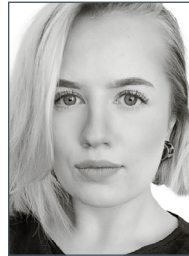
Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 02L19C157 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literatur

- Apt, Wenke; Priesack, Kai (2019): KI und Arbeit. Chance und Risiko zugleich. In: Volker Wittpahl (Hg.): Künstliche Intelligenz. Technologie, Anwendung, Gesellschaft. Berlin: Springer Vieweg, S. 221–238. https://doi.org/10.1007/978-3-662-58042-4_14
- Backhaus, Nils (2017): Nutzervertrauen und -erleben im Kontext technischer Systeme. Empirische Untersuchungen am Beispiel von Webseiten und Cloudspeicherdiensten. Dissertation, Technische Universität Berlin. Online verfügbar unter <https://d-nb.info/1156183804/34>, zuletzt geprüft am 13.10.2021.

- BMWi – Bundesministerium für Wirtschaft und Energie (Hg.) (2019): Technologie-szenario „Künstliche Intelligenz in der Industrie 4.0“. Working Paper. Berlin: Plattform Industrie 4.0. Online verfügbar unter https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/KI-industrie-40.pdf?__blob=publicationFile&v=10, zuletzt geprüft am 13. 10. 2021.
- Bundesregierung (2018): Strategie Künstliche Intelligenz der Bundesregierung. O. O.: o. V. Online verfügbar unter https://www.bmwi.de/Redaktion/DE/Publikationen/Technologie/strategie-kuenstliche-intelligenz-der-bundesregierung.pdf?__blob=publicationFile&v=10, zuletzt geprüft am 13. 10. 2021.
- Davis, Fred (1989): Perceived usefulness, perceived ease of use and user acceptance of information technology. In: MIS Quarterly 13 (3), S. 319–340. <https://doi.org/10.2307/249008>
- Diekhöner, Philipp Kristian (2018): The trust economy. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-662-57459-1>
- Gefen, David; Karahanna, Elen; Straub, Detmar (2003): Trust and TAM in online shopping. An integrated model. In: MIS Quarterly 27 (1), S. 51–90. <https://doi.org/10.2307/30036519>
- Ilmarinen, Juhani; Tempel, Jürgen (2002): Arbeitsfähigkeit 2010. Was können wir tun, damit Sie gesund bleiben? Hamburg: VSA.
- Kollmann, Tobias (1998): Akzeptanz innovativer Nutzungsgüter und -systeme. Konsequenzen für die Einführung von Telekommunikations- und Multimediasystemen. Wiesbaden: Springer Gabler. <https://doi.org/10.1007/978-3-663-09235-3>
- Norman, Donald; Miller, Jim; Henderson, Austin (1995): What you see, some of what's in the future, and how we go about doing it. HI at Apple computer. In: Conference Companion on Human Factors in Computing Systems. New York: Association for Computing Machinery, S. 155. <https://doi.org/10.1145/223355.223477>
- Oswald, Margit (2011): Vertrauen in Organisationen. In: Martin Schweer (Hg.): Vertrauensforschung. State of the art. Frankfurt am Main: Peter Lang, S. 63–85.
- Pavlou, Paul (2003): Consumer acceptance of electronic commerce. Integrating trust and risk with the technology acceptance model. In: International Journal of Electronic Commerce 7 (3), S. 101–134. <https://doi.org/10.1080/10864415.2003.11044275>

- Rammer, Christian; Bertscheck, Irene; Schuck, Bettina; Demary, Vera; Goecke, Henry (2020): Einsatz von Künstlicher Intelligenz in der Deutschen Wirtschaft. Stand der KI-Nutzung im Jahr 2019. Berlin: Bundesministerium für Wirtschaft und Energie.
- Rogers, Everett (1983): Diffusion of Innovations. New York, NY: The Free Press.
- Stowasser, Sascha et al. (2020): Einführung von KI-Systemen in Unternehmen. Gestaltungsansätze für das Change-Management (Whitepaper). München: Lernende Systeme – Die Plattform für Künstliche Intelligenz.
- Venkatesh, Viswanath; Bala, Hillol (2008): Technology acceptance model 3 and a research agenda on interventions. In: Decision Sciences 39 (2), S. 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Venkatesh, Viswanath; Davis, Fred (2000): A theoretical extension of the technology acceptance model. Four longitudinal field studies. In: Management Science 46 (2), S. 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>



MARIA JUNG

studierte Angewandte Sozialwissenschaften an der Hochschule Darmstadt und ist seit Januar 2021 wissenschaftliche Mitarbeiterin im BMBF-geförderten Projekt *Kompetenzzentrum für Arbeit und Künstliche Intelligenz (KomPAKI)*.



PROF. DR. JÖRG VON GARREL

hat eine Professur für Prozess- und Produktinnovationen an der Hochschule Darmstadt inne. Seine Forschungen fokussieren auf eine partizipative sowie effektive und effiziente Gestaltung von Arbeitssystemen und Arbeitsprozessen vor dem Hintergrund aktueller Entwicklungen der Digitalisierung und des demografischen Wandels.

RAUMFORSCHUNG UND RAUMORDNUNG SPATIAL RESEARCH AND PLANNING

Neu im oekom verlag! Die gedruckte Version erhalten Sie auch im Abo. Infos unter www.oekom.de/rur



Gold Open
Access
rur.oekom.de

RESEARCH ARTICLE

“Don’t let me be misunderstood”

Critical AI literacy for the constructive use of AI technology

Stefan Strauß, Institut für Technikfolgen-Abschätzung (ITA), Österreichische Akademie der Wissenschaften,
Apostelgasse 23, 1030 Wien, AT (ss Strauss@oeaw.ac.at)  0000-0003-1877-2415

44

Abstract • Research and development as well as societal debates on the risks of artificial intelligence (AI) often focus on crucial but impractical ethical issues or on technocratic approaches to managing societal and ethical risks with technology. To overcome this, more practical, problem-oriented analytical perspectives on the risks of AI are needed. This article proposes an approach that focuses on a meta-risk inherent in AI systems: deep automation bias. It is assumed that the mismatch between system behavior and user practice in specific application contexts due to AI-based automation is a key trigger for bias and other societal risks. The article presents the main factors of (deep) automation bias and outlines a framework providing indicators for the detection of deep automation bias ultimately triggered by such a mismatch. This approach intends to strengthen problem awareness and critical AI literacy and thereby create some practical use.

„Don’t let me be misunderstood“. Kritische KI-Kompetenz für den konstruktiven Umgang mit KI-Technologie

Zusammenfassung • Gesellschaftlicher Diskurs sowie Forschung und Entwicklung zu Risiken künstlicher Intelligenz (KI) fokussieren oft einseitig entweder auf praxisferne ethische Aspekte oder auf technokratische Ansätze zur Bewältigung gesellschaftlicher Risiken allein durch Technologie. Es bedarf jedoch praktikabler, problemorientierter Perspektiven. Dieser Beitrag konzentriert sich daher auf ein zentrales Meta-Risiko von KI-Systemen: Deep Automation Bias. Es wird davon ausgegangen, dass Diskrepanzen zwischen Systemverhalten und Nutzungspraktiken in bestimmten Anwendungskontexten aufgrund KI-basierter Automatisierung zentrale Auslöser von Bias und gesellschaftlichen Risiken sind. Der Beitrag stellt zentrale Faktoren von (Deep) Automation Bias vor und entwickelt einen analytischen Rahmen mit Indikatoren zur Erkennung von Diskrepanzen in KI-Systemen. Dieser Ansatz will durch Stärkung von Problembewusstsein und kritischer KI-Kompetenz auch praktischen Nutzen erzielen.

Keywords • deep automation bias, AI assessment, machine learning, uncertainty, awareness

Introduction

The hype around artificial intelligence (AI) is yet unbroken. Machine learning (ML) algorithms gain influence on economic, social and political decisions affecting individuals directly and indirectly. Accordingly, there is a scientific and political debate on how to tackle the various ethical risks of a broader use of AI. These discussions are, though, mostly dominated by either general ethical issues such as human versus machine autonomy, matters of trust, fairness, accountability and transparency (FAT) or on technical solutions to avoid algorithmic discrimination. Correspondingly, there is a number of guidelines for “ethical AI” or “trustworthy AI” issued by the EU Commission’s high-level expert group on AI and others (Floridi et al. 2018; HLEG 2019; AlgorithmWatch 2019; Hallensleben 2020). And a growing community deals with developing technical solutions for de-biasing and FAT-ML, for example in the annual ACM-FAT conferences (Selbst et al. 2019; Wieringa 2020; Eid et al. 2021). Without doubt, this involves various relevant research and development activities.

But there is also a certain gap between important but impractical ethical concepts on the one side and technocratic approaches to fix societal problems with algorithms on the other. Not without irony, this situation could even reinforce the myriad of AI-related risks ranging from bias and discrimination, lacking transparency, erosion of privacy and security, loss of autonomy etc. There is thus need for a broader debate and problem-oriented approaches on how to effectively comprehend and conceptualize socio-technical risks related to AI.

A main argument of this paper¹ is that AI-based automation plays a particular role here. To explore the risks of AI thus requires a stronger analytical focus on automation. To facilitate

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>) <https://doi.org/10.14512/tatup.30.3.44>
Received: Jun. 14, 2021; revised version accepted: Oct. 18, 2021;
published online: Dec. 20, 2021 (peer review)

1 Parts of this paper represent a condensed and modified version of Strauß (2021).

this, I identified deep automation bias (DAB) as a meta-risk of the societal use of AI entailing further risks. DAB is a multi-dimensional, wicked problem inherent to AI technology alluding to progress in deep learning and self-optimizing algorithms (Strauß 2018, 2021). The aim is to develop this concept of DAB further and propose it as part of a problem-oriented assessment framework of AI. The premise here is that essentially, AI-based technology represents a socio-technical system that fosters automation at different levels. Bias can result from pre-existing prejudice during technical development, technical issues like poor data quality, insufficient models or inappropriate operation of ML-algorithms; but also from rule conflicts between AI design

lated, where the information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing” (Churchman 1967 cited in Buchanan 1992, p. 15). Wicked problems bear tensions between the *artificial* and the *natural* (ibid.). This basic conflict can reinforce with the use of AI, particularly due to its high degree of automation: AI transforms decision-making and entails risks of reducing *natural* aspects of society to machine-readable data models that are interpretable by *artificial* algorithms.

Bias in ML is a wicked problem inherent to AI. However, unbiasing and fostering FAT is not sufficient to avoid the re-

*Beyond practicability or misleading technocratic
approaches and underestimation of risks,
raising problem-awareness among decision-makers and persons
interacting with AI systems is essential.*

45

and AI application contexts due to complexity gaps between statistical assumptions in the system and user practices. In each case, the common denominator is automation, though, on different socio-technical levels.

To understand how these levels interact requires a multilayer view on the interplay between design and use of AI technology which together shape societal impacts. The main focus of the paper is thus on how to improve the analytical perspective on AI as a socio-technical issue to foster the basic understanding and awareness on the related societal challenges. This is a contribution towards what I call ‘critical AI literacy’ to avoid the fallacy of seeking for technological fixes for societal problems. The paper is structured as follows: after this introduction, section two briefly discusses why AI bears wicked problems which cannot be addressed with technical means only. Section three then sheds light on critical AI literacy and the role of automation. Based on main factors affecting DAB section four sketches a problem-oriented assessment framework. Section five presents a short summary and concluding remarks.

Wicked problems require more than fairness, accountability and transparency

As several scholars argue, there is need for alternative socio-technical approaches to better grasp the societal and ethical issues of AI (Edwards and Veale 2017; Selbst et al. 2019; Tsamados et al. 2020). This is particularly relevant as the use of AI systems can involve and reinforce so-called wicked problems (Strauß 2021). They are a “class of social system problems which are ill-formu-

lated risks of undetected failure, self-fulfilling prophecies and an incremental normalization of AI biases in society. Sheer techno-fixes could even intensify these risks. Research on FAT and bias in ML is dominated by debates on how different types occur, i. e., preexisting, technical or emergent bias and how to avoid that AI and algorithms lead to discrimination and injustice (Friedman and Nissenbaum 1996; Simon et al. 2020; Wieringa 2020). This is important work but there is a tendency to frame this socio-technical issue as a technological one or to get lost in general ethical debates on fairness, justice etc. and seeking technical solutions to ethical problems. This can be counterproductive. Unbiasing approaches, e. g., with adaptive algorithms, may increase complexity and opacity of AI which further reinforce societal risks.

Obviously, not just the technical design of AI is relevant but in particular, how (in-)compatible the technical system is with its socio-technical application contexts. This is crucial to tackle the risks of AI, which requires a broader, problem-oriented perspective that fosters analytical views on both, technical and societal issues of AI systems. To circumvent one-sided views, like ethical debates beyond practicability or misleading technocratic approaches and underestimation of risks, raising problem-awareness among decision-makers and persons interacting with AI systems is essential. However, as of yet, there is a lack of awareness and analytical perspectives in this regard. I thus suggest to focus more on the specific role of automation in AI and how to establish what I call here critical AI literacy. Critical AI literacy here means the ability to comprehend the core features of an AI system and its (in-)compatibility with its particular application contexts in a (necessarily) more complex sociotechnical reality.

Critical AI literacy: understanding AI-based automation (bias)

A crucial question for the use of AI is whether it matches with the requirements of a particular application context. This implies that the contextual environment of an AI system affects the occurrence of bias. Tsamados et al. (2020) discuss context bias on the example of a healthcare system for resource management in hospitals. The system may function properly for one hospital that fits to the model the system uses but may cause problems in others, e.g., rural clinics with different contextual factors. But as argued, at the core, the various risks of AI ultimately derive from conflicts due to different forms of automation. Automation bias (AB) is the general risk of uncritically accepting the outcome of an automated system (Goddard et al. 2012, 2014). AI intensifies this risk and thus DAB represents a meta-risk of AI. The following examples highlight this:

Even very simple forms of automation can cause serious problems as the case of the automated renaming function in Excel tables shows: studies detected failure rates of 20 per cent implying that every third table containing genetic data presents false information as gene names are automatically renamed to dates (e.g., MARCH1 to 1-Mar). Abeysooriya et al. (2021) show that this problem still exists and recommend human workarounds. Thus, even simple errors may create severe impact. Particularly, if these errors remain undetected and are processed further by AI systems.

Imagine an autopilot-system of an airplane, a classical form of automation. Basically, it is a rule-based system which functions with sensors and real-time data on geolocation, weather etc. Hence it needs a plausible data model of the plane's environment and reliable information on its behavior so that the human pilot can monitor if autopilot and plane operate as intended and can intervene immediately in case of problems. Any hidden error like a faulty label in a data table could threaten human lives. Recent cases of military drones autonomously attacking soldiers in Libya highlight that this is not a sheer theoretical risk (Hambling 2021). AB is a known risk of autopilots (Parasuraman et al. 2010; Goddard et al. 2014), mitigated with extensive training and technical features to improve controllability and avoid overreliance on the system. A precondition here is the basic predictability of system behavior and comprehensible rules determining its functionality. Hence system complexity must remain manageable. An autopilot that would permanently try to optimize a flight (e.g. with some predictability algorithm) without effective human intervention would be uncontrollable. The system would lever out human autonomy and agency and the conflict between system behavior and human intervention could escalate at any time. Tackling this risk requires more than transparency, accountability or explicability and is impossible without plausibility, reliability, predictability and effective intervenability to comprehend and correct the automated system.

Further examples are AI systems for job applications which evidently led to discrimination in various cases. As Harwell

(2019) illustrates, the hiring platform HireVue calculated an "employability score" based on various data on job applicants including facial expressions and speech. Critics filed complaint and argued the system is biased, unfair and deceptive as it discriminates, e.g., due to different facial looks and spoken accents. Another system uses background images in applicants' portray photos to predict job qualification (Harlan and Schnuck 2021). For example, a person standing in front of a bookshelf then has higher chances to get a job offer for certain job sectors than a person submitting a photo with plain background. Obviously, skin color, ethnicity and background images have no relevance for a person's qualification. But people of color or persons with lower contrasting background images generally get lower scores. Hence the system reinforces racial and other forms of discrimination. This is an evident issue of various other AI systems, too. Various cases (O'Neil 2016; Borgesius 2018; Obermeyer et al. 2019; Köchling and Wehner 2020) demonstrate, how problematic it can be to automate social domains with AI. They underline the risk of DAB which is inevitable here if neither job applicants nor recruiters are unaware of the problem and no countermeasures to avoid discrimination are set. In any case, the system behaves unfair and unreliable.

Technical fixes like de-biasing to fix deficient image processing do not solve such problems as they are more than just technical issues. A typical technical solution to the above-mentioned bias would be to modify the algorithm so that it excludes image backgrounds when calculating a qualification score. This may ease bias resulting from images but any other problems with criteria the algorithm may process (e.g., ethnic facial features, residential district) would remain unsolved. Also, FAT is ineffective as transparency on the issue would not prevent from discrimination. Moreover, there are various cases of bias or stereotyping in data models and ML approaches with problematic effects on system behavior. Particularly sensitive is the use of AI in the health domain. Several studies reveal problems and unintended effects of decision-making systems here (Goddard et al. 2012; Cabitza et al. 2017; Gianfrancesco et al. 2018; Obermeyer et al. 2019). Gianfrancesco et al. (2018, p.5) analyzed ML algorithms in clinical applications and found serious issues such as "overreliance on automation, algorithms based on biased data, and algorithms that do not provide information that is clinically meaningful". They conclude that easing these problems requires better understanding of AI and their ML approaches and corresponding measures to achieve this.

Towards a problem-oriented assessment framework

To raise problem-awareness, it is essential to understand how AI-based automation operates and how DAB occurs. The factors and framework presented here are meant as an approach to improve critical AI literacy. Basically, AI becomes problematic when there is a mismatch between system behavior and user

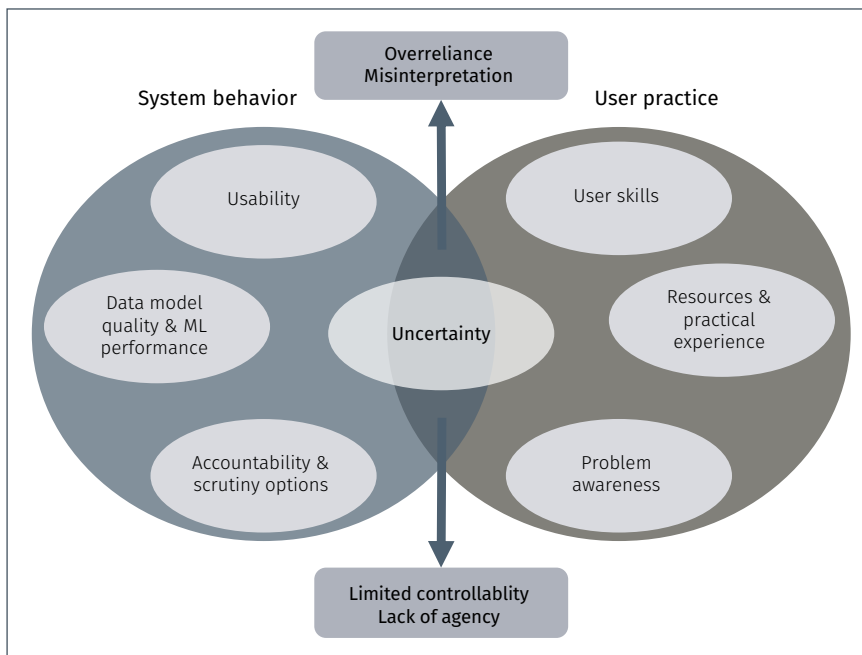


Fig. 1: Factors affecting DAB.

Source: Strauß 2021, p. 8

practice in specific application contexts. To comprehend the meaning of mismatch is a precondition for the assessment of AI-related risks. This implies to understand the peculiarities of AI-based automation. Because irrespective of specific features, every AI system uses some form of automation and bears risks of (D)AB. But as Tsoukiàs (2020) reminds, automation is not inevitable with AI. It is a choice that needs to be legitimated and not an end in itself. It is crucial to scrutinize the automation approach of an AI system when assessing its impact. As a first step to sharpen the analytical lens, I suggest to conceptualize DAB as meta-risk of AI from which other risks derive from.

Main determinants of DAB

DAB addresses the problem of increasing complexity and opacity of AI technology that reinforce AB due to its dynamic, unpredictable and thus potentially uncontrollable behavior (Strauß 2018, 2021). Sophisticated ML approaches with features to self-optimization like deep learning or other forms of unsupervised learning aggravate this problem. It can extend the gap between human propensity to blindly trust AI technology and the limits of technology to match social complexity due to necessarily reductionistic models of society. The wider risk here is that dependence of society on AI systems reaches a stage where automated decisions – no matter if socially acceptable, ethical, correct or false – become uncontrollable. AI then constantly reshapes society and individual lives without effective alternatives. Figure 1 shows basic factors affecting DAB.

DAB bears at least two main risks: at the top is the “classical” risk of overreliance on the behavior of an AI system and misinterpretation; the bottom shows the additional risk of limited

controllability and lack of agency. Both dimensions are interrelated and the severity of DAB depends on the interplay of different factors. The main connecting factor is uncertainty, shaped by technical as well as social issues that can reinforce mutually. System behavior strongly depends on the quality of data models and ML performance, usability, accountability and scrutiny options in the system. The social perspective involves user practices: various studies show that user skills, practical experience, resources (e. g. user knowledge to interpret a system, time and pressure to act), workload and effective options to scrutinize automated procedures affect AB (Goddard et al. 2012, 2014; Lyell and Coiera 2016). DAB further complicates these factors as AI increases system complexity, opacity and decreases options to scrutinize its functionality. Consequently, controllability, agency and options to intervene into automated decision-making can also decrease.

The interplay of all these factors affect the severity of DAB and related risks. It is particularly higher, when the system lacks in options to scrutinize its behavior and/or problem awareness of the human user is low. For example, lacking accountability reduces agency and low problem awareness limits the user’s ability to scrutinize, which again limits agency (Strauß 2021).

How to assess system behavior

Figure 2 sketches a three-level framework with indicators to identify DAB-related risks. The basic idea is to provide a simple checking tool to detect obscurities in system behavior. The focus is on the operational level as DAB risks become most apparent here.

The four main indicators (explicability, validity, plausibility and acceptability) represent a toolbox to check if the system operates properly. The related guiding questions apply to the whole operation process from input, output to action or decision. Any case of doubt or uncertainty triggers a more detailed review at all levels to uncover eventual mismatch between system behavior and user practice, technical flaws, or legal or ethical problems such as violation of human rights, discrimination etc. If all indicators point to normal system behavior, no DAB risk was detected. But regular detailed reviews of system behavior including all levels are advisable to avoid instability or any other issues. Obviously, all assessment levels are intertwined, but the schematic distinction supports comprehension of whether and how DAB occurs. If the system malfunctions because the data model is biased, then the system as a whole is biased. If the data model is OK but the system behavior is implausible there might be a

different reason; and if a system is basically explicable, its outcome is not necessarily ethical. For users interacting or testing AI, ethical reviews are impractical. It makes little sense to ask, for example, whether the system affects autonomy during operation. But it makes sense having some indicators to assess how exposed system behavior is to DAB-related risks. Acceptability is thus drawn at the intersection between operational and ethical level, because an unacceptable outcome during operation may indicate a severe legal or ethical issue which then needs to be analyzed further.

Briefly, applying the framework to the afore mentioned case of AI for applications underlines the necessity of different assessment levels. In a sheer technical sense, there might be no problem observable. Consider a typical ML framework with preexisting external data models embedded in the system from a trustful source. Without technical and operational assessment, revealing bias in the system, for example due to its mode of image processing, is impossible. To effectively assess how the system operates requires knowledge about technical features and how they affect the preselection of job applicants. The assessment would then show that the system is neither explicable nor valid, nor plausible, nor acceptable. Consequently, a deeper, ethical assessment would be triggered which then would reveal that the system approach in total is unethical as it discriminates persons based on irrelevant data. More detailed testing of the frameworks' practicability, as presented in this article, is subject to further research.

Concluding remarks

„I am just a soul whose intentions are good; oh lord please don't let me be misunderstood“. This refrain of the famous song, first interpreted by Nina Simone in 1964, highlights the societal dilemma of the broader use of AI: (mis-)understanding is a key issue. AI bears high potential to transform society and there are many “good” intentions behind AI-based innovation. But good intentions, such as causing no harm and creating benefit, are not enough to tackle the myriad of risks and prevent AI from becoming a severe threat to society. The crux are misconceptions between technology design, specific application contexts and individual, institutional, societal and ethical requirements. Attempts to resolve them by integrating ethics into AI systems is

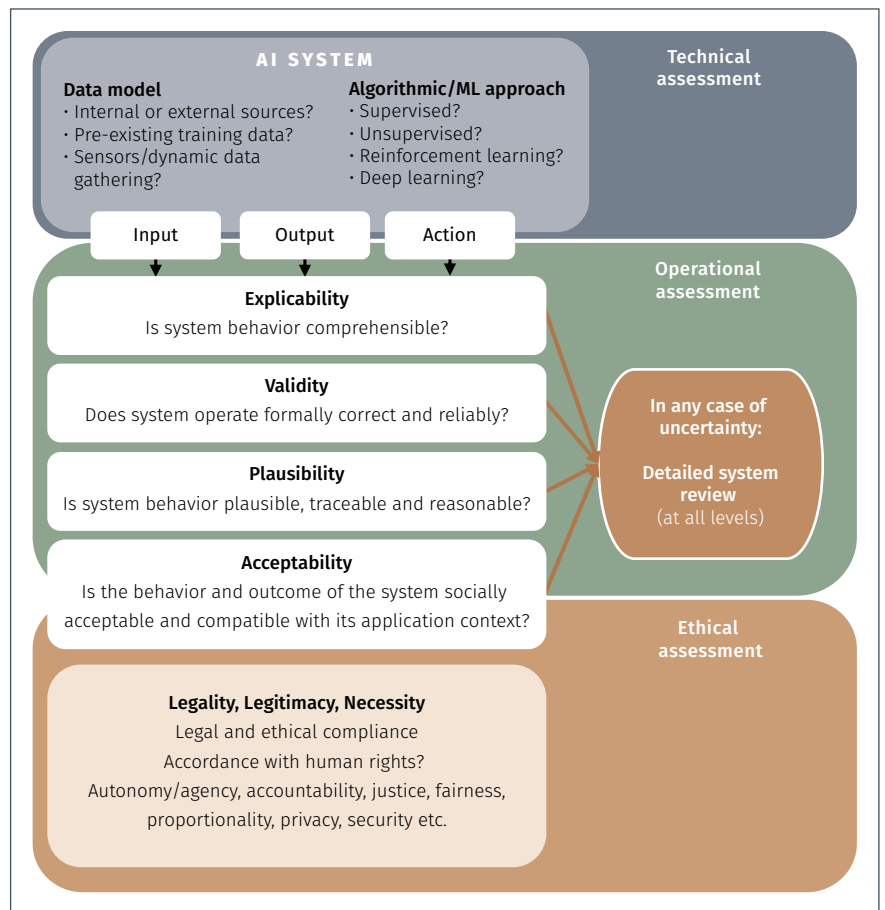


Fig. 2: A problem-oriented assessment framework.

Source: enhanced version of Strauß 2021, p. 9

like trying to explain to an automated vehicle why it should not cause accidents. This is doomed to fail because machines, irrespective of the degrees of their mechanical or digital automation processes, remain machines. Any such attempt aggravates ethical problems. To avoid that humans become “slave to the algorithm” (Edwards and Veale 2017, p. 1) we need more knowledge on the risks of AI and better strategies to cope with them. As a step in this direction, this paper suggests to foster critical AI literacy based on a problem-oriented approach with an explicit focus on DAB-related risks as trigger for further risks of AI. This approach is meant as awareness-raising tool which may also be of some practical use. The intention behind it is rather simple: to envision non-dystopian futures requires novel perspectives on AI to overcome technocratic approaches and revitalize humanistic perspectives on how to deal with AI in a constructive, socially acceptable manner. This can only work if all stakeholders, engineers, designers, policy makers, users and other persons concerned are aware of the factual risks and find ways to reduce them.

Acknowledgement

This research article has not received any external funding.

References

- Abeysooriya, Mandhri; Soria, Megan; Kasu, Mary; Ziemann, Mark (2021): Gene name errors. Lessons not learned. In: *PLoS Computational Biology* 17 (7), p. e1008984. <https://doi.org/10.1371/journal.pcbi.1008984>
- AlgorithmWatch (2019): Automating society. Taking stock of automated decision-making in the EU. Available online at <https://www.algorithmwatch.org/automating-society>, last accessed on 12. 10. 2021.
- Borgesius, Frederik (2018): Discrimination, artificial intelligence, and algorithmic decision-making. Study for the Council of Europe. Strasbourg: DG of Democracy.
- Buchanan, Richard (1992): Wicked problems in design thinking. In: *Design Issues* 8 (2), pp. 5–21. <https://doi.org/10.2307/1511637>
- Cabitza, Federico; Rasoini, Raffaele; Gensini, Gian (2017): Unintended consequences of machine learning in medicine. In: *JAMA* 318 (6), pp. 517–518. <https://doi.org/10.1001/jama.2017.7797>
- Edwards, Lilian; Veale, Michael (2017): Slave to the algorithm? Why a 'right to explanation' is probably not the remedy you are looking for. In: *Duke Law & Technology Review* 16 (1), pp. 18–84. <https://doi.org/10.2139/ssrn.2972855>
- Eid, Fatma-Elzahraa et al. (2021): Systematic auditing is essential to debiasing machine learning in biology. In: *Communications Biology* 4 (183), p. 1–9. <https://doi.org/10.1038/s42003-021-01674-5>
- Floridi, Luciano et al. (2018): AI4People – an ethical framework for a good AI society. Opportunities, risks, principles, and recommendations. In: *Minds & Machines* 28, pp. 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Friedman, Bataya; Nissenbaum, Helen (1996): Bias in computer systems. In: *ACM Transactions on Information Systems*, 14 (3), pp. 330–347. <https://doi.org/10.1145/230538.230561>
- Gianfrancesco, Milena et al. (2018): Potential biases in machine learning algorithms using electronic health record data. In: *JAMA Internal Medicine* 178 (11), pp. 1544–1547. <https://doi.org/10.1001/jamainternmed.2018.3763>
- Goddard, Kate; Roudsari, Abdul; Wyatt, Jeremy (2012): Automation bias. A systematic review of frequency, effect mediators, and mitigators. In: *Journal of the American Medical Informatics Association* 19 (1), pp. 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>
- Goddard, Kate; Roudsari, Abdul; Wyatt, Jeremy (2014): Automation bias. Empirical results assessing influencing factors. In: *International Journal of Medical Informatics* 83 (5), pp. 368–375. <https://doi.org/10.1016/j.ijmedinf.2014.01.001>
- Hallensleben, Sebastian et al. (2020): From principles to practice. An interdisciplinary framework to operationalise AI ethics. Gütersloh: Bertelsmann Stiftung. Available online at https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf, last accessed on 12. 10. 2021.
- Hambling, David (2021): Drones may have attacked humans fully autonomously for the first time. In: *New Scientist*, 27. 05. 2021. Available online at <https://www.newscientist.com/article/2278852-drones-may-have-attacked-humans-fully-autonomously-for-the-first-time/>, last accessed on 12. 10. 2021.
- Harlan, Elisa; Schnuck, Oliver (2021): Objective of biased? On the questionable use of artificial intelligence for job applications. Available online at <https://web.br.de/interaktiv/ki-bewerbung/en/>, last accessed on 12. 10. 2021.
- Harwell, Drew (2019): A face-scanning algorithm increasingly decides whether you deserve the job. In: *Washington Post*, 06. 11. 2019. Available online at www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/, last accessed on 12. 10. 2021.
- HLEG – High-Level Expert Group on Artificial Intelligence (2019): Ethics guidelines for trustworthy AI. Available online at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, last accessed on 28. 09. 2021.
- Köchling, Alina; Wehner, Marius (2020): Discriminated by an algorithm. A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. In: *Business Research* 13 (3), pp. 795–848. <https://doi.org/10.1007/s40685-020-00134-w>
- Lyell, David; Coiera, Enrico (2016): Automation bias and verification complexity. A systematic review. In: *Journal of the American Medical Informatics Association* 24 (2), pp. 424–431. <https://doi.org/10.1093/jamia/ocw105>
- O'Neil, Cathy (2016): Weapons of math destruction. How big data increases inequality and threatens democracy. New York, NY: Crown.
- Obermeyer, Ziad; Powers, Brian; Vogeli, Christine; Mullainathan, Sendhil (2019): Dissecting racial bias in an algorithm used to manage the health of populations. In: *Science* (336), pp. 447–453. <https://doi.org/10.1126/science.aax2342>
- Parasuraman, Raja; Manzey, Dietrich (2010): Complacency and bias in human use of automation. An attentional integration. In: *The Journal of the Human Factors and Ergonomics Society* 52 (3), pp. 381–410. <https://doi.org/10.1177/0018720810376055>
- Selbst, Andrew; Boyd, Danah; Friedler, Sorelle; Venkatasubramanian, Suresh; Vertesi, Janet (2019): Fairness and abstraction in sociotechnical systems. In: *Association for Computing Machinery New York, NY (ed.): FAT* '19, Proceedings of the conference on fairness, accountability, and transparency*, pp. 59–68. <https://doi.org/10.1145/3287560.3287598>
- Simon, Judith; Wong, Pak-Hang; Rieder, Gernot (2020): Algorithmic bias and the value sensitive design approach. In: *Internet Policy Review* (9) 4, p. 1–16. <https://doi.org/10.14763/2020.4.1534>
- Strauß, Stefan (2018): From big data to deep learning. A leap towards strong AI or 'intelligentia obscura'? In: *Big Data and Cognitive Computing* 2 (3), pp. 1–19. <https://doi.org/10.3390/bdcc2030016>
- Strauß, Stefan (2021): Deep automation bias. How to tackle a wicked problem of AI? In: *Big Data and Cognitive Computing* 5 (2), pp. 1–14. <https://doi.org/10.3390/bdcc5020018>
- Tsamados, Andreas et al. (2020): The ethics of algorithms. Key problems and solutions. Available online at SSRN's eLibrary. <https://doi.org/10.2139/ssrn.3662302>
- Tsoukiās, Alexis (2020): Social responsibility of algorithms. An overview. Available online at <https://arxiv.org/pdf/2012.03319.pdf>, last accessed on 12. 10. 2021.
- Wieringa, Maranke (2020): What to account for when accounting for algorithms. A systematic literature review on algorithmic accountability. In: *Association for Computing Machinery New York, NY (ed.): FAT* '19, Proceedings of the conference on fairness, accountability, and transparency*, pp. 1–18. <https://doi.org/10.1145/3351095.3372833>



DR. STEFAN STRAUSS

is senior scientist at the Institute of Technology Assessment (ITA) at the Austrian Academy of Sciences. His main research focus is on the interplay between technology and society, governance of socio-technical systems and the question how digitization affects social practices, human rights, policy and value systems.

RESEARCH ARTICLE

Künstliche Intelligenz parlamentarisch (mit)gestalten

Vergangene technische Zukünfte in den Berichten
der Enquete-Kommissionen des Deutschen Bundestags

50

Christian Vater, Akademie der Wissenschaften und der Literatur Mainz, Digitale Akademie,

Geschwister-Scholl-Straße 2, 55131 Mainz, DE (christian.vater@adwmainz.de)  0000-0003-1367-8489

Eckhard Geitz, Institut für Soziologie, Albert-Ludwigs-Universität Freiburg, Freiburg, DE (eckhard.geitz@soziologie.uni-freiburg.de)

Zusammenfassung • Seit 40 Jahren beschäftigen sich Enquete-Kommissionen des Deutschen Bundestages in ihren Berichten mit Fragen der Auswirkungen „neuer Informations- und Kommunikationstechnologien“ auf die bundesdeutsche Gesellschaft, worunter bereits frühzeitig auch ‚künstliche Intelligenz‘ fällt. Im vorliegenden Aufsatz wird ein erster Überblick über die Entwicklung der inhaltlichen Schwerpunkte wie auch der Präsentationsform in diesem Quellentyp gegeben dessen Aufkommen in den 1980er-Jahren mit dem Anfang der parlamentarischen Technikfolgenabschätzung zusammenfällt. Unsere Forschungsergebnisse zeigen unter anderem, dass aktuelle Berichte weniger konkrete Handlungsempfehlungen an Parlamentarier:innen geben als frühere und somit politische (Mit-)Gestaltungsansprüche weniger stark betonen.

Parliamentarian (co-)design of artificial intelligence.

Past technical futures in the reports of the German Bundestag's committees of inquiry

Abstract • For 40 years, committees of inquiry of the German Bundestag have been dealing in their reports with questions of the effects of “new information and communication technologies” on German society, including artificial intelligence at an early stage. This article gives an initial overview of the development of the main focus areas and the form of presentation in this source type, whose emergence in the 1980s coincides with the beginning of parliamentary technology assessment. Our research findings show that current reports give less concrete policy recommendations to parliamentarians than earlier ones, thus placing less emphasis on political (co-)design.

Keywords • artificial intelligence, enquete, German Federal Parliament, technology assessment, information technology

Einleitung

In Zeiten der Veränderung, in denen die Anzeichen eines Übergangs so zahlreich sind, dass allgemein anerkannt von einer Epochenwelle gesprochen werden kann, steigt der individuelle wie der kollektive Orientierungsbedarf. Diesem kann auf verschiedene, prinzipiell gleichwertige Weisen gedient werden: Man kann sich in der Gegenwart einer Sache versichern und mit Hilfe von Umfragen und Untersuchungen synchron eine Lage kartieren und so die eigene Position feststellen und perspektivieren. Man kann über die Zukunft spekulieren und in die interessante, aber unzugängliche Zeitspanne zwischen dem morgigen Tag und möglichen ‚langen Wellen‘ mathematisch modellierte Prognosen oder heuristisch gefgriffene Szenarien einschreiben, mit dem typischen unsicheren Geltungsanspruch (Koselleck 2003). Man kann aber auch das Archiv konsultieren und das Labyrinth der Akten und Aufzeichnungen betreten, um die Geltungsansprüche der als selbstverständlich gesetzten Vorannahmen der gegenwärtigen Debatte zu befragen: Sind etwa Phänomene wie Digitalisierung und künstliche Intelligenz wirklich so neu, wie es in aktuellen Diskursen bisweilen nahegelegt wird? Sind die damit verknüpften Herausforderungen tatsächlich so dringend, wie sie dargestellt werden? Sind die Vorschläge gegenwärtiger Berater:innen und Expert:innen wirklich singular oder stehen sie in Traditionen? Oder: Gibt es Traditionslinien, die vergessen oder gebrochen wurden? Lassen sich mediale Spuren im kulturellen Gedächtnis aufzeigen (Assmann und Assmann 1994)?

Der vorliegende Aufsatz möchte aufzeigen, dass zumindest eine Textgruppe aus dem Archiv des Deutschen Bundestags –

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.50>
Received: Jun. 27, 2021; revised version accepted: Oct. 19, 2021;
published online: Dec. 20, 2021 (peer review)

dem kollektiven Gedächtnisspeicher der entscheidungstreibenden ‚Herzkammer‘ unseres Parlaments – zur Annäherung an eine Antwort auf die soeben gestellten Fragen unsere Aufmerksamkeit verdient: die Berichte der fachlich einschlägigen Enquete-Kommissionen zu neuen Informations- und Kommunikationstechniken. An jedem dieser Zeitdokumente könnte man innehalten und das darin verwobene, sich aus Wissenschaft, Wirtschaft und Gesellschaft speisende Wissen sortieren und freilegen. Jeder Bericht könnte wiederum zentraler Knotenpunkt eines Akteur-Netzwerks beteiligter Menschen und Nicht-Menschen werden. Diese Detailzeichnung kann hier allerdings aus pragmatischen Gründen nicht vorgenommen werden. Vielmehr geht es darum, in einem ersten Schritt und nach kursorischer Zusammenschau zu verdeutlichen, dass sich weitere Folgeschritte ins Archiv lohnen könnten – um gleichermaßen methodisch und inhaltlich von den handelnden Technologieerklärer:innen der Vergangenheit zu lernen.

Quellen und methodische Überlegungen

Die verarbeiteten Berichte zeigen als Zeitdokumente, neben allen historischen Feinheiten der Entwicklung von Technik(en) und ihrer Debatte(n) auch die Entwicklung der Technikfolgenabschätzung in Deutschland auf: von sachbezogenen Klärungsaufträgen an kleine Gruppen moderierender Parlamentarier:innen über die Institutionalisierung der Technikfolgenabschätzung bis hin zur aufwendig organisierten Großdebatte. Im vorliegenden Text konzentrieren wir uns vornehmlich auf die ersten beiden Berichte von 1983 und 1990, da diese am Anfang der Entwicklung der Technikfolgenabschätzung in Deutschland stehen. Lediglich kurze Erwähnung finden die Berichte von 1998 (Deutschland auf dem Weg in die Informationsgesellschaft), 2013 (Digitale Gesellschaft) und 2020, der sich explizit dem Topos ‚künstliche Intelligenz‘ widmet.

Ein methodischer Befund unserer Forschung soll nicht unerwähnt bleiben: Der Fokus auf die Quellengruppe der Enquete-Kommissions-Berichte – eine Textart mit eigenen Genese- und Rezeptionsbedingungen – hat den Blick auf diachrone Entwicklungen eines ‚weiten‘ Sachfeldes geschärft. Statt also ‚semasiologisch‘ entlang von Schlagwörtern zu suchen, zu finden und zu ordnen, haben wir ‚onomasiologisch‘ den Zusammenhang greifen können, der ‚Informations- und Kommunikationstechnik‘, ‚Expertensysteme‘, ‚Neue Medien‘, ‚Digitale Gesellschaft‘ und ‚künstliche Intelligenz‘ jenseits aller Begriffsarbeit verbindet.

Auf dem Weg ins tiefe Archiv der Technologie-Debatten

Anfang der 1980er-Jahre beginnt die Öffentlichkeit, sich für die Veränderungen ihrer Kommunikationsumgebung und ihrer medialen Kanäle in höchstem Maße zu interessieren. Nicht nur treten mit Marshal McLuhans Kurz-Auftritt in Woody Allans ‚Annie Hall‘ 1977 Medienwissenschaftler:innen als Persönlich-

keiten endgültig in den Populärdiskurs ein, auch das Nachdenken über die Wechselwirkung zwischen medialem Wandel und gesellschaftlicher Wirklichkeit wurde entlang der Rolle des Fernsehens und vor der Anschauung des ‚Globalen Dorfes‘ diskutiert. 1979 hatte der französische Philosoph Jean-François Lyotard für den Universitätsrat Québecks einen Bericht über den Wandel des Wissens in einer technologisch veränderten Welt verfasst. Er stellt fest, dass – auch wenn es „unklug [sei], der Futurologie exzessiven Kredit einzuräumen“ – man nicht umhinkäme, festzuhalten, dass die „Auswirkungen [...] technologischer Transformation auf das Wissen [...] erheblich“ (Lyotard 2019, S. 23 f.) seien. Die technologische Transformation werde getrieben von Kybernetik, Kommunikation und Telematik, von ‚modernen Algorithmen‘, Informatik, Computern, Datenbanken und der ‚Perfektionierung ‚intelligenter‘ Terminals‘“ (Lyotard 2019, S. 30). Dieses ‚postmoderne Wissen‘ löste weltweite Diskussionen aus, die auch in Deutschland anbrannten: Am 9. April 1981 erhält die Enquete-Kommission „Neue Informations- und Kommunikationstechniken“ vom Deutschen Bundestag in Bonn ihren Arbeitsauftrag. Mit diesem Arbeitsauftrag beginnt eine (mittellange) Geschichte der parlamentarischen Beratung von und Debatte über neue Technologien und Techniken, die wir heute unter den Schlagworten „Digitalisierung“ und „künstliche Intelligenz“ weiterführen. Entlang der einzelnen Berichte kann man sie auf ihrem Zeitstrahl folgendermaßen gliedern:

1983 sehen einige Experten deutlich, dass tiefgreifende Entwicklungen eintreten werden und es wird der Bedarf an Aufklärung der technischen Grundlagen im Dienst einer informierten Forschungs-, Wirtschafts-, Bildungs- und Sozialpolitik angemeldet. Fragen nach Netzinfrastruktur, Hardwareentwicklung und Erhebungsmethoden einer wissenschaftlichen Technikfolgenabschätzung stehen im Mittelpunkt (Deutscher Bundestag 1983). Die Berichtslegung fällt in ein für Deutschland folgenreiches Jahr. Auf dem Höhepunkt der Friedensbewegung beginnt die Kanzlerschaft Helmut Kohls, dessen Koalitionsregierung weitreichende marktliberale Reformen auf den Weg bringt.

1990, im Jahr der Wiedervereinigung und ein Jahr nachdem mit der Privatisierung der Deutschen Bundespost begonnen wurde, konkretisieren sich die Anwendungsfälle und Orientierungsansprüche, technologische Durchbrüche werden von Fachleuten genauso allgemeinverständlich aufbereitet wie deren allgemeine Grundlagen und vermuteten Folgen, wobei sich die Darstellung in klaren Handlungsoptionsangeboten verdichtet. Der Bericht ist zudem infographisch auf höchstem Niveau durchgestaltet. Ein Schwerpunkt der Fragen liegt auf Softwareentwicklung und -einsatz (Deutscher Bundestag 1990).

1998, zum Ende der Ära Kohl, wird die gesellschaftliche Aneignung der technischen Innovation verhandelt, strategische Abstraktionen im Dienst konkreter Politik- und Unternehmensberatung greifen Raum. (Technik-)soziologische Fragestellungen werden formuliert. Der Bundestag pflegt eine Webseite und stellt Infomaterial zur Verfügung (Deutscher Bundestag 1998).

2013 markieren konkrete rechtliche Klärungsbedarfe die technikbedingten Reibungen und Nutzungsansprüche der Zivil-

gesellschaft mit den Unternehmungen einer global verschalteten, digitalen und multimedialen Kommunikation. Der Bericht selbst ist hinsichtlich Schreibtechnik, multimedialer Netzpräsentation und Bürgerbeteiligungsangeboten ein Medienexperiment, inklusive webarchivierter Videostreams einiger Anhörungen und Sitzungen sowie abgedruckter qualitativer Expert:inneninterviews (Deutscher Bundestag 2013).

2020 wird eine Geschichte möglicher Zukünfte verfasst – nun in groben Linien entkoppelt von konkreten technischen Materialitäten und Fragen, ihrer (Er-)Klärung oder Vermittlung – und im Modus der gemeinsamen Entwicklung einer ziel- und nicht sachstandsorientierten Narration werden technikbasierte

schen, sozialen und kulturellen Bereiche der BRD zu Beginn der 1980er-Jahre umfangreiche Desiderata hinsichtlich steuernder und regulierender technopolitischer Maßnahmen – etwa den Ausbau des Glasfasernetzes bis Ende der 1980er-Jahre, siehe den abschließenden Diskussionsstand der Enquete-Kommission (Deutscher Bundestag 1983, S. 9–13 b). Ob eben dieser Steuerungsanspruch der Grund dafür war, dass kein weiteres Interesse an seiner Fertigstellung bestand, steht dahin. Außerordentlich vielversprechend und gehaltvoll ist jedenfalls der Zwischenbericht, der am 28.03.1983 veröffentlicht wird. Der Zwischenbericht beginnt mit zwei Thesen, auf die sich die Mitglieder ausdrücklich einigen konnten:

*Im April 1981 erhielt die Enquete-Kommission
,Neue Informations- und Kommunikationstechniken‘ vom
Deutschen Bundestag in Bonn ihren Arbeitsauftrag.*

Hoffnungen und Wünsche eines breiten Spektrums von Interessensgruppen gesammelt. ‚Menschenzentrierte KI‘ und ‚KI Made in Germany‘ sind die Brennpunkte und Zielvorstellungen eines Aushandlungsprozesses, an dessen Ende das mit Abstand umfangreichste Dokument dieser Reihe als Bundesdrucksache verfertigt wird. Das infographische Bildvokabular von Unternehmensberater:innen ist mit den Kreuztabellen der SWOT-Analyse (englisch für Stärken, Schwächen, Chancen, Risiken) genauso parlamentsfähig geworden wie das überreiche Verzeichnis ephemerer Onlinequellen, welches das wissenschaftliche Literaturverzeichnis abgelöst hat. Ein Onlinebefragungsportal zur niedrigschwelligen Bürgerbeteiligung begleitet die Kommissionsarbeit, wird allerdings nur in geringem Umfang genutzt (Deutscher Bundestag 2020).

Im Gegensatz zum letzten vorliegenden Enquete-Bericht beginnt die Reihe mit dem aus dem Jahr 1983 sehr konkret: mit Satelliten und Glasfaserkabeln, mit den materiellen Schaltungen und den erwartbaren Anwendungen der Informations- und Telekommunikationstechnologie.

1983 – Hardware. Technische Präzision und Bemühen um technologische Aufklärung: Vernetzte Informations- und Kommunikationstechnik als technologische Bedingung. Techniker:innen vermitteln

Die parlamentarische Auf- und Abklärung der Möglichkeiten und Grenzen digitaler Neuer Informations- und Kommunikationstechnik (IuK) muss ihren ersten, vielversprechenden Schritt leider verstoßen: Mitten während des Arbeitsprozesses werden nach dem Bruch der Regierungskoalition außerplanmäßige Wahlen ausgerufen, ein Abschlussbericht der am 9. April 1981 eingerichteten Enquete-Kommission kann nicht mehr ausgearbeitet werden. Gleichwohl ergeben sich aus dem dezidierten Transfer technologischer (Un-)Möglichkeiten auf die relevanten politi-

„These 1: Die Entwicklung neuer Technologien, einschließlich der IuK-Techniken und deren Nutzung, ist von den jeweils gegebenen politischen, ökonomischen, sozialen und rechtlichen Bedingungen abhängig.“ (Deutscher Bundestag 1983, S. 8)

„These 2: Die Analyse der technologischen, ökonomischen, sozialen und rechtlichen Aspekte der Kommunikationstechnologien und die Folgenabschätzung treffen auf Schwierigkeiten bei der Erfassung der Tatsachenbasis, auf Unsicherheiten der theoretischen Annahmen über Wirkungszusammenhänge sowie auf Kontroversen bei der Bestimmung von Kriterien für die Auswahl entscheidungserheblicher Gegenstände und die Bewertung der Befunde.“ (Deutscher Bundestag 1983, S. 8)

Wenn also, so die Kommission, neue Technologien entwickelt und in einem zweiten Schritt auch genutzt werden sollen, hat dies komplexe gesellschaftliche und kulturelle Voraussetzungen. Außerdem seien zur Zeit der Abfassung dieses Zwischenberichtes sowohl praktisches Technikverständnis und theoretisches Systemmodell als auch Auswahl- und Bewertungskriterien von Sachverhalten und Befunden unzureichend.

Ausgehend von einer detaillierten Darstellung des technischen Aufbaus der Hardware der Kommunikationsnetze (Kapitel 1) geht der Zwischenbericht der Kommission über den aktuellen Stand von Hersteller- und Anwenderbranchen sowie einem kurzen Ausblick auf Verbraucher (Kapitel 2) über zur mit 100 Seiten umfassenden Abschätzung der Auswirkungen auf Wirtschaft und Gesellschaft (Kapitel 3). Es folgt abschließend die Behandlung besonders hervorgehobener Teilaspekte (Kapitel 4), wozu Normung, Forschungsförderung, Datenschutz, Verbraucherschutz und Urheberrecht gehören. Der detaillierte Text

selbst bettet Fachausdrücke der Verständlichkeit halber in zugängliche und konkrete Anwendungszusammenhänge ein (z. B. hinsichtlich der Schlüsseltechnologie ISDN, Deutscher Bundestag 1983, S. 30) und liefert graphisch abgesetzte und kognitiv leicht zugänglich gestaltete Kurzzusammenfassungen (z. B. hinsichtlich relevanter technischer Innovationen, Deutscher Bundestag 1983, S. 30).

Einige der diskutierten Problemstellungen wirken heute allzu bekannt, so etwa die Frage, ob „Kreativität, Intelligenz, Empathie, Bindungsfähigkeit, Solidarität und Verantwortungsbereitschaft“ durch ‚neue Medien‘ leiden würden (Deutscher Bundestag 1983, S. 154). Andere Kapitel, etwa „3. Bedeutung und Aus-

„Wesentliche Grundlage der Kultur in einem weiteren – über den engen Bereich der Künste und der Kunstpflege hinausweisenden Sinn – ist die Kommunikation. Die neuen IuK-Techniken erweitern die Möglichkeiten technisch vermittelter Kommunikation. Sie können damit nicht nur Veränderungen im Denken, Erleben und Verhalten des einzelnen, sondern auch im allgemeinen gesellschaftlichen Kulturprozeß bewirken.“

Der Darstellung folgen ein Anmerkungsteil mit namentlichen Kommentaren und Änderungsvorschlägen auf Stand der Diskussion mit frühzeitiger Angabe (Deutscher Bundestag 1983,

Gerade die frühzeitige Beschäftigung mit einer Technik soll das Parlament in die Lage versetzen, sie zum Wohl des Ganzen zu gestalten, statt nachträglich unerwünschte Folgen zu beseitigen.

wirkungen der IuK-Techniken in Wirtschaft und Gesellschaft“ wirken wiederum brandaktuell und haben sogar noch immer eine utopische Komponente, wenn es um die „Demokratisierung des Informationszugriffs“ geht, und zwar „für alle Bürger“ (Deutscher Bundestag 1983, S. 166).

Zur Gestaltung des Produktiveinsatzes der neuen Technologien findet sich im Infokasten 2 des Unterkapitels 3.2.3 „Arbeitsplatz, Heimarbeit, Gesundheitsschutz, Mitbestimmung“ (Deutscher Bundestag 1983, S. 106 f.) eine heute ungewohnte Haltung partizipativer betrieblicher Mitbestimmung:

„Die Nutzung der IuK-Technologien (Geräte, Systeme, Dienste) zu arbeitsplatzrelevanten Innovationen ermöglicht eine starke Erhöhung (Ausweitung) des sozialorganisatorischen Gestaltungsspielraums hinsichtlich der Arbeitsbedingungen und der Qualifikationsstrukturen. Welche der möglichen Gestaltungsoptionen sich durchsetzen können, hängt weniger von den neu eingesetzten Techniken und den daraus folgenden technischen Restriktionen ab, sondern vielmehr von der organisatorischen Konzeption des betrieblichen Ablaufes, der Lern- und Anpassungsfähigkeit aller am Arbeitsprozeß Beteiligten, sowie der politischen und vertraglichen Durchsetzungskraft der Tarifvertragsparteien.“

Die hier hinterlegte Einordnung des Verhältnisses neuer Lehr-Technologien (heute E-Learning) zur Mitbestimmung und Berufsfachlichkeit ist selbst in den Pandemie- und Digitalisierungsjahren 2020 und 2021 noch kein Allgemeinut.

Die Sphäre des Kulturellen wird gleichzeitig als genauso betroffen gedacht, da ihr eine grundsätzliche Medialität zugeschrieben wird, so im Infokasten 1 im Unterkapitel 3.2.4 „Ausbildung, Bildung, Wissenschaft, Kultur“ (Deutscher Bundestag 1983, S. 122):

S. 217–237) sowie ein brauchbares Glossar. Es wurden sieben Sachverständige berufen, drei Anhörungen veranstaltet, vier Unterkommissionen eingerichtet, 150 schriftliche Stellungnahmen eingearbeitet, Informationsreisen ins Ausland getätigt und als Partner das Deutsche Institut für Wirtschaftsforschung, das Heinrich-Hertz-Institut, und die PROGNOSE AG eingebunden.

1990 – Software. Zentrale Innovationsfelder werden abgesteckt und kartiert: ‚Expertensysteme‘ in industrieller Fertigung und Medizintechnik im Detail. Entwickler:innen wollen gestalten

Schon der Bundestag der zehnten Wahlperiode hatte eine Enquete-Kommission „Technikfolgen-Abschätzung und -Bewertung“ eingerichtet. Dieser Ansatz war neu, da erst 1989 der erste Bericht dieser Kommission vorgelegt wurde, der Technikfolgenabschätzung (TA) überhaupt als Arbeitsfeld ordnete – das Anwendungsfeld der ‚Expertensysteme‘ musste als nur teilbehandelt an die Nachfolge in der elften Wahlperiode übergeben werden (Deutscher Bundestag 1990, S. 9). Hier wurden nun zwei methodische Zugänge der TA exemplarisch erprobt: Ein „probleminduzierter“ mittels des Sachfeldes „Alternativen landwirtschaftlicher Produktionsweisen“ und ein „technikzentrierter“ mittels der „Expertensysteme“. Das Parlament sollte anhand des Vergleichs der Ergebnisse beurteilen, welcher der erprobten methodischen Zugänge angemessener oder brauchbarer für die Bedarfe parlamentarischer Praxis sei. „Außerdem“, so wurde argumentiert, könne „man mit dieser Untersuchung einen Beitrag zur ‚Entmystifizierung‘ der künstlichen Intelligenz leisten.“ (Deutscher Bundestag 1990, S. 9) Der Bericht beginnt so plastisch wie deutlich:

„Ist der Mensch ein Roboter?“, „Sind Computer lebendig?“, dies sind die Titel von nur zwei der vielen populärwissen-

schaftlichen Bücher, die in den letzten Jahren zur ‚Künstlichen Intelligenz‘ erschienen sind. Dabei ist die Bezeichnung ‚Künstliche Intelligenz‘ bereits mehr als dreißig Jahre alt, und die Ideen, die den damit bezeichneten Forschungsarbeiten zugrunde liegen, beschäftigen die Philosophen schon seit der Antike.“ (Deutscher Bundestag 1990, S. 7)

Nach dieser dezidiert philosophischen Einordnung in einen historischen Prozess, der sich mindestens bis ins klassische Athen zurückverfolgen lasse, folgt ein abklärender, versachlichender und gleichzeitig phantastisch fabulierender, einhegender Befund:

„Um es gleich vorwegzunehmen: dieser Bericht kann auf die Frage, ob der Mensch eine Maschine sei oder ob uns eines Tages Maschinen als Träger der höchsten Intelligenz beerben werden, keine Antwort geben. Untersucht werden ‚Expertensysteme‘, Computerprogramme, die aus der Forschung zur ‚Künstlichen Intelligenz‘ hervorgegangen sind.“ (Deutscher Bundestag 1990, S. 7)

Die Kommission will also kein Urteil vorwegnehmen, sondern möglichst zugänglich und verständlich Material für das Parlament vorbereiten, und zwar rechtzeitig, um politische Gestaltung zu ermöglichen. Diese soll nach engagierter und kritischer Diskussion „über den Bericht und insbesondere über die von ihr zusammengetragenen Handlungsoptionen“ (Deutscher Bundestag 1990, S. 3) durch die Parlamentarier:innen stattfinden. Das Ziel, Handlungs- und Gestaltungsoptionen aufzuzeigen, verdeutlicht sich in ausführlichen Beschreibungen und Ausführungen im Fließtext, durchbrochen von 19 Infokästen und 19 Einfassungen mit klar ausgearbeiteten Optionen für Handlungsfelder, Übersichtstabellen erleichtern zudem den Zugang. Die Behandlung des Themas beginnt bei Geschichte, Begriffsklärung (z. B. ‚künstliche Intelligenz: Eine mißratene Bezeichnung?‘), und technischen Grundlagen, behandelt mit dem Einsatz in der Produktion und dem Einsatz in der Medizin zwei epistemisch unterschiedliche Praxisfelder und endet bei Querschnittsthemen wie Risiken (z. B. „Fehlerquellen bei Expertensystemen“) oder Datenschutz.

Die Optionen der Handlungsfelder sind jeweils als ganz konkrete parlamentarische Handlungsmöglichkeiten gestaltet. Ein spannendes Beispiel findet sich im Handlungsfeld 16 „Haftung bei Expertensystemen“: „Die Bundesregierung auffordern, die ärztliche Haftpflicht auf ihre Anwendbarkeit im Zusammenhang mit der Verwendung von Computerprogrammen (bei fehlerhaftem Einsatz) sowie bei Nichtanwendung zu überprüfen und dabei die Relevanz des Begriffs ‚ärztlicher Kunstfehler‘ zu erörtern.“ (Deutscher Bundestag 1990, S. 75) Ein weiteres im Handlungsfeld 18 „Mitbestimmung“: „Die Bundesregierung auffordern, eine Untersuchung über Ansätze, Formen und Erfahrungen mit der ‚partizipativen Entwicklung und Gestaltung‘ informationstechnischer Systeme durchführen zu lassen.“ (Deutscher Bundestag 1990, S. 84)

Die Kommission von 1990 griff auf Materialien ihrer Vorgängerin zurück, die unter dem Titel „Chancen und Risiken des Einsatzes von Expertensystemen in Produktion, Verwaltung, Handwerk und Medizin“ vom Battelle-Institut und dem Institut für Medizinische Informatik und Systemforschung der Gesellschaft für Strahlen- und Umweltforschung geliefert worden waren. Sie beauftragte selbst das Fraunhofer Institut für Arbeitswirtschaft und Organisation (IAO), das Institut für Sozialwissenschaftliche Forschung (ISF) und das Institut für Medizinische Informatik und Systemforschung (medis) der Gesellschaft für Strahlen- und Umweltforschung. Achtzehn Sachverständige wurden gehört, neunzehn weitere um schriftliche Stellungnahmen gebeten. Es folgen neben den erwähnten Tabellen zu „Kästen“ und zu „Optionen“ (also „Wissen“ und „Handlungsaufforderungen“) ein auch heute noch belastbares Literaturverzeichnis, ein brauchbares Glossar und ein Verzeichnis der mit dem Bericht verbundenen Kommissionsvorlagen. Die vorbildliche graphische Gestaltung und Aufbereitung dieses Berichtes sollten folgende Berichte nicht wieder erreichen.

Können ‚Technikzukunft‘ ohne ‚Technikvergangenheiten‘ gelingen? – ein (Zwischen-)Fazit

Der aktuelle Enquete-Bericht zur ‚Künstlichen Intelligenz‘ hat vom Präsidenten des Bundestages einen klaren Auftrag erhalten: Da „die Dynamik der Digitalisierung [...] mit der Forschung zur Künstlichen Intelligenz eine neue Dimension erreicht“ habe, müsse geklärt werden, „was KI eigentlich bedeute, was sie leisten könne und welche Chancen und Herausforderungen für Staat, Gesellschaft und Recht entstünden.“ (Deutscher Bundestag 2020, S. 43) Man könnte auch festhalten: Noch immer! Da es sich – im eskalativen Extremfall – nicht nur um Wirtschaft, sondern auch um Waffen handle, wären die Fragen dringlich, „wie diese Entwicklung so gestaltet werden könne, dass KI den Menschen diene“. (ibid.) Dass ein nachhaltiger und wohlfstandsorientierter Einsatz KI Made in Germany gefördert werden soll, ist jedenfalls Zielvorstellung – auch, wenn Bundestagspräsident Schäuble selbst und bezeichnender Weise von einer „neue[n] Zauberformel des technischen Fortschritts“ spricht.

Lässt sich aber ein so großes wie gewichtiges Projekt ohne Blick in die Geschichte – und die Genese der eigenen Grundlagen – belastbar abschließen?

Blickt man nach mehr als 30 Jahren in die hier vorgestellten Berichte, muss man der Technikfolgeabschätzung zubilligen, dass sie hier ausgesprochen treffsicher Technikzukunft antizipiert hat – im ersten Fall mit Implikationen für politische Steuerung, im zweiten Fall mit klarem Blick für künftige Diskurse über Potenziale und Probleme der künstlichen Intelligenz. Wenn heute die etwas lapidare Rede davon ist, Deutschland habe die Digitalisierung verschlafen, bedarf diese Einordnung einer Richtiger Verspätung allmählich auch im ländlichen Raum ankommt,

war bereits zu Beginn der 1980er-Jahre absehbar, wozu dessen Vernachlässigung führt und was sie verhindert. Technikfolgenabschätzer:innen haben mit den beiden Berichten zwei starke Argumente, mit denen sie zeigen können, dass Technikzukunft kein Produkt phantasiereicher Spekulationen sein müssen, sondern Szenarien sein können, die politische Steuerung ermöglichen und Versäumnisse dingfest machen. Denn auch die Grundsatfrage nach der ‚Planung‘ auf Grundlage verlässlicher und belastbarer Prognosen begleitet die deutschsprachige Technikfolgenabschätzung seit ihren Anfängen, mal als ‚Reizvokabel‘, mal in Euphorie (Brinkmann 2006, S. 177 ff.).

Eine Rückkehr zum politischen Gestaltungsanspruch der ersten, hier aus dem Archiv gehobenen Berichte halten wir – auch als historiographische Erweiterung eines ‚problemorientierten‘ Ansatzes – für ratsam. Methodisch sehen wir im hier greifbaren Bemühen um Grundlagenvermittlung bei gleichzeitiger Allgemeinverständlichkeit auf dem technischen Stand der (jeweiligen) Gegenwart ein Vorbild, auch weil alle Mittel des Infographikdesigns in ihrer Übersetzungsleistung genutzt wurden, um der Kommunikation wissenschaftlicher Erkenntnisse in den als partizipativ gedachten politischen Raum zu dienen. Es könnte in einem nächsten Schritt um Mitgestaltung und Mitbestimmung gehen, um die zwei Leitgedanken teilhabender gemeinschaftsorientierter Technologiebesprechung – und um Umnutzung und Weiternutzung, die zwei Leitgedanken emanzipatorischer individueller Aneignungsansprüche. Mit Sicherheit sind jedenfalls die Arbeiten von Enquete-Kommissionen und TA zu künstlicher Intelligenz – als Begriff, als Technik, als Zaubertrick, als Traum – mit dem aktuellen Bericht nicht zu einem befriedigenden Ende gekommen.

Es wird sich also lohnen, die ‚technischen Zukünfte der Vergangenheit‘, wie sie aus den Berichten der Enquete-Kommissionen des Bundestags hervorgehen, zur Kenntnis zu nehmen, um so der „Struktur solcher ‚Vorstellungen‘ selbst und ihre[r] realitätsprägenden Funktionen“ (Popplow 2020, S. 43) auf die Spur zu kommen – und zwar von den Begriffen über Diskurspuren bis zur Gestaltung ihrer Präsentation und den Blick auf zugrunde liegende Sachverhalte.

Angabe von Finanzierungsquellen

Dieser Forschungsartikel hat keine Förderung erhalten.

Literatur

- Assmann, Aleida; Assmann, Jan (1994): Das Gestern im Heute. Medien und soziales Gedächtnis. In: Klaus Merten, Siegfried Schmidt und Siegfried Weischenberg (Hg.): Die Wirklichkeit der Medien. Eine Einführung in die Kommunikationswissenschaft. Wiesbaden: Springer, S. 114–140. https://doi.org/10.1007/978-3-663-09784-6_7
- Brinkmann, Andrea (2006): Wissenschaftliche Politikberatung in den 60er Jahren. Die Studiengruppe für Systemforschung, 1958 bis 1975. Berlin: edition sigma. <https://doi.org/10.5771/9783845271095>
- Deutscher Bundestag (1983): Zwischenbericht der Enquete-Kommission „Neue Informations- und Kommunikationstechniken“ gemäß Beschluß des Deutschen Bundestages vom 9. April 1981. Bonn: Bundesdrucksache 09/2442.

Online verfügbar unter <https://dserver.bundestag.de/btd/09/024/0902442.pdf>, zuletzt geprüft am 18.10.2021.

- Deutscher Bundestag (1990): Bericht der Enquete-Kommission „Gestaltung der technischen Entwicklung; Technikfolgen-Abschätzung und -Bewertung“ gemäß Beschluß des Deutschen Bundestages vom 5. November 1987. Bonn: Bundesdrucksache 11/7990. Online verfügbar unter <http://dipbt.bundestag.de/doc/btd/11/079/1107990.pdf>, zuletzt geprüft am 18.10.2021.
- Deutscher Bundestag (1998): Schlußbericht der Enquete-Kommission Zukunft der Medien in Wirtschaft und Gesellschaft – Deutschlands Weg in die Informationsgesellschaft (eingesetzt durch Beschluß des Deutschen Bundestages vom 7. Dezember 1995) zum Thema Deutschlands Weg in die Informationsgesellschaft. Bonn: Bundesdrucksache 13/11004. Online verfügbar unter <http://dipbt.bundestag.de/doc/btd/13/110/1311004.pdf>, zuletzt geprüft am 18.10.2021.
- Deutscher Bundestag (2013): Schlußbericht der Enquete-Kommission „Internet und digitale Gesellschaft“ eingesetzt durch Beschluss des Deutschen Bundestages vom 4. März 2010. Berlin: Bundesdrucksache 17/12550. Online verfügbar unter <https://dserver.bundestag.de/btd/17/125/1712550.pdf>, zuletzt geprüft am 18.10.2021.
- Deutscher Bundestag (2020): Unterrichtung der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale (eingesetzt durch Beschluss des Deutschen Bundestages vom 26. Juni 2018). Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Berlin: Bundesdrucksache 19/23700. Online verfügbar unter <https://dserver.bundestag.de/btd/19/237/1923700.pdf>, zuletzt geprüft am 18.10.2021.
- Koselleck, Reinhart (2003): Zeitschichten. Studien zur Historik. Frankfurt am Main: Suhrkamp.
- Lyotard, Jean-François (2019): Das postmoderne Wissen. Ein Bericht. Berlin: Passagen.
- Popplow, Marcus (2020): Zur Erforschung von Technikzukunft aus technikhistorischer Perspektive. In: Paulina Dobroć und Andie Rothenhäusler (Hg.): 2000 revisited. Rückblick auf die Zukunft. Karlsruhe: KITopen, S. 41–58. <https://doi.org/10.5445/KSP/1000117728>



CHRISTIAN VATER

ist Wissenschaftlicher Mitarbeiter an der Digitalen Akademie der Akademie der Wissenschaften und der Literatur Mainz sowie Lehrbeauftragter am Department für Geschichte des Instituts für Technikzukunft (ITZ) des Karlsruher Instituts für Technologie (KIT). Als Wissenschaftstheoretiker und Technikphilosoph befasst er die Wissensgeschichte der künstlichen Intelligenz.



ECKHARD GEITZ

ist Leiter des Bildungsinstituts im Gesundheitswesen (BIG) in Essen und arbeitet am Institut für Soziologie der Universität Freiburg an einem Promotionsprojekt zur Digitalisierung der Psychiatrie.

RESEARCH ARTICLE

Rechtliche Rahmenbedingungen für KI-Systeme

Immanente Herausforderungen und mögliche Lösungen durch Control by Design

Thomas Wilmer, Institut für Informationsrecht, Hochschule Darmstadt, Haardtring 100, 64295 Darmstadt, DE (thomas.wilmer@h-da.de)

56

Zusammenfassung • Der Autor stellt die komplexen rechtlichen Rahmenbedingungen für künstliche Intelligenz aus den Bereichen des Dateneigentums, des Datenschutzes und des Urheberrechts dar und erläutert, weswegen die unübersichtliche Rechtslage nicht dazu geeignet ist, auf Rechtssicherheit durch neue Regulierungen auf EU- oder nationaler Ebene zu hoffen. Stattdessen werden ein Regulierungsrahmen für zulässige Vertragsvereinbarungen sowie dazu passende Control by Design-Technikeinstellungen vorgeschlagen.

Legal framework conditions for AI systems. Immanent challenges and possible solutions (control by design)

Abstract • The author presents the complex legal framework for artificial intelligence in data ownership, data protection, and copyright and explains why the confusing legal situation is not suitable for hoping for legal certainty through new regulations at the EU or national level. Instead, he proposes a regulatory framework for permissible contractual agreements and matching control-by-design settings.

Keywords • data ownership, data governance act, copyright, artificial intelligence

Künstliche Intelligenz (KI) stellt aus rechtlicher Sicht verschiedene Herausforderungen an die Einordnung in das vorhandene Regulierungssystem: KI kann als Software mit vorgefertigten Programm-Elementen und von der Software selbst geschaffenen neuen Programmanteilen verstanden werden, was die Frage nach dem Geistigen Eigentum an den immateriellen Werten stellt, welche von der KI geschaffen werden. Daneben basiert KI immer auch auf einem (dann fortlaufend aktualisierten) Datensatz, welcher in aller Regel der Manifestation menschli-

cher oder maschineller Erfahrungswerte entspricht. Zu klären ist, wem dieser Datensatz zusteht. Soweit es sich hierbei um personenbezogene Daten handelt, gelten innerhalb der Europäischen Union strikte Regelungen zur Verwendbarkeit dieser Daten. Insgesamt stellt sich Frage nach Folgen des KI-Einsatzes für die Gesellschaft und nach der Transparenz der zugrunde gelegten oder neu geschaffenen Algorithmen sowie der Kontrolle der Einsatzbedingungen.

Bei kritischer Betrachtung der gängigen Geschäftsmodelle kann KI daneben auch als Mittel dienen, um Daten und Know-how der KI-Nutzer zu zentralisieren und dann wiederum allen zur Verfügung zu stellen, wobei die Wertschöpfung bei den KI-Betreibern verbleibt.

KI steht aktuell im Spannungsfeld zahlreicher neuer Regulierungsversuche auf nationaler und europäischer Ebene.

KI steht aktuell im Spannungsfeld zahlreicher neuer Regulierungsversuche auf nationaler und europäischer Ebene. Neben gesetzlichen Begrenzungen der Einsatzbereiche der KI und der Regelung der Verfügungsgewalt über Datensätze sollte verstärkt auch an die zentrale Frage gedacht werden, welche Möglichkeiten es gibt, vertragliche Vereinbarungen mit Endnutzern zur Nutzung der KI und Übertragung von Daten zu regulieren. Solche vertraglichen Vorgaben könnten dann in das Design der KI-Systeme einfließen und den Rahmen für eine Control by Design-Regelung bilden. Diese könnte eine übergeordnete datenschutzfreundliche Gesamteinstellung sein, welche sich an den Regelungen des Art. 25 EU-Datenschutzgrundverordnung (DSGVO) für Privacy-by-Design (datenschutzfreundliche Systemkonzepte)

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>) <https://doi.org/10.14512/tatup.30.3.56>
Received: Jun. 22, 2021; revised version accepted: Oct. 21, 2021;
published online: Dec. 20, 2021 (peer review)

tion) und den Neuregelungen des Telekommunikation-Telemedien-Datenschutzgesetzes (TTDSG) orientiert.

Die Bedeutung der KI für Gesellschaft und Wirtschaft findet ihren Niederschlag in internationalen Bemühungen, die Entwicklung zu analysieren und zu regulieren. Nach Auffassung der deutschen Datenethikkommission (Datenethikkommission 2019) ist eine möglichst europäische Regulierungsreichweite analog zu den datenschutzrechtlichen Regelungen anzustreben. Das europäische Weißbuch zur künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen fordert „angesichts der erheblichen Auswirkungen, die KI auf unsere Gesellschaft und die notwendige Vertrauensbildung haben kann [...], dass die europäische KI auf unseren Werten und Grundrechten wie Menschenwürde und Schutz der Privatsphäre fußt“ (Europäische Kommission 2020 a, S. 2).

Umsetzungsbedarf

Fraglich ist, inwiefern diese Ziele in einer funktionsfähigen, den Interessen der Betroffenen und der Wirtschaft angemessenen Form berücksichtigt werden können. Aus juristischer Sicht stellen sich eine ganze Reihe von Fragen aus verschiedenen Rechtsgebieten an den Einsatz der KI, einschließlich sämtlicher unscharf definierter Formen smarter und selbstlernender Produkte oder des Einsatzes sogenannter ‚Legal Tech‘ (Herberger 2018). Es bleibt zu klären, ob die neuen Regulierungsansätze und Absichtserklärungen mit bestehenden rechtlichen Rahmenbedingungen umzusetzen sind oder ob es einer grundsätzlich neuen Gesetzgebung bedarf, welche die Zulässigkeit des KI-Einsatzes in bestimmten Einsatzfeldern (etwa der Personal- oder Einstellungsbeurteilung oder des Kampfdrohneinsatzes) oder in

verlieren das Interesse an der Wahrnehmung der zahllosen Informationen und sind bereit, Nutzungen zu akzeptieren, ohne überhaupt noch zu lesen, welchen Inhalt die Informationen haben, selbst wenn es eine transparente Zusammenfassung geben sollte. Als Musterbeispiel dafür können die ‚Cookie-Pop-Ups‘ gelten, die von vielen Nutzern nicht mehr gelesen, sondern nur noch akzeptiert werden. Nach einer Statista-Umfrage vom 04. 06. 2020 gaben 41 Prozent der Befragten in Deutschland an, sich grundsätzlich die Inhalte der Cookie-Hinweise nicht durchzulesen und einfach auf ‚Okay‘ oder ‚Cookies akzeptieren‘ zu klicken (Statista Research Department 2020).

Im Folgenden soll vor der Darstellung der auf deutscher und EU-Ebene geplanten Regelungen aufgezeigt werden, welche rechtlichen Rahmenbedingungen grundsätzlich zu beachten sind, die sich mit den Rechten an den durch KI betroffenen Daten und der zugrundeliegenden Software befassen.

Fragen des Geistigen Eigentums

Geistiges Eigentum an den in Programm- und Datenform vorliegenden Ergebnissen der KI wird vor allem an urheberrechtlichen Fragen auszurichten sein. Urheberrechtlich ist zu klären, wer der Schöpfer der Ergebnisse der KI ist, sowohl was neu generierten Programmcode als auch Datensätze betrifft. Handelt es sich um den Urheber des Programms, da dieser typischerweise bereits vorgefertigte Routinen in die KI implementiert hat, welche dann – ebenfalls nach vom Urheber vorgegebenen Algorithmen – zusammengesetzt werden (siehe Abb. 1)? Ist das Ergebnis der KI daher nichts weiter als eine verlängerte Schöpfung durch einen Programmierer? Falls dies so wäre, würde sich jedoch die Frage stellen, ob KI dann selbst eine ausreichende Kreati-

Unvorhersehbarkeit und Intransparenz dürfen nicht dazu führen, dass das Recht den Innovationszyklen der KI hinterherhinkt.

bestimmten Auswertungsbereichen (Big Data/gläserner Bürger) beschränkt. Anzustreben wäre eine homogene Regulierung sowohl der Fragen der Inhaberschaft der von der KI geschaffenen Softwareroutinen, der Zuordnung der ausgewerteten und erzeugten Datensätze, der Begrenzung der Einsatzgebiete und der Haftung für die Einsatzfolgen.

Diese Regelung müsste sich zugleich in KI-Systemen transparent abbilden lassen, so wie dies etwa beim Risikomanagement im Datenschutzbereich nach der DSGVO vorgesehen ist. Andernfalls droht eine Zersplitterung der gesetzlichen Lage, welche es den Betroffenen der KI-Nutzung erschwert, überhaupt einen Überblick über die KI-Nutzung zu erhalten. Je mehr Einzelregulierungen vorhanden sind, welche umfangreiche Informationspflichten enthalten, umso größer ist die Gefahr der sogenannten ‚Consent Fatigue‘ (Rauer und Ettig 2021): Nutzer

vität beinhaltet und so wiederum die Schöpfungshöhe des Programms in Frage stellen. Der KI selbst Rechte an den Ergebnissen einzuräumen, würde offenlassen, wer über die Rechteeräumung an Dritte entscheiden soll (Legner 2019): Die KI oder die Schöpfer der KI? Oder auch die Nutzer?

Grundsätzlich erfordert der Investitionsschutz nach dem geltenden Urheberrechtsgesetz eine menschliche Schöpfung. Je weiter sich das Schöpfungsergebnis durch den Einsatz von KI von den Vorgaben der Programmierung entwickelt, umso weniger kann der Schutz dem Urheber als menschlichem Schöpfenden zugerechnet werden (Specht-Riemenschneider 2021). Um die Rechtsunsicherheiten beim Investitionsschutz zu beseitigen, wird für die Softwareerzeugnisse der KI u. a. ein Schutzrecht für Algorithmen erzeugnisse gefordert (Specht-Riemenschneider 2021).

Eigentumsrechtliche Zuordnung der Datensätze

Neben den neu geschaffenen Software-routinen ist zu fragen, wer Inhaber der durch die KI geschaffenen Daten sein kann. Grundsätzlich geht die Rechtsprechung davon aus, dass Software eine Sache darstellt, da sie einen binären physikalischen Ladungszustand auf einem Datenträger repräsentiert. Folgt man dieser Auffassung und wendet sie auch auf sonstige Daten an, könnten die Inhaber der Hardware Ausschlussrechte gegenüber Dritten an den Ergebnissen der KI geltend machen. Gegen die dingliche beziehungsweise eigentumsrechtliche Zuordnung von Datensätzen wendet sich der Bundesbeauftragte für Datenschutz und Informationssicherheit mit der Forderung, dass statt „des verdinglichenden Datenbegriffes im Sinne eines ‚Dateneigentums‘ [...] eine Datenwirtschaft den Leitbegriff der Information und damit auch die Wissensperspektive“ (BfDI 2021, S. 2) betont werden sollten, damit „die gesellschaftlichen Herausforderungen [...] mit Blick auf Daten als öffentliches Gut und die Potenziale von Open Source, Open Data und Stärkung von Demokratiestrukturen besser sichtbar“ (ebd.) würden. Teils werden auch Rechte der Skribenten der Daten (also derjenigen, welche faktisch, etwa durch das Festlegen und Befahren einer Navigationsroute im Pkw, die Daten-spur ‚schreiben‘) und der Datenbesitzer als Anhaltspunkt formuliert (Hoeren 2019).

Allein aus dem Versuch einer eigentumsrechtlichen Zuordnung lassen sich keine adäquaten Lösungen finden, da beispielsweise beim Wechsel des Geschäftsmodells vom Verkauf eines smarten Produkts (samt KI) zu einer Vermietung der Systembetreiber die Zuordnung des Eigentums bzw. des Besitzes allein bestimmen könnte. Auch der Besitz als Anhaltspunkt ist manipulierbar (etwa durch die Verlegung der Daten in die Cloud), so dass die Skribenteneigenschaft noch am ehesten eine Rechtezuordnung an den Nutzer der KI erlaubt.

Schutz personenbezogener Daten

Datenschutzrechtlich können die Ergebnisse der KI nur relevant sein, wenn sie auf eine natürliche Person bezogen oder beziehbar sind. Ob dies zutrifft, kann nur anhand des Einsatzes der KI beurteilt werden. Liegt ein Personenbezug der auszuwertenden Daten vor, können KI-Einsätze nur auf Basis einer Rechtsgrund-

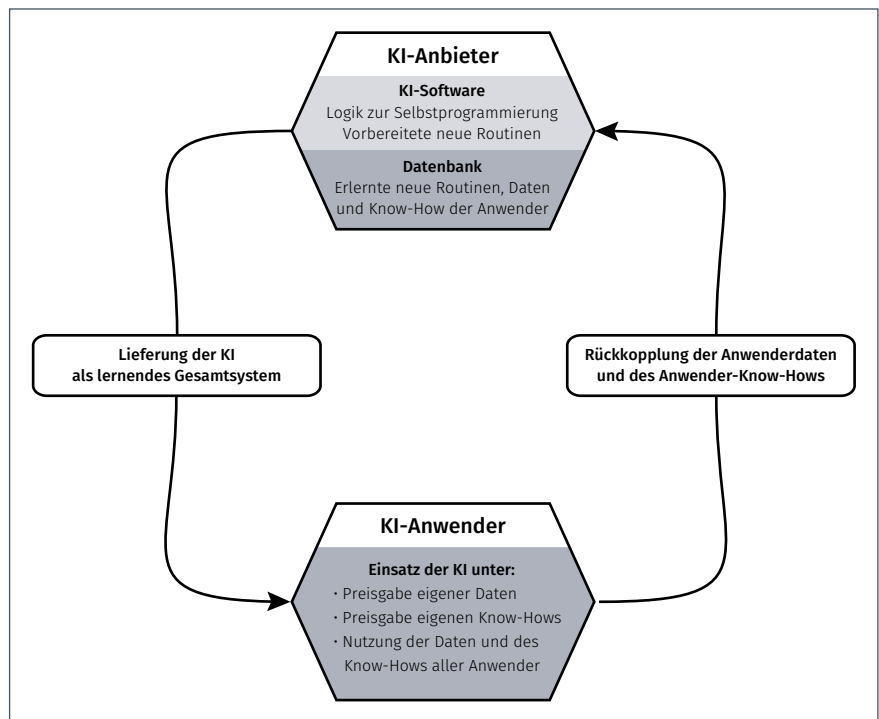


Abb. 1: Schematische KI-Darstellung für KI mit fortlaufender Rückkoppelung der Ergebnisse.

Falls man den Urheber des Programms nicht als Schöpfer des Ergebnisses der KI betrachtet, würde es sich bei den Ergebnissen (einschließlich neuer Algorithmen) nicht mehr um eine ‚menschliche‘ Schöpfung im Sinne des Urhebergesetzes handeln (Ory und Sorge 2019). Gleiches würde für den nach dem Urheberrecht möglichen Datenbankschutz gelten. In der Konsequenz wären die Ergebnisse dann gemeinfrei und würden der Allgemeinheit zur Verfügung stehen. Die Nichtzuordnung einer kreativen Leistung zu einem Menschen bedeutete damit, dass die Schöpfung von Ergebnissen durch KI nach urheberrechtlichen Maßstäben nicht zu kommerzialisieren wäre.

Quelle: eigene Darstellung

lage (u. a. Einwilligung, Vertragserfüllung, berechtigtes Interesse etc.) erfolgen. Bereits 15,9 % der 2019 vom Bundesverband der Personalmanager im März 2019 befragten Unternehmen setzen KI-Anwendungen in der Personalarbeit ein (BPM 2019). Nach den europäischen Maßstäben der DSGVO erfordert ein solcher Einsatz neben einer Rechtsgrundlage die Offenlegung der Scoring-Berechnungen und die Anwendung der Antidiskriminierungsgrundsätze des Allgemeinen Gleichbehandlungsgesetzes. Zu den Risiken im Datenschutzbereich wird auf die ‚Opazität‘ der KI verwiesen, welche keine transparenten Vorhersagen über das Verhalten der KI zulasse (Europäische Kommission 2020 a, S. 14). Als KI-immanente Risiken gelten Unvorhersehbarkeit und Intransparenz (Meyer 2018), diese dürfen jedoch nicht dazu führen, dass das Recht den Innovationszyklen der KI hinterherhinkt.

Problematisch kann der KI-Einsatz in Arbeitsverhältnissen sein, in welchem KI-Einsatz aufgrund einer Einwilligung – trotz des möglichen Machtungleichgewichts – denkbar ist. Nach einem datenschutz- und grundrechtlich kritisierten Urteil des Landesarbeitsgerichts München¹ kann sogar die Verarbeitung personenbezogener Beschäftigendaten durch eine KI-Software

¹ Beschluss vom 23. 07. 2020 – 2 Tabakverordnung 126/19.

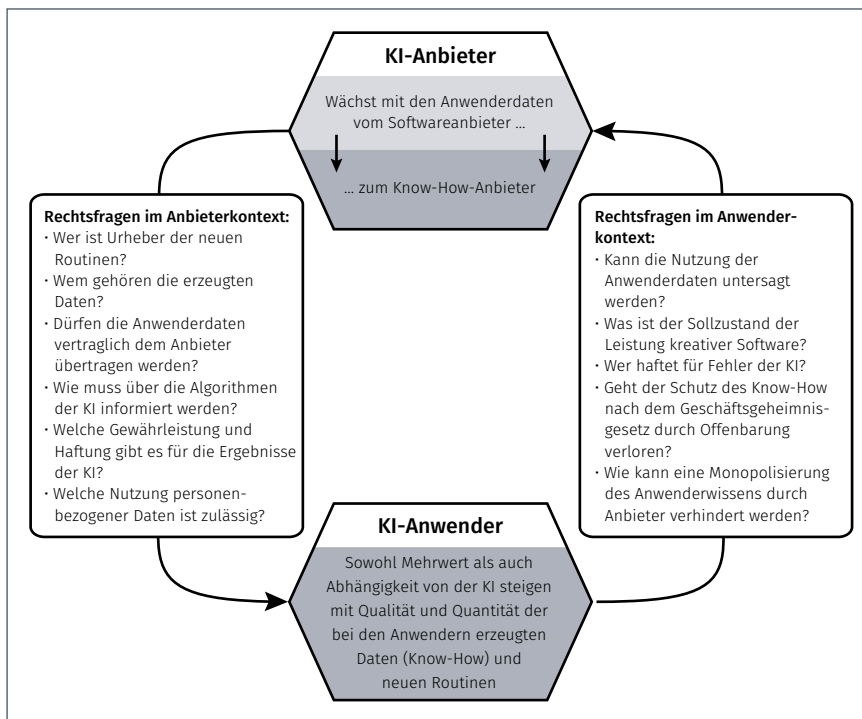


Abb.2: Formaler Knowhow-Verlust durch Offenbarung des Knowhows an KI.

Quelle: eigene Darstellung

für präventive Zwecke zulässig sein, wenn es darum geht, Auffälligkeiten des individuellen Arbeitsverhaltens zu erkennen (Wedde 2021).

Eine Einwilligung in die KI-Nutzung und entsprechende Big-Data-Anwendungen wird allerdings in aller Regel an den erforderlichen datenschutzrechtlichen Transparenzvorgaben scheitern. Die Zwecke der durch eine Einwilligung erlaubten Datenverarbeitung nach Art. 6 I a/Art. 7 DSGVO müssen konkret benannt werden. Dies gebietet auch das Prinzip der Zweckbindung nach Art. 5 I b DSGVO, was bei einer echten Big Data-Datenaggregation naturgemäß nicht der Fall sein wird, da es gerade die Idee einer Datenaggregation ist, noch unbekannte spätere Analyse und Netzwerkeffekte zu generieren (Holthausen 2021).

Diese Widersprüche zum Wunsch nach einer Big-Data-Nutzung in Europa lassen sich schwer auflösen. Nachdem sich die deutsche Bundesregierung mit den Eckpunkten einer Datenstrategie bemüht hat, eine Kommerzialisierung der Datenauswertung nicht grundsätzlich abzulehnen, wird diese Strategie vom Bundesbeauftragten für Datenschutz kritisch hinterfragt (BfDI 2021, S. 2). Auch die Idee, dem Konflikt mit dem Datenschutz durch eine Anonymisierung personenbezogener Daten zu entgehen, welche die Daten dem Anwendungsbereich der DSGVO entziehen würde, ist umstritten. Teilweise wird nämlich vertreten, dass die Anonymisierung selbst ebenfalls einer datenschutzrechtlichen Rechtsgrundlage bedarf (Hornung und Wagner 2020). Datenschutz kann jedenfalls dann Big-Data-Anwendungen entgegenstehen, soweit keine vollständige Anonymisierung der Daten vorgenommen werden kann (Roßna-

gel 2013). Selbst wenn eine Anonymisierung gelingt, sind Beeinträchtigungen der Grundrechtsausübung nicht ausgeschlossen. Roßnagel (2013, S. 566) weist für bestimmte Einsatzszenarien darauf hin, dass KI-Analysen etwa dazu führen können, dass „durch eine Big-Data-Analyse bekannt wird, dass Angehörige einer politischen Gruppe sich überdurchschnittlich oft an einem bestimmten Ort aufhalten, in einem bestimmten Geschäft einkaufen und ÖPNV fahren“. Dies könne wiederum dazu führen, dass „bei vielen, die vermeiden wollen, dieser Gruppe zugerechnet zu werden, [...] dass sie den Ort, das Geschäft und vielleicht sogar das Verkehrsmittel meiden, um nicht in einen ihnen unangenehmen Verdacht zu geraten“ (ebd.). Damit verstärkten solche Analysen „die Normativität der Normalität“ (ebd.) und reduzierten die für die Demokratie notwendige „Soziodiversität“.

Das Datenschutzrecht bietet somit keine befriedigende Auskunft über die Grenzen der Nutzungsmöglichkeit der KI, da

der von dieser generierte Datensatz – durch Anonymisierung den Regelungen der Datenschutzgrundverordnung und weiterer in Verabschiedung befindlicher E-Privacy-Regelungen (Council of the European Union 2021) – der Kontrolle der Betroffenen entzogen werden kann.

Versuche der Monopolisierung der KI-Ergebnisse

Vertragsrechtlich sind in der Praxis Versuche zu verzeichnen, die Nutzern der KI zu verpflichten, die KI-Ergebnisse den Herstellenden zu übertragen, damit diese dann Zugriff auf die Ergebnisse erhalten und entsprechende Big-Data-Anwendungen erstellen, um die Daten anderweitig nutzen zu können. Grenzen dieser Rückübertragungsversuche finden sich im Recht der Allgemeinen Geschäftsbedingungen und des Kartellrechts.

De lege lata ist mithin festzuhalten, dass die KI-Ergebnisse nicht ohne Weiteres den Herstellenden zustehen. Da jedoch die KI vertragliche Rahmenbedingungen voraussetzt, wird in aller Regel auch die Rückkopplung der KI-Ergebnisse den Herstellenden zu Gute kommen, siehe Abb. 2. Gesetzliche Haftungsregelungen wiederum werden den multikausalen Zusammenhängen bei KI-Ergebnissen und der Zersplitterung der Haftungsordnung angesichts der Beweismöte der Anwendern nicht gerecht.

Daher bedarf es einer Regelung der Datensouveränität der Betroffenen, die die Wissensagglomeration der Betreiber der KI-Systeme angemessen einbezieht.

Gesellschaftliche Folgen des KI-Einsatzes und Lösungsansätze

Angesichts der unübersichtlichen und zu komplexen rechtlichen Situation (Kühling und Sackmann 2020) besteht weiterer Handlungsbedarf zum Umgang mit der KI, eine lediglich beobachtende Haltung wird nicht genügen.

De lege ferenda werden Versuche unternommen, in einzelnen marktrelevanten Bereichen die Offenlegung von Algorithmen vorzugeben, welche beispielsweise die Funktionsweise digitaler Marktplätze und die Information über KI-Funktionsweisen betreffen:

Deutsche Gesetzgebung

Im Gesetzentwurf der Bundesregierung „zur besseren Durchsetzung und Modernisierung der Verbraucherschutzvorschriften der Union“ (Deutscher Bundestag 2021) werden in Artikel 246 d § 1 BGB-neu allgemeine Informationspflichten für Betreiber von Online-Marktplätzen festgelegt, etwa über die „Hauptparameter zur Festlegung des Rankings und die relative Gewichtung der Hauptparameter zur Festlegung des Rankings im Vergleich zu anderen Parametern“ zu informieren. Die Reform des Netzwerkdurchsetzungsgesetzes (NetzDG), welches sich mit der Bekämpfung von ‚Hate Speech‘ auf sozialen Plattformen befasst, greift die KI-Problematik auf, indem eine in § 2 Absatz 2 NetzDG festgelegte neue Informationspflicht vorsieht, dass über „Art, Grundzüge der Funktionsweise und Reichweite von gegebenenfalls eingesetzten Verfahren zur automatisierten Erkennung von Inhalten“ Auskunft gegeben werden muss.

Weiterhin wird in einer Neuregelung des Kartellrechts (§ 19 Abs. 2 Nr. 1 GWB) vorgesehen, dass unter bestimmten Umständen Zugang zu Daten und gegebenenfalls auch KI-Datenbanken eröffnet werden muss (Huerkamp und Nuys 2021). Weiterhin wird im Betriebsrätemodernisierungsgesetz zur Stärkung der Stellung des Betriebsrats bei der Beurteilung von KI-Einsätzen die Hinzuziehung einer sachverständigen Person ermöglicht.

Europäische Ansätze

Umfassender als diese zersplitterten nationalen Ansätze fallen, wie im Folgenden dargestellt, europäische Regelungen zur KI-Kontrolle aus:

Artificial Intelligence Act

In diesem Verordnungsvorschlag (COM/2021/206 final) sollen unter anderem KI-Systeme verboten werden (Art. 5 des Verordnungsentwurfs), die Techniken der unterschweligen Beeinflussung außerhalb des Bewusstseins einer Person einsetzen oder eine Schwäche beziehungsweise Schutzbedürftigkeit einer bestimmten Gruppe von Personen ausnutzen sowie Systeme zur Klassifizierung der Vertrauenswürdigkeit natürlicher Personen und biometrische Echtzeit-Fernidentifizierungssysteme. Daneben sind Regelungen zur transparenten Information, zum Risikomanagement, zu Daten und Daten-Governance, zur mensch-

lichen Aufsicht und zu weiteren Sicherheitsmaßnahmen bei ‚Hochrisiko‘-KI-Systemen vorgesehen.

Data Governance Act, Data Act

Mit dem Data Governance Act (Europäische Kommission 2020b) soll generell die Bereitstellung von Daten geregelt werden. Dies gilt für Daten des öffentlichen Sektors zur Weiterverwendung in Fällen, in denen diese Daten den Rechten anderer unterliegen, die gemeinsame Datennutzung durch Unternehmen gegen Entgelt in jedweder Form, die Ermöglichung der Nutzung personenbezogener Daten mithilfe eines „Mittlers für die gemeinsame Nutzung personenbezogener Daten“ (Europäische Kommission 2020b), der Einzelpersonen bei der Ausübung ihrer Rechte gemäß der Datenschutz-Grundverordnung (DSGVO) unterstützen soll, sowie die Ermöglichung der Nutzung von Daten aus altruistischen Gründen. Er soll nicht darauf abzielen, „wesentliche Rechte auf den Zugang zu Daten und deren Nutzung zu gewähren, zu ändern oder zu beseitigen“ (Europäische Kommission 2020b). Es ist geplant, diese Art von Maßnahmen in einen möglichen Rechtsakt über Daten (2021) aufzunehmen (European Parliament 2021).

In Erwägungsgrund 6 des Data Governance Act wird darauf verwiesen, dass es für Datenbanken, die personenbezogene Daten enthalten, Techniken gibt, „die datenschutzfreundliche Analysen ermöglichen, zum Beispiel Anonymisierung, Pseudonymisierung, differentielle Privatsphäre, Generalisierung oder Datenunterdrückung und Randomisierung“ (Europäische Kommission 2020b). Mithilfe dieser soll eine sichere Weiterverwendung personenbezogener Daten und vertraulicher Geschäftsdaten für Forschung, Innovation und statistische Zwecke gewährleistet werden können. Schließlich heißt es, dass die Verarbeitung personenbezogener Daten stets auf einem der Rechtsgründe beruhen solle, die in Artikel 6 DSGVO aufgeführt sind.

Die Abgrenzung der DSGVO zum Data Governance Act bleibt leider unklar (Graef et al. 2020), da nicht eindeutig geklärt ist, ob der Data Governance Act auch pseudonyme Daten erfasst. Die neue Gesetzgebung „wird die im Rahmen der Richtlinie über offene Daten noch zu erlassenden Vorschriften über hochwertige Datensätze ergänzen, die EU-weit den kostenlosen Zugang zu bestimmten Datensätzen in maschinenlesbarem Format und über genormte Anwendungsprogrammierschnittstellen [...] regeln sollen. [...] Für 2021 sind weitere konkrete Vorschläge für besondere Datenräume geplant, beispielsweise für einen europäischen Gesundheitsdatenraum und einen Datenraum für den Grünen Deal. All dies wird ergänzt durch einen neuen Rechtsakt über Daten, der den Bürgerinnen und Bürgern sowie den Behörden einen besseren Zugang zu Daten aus dem sogenannten Internet of Things der Industrie und zu Massendaten im Besitz von Unternehmen sichern und eine bessere Kontrolle über solche Daten verschaffen soll, um eine gerechtere Wirtschaft aufzubauen und Vorteile für die Gesellschaft insgesamt zu erzielen.“ (Europäische Kommission 2020c, 4) Soweit diese Regelungen parallel nebeneinander bestehen bleiben, wird der Wunsch nach einer transparenten Regelung jedoch unerhört bleiben.

Vorschläge zur vertraglichen Regulierung, zu Control by Design und Technikgestaltung

Politisch wünschbare Regulierungsansätze – und insbesondere der geplante europäische ‚Data Act‘ sollten daher Grenzen der Auswertbarkeit der KI-Ergebnisse durch die Systembetreiber festlegen, einen transparenten Einsatz der KI insbesondere in sensiblen persönlichkeitsrechtlichen Bereichen der Analyse (Medizin und Arbeitswelt) festlegen und arbeitsweltliche Grenzen des Ersatzes menschlicher Arbeitsleistung durch KI zumindest anstreben.

Denkbar wäre eine übergreifende Regulierung auf europäischer Ebene, welche technikorientiert, ähnlich wie die aktuell (erneut) in Verabschiedung befindliche E-Privacy-Verordnung, und szenariobasiert die Risiken der KI in den Blick nimmt und die erwähnten rechtliche Lücken schließt. Zugleich muss sie jedoch das Verhältnis zur Datenschutzgrundverordnung klarer als im Data Governance Act regeln und auch eine Einwilligung in noch nicht bekannte Auswertungsarten und Daten mit verschiedenen Stufen des Personenbezugs in bestimmten Fällen zulassen, wenn KI und Big Data ernsthaft betrieben werden sollen.

Hierzu kämen zwei Ansätze in Betracht:

Klauselregelungen für Datenübertragungen

Da die dargestellten gesetzlichen Regulierungsansätze völlig heterogene Fälle der Information der Nutzer über den KI-Einsatz (abhängig von der Art der Daten) vorsehen, sollte eine Rahmenregelung für zulässige Datenauswertungen sowohl für personenbezogene als auch für anonymisierte Daten vorgesehen werden (Schur 2020). Diese Klauselregelungen könnten sich an den AGB-Klauselregelungen der §§ 308, 309 BGB oder den sog. ‚grauen‘ und ‚schwarzen‘ Klauseln der EU-Gruppenfrei stellungsverordnungen orientieren. So könnten – orientiert an den bekannten Technikfolgen der KI – die Grenzen einwilligungsfähiger personenbezogener Datenauswertungen und nicht personenbezogener Datenauswertungen unterschieden werden. Dies würde zugleich einen klaren Rahmen für die Ersteller und Betreiber von KI-Systemen darstellen, die Rechtssicherheit über die wirksamen Vereinbarungen mit den Nutzern erhalten. Denkbar wäre hier auch eine Beteiligung der Nutzer an der Wertschöpfung abhängig vom Umfang der Einwilligung in die Datenauswertung.

Control by Design

Analog zu den Regelungen der DSGVO für Privacy by Design könnte auch für nicht personenbezogene Daten – begrenzt durch die und angebunden an die oben erwähnten Klauselregelungen – eine Control by Design-Regelung eingeführt werden (der Begriff ‚Privacy‘ träfe hier nicht zu, da auch anonyme Auswertungen von der Datensouveränität umfasst wären).

So wie das im Entwurf vorliegende Gesetz über den Datenschutz und den Schutz der Privatsphäre in der Telekommunikation und bei Telemedien (TTDSG) in seinen §§ 25, 26 den

Einsatz von ‚Personal Information Management Services‘ regelt und eine transparente Einwilligung der Freigabe von Endnutzerdaten (etwa bei Cookies) ermöglicht, könnte auch im Data Act die im TTDSG vorgesehene Einschaltung von „Anerkannten Diensten zur Einwilligungsverwaltung“ (§ 26 TTDSG) erfolgen.

In der konkreten Ausprägung der Einwilligung könnten auch banale Fragen wie die Haftung für Ergebnisse der KI-Aktivitäten gegenüber den beteiligten Akteuren (Hacker 2020; Zech 2019) oder die Übertragung der Scoring-Transparenz der DSGVO für nicht-personenbezogene KI-Daten berücksichtigt werden.

Die EU versucht in ihren Regulierungsansätzen, der Marktmonopolisierung im KI-Umfeld und der Monetarisierung der Nutzerdaten entgegenzuwirken, um auch altruistische Datenverwendungen zu ermöglichen und bei der Anwendung der KI die Grundrechte der Nutzer zu beachten. Eine entsprechende Umsetzung in der Praxis müsste sich jedoch am Kriterium der Umsetzbarkeit messen lassen, wenn sie erfolgreich sein soll.

Mit einer solchen – über den Data Governance Act hinausgehenden – Regelung des Control by Design wäre es denkbar, der Datensouveränität der Nutzer gerecht zu werden und zugleich eine sichere Rechtsgrundlage für die Betreiber der Systeme zu schaffen.

Angabe von Finanzierungsquellen

Der vorliegende Forschungsartikel hat keine Förderung erhalten.

Literatur

- BfDI – Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (2021): Stellungnahme des Bundesbeauftragten für den Datenschutz und die Informationsfreiheit zur öffentlichen Anhörung des Ausschusses Digitale Agenda am 24. Februar 2021. Online verfügbar unter https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Stellungnahmen/2021/StgN_StgN_Anh%C3%B6rung-Datenstrategie-Bundesregierung.pdf;jsessionid=7D1B64827D3B8DA4D33C87024D824D47.intranet242?__blob=publicationFile&v=4, zuletzt geprüft am 21.10.2021, S. 1.
- BPM – Bundesverband der Personalmanager (2019): Künstliche Intelligenz in der Personalarbeit. Online verfügbar unter https://www.bpm.de/sites/default/files/20190429_auswertung_bpm-pressemitteilung_final_0.pdf, zuletzt aufgerufen am 18.10.2021.
- Council of the European Union (2021): Proposal for a regulation of the European Parliament and of the council concerning the respect for private life and the protection of personal data in electronic communications and repealing directive 2002/58/EC (regulation on privacy and electronic communications). Online verfügbar unter <https://data.consilium.europa.eu/doc/document/ST-6087-2021-INIT/en/pdf>, zuletzt geprüft am 21.10.2021.
- Datenethikkommission (2019): Gutachten der Datenethikkommission der Bundesregierung, Potsdam, S. 224. Online verfügbar unter https://datenethikkommission.de/wp-content/uploads/191015_DEK_Gutachten_screen.pdf, zuletzt geprüft am 26.09.2021.
- Deutscher Bundestag (2021): Drucksache 19/27655. Berlin: H. Heenemann. Online verfügbar unter <https://dserver.bundestag.de/btd/19/276/1927655.pdf>, zuletzt geprüft am 21.10.2021.
- Europäische Kommission (2020 a): Weißbuch zur Künstlichen Intelligenz. Ein europäisches Konzept für Exzellenz und Vertrauen. Online verfügbar unter

https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf, zuletzt geprüft am 10. 11. 2021.

Europäische Kommission (2020 b): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über europäische Daten-Governance (Daten-Governance-Gesetz). Online verfügbar unter <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52020PC0767&from=EN>, zuletzt geprüft am 08. 11. 2021.

Europäische Kommission (2020 c): Verordnung über Daten-Governance – Fragen und Antworten. Online verfügbar unter https://ec.europa.eu/commission/presscorner/api/files/document/print/de/qanda_20_2103/QANDA_20_2103_DE.pdf, zuletzt geprüft am 25. 11. 2021.

European Parliament (2021): Legislative Train 09.2021. 2 a Europe fit for the digital age. Online verfügbar unter <https://www.europarl.europa.eu/legislative-train/api/stages/report/current/theme/a-europe-fit-for-the-digital-age/file/data-act>, zuletzt geprüft am 21. 10. 2021.

Graef, Inge; Husovec, Martin; van den Boom, Jasper (2020): Spill-overs in data governance. Uncovering the uneasy relationship between the GDPR's right to data portability and EU sector-specific data access regimes. In: *Journal of European Consumer and Market Law* 9 (1), S. 3–16. <https://doi.org/10.2139/ssrn.3369509>

Hacker, Philipp (2020): Europäische und nationale Regulierung von Künstlicher Intelligenz. In: *Neue Juristische Wochenschrift* 73 (30), S. 2142–2147. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Ffnjw%2F2020%2Fcont%2Ffnjw.2020.2142.1.htm&pos=1&hlwords=on>, zuletzt geprüft am 21. 10. 2021.

Herberger, Maximilian (2018): „Künstliche Intelligenz“ und Recht. In: *Neue Juristische Wochenschrift* 71 (39), S. 2825–2829. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Ffnjw%2F2018%2Fcont%2Ffnjw.2018.2825.1.htm&pos=2&hlwords=on>, zuletzt geprüft am 21. 10. 2021.

Hoeren, Thomas (2019): Datenbesitz statt Dateneigentum. In: *Multimedia und Recht* 22 (1), S. 5–8. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fmrmr%2F2019%2Fcont%2Fmrmr.2019.5.1.htm&pos=1&hlwords=on>, zuletzt geprüft am 21. 10. 2021.

Holthausen, Joachim (2021): Big data, people analytics, KI und Gestaltung von Betriebsvereinbarungen. Grund-, arbeits- und datenschutzrechtliche An- und Herausforderungen. In: *Recht der Arbeit* 74 (1), S. 19–32. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Frda%2F2021%2Fcont%2Frda.2021.19.1.htm&pos=3>, zuletzt geprüft am 21. 10. 2021.

Hornung, Gerrit; Wagner, Bernd (2020): Anonymisierung als datenschutzrelevante Verarbeitung? In: *Zeitschrift für Datenschutz* 10 (5), S. 223–228. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fzd%2F2020%2Fcont%2Fzd.2020.223.1.htm&anchor=Y-300-Z-ZD-B-2020-S-223-N-1>, zuletzt geprüft am 21. 10. 2021.

Huerkamp, Florian; Nuys, Marcel (2021): Datenzugang nach § 19 Abs. 2 Nr. 4 GWB n.F. – Geglückte „Klarstellung“? In: *Neue Zeitschrift für Kartellrecht* 9 (7), S. 327–329. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fnzkart%2F2021%2Fcont%2Fnzkart.2021.327.1.htm&pos=7>, zuletzt geprüft am 21. 10. 2021.

Kühling, Jürgen; Sackmann, Florian (2020): Irrweg Dateneigentum. In *Zeitschrift für Datenschutz* 10 (1), S. 24–30. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fzd%2F2020%2Fcont%2Fzd.2020.24.1.htm&pos=16>, zuletzt geprüft am 21. 10. 2021.

Legner, Sarah (2019): Erzeugnisse Künstlicher Intelligenz im Urheberrecht. In: *Zeitschrift für Urheber und Medienrecht* 63 (11), S. 807–812. Online verfügbar

unter https://www.researchgate.net/profile/Sarah-Legner-2/publication/343510297_Erzeugnisse_Kunstlicher_Intelligenz_im_Urheberrecht/links/5fb4d44fa6fdcc9ae05ef8c8/Erzeugnisse-Kuenstlicher-Intelligenz-im-Urheberrecht.pdf, zuletzt geprüft am 21. 10. 2021.

Meyer, Stephan (2018): Künstliche Intelligenz und die Rolle des Rechts für Innovation. In: *Zeitschrift für Rechtspolitik* 51 (8), S. 233–237. Online verfügbar unter <https://beck-online.beck.de/?vpath=bibdata%2Fzeits%2Fzrp%2Fcont%2Fzrp%2ehtm>, zuletzt geprüft am 21. 10. 2021.

Ory, Stephan; Sorge, Christoph (2019): Schöpfung durch Künstliche Intelligenz? In: *Neue Juristische Wochenschrift* 72 (11), S. 710–712. Online verfügbar unter <https://beck-online.beck.de/?vpath=bibdata%2Fzeits%2Ffnjw%2F2019%2Fcont%2Ffnjw%2E2019%2E710%2E1%2Ehtm>, zuletzt geprüft am 21. 10. 2021.

Rauer, Nils; Ettig, Diana (2021): Update Cookies 2020. Aktuelle Rechtslage und Entwicklungen. In: *Zeitschrift für Datenschutz* 11 (1), S. 18–24. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fzd%2F2021%2Fcont%2Fzd.2021.18.1.htm&pos=2&hlwords=on>, zuletzt geprüft am 21. 10. 2021.

Roßnagel, Alexander (2013): Big Data – Small Privacy? Konzeptionelle Herausforderungen für das Datenschutzrecht. In: *Zeitschrift für Datenschutz* 3 (11), S. 562–567. Online verfügbar unter <https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fzd%2F2013%2Fcont%2Fzd.2013.562.1.htm&anchor=Y-300-Z-ZD-B-2013-S-562-N-1>, zuletzt geprüft am 21. 10. 2021.

Schur, Nico (2020): Die Lizenzierung von Daten. Einordnung, Grenzen und Möglichkeiten von vertraglichen Zugangs- und Datennutzungsrechten in der digitalen Ökonomie. In: Jeanette Hofmann; Ingolf Pernice; Thomas Schildhauer; Wolfgang Schulz (Hg.): *Internet und Gesellschaft. Schriften des Alexander von Humboldt Institut für Internet und Gesellschaft*. Tübingen Mohr Siebeck.

Specht-Riemenschneider, Louisa (2021): Urheberrechtlicher Schutz für Algorithmen-erzeugnisse? Phasenmodell de lege lata, Investitionsschutz de lege ferenda? In: *Wettbewerb in Recht und Praxis* 18 (3), S. 273–275.

Statista Research Department (2020): Umfrage zum Umgang mit Cookie-Hinweisen in Deutschland 2020. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/1121071/umfrage/umgang-mit-cookie-hinweisen-in-deutschland/>, zuletzt geprüft am 13. 10. 2021.

Wedde, Peter (2021): Anmerkung – Streit um Einigungsstellenspruch zur Einführung eines IT-Sicherheitssystems. Anlasslose präventive Verarbeitung von Beschäftigtendaten durch KI-Software zulässig, LArBG München v. 23. 07. 2020–2 TaBV 126/19, jurisPR-ArbR 17/2021 Anm. 6. In: *Juris – Das Rechtsportal*.

Zech, Herbert (2019): Künstliche Intelligenz und Haftungsfragen. In: *Zeitschrift für die gesamte Privatrechtswissenschaft* 5 (2), S. 198–219.



PROF. DR. THOMAS WILMER

ist seit 2002 Professor für Informationsrecht an der Hochschule Darmstadt und leitet dort das Institut für Informationsrecht. Er befasst sich mit Fragen des Lizenz- und Datenschutzrechts und verwandter Bereiche des Informationsrechts.

RESEARCH ARTICLE

Bürgerwissenschaftliche Forschungsansätze in Medizin und Gesundheitsforschung

Ausgewählte Begriffe mit Fokus auf den Beteiligungsgrad

63

Gertrud Hammel, Institut für Umweltmedizin, Helmholtz Zentrum München, Standort Augsburg,

Neusässer Str. 47, 86156 Augsburg, DE (gertrud.hammel@helmholtz-muenchen.de)  0000-0002-1800-7679

Silvia Woll (silvia.woll@kit.edu)  0000-0003-0485-2246, Martina Baumann (martina.baumann@kit.edu)  0000-0003-0487-9975,

Constanze Scherz (constanze.scherz@kit.edu)  0000-0003-4488-439X, Maria Maia (maria.maia@kit.edu)  0000-0002-3501-6876,

alle Institut für Technikfolgenabschätzung und Systemanalyse (ITAS), Karlsruher Institut für Technologie (KIT), Karlsruhe, DE

Birgit Behrisch, Katholische Hochschule für Sozialwesen Berlin, Berlin, DE (birgit.behrisch@khsb-berlin.de)  0000-0001-9678-7554

Sandra Olivia Borgmann, Institut für Versorgungsforschung und Gesundheitsökonomie, Deutsches Diabetes-Zentrum, Leibniz-Zentrum für Diabetes-Forschung

an der Heinrich-Heine-Universität Düsseldorf, Düsseldorf, DE (sandra.borgmann@med.uni-duesseldorf.de)  0000-0001-5843-0730

Michael Eichinger, Mannheimer Institut für Public Health, Sozial- und Präventivmedizin, Medizinische Fakultät Mannheim der Universität Heidelberg,

Mannheim, DE (michael.eichinger@medma.uni-heidelberg.de)  0000-0003-0064-7640

Johanna Gardecki, Universitätsklinikum Frankfurt, Goethe-Universität Frankfurt, Frankfurt am Main, DE (johanna.gardecki@kgu.de)  0000-0003-1619-0059

Nils B. Heyen, Fraunhofer-Institut für System- und Innovationsforschung ISI, Karlsruhe, DE (nils.heyen@isi.fraunhofer.de)  0000-0002-9354-1388

Andrea Icks, Institut für Versorgungsforschung und Gesundheitsökonomie, Centre for Health and Society, Medizinische Fakultät und Universitätsklinikum Düsseldorf,

Heinrich-Heine-Universität Düsseldorf, Düsseldorf, DE (andrea.icks@uni-duesseldorf.de)  0000-0002-4882-969X

Monika Pobiruchin, GECKO Institut für Medizin, Informatik und Ökonomie, Hochschule Heilbronn, Heilbronn, DE

(monika.pobiruchin@hs-heilbronn.de)  0000-0002-9925-2173

Sarah Weschke, QUEST Center for Responsible Research, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, DE

(sarah.weschke@bih-charite.de)  0000-0003-2069-9245

Zusammenfassung • In den Bürgerwissenschaften, auch bekannt unter dem englischen Begriff Citizen Science, existiert eine Vielzahl an Forschungsansätzen und Methoden. Während diese in vielen wissenschaftlichen Disziplinen gut etabliert sind, finden sich augenscheinlich relativ wenige davon in der medizinischen und Gesundheitsforschung. Allerdings zeigt ein Blick in die Praxis, dass bürgerwissenschaftliche Ansätze in der Medizin und Gesundheitsforschung durchaus praktiziert werden, jedoch häufig unter anderen Namen. Der Artikel bietet aus interdisziplinärer Perspektive einen (selektiven) Überblick über Begriffe, reflektiert diese und die dahinterstehenden Methoden und diskutiert sie vergleichend. Im Fokus steht dabei der Grad der Beteiligung der Bürger*innen bzw. Patient*innen an wissenschaftlicher Forschung.

Citizen science approaches in medical and health research.

Selected terms with focus on the degree of participation

Abstract • In citizen science, a variety of research approaches and methods exist. While these are well established in many scientific disciplines, there are apparently relatively few of them in medical and health research. However, a glance at practice shows that citizen science approaches are indeed practiced in medical and health research, but often under different names. The paper therefore provides a (selective) overview of these terms, reflects on them and the methods behind them, and discusses them comparatively from an interdisciplinary perspective. The focus is on the degree of active participation of citizens and patients in research.

Keywords • citizen science, participation, medical and health research, terminology

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>) <https://doi.org/10.14512/tatup.30.3.63>
Received: Jul. 26, 2021; revised version accepted: Nov. 02, 2021;
published online: Dec. 20, 2021 (peer review)

Einleitung

Bürgerwissenschaftliche Forschungsansätze sind in der medizinischen Forschung noch nicht in dem Maße etabliert, wie das etwa in der Biologie (NABU 2013) oder Astronomie (FVN 2021) der Fall ist. Ein Blick in die Praxis lässt jedoch vermuten, dass bürgerwissenschaftliche Ansätze in der Medizin und Gesundheitsforschung durchaus praktiziert werden, allerdings häufig unter Verwendung anderer Begriffe.

Der angloamerikanische Begriff Citizen Science (CS) wird im deutschen Sprachgebrauch mit Bürgerwissenschaften gleichgesetzt. Im GEWISS-Grünbuch wird CS beschrieben als „die Beteiligung von Personen an wissenschaftlichen Prozessen, die nicht in diesem Wissenschaftsbereich institutionell gebunden sind. Dabei kann die Beteiligung in der kurzzeitigen Erhebung von Daten bis hin zu einem intensiven Einsatz von Freizeit bestehen [...]“ (Bonn et al. 2016, S. 13). Die European Citizen Science Association (ECSA) beschreibt in ihren „10 Prinzipien von Citizen Science – Bürgerwissenschaften“, dass Bürger*innen dabei als Beitragende, Mitarbeitende, Projektleitende oder in anderen relevanten Rollen agieren können (ECSA 2015).

In der Medizin und Gesundheitsforschung haben sich einige Begriffe zur Beteiligung von Patient*innen am Forschungsinhalt etabliert, allerdings lassen sich diese häufig nicht trennscharf differenzieren. So beschreiben etwa Buyx et al. (2017), dass Begriffe wie ‚citizen science‘, ‚participatory research‘ und ‚patient-led research‘ häufig synonym verwendet werden. Weiter bleibt unklar, inwieweit die Konzepte hinter diesen Begriffen tatsächlich für die Patient*innen mehr Möglichkeiten bieten, ihre eigene Perspektive in die Forschung einzubringen, als dies in klinischen Studien der Fall ist (Mahr und Strasser 2021). Bürger*innen bzw. Patient*innen können gerade die Medizin und Gesundheitsforschung bereichern, indem sie unter anderem ihr eigenes (Erfahrungs-)Wissen über ihre Erkrankung, die aus ihrer Sicht relevanten Forschungsfragen sowie die für sie akzeptablen ethischen Standards beitragen (Mahr und Strasser 2021; Vayena 2021). Jedoch werden die meisten CS-Projekte in der Medizin von akademischen Forschenden geleitet und durchgeführt, während die Partizipation der Bürger*innen bzw. Patient*innen im (passiven) Spenden von Daten und Proben besteht (Buyx et al. 2017; Del Savio et al. 2016).

Die sicherlich nicht vollständige Auswahl der im Folgenden beschriebenen Begriffe ist im Rahmen eines interdisziplinären webbasierten Workshops zum Thema ‚Bürgerwissenschaften in der Medizin: Terminologie‘ im Juli 2020 entstanden. Ausgangspunkt des Workshops war die Frage nach der Notwendigkeit neuer deutschsprachiger Begriffe in der medizinischen CS-Forschung. Durch die Darstellung bekannter Begriffe aus dem angloamerikanischen Sprachraum zeigte sich ein wesentliches Unterscheidungsmerkmal der Konzepte: der Grad der Beteiligung der Patient*innen bzw. Bürger*innen an der Forschung.

Die ausgewählten Begriffe werden im Folgenden anhand dieses Fokus vorgestellt und diskutiert mit dem Ziel, einen selektiven Überblick aus interdisziplinärer Forschungsperspektive zu

generieren, und zwar durch Autor*innen, die unter Verwendung der jeweiligen Begriffe in den verschiedenen Bereichen von CS praktizieren.

Klinische und epidemiologische Studien unterliegen den international anerkannten Regeln von „Good clinical practice“ (ICH 2021) bzw. „Good epidemiological practice“ (Hoffmann et al. 2019). Diese prägen die medizinische Forschung durch Standardisierung und hohe Qualitätsstandards bei der Planung, Durchführung, Auswertung und Veröffentlichung von Forschungsergebnissen. Der Beteiligungsgrad von Studienteilnehmer*innen an Forschungsfragen ist dabei niedrig. Studienteilnehmer*innen stellen ihre Daten zur Verfügung; dies können die Ergebnisse klinischer Untersuchungen und Biomaterialien, aber auch selbst berichtete Symptome und Einschätzungen (*Patient Reported Outcomes*) sein. Bei Letzterem wird die Sichtweise der Teilnehmenden zwar berücksichtigt, die Kontrolle über den Forschungsprozess obliegt aber den Forschenden. Ein solcher Beteiligungsgrad wird auch als „Vorstufe der Partizipation“ bezeichnet (Wright 2010). Zunehmend gibt es medizinische Forschungsprojekte, die einen CS-Ansatz einbeziehen. Als aktuelles Beispiel sei das Online-Formular des Paul-Ehrlich-Instituts genannt, mithilfe dessen Geimpfte gegen das SARS-CoV-2 (Neben-)Wirkungen erfassen können (PEI 2021). Der Grad der Beteiligung ist dabei sehr gering und würde im angloamerikanischen Raum als ‚Crowdsourcing‘ bezeichnet werden (Haklay 2013).

Versorgungsforschung ist „ein fachübergreifendes Forschungsgebiet, das ausgehend von der Patienten- und Populationsperspektive und vor dem Hintergrund komplexer Kontextbedingungen die Versorgungsstrukturen und -prozesse der Gesundheitsversorgung untersucht, den Outcome auf Ebene der Alltagsversorgung beschreibt und komplexe Interventionen zur Verbesserung der Versorgung evaluiert“ (Schrappe und Pfaff 2017, S. 11). Nach dem Aktionsplan Versorgungsforschung des Bundesministeriums für Bildung und Forschung (BMBF) von 2014 wird Patient*innenorientierung als wichtiger inhaltlicher Aspekt der Forschungsprojekte angesehen und die Einbindung von Patientenvertreter*innen in jeder Phase der Forschungsprojekte gefordert (BMBF 2016). Jedoch fokussieren die ausgeführten Fördermaßnahmen meist auf den etablierten akademischen Wissenschaftsbetrieb. Die Förderung von Projekten, die direkt von Patient*innen benannt und ins Leben gerufen werden, ist aus Sicht der Autor*innen derzeit nicht zu sehen.

Partizipative Ansätze

Partizipation in der akademischen Forschung ist mehr als jemanden ‚mitmachen lassen‘. Partizipative Verfahren unterschiedlichster Art haben die traditionellen Rollen von Expert*innen und Laien wechselseitig verändert und dazu geführt, dass die Laienperspektive als eigenständiges Element (neben der wissenschaftlichen Expertise) in der Bewertung der Relevanz wissenschaftlichen Wissens für anstehende Entscheidungen anerkannt

wurde. Teilhabe im Sinne einer vollumfänglichen Beteiligung am Forschungsprozess sehen partizipative Verfahren *qua definitionem* erst einmal nicht zwingend vor; jedoch zielt das durch partizipative Verfahren „realisierte Neuarrangement der Rollen von Wissenschaft, Politik und Öffentlichkeit [...] darauf ab, Foren für einen gleichberechtigten Dialog zwischen Laien und Experten zur Klärung umstrittener empirischer und normativer Fragen zu schaffen“ (Hennen et al. 2004, S. 53). In der Medizin wird der Begriff der Partizipation häufig im Zusammenhang mit therapeutischen Entscheidungen verwendet. Patient*innen sollen in unterschiedlichen Phasen einer Therapie gemeinsam mit den Behandelnden die richtigen Schritte auf dem Weg zur Verbesserung des Gesundheitszustandes gehen.

ginalisierte Personengruppen (z. B. indigene Gemeinschaften) und Sachverhalte, von denen die beteiligten Gemeinschaften direkt betroffen sind. Ziel ist die gleichberechtigte Beteiligung von Wissenschaftler*innen und Mitgliedern der Gemeinschaft am gesamten Forschungsprozess (Jull et al. 2017). Forschung findet damit nicht *in* Gemeinschaften, sondern *mit* Gemeinschaften statt. In der Literatur zeigen sich jedoch Hinweise, dass Vertreter*innen der beteiligten Gemeinschaften in einige Aspekte weniger häufig eingebunden sind (z. B. Entwicklung von Forschungshypothesen) als in andere. Ziele von CBPR sind (1) das Gewinnen von Erkenntnissen, die für die beteiligten Gemeinschaften relevant sind, (2) das Anstoßen sozialen Wandels und/oder gesetzlicher Änderungen, um die Gesundheit und das

Medizinische Forschung durch Bürger*innen kann auch unabhängig von der akademischen Wissenschaft geschehen.

*Patient*innenzentrierte Versorgung* wird vom Institute of Medicine definiert als eine Versorgung, die respektvoll individuelle Präferenzen, Bedürfnisse und Werte der*s Patient*in berücksichtigt und sicherstellt, dass diese in alle klinischen Entscheidungen einfließen (Institute of Medicine 2001). Die partizipative Entscheidungsfindung (*shared decision-making*) ist dabei ein Kernelement. Sie umfasst ein Modell der Patient*innen-Behandelnden-Interaktion, die auf einer „gemeinsam verantworteten Übereinkunft“ (Loh et al. 2007, S. A-1484) beider Parteien beruht. Durch einen bidirektionalen Informationsaustausch und unter Berücksichtigung der Präferenzen der Betroffenen findet eine Diskussion über die optimale Behandlung statt (Stigglebout et al. 2015). Hierbei können Betroffene darin gestärkt werden, eine aktivere Rolle in der Behandlung einzunehmen und eine informierte Entscheidung zu treffen (Légaré und Witteman 2013). Die Aktualität und Relevanz der Thematik wurde in Deutschland durch das Bundesministerium für Bildung und Forschung (BMBF) aufgegriffen (BMBF 2018, 2016). Das BMBF hebt hervor, dass „Gesundheitsforschung [...] nur erfolgreich sein [kann], wenn sie sich am Bedarf der Patientinnen und Patienten orientiert“ (BMBF 2018, S. 8). Zusätzlich formuliert es als eines ihrer zentralen Ziele, Gesundheitsinnovationen stärker an den Bedürfnissen der Menschen zu orientieren (BMBF 2018).

Bürgerwissenschaftliche Ansätze

Community-based Participatory Research (CBPR) ist ein Forschungsansatz, der sich durch eine partnerschaftliche Zusammenarbeit zwischen Wissenschaftler*innen und Gemeinschaften auszeichnet (LaVeaux und Christopher 2009). Gemeinschaften werden primär als soziale Gefüge verstanden, die sich durch gemeinsame Interessen und Bedarfe charakterisieren lassen (George et al. 2006). CBPR fokussiert dabei häufig auf mar-

Wohlergehen der beteiligten Gemeinschaften zu stärken und gesundheitliche Ungleichheiten zu reduzieren und (3) der Aufbau von Kapazitäten, die Gemeinschaften in die Lage versetzen, Herausforderungen eigenständig zu überwinden (Viswanathan et al. 2004).

Patient Engagement und Patient Involvement unterliegen keiner genauen Definition und werden häufig synonym, aber nicht immer konsistent verwendet. Das mit einem Begriff verbundene Maß an (erwünschter) Forschungsbeteiligung der Patient*innen ist teils länderspezifisch oder variiert sogar je nach Institution. In Großbritannien engagiert sich das National Institute of Health Research (NIHR) für die aktive Beteiligung von Patient*innen und der Öffentlichkeit in der Forschung (*Patient and Public Involvement*) (NIHR 2021). Der Begriff *engagement* wird als etwas verstanden, das wir im Deutschen ‚Wissenstransfer in die Gesellschaft‘ oder ‚Wissenschaftskommunikation‘ nennen würden: die eher unidirektionale Verbreitung bzw. Kommunikation von wissenschaftlichen Ergebnissen oder Wissen über Forschung in die Gesellschaft. Das Patient-Centered Outcomes Research Institute (PCORI) aus den USA definiert *engagement in research* als „meaningful involvement of patients, caregivers, clinicians, and other healthcare stakeholders throughout the entire research process [...]“ (PCORI 2018). Am QUEST Center for Responsible Research des Berlin Institute of Health at Charité wird der Begriff *Patient & Stakeholder Engagement* im Sinne der Definition von PCORI genutzt und übersetzt als „aktive Patient*innenbeteiligung in der Gesundheitsforschung“ (BIH 2021). Sowohl bei PCORI als auch beim NIHR ist mit dieser Art des *engagement* oder *involvement* nicht die Beteiligung als Proband*in in Studien gemeint, sondern Forschung, die *mit* oder *von* Menschen aus der Bevölkerung gestaltet und nicht *über* oder *für* sie durchgeführt wird.

Die Herangehensweise der *Partizipativen Gesundheitsforschung* (PGF, im Englischen zumeist *Participatory Health Re-*

search) stellt einen sehr weitreichenden Beteiligungsansatz dar: Patient*innen werden als Ko-Forschende verstanden. Beginnend mit der Aktionsforschung (Lewin 1946) in den 1940er-Jahren wird in den Sozial- und Gesundheitswissenschaften die gleichberechtigte Beteiligung und Mitwirkung von Nichtwissenschaftler*innen an Forschungsprozessen thematisiert. Dieser Forschungsansatz gründet auf dem Anspruch, „nicht nur zu forschen, sondern auch zu handeln, und dadurch soziale Wirklichkeit nicht nur zu verstehen, sondern auch zu verändern“ (Unger 2014, S. 35) und somit Befähigungs- und Ermächtigungsprozesse von nichtakademischen Ko-Forschenden zu befördern. Wichtige definitorische Aspekte von PGF beziehen sich darauf, Forschung als Koproduktion verschiedener Akteur*innen zu verstehen sowie Forschungsprozesse partnerschaftlich, machtkritisch und möglichst mit maximaler Mitgestaltung derjenigen Menschen zu organisieren, deren Lebensbereiche erforscht werden. Ziel der PGF ist die Förderung von Gesundheit und Wohlbefinden der Menschen und die Stärkung gesundheitlicher Chancengleichheit (Peter et al. 2020).

Patient Science wurde im Zuge eines vom BMBF geförderten Pilotprojekts am Beispiel der Seltenen Erkrankung Mukoviszidose entwickelt (Heyen et al. 2021). Das Projekt wollte dezidiert ein Format erproben, das ein höchstes Maß an Partizipation für Patient*innen in allen Phasen des Forschungsprozesses ermöglicht (Gardecki et al. 2021). Das Wissen, über das Patient*innen und ihre Angehörigen mit Blick auf eine bestimmte (in der Regel chronische) Erkrankung verfügen, wird hier als Expertise anerkannt und in allen Phasen des epistemischen Prozesses systematisch genutzt (Heyen et al. 2021). Patient Science geht damit hinsichtlich des Beteiligungsgrads weit über die meisten partizipativen Ansätze hinaus. Gleichzeitig unterscheidet sich Patient Science von etablierten Ansätzen wie PGF oder CBPR insofern, als es weniger auf die Transformation von bestimmten (oft lokalen) gesundheitsbezogenen Lebensverhältnissen bzw. auf ein diesbezügliches Empowerment als Förderung der Fähigkeit für selbständiges bzw. selbstbestimmtes Handeln der Betroffenen zielt, sondern ganz im Sinne von CS primär auf ein Empowerment bei der Produktion wissenschaftlichen Wissens. Dabei weist das generierte Wissen eine große Nähe zur Lebenswirklichkeit der Betroffenen und eine hohe Praxisrelevanz auf und hat das Potenzial, entscheidend zur Verbesserung der Gesundheitsversorgung beizutragen.

Patient Innovation ist ein Ansatz, der sich insbesondere durch seine Loslösung von der akademischen Forschung auszeichnet und damit – im Vergleich zu den bislang vorgestellten Begriffen – der *bürgerwissenschaftlichen* Idee am stärksten entspricht. Das Gesundheitssystem bietet oft keine adäquaten therapeutischen Optionen für Patient*innen mit (häufig seltenen, chronischen) Erkrankungen, die meist mit starken gesundheitlichen Einschränkungen und niedriger Lebensqualität konfrontiert sind. Traditionell geht man im Gesundheitswesen davon aus, dass medizinische Fachkräfte eher als Patient*innen das Wissen und die Fähigkeiten zur Innovation haben, jedoch entwickeln manche der Betroffenen bzw. ihre nicht-professio-

nellen Betreuenden eigene Lösungen für ihre spezifischen Probleme (Habicht et al. 2012). Die Innovationen durch Patient*innen (*patient innovators*), von denen über 1.600 alleine auf der Plattform patient-innovation.com zu finden sind, reichen von einfachen nicht-technologischen Alltagspraktiken über die Erstellung komplexer Automatisierungsalgorithmen zur Verabreichung von Insulin bis hin zur Entwicklung chirurgischer Methoden und Implantate (Patient Innovation 2021). Diese Innovationen bieten das Potenzial, wesentlich zu einer Vermehrung von behandlungsrelevantem Wissen im Gesundheitswesen beizutragen (Canhao et al. 2016). Patient Innovation steht in der Tradition der Do-It-Yourself- und Open-Source-Bewegung (z. B. OpenAPS 2021) sowie der sozialen Gesundheitsbewegungen, wobei letztere ihren Ursprung bereits in den 1970er-Jahren haben. Sie zeigten schon lange vor der Etablierung (des Begriffs) der Bürgerwissenschaften, dass medizinische Forschung durch Bürger*innen bzw. Patient*innen auch unabhängig von der akademischen Wissenschaft geschehen kann (Mahr und Strasser 2021; Rabeharisoa und Callon 2002).

Diskussion und Fazit

Wie viel Partizipation in den verschiedenen Ansätzen tatsächlich angestrebt wird, differiert stark. Während Versorgungsforschung bereits die Patient*innenperspektive berücksichtigt, wird jedoch noch primär Forschung *an*, nicht Forschung *mit* Patient*innen durchgeführt. Im Sinne einer patient*innenzentrierten Forschung werden Fragen und Ergebnisse (*outcomes*) evaluiert, die Patient*innen selbst als sinnvoll und wichtig erachten, sowie das Empowerment von Patient*innen thematisiert. Auch die Möglichkeiten der partizipativen Entscheidungsfindung werden untersucht, in der Regel erfolgt jedoch kein Einbezug in die Forschung. Allerdings finden sich in jüngerer Zeit auch Ansätze, die stärker partizipativ angelegt sind (Farin 2017), wie etwa das Projekt GestDiNa_basic (Greiner et al. 2021).

Um dem Anspruch von CS zu genügen, ist in mindestens einer Phase des Forschungsgeschehens eine wissensgenerierende Beteiligung von Patient*innen oder Betroffenen notwendig. Dieses Kriterium erfüllen die hier genannten bürgerwissenschaftlichen Ansätze. CBPR und Patient Engagement/Patient Involvement gehen dabei von einer partnerschaftlichen Zusammenarbeit zwischen Bürgerwissenschaftler*innen und akademisch Forschenden aus. Die Zielgruppen sind in beiden Ansätzen unterschiedlich: Während CBPR soziale Gemeinschaften beteiligt, sind bei Patient Engagement und Involvement vor allem Patient*innen angesprochen. Offen bleibt, wie intensiv die Zusammenarbeit bei der Wissensgenerierung ist. PGF steht in der Traditionslinie von Sozial- und Gesundheitsforschung (*public health*), versteht Forschung als Ko-Produktion verschiedener Akteur*innen, ist jedoch in der medizinischen Forschung weniger implementiert.

Im Patient-Science-Ansatz werden die Bürgerforschenden stärker in den Forschungsprozess eingebunden und können die

Forschungsinhalte mitbestimmen. Dies ist vor allem bei chronischen Erkrankungen gut einsetzbar, wenn Patient*innen und deren Umfeld eigene Erfahrungen und großes Vorwissen mitbringen. Patient Innovation hingegen umfasst selbstständig von Patient*innen und Betroffenen kreierte Entwicklungen, die (auch) ohne Einfluss und ohne Beteiligung der akademischen Forschung entstehen. Die Entwickler*innen sind in der Forschung und Wissensgenerierung vollkommen autonom. Diese Innovationen gehen den umgekehrten Weg: Erst nach der Entwicklung durch Bürgerforschende wird (manchmal) von akademischer Seite geforscht, wie im Projekt TeQfor1 (TeQfor1 2021). Bei den hier beschriebenen bürgerwissenschaftlichen Ansätzen – Patient Involvement/Engagement – CBPR – PGF – Patient Science – Patient Innovation – ist ein hohes Maß an Beteiligung von Patient*innen und Betroffenen in den einzelnen Forschungsphasen gegeben. Sie unterscheiden sich bezüglich Zielgruppe und Wissensgenerierung.

Die hier beschriebenen Begriffe sind in unterschiedlichen historischen und länderspezifischen Entwicklungen in der medizinischen Forschung entstanden. Angloamerikanische Begriffe mit ihrer meist älteren Tradition können auch deshalb nicht eins zu eins ins Deutsche mit einer noch jüngeren Tradition übernommen werden. Auch die unterschiedlichen Gesundheitssysteme der verschiedenen Länder spielen bei der Entwicklung von CS-Ansätzen eine Rolle. Eine Übersetzung der englischen Begriffe ins Deutsche ist somit nicht sinnvoll, da die dahinter liegenden Konzepte berücksichtigt werden müssen.

Ethische Aspekte, wie etwa die durch CS häufig versprochene Demokratisierung der Wissenschaft versus die Verstärkung der Exklusion benachteiligter Gruppen (Fiske et al. 2019), ebenso wie Aspekte der Datenqualität und -validität oder der Wissensgenerierung, haben in der medizinischen und CS-Forschung eine große Bedeutung, können jedoch im Rahmen dieses Artikels nicht behandelt werden. Jede einzelne dieser Fragen hinsichtlich der dargestellten Begriffe bedarf einer eingehenden Analyse und Diskussion. Es braucht interdisziplinäre Zusammenarbeit, um eine gemeinsame Sprache und transdisziplinäre Konzepte für bürgerwissenschaftliche Ansätze in Medizin und Gesundheitsforschung entwickeln zu können. Wir hoffen daher, mit diesem Beitrag eine interdisziplinäre Diskussion anregen zu können, an der sich ganz im Sinne eines CS-Verständnisses auch Bürgerforschende beteiligen.

Angabe von Finanzierungsquellen

Der vorliegende Forschungsartikel hat keine Förderung erhalten.

Literatur

- BIH – Berlin Institute of Health @ Charité (2021): Patient & stakeholder engagement. Online verfügbar unter <https://www.bihealth.org/de/translation/innovationstreiber/quest-center/mission-ansaezte/stakeholder-engagement>, zuletzt geprüft am 14. 10. 2021.
- BMBF – Bundesministerium für Bildung und Forschung (2016): Aktionsplan Versorgungsforschung. Forschung für ein patientenorientiertes Gesundheitswesen des Bundesministeriums für Bildung und Forschung. Berlin: BMBF. Online verfügbar unter https://www.gesundheitsforschung-bmbf.de/files/Aktionsplan_Versorgungsforschung_barrierefrei.PDF, zuletzt geprüft am 28. 05. 2021.
- BMBF (2018): Rahmenprogramm Gesundheitsforschung der Bundesregierung. Berlin: BMBF. Online verfügbar unter https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/pdf/rahmenprogramm-gesundheitsforschung-der-bundesregierung.pdf?__blob=publicationFile&v=2, zuletzt geprüft am 30. 10. 2021.
- Bonn, Aletta et al. (2016): Grünbuch Citizen Science Strategie 2020 für Deutschland. Berlin: GEWISS.
- Buyx, Alena; Del Savio, Lorenzo; Prainsack, Barbara; Völzke, Henry (2017): Every participant is a PI. Citizen Science and participatory governance in population studies. In: *International Journal of Epidemiology* 46 (2), S. 377–384. <https://doi.org/10.1093/ije/dyw204>
- Canhao, Helena; Oliveira, Pedro; Zejinilovic, Leid (2016): Patient innovation. Empowering patients, sharing solutions, improving lives. In: *NEJM Catalyst*. Online verfügbar unter <http://catalyst.nejm.org/patient-innovation-empowering-sharing-improving-lives/>, zuletzt geprüft am 30. 10. 2021.
- Del Savio, Lorenzo; Prainsack, Barbara; Buyx, Alena (2016): Crowdsourcing the human gut. Is crowdsourcing also citizen science? In: *Journal of Science Communication* 15 (03), S. 1–16. <https://doi.org/10.22323/2.15030203>
- European Citizen Science Association (ECSA) (2015): Ten principles of Citizen Science. <http://doi.org/10.17605/OSF.IO/XPR2N>
- Farin, Erik (2017): Partizipative Versorgungsforschung. In: Holger Pfaff, Edmund Neugebauer, Gerd Glaeske und Matthias Schrappe (Hg.): *Lehrbuch Versorgungsforschung. Systematik – Methodik – Anwendung*. Stuttgart: Schattauer, S. 182–186.
- Fiske, Amelia; Prainsack, Barbara; Buyx, Alena (2019): Meeting the needs of underserved populations. Setting the agenda for more inclusive citizen science of medicine. In: *Journal of Medical Ethics* 45 (9), S. 617–622. <https://doi.org/10.1136/medethics-2018-105253>
- FVN – Forschungsverbund „Verlust der Nacht“ (2021): Weißt du wie viel Sternlein stehen? Online verfügbar unter http://www.verlustdernacht.de/Loss_of_the_Night_App/articles/loss-of-the-night-app.html, zuletzt geprüft am 28. 05. 2021.
- Gardecki, Johanna et al. (2021): Patient Science. A new citizen science approach for health and medical research involving people affected by cystic fibrosis. In: *Journal of Cystic Fibrosis* 20 (1), S. S115. [https://doi.org/10.1016/S1569-1993\(21\)01273-X](https://doi.org/10.1016/S1569-1993(21)01273-X)
- George, Anne; Daniel, Mark; Green, Lawrence (2006): Appraising and funding participatory research in health promotion. In: *International Quarterly of Community Health Education* 26 (2), S. 171–187. <https://doi.org/10.2190/R031-N661-H762-7015>
- Greiner, Gregory et al. (2021): Study protocol for a mixed methods exploratory investigation of aftercare services for gestational diabetes in women to develop a new patient-centred model in Germany. The GestDiNa_basic study. In: *BMJ Open* 11 (8), S. e046048. <https://doi.org/10.1136/bmjopen-2020-046048>
- Habicht, Hagen; Oliveira, Pedro; Shcherbatiuk, Viktoriia (2012): User Innovators. When patients set out to help themselves and end up helping many. In: *Die Unternehmung* 66 (3), S. 277–295. <https://doi.org/10.5771/0042-059X-2012-3-277>
- Haklay, Muki (2013): Citizen Science and volunteered geographic information. Overview and typology of participation. In: Daniel Sui, Sarah Elwood und Michael Goodchild (Hg.): *Crowdsourcing geographic knowledge*. Springer: Dordrecht, S. 105–122. https://doi.org/10.1007/978-94-007-4587-2_7

- Hennen, Leonhard; Petermann, Thomas; Scherz, Constanze (2004): Partizipative Verfahren der Technikfolgenabschätzung und parlamentarische Politikberatung. TAB-Arbeitsbericht Nr. 96. Berlin: Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag. Online verfügbar unter <https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Arbeitsbericht-ab096.pdf>, zuletzt geprüft am 30. 10. 2021.
- Heyen, Nils B. et al. (2021): Patient Science als bürgerwissenschaftliches Format. Erkenntnisse aus dem Pilotprojekt und Empfehlungen für zukünftige Projekte. Karlsruhe: Fraunhofer ISI. Online verfügbar unter https://www.buergerschaffenwissen.de/sites/default/files/assets/projekte/user-1719/pdf/Patient%20Science_Empfehlungspapier.pdf, zuletzt geprüft am 30. 10. 2021.
- Hoffmann, Wolfgang et al. (2019): Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP). A guideline developed by the German Society for Epidemiology. In: *European Journal of Epidemiology* 34 (3), S. 301–317. <https://doi.org/10.1007/s10654-019-00500-x>
- ICH – Good clinical practice Guidelines (2021): ICH harmonised guideline integrated addendum to ICH E6(R1). Guideline for Good Clinical Practice ICH E6(R2) ICH Consensus Guideline. Online verfügbar unter <https://ichgcp.net/de>, zuletzt geprüft am 18. 07. 2021.
- Institute of Medicine (2001): Crossing the quality chasm. A new health system for the 21st century. Washington, DC: National Academy Press. <https://doi.org/10.17226/10027>
- Jull, Janet; Giles, Audrey; Graham, Ian (2017): Community-based participatory research and integrated knowledge translation. Advancing the co-creation of knowledge. In: *Implementation Science* 12 (1), S. 150. <https://doi.org/10.1186/s13012-017-0696-3>
- LaVeaux, Deborah; Christopher, Suzanne (2009): Contextualizing CBPR. Key principles of CBPR meet the Indigenous research context. In: *Pimatisiwin* 7 (1), S. 1.
- Légaré, France; Witteman, Holly (2013): Shared decision-making. Examining key elements and barriers to adoption into routine clinical practice. In: *Health affairs* 32 (2), S. 276–284. <https://doi.org/10.1377/hlthaff.2012.1078>
- Lewin, Kurt (1946): Action research and minority problems. In: *Resolving social conflicts. Selected papers on group dynamics*. New York: Harper & Brothers, S. 201–216. <https://doi.org/10.1111/j.1540-4560.1946.tb02295.x>
- Loh, Andreas; Simon, Daniela; Kriston, Levente; Härter, Martin (2007): Patientenbeteiligung bei medizinischen Entscheidungen. Effekte der Partizipativen Entscheidungsfindung aus systematischen Reviews. In: *Deutsches Ärzteblatt* 104 (21), S. A-1483–8.
- Mahr, Dana; Strasser, Bruno (2021): Citizen science and biomedical research. In: *The Lancet Child & Adolescent Health* 5 (10), S. 682–683. [https://doi.org/10.1016/S2352-4642\(21\)00237-6](https://doi.org/10.1016/S2352-4642(21)00237-6)
- National Institute of Health (NIHR) (2021): Supporting patient and public involvement in research. Online verfügbar unter <https://www.nihr.ac.uk/explore-nihr/campaigns/supporting-patient-and-public-involvement-in-research.htm>, zuletzt geprüft am 14. 10. 2021.
- Naturschutzbund Deutschland e. V. (NABU) (2013): Auch in den USA ist im Winter Vogelzählzeit – „Christmas Bird Count“. Eine Mitmachaktion fast so alt wie der NABU. Online verfügbar unter <https://www.nabu.de/tiere-und-pflanzen/aktionen-und-projekte/stunde-der-wintervogel/ueber-die-aktion/13276.html>, zuletzt geprüft am 28. 05. 2021.
- OpenAPS – Open Artificial Pancreas System project (2021). Online verfügbar unter <https://openaps.org/>, zuletzt geprüft am 14. 10. 2021.
- PCORI – Patient-Centered Outcomes Research Institute (2018): The value of engagement. Online verfügbar unter <https://www.pcori.org/engagement/value-engagement>, zuletzt geprüft am 02. 11. 2021.
- Patient Innovation (2021): Patient innovation. Sharing solutions, improving lives. Online verfügbar unter <https://patient-innovation.com/>, zuletzt geprüft am 14. 10. 2021.
- PEI – Paul-Ehrlich-Institut (2021): Meldeformular für Nebenwirkungen unter Berücksichtigung der Nationalen Impfkampagne gegen die COVID-19 Erkrankung. Online verfügbar unter https://nebenwirkungen.bund.de/nw/DE/home/home_node.html, zuletzt geprüft am 02. 11. 2021.
- Rabeharisoa, Vololona; Callon, Michel (2002): The involvement of patients associations in research. In: *International Social Science Journal* 54 (171), S. 57–63. <https://doi.org/10.1111/1468-2451.00359>
- Peter, Sebastian von et al. (2020): Partizipative Gesundheitsforschung in Deutschland – quo vadis? In: *Das Gesundheitswesen* 82 (4), S. 328–332. <https://doi.org/10.1055/a-1076-8078>
- Schrappe, Matthias; Pfaff, Holger (2017): Einführung in Konzepte und Grundlagen der Versorgungsforschung. In: Holger Pfaff, Edmund Neugebauer, Gerd Glaeske und Matthias Schrappe (Hg.): *Lehrbuch Versorgungsforschung. Systematik – Methodik – Anwendung*. Stuttgart: Schattauer, S. 1–67.
- Stiggelbout, Anne; Pieterse, Arwen; de Haes, Johanna (2015): Shared decision making. Concepts, evidence, and practice. In: *Patient Education and Counseling*, 98 (10), S. 1172–1179. <https://doi.org/10.1016/j.pec.2015.06.022>
- TeQfor1 (2021): Auswirkungen technischer Systeme auf die eigene Lebensqualität von Menschen mit Typ-1-Diabetes. In: *Bürger schaffen Wissen. Die Citizen Science Plattform*. Online verfügbar unter <https://www.buergerschaffenwissen.de/index.php/projekt/teqfor1-auswirkungen-technischer-systeme-auf-die-eigene-lebensqualitaet-von-menschen-mit>, zuletzt geprüft am 14. 10. 2021.
- Unger, Hella von (2014): Partizipative Forschung. Einführung in die Forschungspraxis. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-01290-8>
- Vayena, Effy (2021): Value from health data. European opportunity to catalyse progress in digital health. In: *The Lancet* 397, S. 652–653. [https://doi.org/10.1016/S0140-6736\(21\)00203-8](https://doi.org/10.1016/S0140-6736(21)00203-8)
- Viswanathan, Meera et al. (2004): Community-based participatory research. Assessing the evidence. In: *Evidence Report/Technology Assessment* (99), S. 1–8. <https://doi.org/10.1037/e439622005-001>
- Wright, Michael (Hg.) (2010): Partizipative Qualitätsentwicklung in der Gesundheitsförderung und Prävention. Bern: Hans-Huber.



GERTRUD HAMMEL

ist Soziologin und leitet seit 2016 das klinische Studienzentrum am Institut für Umweltmedizin. Arbeitsschwerpunkte sind methodische, biostatistische sowie bürgerwissenschaftliche Ansätze in der klinischen Forschung.

**SILVIA WOLL**

ist seit 2015 Wissenschaftlerin mit Schwerpunkt CS sowie Gesundheit und Technisierung des Lebens am Institut für Technikfolgenabschätzung und Systemanalyse am Karlsruher Institut für Technologie.

**DR. MICHAEL EICHINGER**

leitet die AG Planetary Health am Mannheimer Institut für Public Health, Sozial- und Präventivmedizin der Medizinischen Fakultät Mannheim der Universität Heidelberg.

**MARTINA BAUMANN**

ist seit 2017 Wissenschaftlerin im Bereich Gesundheit und Technisierung des Lebens am ITAS des KIT und arbeitet schwerpunktmäßig zu ethischen und sozialen Aspekten von Bio- und Medizintechnologien.

**JOHANNA GARDECKI**

ist seit 2017 klinische Psychologin und wissenschaftliche Mitarbeiterin am Universitätsklinikum Frankfurt mit den Schwerpunkten CS in Medizin und Gesundheitsforschung und chronische Erkrankungen.

**CONSTANZE SCHERZ**

ist Sozialwissenschaftlerin und forscht u. a. zum Wissenstransfer in unterschiedlichen Anwendungskontexten. Sie ist stellvertretende Institutsleiterin des ITAS/KIT.

**DR. NILS B. HEYEN**

ist Soziologe und seit 2011 Projektleiter am Fraunhofer ISI. Schwerpunkte sind technische und soziale Innovationen, v. a. in den Bereichen Medizin und Gesundheit, sowie TA-Methoden und Citizen Science.

**DR. MARIA MAIA**

ist seit 2016 Wissenschaftlerin am ITAS des KIT. Mit einem Hintergrund in Gesundheitswissenschaften fokussiert sie ihre Forschung auf gesundheitsbezogene Themen.

**PROF. DR. DR. ANDREA ICKS**

leitet das Institut für Versorgungsforschung und Gesundheitsökonomie des Deutschen Diabetes-Zentrums und ist Direktorin des gleichnamigen Institutes an der Heinrich-Heine-Universität Düsseldorf.

**PROF. DR. BIRGIT BEHRISCH**

ist seit 2019 Professorin für Erziehungswissenschaft/Pädagogik der Kindheit. Sie ist eine der Sprecher:innen des „Netzwerk Partizipative Gesundheitsforschung“ (Partnet).

**DR. MONIKA POBIRUCHIN**

forscht im Bereich der Consumer Health Informatics, wie Bürger*innen ihre Gesundheit zunehmend selbst in die Hand nehmen. Seit 2017 ist sie Leiterin der gleichnamigen Arbeitsgruppe der GMDS e. V.

**SANDRA OLIVIA BORGMANN**

ist seit 2020 stellv. Leiterin der Nachwuchsforscherguppe Patientenbedürfnisse und -präferenzen am Deutschen Diabetes-Zentrum. Arbeitsschwerpunkt ist die personenzentrierte Versorgungsforschung.

**DR. SARAH WESCHKE**

ist Psychologin und seit 2019 Referentin für Patient & Stakeholder Engagement am QUEST Center des Berlin Institute of Health at Charité.

In this interview, Karim Eid-Sabbagh and Ulrich Ufer discuss how the case of the public infrastructure crisis in Lebanon highlights the importance of including analytical dimensions of critical political economy and global financial dynamics in technology assessment alongside a technology-society-governance perspective – in particular when focusing on the Global South. The Lebanese crisis has built up through long-term structural problems that include the legacies of colonialism, the country's peripheral position in global capital relations, elite nepotism, sectarian strife, and the state's dependency on international donor funding to build and maintain public infrastructure. These have coincided with short-term disintegration and disaster events over the past two years: mass migration, countrywide anti-government protests in fall 2019, the outbreak of the Covid-19 pandemic in early 2020, the destruction of large parts of the country's capital by the devastating explosion in the port of Beirut in August 2020, and the spiraling devaluation of the Lebanese currency.

TATuP: *You have analyzed and lived through many years of structural and recently accelerating national crisis in Lebanon. Let's enter this discussion with a look at the current situation in Lebanon regarding the disintegration of public infrastructure.*

Karim Eid-Sabbagh: The crisis is of almost unimaginable proportion, and its immense dynamics are frightening. It wreaks havoc on public infrastructure: Roads, telecommunications, electricity, health, education – there is progressive collapse and disintegration everywhere. Greater Beirut, the capital with almost two million inhabitants, is threatened by failure of public water supply. Some Lebanese villages have been literally left dry for some time.

Is this a technology problem, or a government and governance problem?

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>) <https://doi.org/10.14512/tatup.30.3.70>

INTERVIEW

Critical political economy of the public infrastructure crisis in Lebanon

How can critical political economy perspectives contribute to technology assessment for countries in the Global South?

Can one separate them? It is a crisis of both, although I see government more political and distinct from rather managerial notions of “governance.” The current crises have led to a near collapse of public revenues. Fee collection rates of the public water utilities and agencies have been weak in the past and have continued to decline as hyperinflation has destroyed the value of both private and public Lebanese lira revenues and reserves. This means that operation and maintenance of supply infrastructure, e.g., fuel and spare parts, have become far too costly.

How have Lebanese households dealt with water supply shortages or failure?

Over the past decades, alternative forms of water supply have come into existence because of the failure to maintain and expand public water supply. Public networks still supply most of the water for domestic use, but supply is intermittent, i.e., available only for a certain period during the day. For supply during intermittent periods, Lebanese buildings are usually equipped with rooftop or ground level tanks for intermediate storage of water. Many households now rely on drinking water delivered in plastic containers and by bulk truck delivery, or they draw it from springs and wells where such are available. In a sense, this shows the resourcefulness of the Lebanese people to find alternative, socially organized solutions where large technical systems fail. Though, obviously this comes at considerable social and ecological cost. The cost burden is considerably higher for poorer households, and public infrastructure also needs to deal with waste. Together with colleagues, I have documented in a film the interaction between the water and waste crises, which is largely connected to the hyper-consumption of plastic containers for water supply.¹

If we look at other sectors of public infrastructure, experts such as from the United Nations Development Programme consider photovoltaics an answer to Lebanon's power supply challenges. With well over two hundred hours of sunshine on average per month, this looked like a promising solution for distributed power generation in the face of failing centralized networks.

Right, there is a huge potential for innovative distributed energy supply in Lebanon, and one does see growth of photovoltaics at the private level. This makes sense because, much like the water sector, power supply is heavily dependent on non-public supply. Public supplies meet only half of the country's estimated en-

¹ “We made every living thing from water” is available for free online viewing on YouTube and Vimeo in English (<https://youtu.be/reJEh0jM5-8>; <https://vimeo.com/259600292>) and Arabic (<https://youtu.be/wO9yfG6H1Ws>; <https://vimeo.com/285329026>).

ergy needs, and the intermittent nature of service meant power cuts for three to 12 hours per day depending on location; around half the population had to manage power cuts of about 12 hours per day even before the current crisis. But, Lebanon's power generation is nowhere near a transition to photovoltaics; power plants primarily run on fuel oil. Throughout the country, intermittent public supply is compensated for by diesel generators that supply electricity to neighborhoods or individual buildings. Civil war as well as recurrent attacks on public infrastructure by Israel have repeatedly set back development and modernization. And similar to water infrastructure, increasing public debt and nepotism in the public sector have further crippled the effectiveness of Électricité du Liban, the public utility company. In fact, the power supply system is degrading and has never caught up with demand. Instead of necessary modernization, floating oil fuel power plants along the Lebanese coast have supplied additional power over the past ten years, but these 370 megawatts, or about a quarter of the country's total production capacity, have just been disconnected from the grid as the contract with the Turkish company ended with the state owing over 100 million U.S. dollars of unpaid dues.

This must surely lead to frequent blackouts and brownouts.

Brownout has been the ordinary network state for decades, but it has all become worse through the crisis. In mid-October 2021, the country's power grid totally collapsed for days due to a lack of oil fuel in the two main power plants, which reduced national production to below 200 megawatts. Grid rebuilding efforts were complicated because the national control center had been destroyed in the Beirut port explosion of 2020. And here we are back to financial, economic, and governmental aspects of the public infrastructure crisis as well as their social consequences. The collapse of the Lebanese currency and decreasing foreign reserves have put a hard brake on all sorts of imports, but importantly on oil deriv-

atives, which are essential to keep public power stations or private generators running. Accordingly, households, businesses, hospitals etc. have had to ration electricity drastically.

In the highly industrialized countries of the Global North, the functioning of "public infrastructure" is taken for granted, a constantly humming background noise to which hardly anyone

Right now, many have access to electricity for no more than a couple of hours per day, and the poorest lack access for days. Currently, the power situation has improved due to recent oil imports, but when the energy crisis peaked during the summer of 2021, the effects were massive. Even telecommunications suffered local interruptions, e. g., because the antennas of mobile network operators depend on backup generators that run on

Brownout has been the ordinary network state for decades, but it has all become worse through the crisis.

pays attention, while in Lebanon, and in countries of the Global South more generally, its precariousness creates constant alarms and makes public infrastructure awareness an important aspect of the organization of society and everyday life.



Dr. Karim Eid-Sabbagh

is an independent researcher and documentary filmmaker based in Lebanon. His research focuses on political ecology, imperialism and sovereign development, water resource management, and agrarian transformation in the Global South. He holds degrees in civil engineering, urban planning and development studies.

fuel oil when not connected to the public power grid in remote regions. Also, due to power supply failures, even charging a mobile phone became a challenge at times. To illustrate the extent of power shortage in households due to lack of supply or purchasing power: Before the crisis, five amperes was the minimum subscription to a private generator, now this has been reduced to three amperes – just enough to run a fridge and a couple of light bulbs, but certainly not hot water or heating. Work schedules and household chores have necessarily been adjusted to the rhythm of electricity supply. Overall, the rhythms of work, schooling, social networks, or any other domains of everyday life have been interrupted and toppled over.

I think this illustrates very well how cascading effects of a public infrastructure crisis bear on the social lifeworld. What about their impacts on other central public services, such as mobility, health, or education?

Mobility has been dramatically affected by hyperinflation and collapse of the Lebanese lira. Twenty liters of fuel currently cost about half the monthly minimum salary, so even bus rides and shared taxi rides become unaffordable to many. In terms of public infrastructure, there is

a slow but noticeable decay of roads and road safety due to years of lack of maintenance, and with the state in default now even more so. Tunnels without lighting are eerie and dangerous places. Increasing numbers of road casualties seem to confirm this.

The health sector has been devastated through shortages of medicine, lack of power, or the exodus of trained and qualified hospital staff in times of pandemic. To give but one example, Beirut's public

global levels of analysis. They can help focus on how local individual experiences of crisis and local failures of public infrastructure relate to historical structures in international relations and to larger dynamics in the global capitalist system.

This perspective is important. The peripheral position of a country like Lebanon in unequal global capital relations and political relations of domination is also an effect of historical structures. Pu-

sis. Lebanon has a sectarian system that works through nepotism and clientelism, wheeling and dealing, without any accountability or oversight. But these are inherited structures, Lebanon has always been a place of many sects, but it was only in the 1860s that these aspects became increasingly instrumentalized through collaboration between local elites and the increasingly influential French regime, which concentrated capital and economic capacity among Ortho-

Public infrastructure failure and crisis in Lebanon must be understood not as a purely technical phenomenon of the present but as economically, socially, and historically embedded.

hospital, the main recipient for Covid-19 patients repeatedly had to turn off air conditioning in some sections of the hospital at way above 30 degrees Celsius in order to supply the necessary power to intensive care units. But beyond the provision of health services, mental illnesses such as depression are on the rise as a psychological consequence of crisis.

As for schools, they have barely been open over the last two years due to anti-government protests in 2019 and the ensuing pandemic. Official exams were cancelled because the ministry could not afford exam papers, which were later donated by UNICEF. Given the problems with power supply and telecommunication services, online education remains a challenge for many. In Beirut specifically, the port explosion did considerable damage to public and private schools, and now some schools have delayed reopening because of the fuel crisis.

I think that our understanding of national crises such as the current one in Lebanon is often hampered by historical myopia and by insufficient understanding of crisis contexts. To go beyond reductionisms such as "erosion of governmental power" or "failed states," we should also talk about critical political economy approaches that link local, national, and

blic infrastructure failure and crisis in Lebanon must be understood not as a purely technical phenomenon of the present but as economically, socially, and historically embedded. For Lebanon, as for many other social formations in the Global South, one can ask whether given technologies and emerging technological innovations are in fact appropriate to their political economic reality. I think that the technological and socio-technological models exported from the world's leading industrial countries are often inappropriate as they are predicated on the "de-development" or "development of underdevelopment" of the Global South, i.e., the drain of resources and value flows. They function where enough wealth has already been amassed, i.e., in the global centers of power, and they rely on peace, social stability, and a country's economic and financial capacity and self-determination. Most of these conditions do not apply to countries of the Global South.

Could you please elaborate on the historical relationship between capital, technology, and society in Lebanon?

Where or when does one start ...? I think here I can only give a very short outline of the structural contradictions that are coming to a head in the present cri-

sis. Lebanon has a sectarian system that works through nepotism and clientelism, wheeling and dealing, without any accountability or oversight. But these are inherited structures, Lebanon has always been a place of many sects, but it was only in the 1860s that these aspects became increasingly instrumentalized through collaboration between local elites and the increasingly influential French regime, which concentrated capital and economic capacity among Ortho-

dox and Maronite Christians as they were displacing the Ottoman rulers. The colonial French and English division of the Arab world in 1920 drew lines of conflict that continue to exist and eventually institutionalized political sectarianism. After independence in 1943, the forced displacement of Palestinians to Lebanon after 1948 and into the 1970s, the civil war (1975 to 1990), as well as the continuing state of war with Israel and recurrent military clashes destabilized the dominance of Christian capital – that is, the network of bankers, traders, and politicians that largely controlled the fate of the country. With the Taif Agreement of 1989, a new sectarian balance was instituted, giving Christian, Sunni, and Shia leaders each one of the highest offices in the state. But even the post civil war political settlement did not take effect until U.S. interest in Syrian cooperation during the first Gulf War produced an entente. This sectarian arrangement was predicated on consensus, which in turn was either achieved through a division of the spoils of government and reconstruction funding or resulted in political gridlock. Political gridlock has wreaked havoc on the development of public infrastructure by hampering the functioning of institutions, governmental agreements on loans, public spending, and so on.



A private diesel generator at ground-level for alternative power supply at a street corner of Tyre, southern Lebanon. It runs and emits exhaust fumes for about 18 hours a day. Photo: Karim Eid-Sabbagh

Can you provide some examples of how international capital flows and the political system interlink?

Lebanon has historically been locked into unequal capital flows from and to Western countries, even before the civil war. Investors from Western countries used Lebanon as a gateway to the Middle East – many of the goods, services, and financial capital flowed through Beirut. Local industry or agriculture, while expanding, have therefore never been sectors with much support among elites or international investors. With their focus on trade and financial services, Lebanese elites have basically been milking the state over the past decades, in close cooperation with the international institutions that provided the ideological backbone. Neoliberal governance prescriptions focusing on privatization, private sector profit, and full cost recovery of public infrastructure investment from user fees have been the central mantra of Lebanese development policy, if one can call it that. Central to this are the workings of the Lebanese financial system. Post civil war reconstruction plans promised to remake Lebanon into the Switzerland of the Middle East, with unrealistic expected GDP annual growth rates of 8% over 25 years – promises that appear as ideological ob-

fuscation considering that Israeli forces continued to occupy South Lebanon until 2000 and Israeli air strikes on Lebanese public infrastructure were a recurrent phenomenon. To perform this miracle, banks had to be recapitalized through public debt at outrageous interest rates in the mid-1990s, i. e., elites were lending their money to the state through banks at around 40% interest. And in order to stabilize the currency, the Lebanese lira was progressively pegged to the U.S. dollar at a rate of 1,507 to one. However, this arrangement necessitated the attraction of U.S. dollars to build reserves to maintain the peg. To achieve this, again, the central bank has been offering interest rates on U.S. dollar deposits of well above 10% for the last 25 years.

This looks very much like an elaborate international and large-scale Ponzi scheme. What have been its effects on the present public infrastructure crisis?

Obviously, the productive sector could not compete with profit rates offered in the financial sector. So, Lebanon's lack of maintenance and modernization of public infrastructure is closely related to international capital flows and the ways in which the country has been turned into a vehicle that has long allowed elites and interna-

tional investors to reap high and relatively risk-free profits from their investments in government bonds and other financial products. By pegging the local currency to the U.S. dollar, the local economy became dollarized and, just like in many other countries of the Global South, foreign investment did not benefit local industrial production or agriculture but instead was largely funneled into short-term profit-making ventures, in particular financial products and real estate speculation. Speculative growth, e. g., in urban real estate, served to maintain the illusion of real GDP growth. But any Ponzi scheme will come to an end at some point when inflows can no longer cover the outflows. By 2016–2017, the central bank resorted to ever more desperate tricks to keep the system afloat, arbitrarily giving away money to banks to keep them from going under. Also, there were some desperate attempts to increase state revenues, such as a tax on WhatsApp messages in 2019, which led to public demonstrations that within a day morphed into a general uprising. International institutions such as the World Bank and the International Monetary Foundation had known at least since 2016 that the system was at breaking point, and international financial capital started divesting from Lebanese banks around that time. Deutsche Bank, which owned 20% shares in one of the largest Lebanese banks, offloaded these after having reaped considerable profits. Wall Street banks did the same.

At the end of this long-term Ponzi scheme is the current dramatic devaluation of the Lebanese lira, which has fallen from its accustomed peg at 1,507 LL for one U.S. dollar to just below 25,000 LL in November 2021 and is expected to continue its downward spiral. What was the short-term cause of this collapse?

International donor conferences in 1999, 2001, and 2007 repeatedly stabilized the state by providing funds for public infrastructure – e. g., for wastewater treatment plants – thus allowing the system's beneficiaries to maintain the profit bonanza with little care for infrastructure develop-

ment. Also, the 2008 global financial crisis provided respite by turning Lebanon into a short-term safe haven for panicked international capital, with significant amounts of capital coming in also from the oil producing states of the Arabian Peninsula. But in 2019 it became clear that the Cedar Conference would end without an agreement on further international donor funds, because Western and Gulf Arab donor countries in particular

within the global economy. Remittances in U.S. dollars from Lebanese expats now represent the primary capital inflow to the country and mainly finance household consumption. Many now can barely afford food, electricity from generators is increasingly becoming a luxury, and health problems mostly spell financial disaster.

As mentioned before, spare parts or oil derivatives – all priced in U.S. dollars

major driver of wealth and comfort in the Global North, but also of climate change, which has brought human civilization to the brink of the abyss and which already negatively affects Lebanon, e. g., in the form of water shortages during excessively dry periods. The negative effects of energy stability in the world's industrialized countries are thus socialized on a global scale. Reparations for climate debt will be necessary for sovereign and sustainable social and technological development in the Global South.

What can be learned is that dependency on external energy and capital inflows, whether the latter come from donors or not, makes any country vulnerable to the vagaries of the “free market,” to foreign powers, and to structural exploitation. For countries of the Global South, this means that foreign direct investment should not be considered the only or sufficient path for the development of national production, public infrastructure, or agriculture. Rather, technological development can only be sustainable on a sovereign and people-centered economic basis. A critical political economy perspective that is able to coherently analyze global and local structures of wealth distribution remains crucial to decipher and address the structural imbalances that affect socio-technological development. From this perspective, what is needed is accumulation from below, indigenous development, technologies that are easy to repair and appropriate for local conditions, and distributed public infrastructures that benefit populations in socially just ways and that minimize the risk of cascading collapses that in many ways characterizes centralized networks.

Reparations for climate debt will be necessary for sovereign and sustainable social and technological development in the Global South.

were reluctant to finance a state in which the *de facto* power balance had shifted in favor of a political alliance including Hezbollah and opposed to U.S.-Israeli dominance in the region. Together with the instability brought by the Syrian war and U.S. sanctions on Hezbollah and Syria all this affected investor confidence and put a break on capital inflows so that the ever-growing need for U.S. dollar inflow could simply not be met anymore.

The country's dependence on a global market and on global institutions to provide the means of sustenance – from energy to consumer goods and medication – clearly indicates the complexity of local-global dynamics and of the social, financial, and political factors that drive technological disintegration.

Eighty percent of Lebanon's food needs, virtually its total energy supplies, any medicines, and almost any other consumption goods are met through imports. With the dramatic decline in U.S. dollar inflows, necessary imports have drained the central bank reserves and caused hyperinflation, thus devaluing both Lebanese currency and labor. In other terms, people living in Lebanon have been rendered worthless in an economic sense

on the world market – have become unaffordable for state institutions, and degradation is accelerating. And with no improvement in sight, the destruction this entails will be of gigantic proportions and enduring consequences. Probably the worst aspect is the brain drain. Trained professionals are leaving in droves, particularly in the medical sector, where the Covid-19 pandemic has created considerable demand in Europe, North America, and the rich Arab countries. So, Lebanon's main export today, as it has been for the last 50 years, is labor through emigration, and the remittances this generates are the last lifeline for those who remain behind. In many ways, the Lebanese are exposed to global systemic violence and suffering, and having little alternative, people find that their capacity for suffering can be expanded – this is very much what “adaptation” can mean in the Global South.

What are the lessons to be learned from the Lebanese public infrastructure crisis for other countries of the Global South as well as Europe?

What is taken for granted in highly industrialized countries, e. g., energy stability and energy abundance, has been the ma-

BOOK REVIEW

Update und Kompendium

Technikfolgenabschätzung – Handbuch für Wissenschaft und Praxis

Georg Kamp, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße,
52428 Jülich, DE (g.kamp@fz-juelich.de)

Seit die Technikfolgenabschätzung (TA) in den 1970er-Jahren eine breitere fachliche Aufmerksamkeit erfährt, ist ein kaum noch überschaubarer Bestand an Übersichts- und Diskussionsbänden zu Konzepten, Methoden und Verfahren angewachsen. Dabei wird nicht zuletzt im deutschsprachigen Raum die Debatte intensiv geführt. Die Bibliotheksdatenbanken zählen die Buchbestände zum Stichwort ‚Technikfolgenabschätzung‘ nach Hunderten; nach einer groben Durchsicht halten sich dabei Beiträge zur Praxis und Beiträge zur Methodenreflexion in etwa die Waage. Ein Band von 1978 erkundet, noch Orientierung suchend, „Ziele, methodische und organisatorische Probleme, Anwendungen“ von TA (Paschen et al. 1978), ein bereits dreibändiges Handbuch ‚Technikfolgenabschätzung‘ hält den Diskussionsstand zu Grundlagen, Methoden und praktischer Umsetzung von 1999 fest (Bröchler et al. 1999), ein eben erst erschienener Tagungsband diskutiert „Möglichkeiten und Grenzen der Technikfolgenabschätzung im Kontext gesellschaftlicher Transformationsprozesse“ (Lindner et al. 2021). Das Publikationsaufkommen ist Ausdruck eines sich dynamisch entwickelnden Feldes, das der fortlaufenden Anpassung an eine sich ändernde Bedarfslage und eine sich immer wieder wandelnde Handlungsumgebung, aber auch der ständigen Selbstvergewisserung und der Abgrenzung gegenüber benachbarten Feldern bedarf. Das jetzt von Stefan Bösch, Armin Grunwald, Bettina-Johanna Krings und Christine Rösch herausgegebene Handbuch ‚Technikfolgenabschätzung: Handbuch für Wissenschaft und Praxis‘ will nun den aktuellen Diskussionsstand strukturiert und praxisorientiert aufbereiten, aber auch Perspektiven für die künftige Weiterentwicklung von TA aufzeigen. Die Einleitung der Herausgeber*innen setzt dafür den Ton: Unter der Überschrift ‚Technikfolgenabschätzung – neue Zeiten, neue Aufgaben‘ als eigenständiges Essay

angelegt, bieten die Herausgeber*innen darin einen Überblick über die Rahmenbedingungen, die Selbstverständnisse und die Entwicklungslogiken von TA. Dass sie dabei die Verhältnisse in westlichen Demokratien und insbesondere im deutschsprachigen Raum vor Augen haben, verrät schon das Literaturverzeichnis, vor allem aber die Selbstverständlichkeit, mit der sie das Bemühen um demokratische Selbstaufklärung neben dem Ziel nachhaltiger Technikgestaltung zum Wesensmerkmal von TA erklären. Die definitorische Lässlichkeit, die eine z. B. ökologisch motivierte TA in nicht-demokratischen Industrie- und Schwellenländern ausschließt, macht aber das Handbuch für seinen intendierten Adressaten gerade praxisrelevant: Entsprechend dem in der Einleitung gesetzten Fokus thematisieren die nachfolgenden Beiträge konkret und praxisnah die Handlungsumgebung ihrer deutschsprachigen Leser*innen. In aller Regel wird es diese Handlungsumgebung sein, für die sie Antworten auf ihre methodologischen Fragen suchen und in der sie ggf. TA-Projekte planen, entwickeln und umsetzen müssen. Über 50 Autor*innen, die in weiten Teilen über das deutschsprachige Netzwerk TA (NTA) in fachlichem Austausch miteinander verbunden und durchweg praxiserfahren sind, blicken dafür in 29 Beiträgen zurück und voraus und beschreiben aktuelle Entwicklungen und Kontroversen.

Ein erster Teil („TA im Wandel der Zeit“, S. 43 ff.) setzt die schon in der Einleitung unternommene Standortbestimmung fort und führt über historische Übersichten an die gegenwärtigen Herausforderungen heran, zunächst mit einem Überblick über die wechselnden ‚Paradigmen‘ der TA und einer konzisen Dogmengeschichte, ausgehend vom planungsoptimistischen TA-Verständnis des US-Office of Technology Assessment (OTA) in den 1960er Jahren über eine ‚partizipative‘ Phase ab ca. 1980 bis hin in eine Gegenwart, die der Autor vor allem von einem pragmatischen, undogmatischen Geist geprägt sieht. Die nachfolgenden Beiträge stützen dies mit einem besonderen Blick auf die jüngeren Entwicklungen in repräsentativen Anwendungsfeldern und zeichnen die sich wandelnden Herausforderungen und die sich mit entwickelnden Antwortversuche nach. Bei der Rückschau auf die andauernden Debatten zur Gentechnik, zur Digitalisierung, zur Mobilität und zur Energieversorgung wird dabei deutlich, dass nicht eine und immer dieselbe methodologische Antwort auf alle sich stellenden Fragen gegeben werden kann:

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>) <https://doi.org/10.14512/tatup.30.3.75>
Received: Oct. 25, 2021; revised version accepted: Nov. 07, 2021;
published online: Dec. 20, 2021 (editorial peer review)



Bösch, Stefan; Grunwald, Armin; Krings Bettina-Johanna; Rösch, Christine (Hg.) (2021): **Technikfolgenabschätzung.** Handbuch für Wissenschaft und Praxis. Baden-Baden: Nomos. 498 S., 58 Euro, ISBN 9783848760701

Während partizipative Verfahren aufgrund der sehr unterschiedlichen normativen Orientierungen in lebenswissenschaftlichen Fragen weniger aussichtsreich erscheinen und TA hier vor allem durch kritische Aufbereitung der Argumente den gesellschaftlichen Diskurs vorbereiten kann, sind bei den auch von politischen Agenden getriebenen Entwicklungen im Bereich der Mobilität und der Energieversorgung TA-Projekte ohne partizipative Verfahren wenig erfolgversprechend.

Der zweite von insgesamt fünf Teilen des Handbuchs beleuchtet „Institutionelle Settings von TA“ (S. 133 ff.). Die Beiträge arbeiten die besonderen Anforderungen heraus, die an eine direkt dem Parlament zuarbeitende TA gestellt sind, analysie-

Die angestellten Überlegungen zum Verhältnis von TA und Wissenschaft finden ihre Fortsetzung im dritten Teil, der unterschiedliche „Theoretische Perspektiven in der TA“ (s. 193 ff.) ausleuchten will. Ein erster Beitrag ‚TA und unterschiedliche Formen des Wissens‘ betont erneut den Anspruch der Wissenschaftlichkeit von TA und versteht dabei wohl – anders als der vorherige Beitrag – TA durchaus auch als wissenschaftliche Disziplin, die sich von anderen aber wesentlich unterscheidet. Dies nicht zuletzt dadurch, dass sie als praxisbezogene, auf gesellschaftliche Problemlagen reagierende Wissenschaft notwendigerweise interdisziplinär ausgerichtet ist und dabei als Technikfolgen beurteilende Wissenschaft immer auch „im Horizont

*Die kollektivistische Ausdrucksweise, welche ‚die TA‘ wie einen Akteur auftreten lässt, verstellt den Blick darauf, dass die Akteure hinter der TA Wissenschaftler*innen sind.*

ren die Rolle der TA in öffentlichen Technikdebatten und stellen die an TA gestellten Herausforderungen dar, wenn TA im Rahmen öffentlich geförderter Projektarbeit, der Auftrags- und Ressortforschung stattfindet. Ein weiterer Blick gilt dem noch wenig erschlossenen Tätigkeitsfeld der TA in und für Unternehmen. Der Beitrag „Technikfolgenabschätzung und Wissenschaft“ (S. 178 ff.) setzt dann den Fokus nicht so sehr auf das Setting praktischer TA-Arbeit, sondern die ‚Interaktionskonstellationen‘ zwischen TA und Wissenschaft. Nach verbreiteter Auffassung sieht die TA eine ihrer zentralen Aufgaben darin, eine nach gesellschaftlichen Maßstäben langfristig verantwortbare Innovationsgestaltung zu unterstützen. In dieser Perspektivierung scheint daher mindestens seit der in den 1980er Jahren in einiger Breite vollzogenen ‚partizipativen Wende‘ eine kritische Distanz zur Wissenschaft als Innovationstreiber erforderlich. Zugleich tritt TA aber ihrerseits mit dem Anspruch von Wissenschaftlichkeit auf, unterstellt sich den wissenschaftlichen Anforderungen und Kriterien und leitet – nicht allein, aber auch nicht zuletzt – daraus ihre Geltungsansprüche ab. TA ist im Verständnis des Autors daher eine „heterogene, hybride Praxis“ (S. 178), die außerhalb des Wissenschaftssystems angesiedelt, aber über Schnittstellen mit diesem verbunden ist und ihren Anspruch an wissenschaftlicher Fundiertheit immer wieder neu mit ihrem gesellschaftlichen Auftrag ausbalancieren muss.

Die kollektivistische Ausdrucksweise, die in diesem Fazit ‚die TA‘ wie einen Akteur auftreten lässt, verstellt dabei allerdings den Blick darauf, dass die Akteure hinter der TA in der Regel Wissenschaftler*innen sind, die mit einem Bein in ihren Disziplinen, mit dem anderen in interdisziplinären und transdisziplinären Beratungskontexten stehen. Hier wären gerade für Neulinge im Forschungsfeld wie für den institutionellen Gestalter von TA eine fruchtbare Weiterführung des sonst aufschlussreichen und überzeugenden Beitrags möglich gewesen.

des Normativen“ (S. 195) agiert. Neben den wissenschaftlichen Rechtfertigungs-Maßstäben muss TA ihren Erfolg auch am Maßstab gesellschaftlicher Akzeptanz und politischer Durchsetzbarkeit messen. In diesem Spannungsfeld sieht der Autor die TA anders – oder doch zumindest stärker – als andere wissenschaftliche Disziplinen vor allem immer auch vor die Aufgabe gestellt, ihre eigene Meta-Reflexion zu betreiben. Beispiele davon bieten die weiteren Beiträge zum dritten Teil des Handbuchs, die über ‚TA als konstruktives Element im Innovationsprozess‘, über ‚TA als Zeitdiagnose‘ oder die ‚Normativität in der TA‘ theoretisieren.

Der vierte Teil („TA als Analyse- und Bewertungsprozess“, S. 269 ff.) nimmt vor allem Instrumente und Methoden in den Blick. Dass hier keine systematische Vollständigkeit zu erwarten ist, liegt auf der Hand. Der Inhalt eines Werkzeugkastens bestimmt sich schließlich nicht aus einer inneren Logik, sondern nach den Mannigfaltigkeiten der handwerklichen Aufgaben. Tatsächlich zeigt der hier gewährte Blick in den Werkzeugkasten der TA durchaus eine beachtliche Breite unterschiedlichster Instrumente: Von der Systemanalyse über die multikriterielle Analyse, das Vision Assessment bis hin zum Reallabor – um nur einige zu nennen. Wer insbesondere im Bereich der Lebenswissenschaften arbeitet, wird vielleicht dennoch ein paar Ausführungen zur ethischen Reflexion vermissen; wer die Politik in Fragen der Energiewende berät, hätte sich vielleicht eine Erwähnung simulatorischer Verfahren gewünscht und Big Data findet sich im Handbuch lediglich als Gegenstand von TA erwähnt, nicht als Werkzeug, dessen Chancen und Risiken es noch auszu-leuchten gilt. Die am Ideal systematischer Vollständigkeit maßnehmenden Erwartungen sind dann auch eher Indiz für die Vielfalt und die innere Komplexität der Aufgaben, derer sich TA inzwischen angenommen hat. Wer in zunehmend vielen Bereichen Technikfolgen antizipieren und Entwicklungsverläufe prognos-

tizieren will, wer darauf praktisch relevante Beratungsleistungen stützen und diese an den vielfältigen und diversen Bedarfen oder Wünschen gegenwärtiger oder künftiger Adressatenkollektive ausrichten will, der wird, was immer hilfreich scheint, erproben und in seine Instrumentenwahl mit einbeziehen.

Naturgemäß lassen sich auch die „Perspektiven einer künftigen TA“ (S. 391 ff.) im fünften Teil nur punktuell ausleuchten. Dazu wird man entweder Ziele definieren, auf die hin eine weitere Entwicklung der TA ausgerichtet werden sollte, oder man wird Herausforderungen benennen, auf die es eine Antwort zu finden gilt. Solche Herausforderungen werden etwa in dem Alternativangebot der „Responsible Research & Innovation“ (RRI, S. 403 ff.), in der Postwachstumsökonomie oder der Wissenschaftskommunikation gesehen. Hier setzen die Autor*innen vor allem auf Kooperation und sehen viel Raum für fruchtbaren Austausch. Weiter in die Zukunft schauen die Beiträge, die die Fortentwicklung der TA vom ambitionierten Ziel ihrer Globalisierung erwägen. Dass TA als „Korrektiv gegen technologische Entwicklungen, die ein inhumanes und antidemokratisches Potenzial in sich bergen“ (S. 391) verstanden und entsprechend fortentwickelt werden kann, ist dabei noch ein eher naheliegender Gedanke. Ebenfalls gut nachvollziehbar ist die Empfehlung, dass sich TA stärker an die Anforderungen einer internationalen inter-institutionellen Zusammenarbeit herantasten sollte – etwa zur Unterstützung einer nachhaltigen internationalen Ernährungspolitik. Die dafür bestehenden Herausforderungen präpariert einer der Beiträge an konkreten Beispielen überzeugend heraus. Weite, aber doch nicht mehr in weiter Zukunft liegende Ziele steckt sich der letzte Beitrag, der die Motivation, die Herausforderungen und die Wege zu einer globalen TA umreißt. Mit dem 2019 gegründeten ‚globalTA Network‘, dem 29 Mitglieder von allen Kontinenten beigetreten sind, ist ein erster Schritt dahin schon getan.

Trotz oder wegen der Fokussierung des Bandes und obwohl oder gerade weil die Autor*innen in engen Diskussionszusammenhängen stehen, bietet das Handbuch keine homogene und entsprechend sterile Bestandsaufnahme. Gerade in den Zielformulierungen und methodologischen Selbstbestimmungen bleiben Kontroversen stehen. Dies ist im Plural der Überschrift von Teil 3 „Theoretische Perspektiven der TA“ (S. 193 ff.) auch bereits explizit angelegt. Um noch einmal auf den ersten Beitrag des Bandes zurückzukommen: Die dort gebotene analytische Darstellung der „Paradigmen der Technikfolgenabschätzung im Wandel der Zeit“ (S. 43 ff.) mündet in der Beschreibung einer Gegenwart, die der Autor durch „eine große Vielfalt unterschiedlicher Spielformen“ von TA bestimmt sieht (S. 56). Um auch diese Phase aufgehobener oder – je nach Perspektive – pluraler Dogmen unter einen Paradigmentitel subsumieren zu können und den „neuen Geist der TA auf den Punkt zu bringen“ (S. 55), fasst er sie unter dem Begriff ‚pragmatische TA‘ zusammen. Das Paradigma grenzt er dabei nicht nur extensional ab, sondern macht bei aller Vielfalt der Akteure durchaus spezifische Gemeinsamkeiten aus: Das gemeinsame Ziel einer verantwortlichen Innovationsgestaltung, die Offenheit für unter-

schiedliche Formen und Quellen des Wissens, und ein Technikverständnis, das Technik vornehmlich als Problemlösung und nicht ausschließlich als Gefahrenquelle sieht. Nicht zuletzt aber auch ein instrumentalistisches Wissenschaftsverständnis, wonach sie weder unumstößliche Autorität für sich beansprucht noch sich auf eine Stimme im Chor der Meinungen reduzieren lässt, sondern nützliches Werkzeug für die Bewältigung der anstehenden Aufgaben ist. Auch wenn mancher die Frage, inwieweit die Spannbreite der Beiträge des Bandes als repräsentativ für die Diskussion um Technikfolgenabschätzung in ihrer gesamten Breite gelten kann, kontrovers diskutieren mag: Die Beiträge bestätigen überzeugend diesen Befund. Ob hingegen der Anspruch der Herausgeber, mit dem Handbuch eine „Anleitung zur gezielten Orientierung bei der Produktion von TA-Wissen“ und eine „spezifische Methodologie problemorientierter Forschung“ (S. 15) bereitgestellt zu haben, eingelöst werden konnte, ist fraglich – zum Glück! TA wird auch weiterhin in einem sich wandelnden Raum agieren und auf unerwartete Herausforderungen weiterführende Antworten finden müssen. Methoden und Instrumente müssen sich in der Praxis bewähren und ihre Fruchtbarkeit beweisen, neue Optionen sind zu prüfen und zu erproben, durchgeführte Projekte kritisch zu hinterfragen. Dafür braucht es eine Vielfalt institutioneller Settings und methodischer Orientierungen. Vor allem aber ist auch weiterhin eine offene, plurale und kontroverse, aber immer lebhaft diskussion um Standpunkte und Perspektiven unverzichtbar. Dazu hat das Handbuch einen bedeutenden und wirksamen Beitrag geleistet. Die durchweg gut lesbaren und angenehm proportionierten Beiträge verschaffen dabei ihren hoffentlich zahlreichen Leser*innen auch bei geringer Vorinformation schnell einen (oft diskurs-historisch geleiteten) Überblick über das jeweilige Themenfeld und dessen angrenzenden Bereiche.

Literatur

- Bröchler, Stephan; Simonis, Georg; Sundermann, Karsten (Hg.) (1999): Handbuch Technikfolgenabschätzung. 3 Bände. Berlin: Edition Sigma.
- Lindner, Ralf; Decker, Michael; Ehrensperger, Elisabeth; Heyen, Nils; Lingner, Stephan; Scherz, Constanze; Sotoudeh, Mahshid (Hg.) (2021): Gesellschaftliche Transformationen. Gegenstand oder Aufgabe der Technikfolgenabschätzung? Baden-Baden: Nomos.
- Paschen, Herbert; Gresser, Klaus; Conrad, Felix (1978): Technology Assessment. Technologiefolgenabschätzung. Ziele, methodische und organisatorische Probleme, Anwendungen. Frankfurt am Main: Campus-Verlag.

BOOK REVIEW

Value sensitive design

A pragmatic approach to ethically guided technology design

Pascal Vetter, *Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT), Karlstr. 11, 76133 Karlsruhe, DE*
(pascal.vetter@kit.edu) 0000-0002-1588-0130

78

When designing technology, ethical issues should also be considered. This demand is discussed repeatedly in politics, practice and science and is therefore frequently subject in works of technology assessment (TA). Eventually, the question arises as to how ethically reflected technology design can be implemented, meaning how the results of theoretical work can be transferred into practical processes. The value sensitive design approach (VSD) is a proposal to answer this question. In their book “Value Sensitive Design. Shaping Technology with Moral Imagination”, VSD pioneer Batya Friedman and David G. Hendry provide a first-hand, concise overview of VSD by conveying the VSD spirit and reconstructing the nearly thirty-year-old history of the approach’s development.

Goals and hard problems

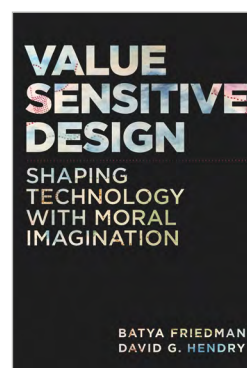
The four overarching goals of VSD are described as “(1) proactive orientation toward influencing design, (2) carrying critical analyses of human values into the design and engineering process, (3) enlarging the scope of human values and (4) broadening and deepening methodological approaches” (pp. 18–19). These ambitious goals do not prevent Friedman and Hendry from being aware of the challenges that prevail in the complex field of ethically guided technology design. Throughout the book, the reader is reminded that critical self-reflection has been and continues to be an important part of VSD’s development. The authors refer to the two biggest of these challenges as “two hard problems” (p. 20). The first hard problem is that by applying VSD, design practitioners may already begin to design technology, even while ethical, legal and social questions may still be unresolved. Thus, the authors do not claim to be able to answer

ethical questions with VSD. Rather they would like to enable design practitioners to achieve progress in the design of technology *despite* these unresolved questions. The second hard problem is gaining robustness for VSD. To meet this challenge, they choose not to commit to a certain context of use, value, or technology in or for which VSD can be applied. Instead, they build on existing methods, adapt or invent new methods on that basis and apply these in practice. From this approach, they hope to identify methodological “blind spots” and advance VSD’s work on theory (p. 21).

Theoretical core elements

VSD’s underlying theory draws mainly on thoughts and discussions that were present in the four scientific communities of applied moral philosophy, social informatics, computer-supported cooperative work, and participatory design in the 1980s and 1990s (pp. 23–24). On the basis of this “intellectual landscape” (p. 24), Friedman and Hendry explain the necessity of defining the term “value” in the context of their approach and reflect on problems and critique that arise by doing so. Unsurprisingly, they provide the pragmatic working definition “what is important to people in their lives, with a focus on ethics and morality” (p. 31) for this purpose. In addition, they also provide working definitions of twelve specific values (e. g. human welfare, privacy and trust) that, in their experience, play an important role in technology design processes. Consistently, the authors explain that values must not be regarded as isolated, but are connected with other values similar to a spider’s web, which can lead to “value tensions” (p. 44). Thus, values can sometimes be in competition or contradiction to each other, which is why a balance must often be found between them.

Another important element of VSD theory is also the fundamental attitude of the approach called “interactional stance”. This refers to the conception of human beings that comes to bear in VSD and understands people in an interacting duality: On the one hand, human beings design tools and technical artefacts, which on the other hand shape human life (p. 35). A very welcome aspect of the interactional stance from the perspective of TA is the awareness of intended and unintended effects in technology design processes. Those responsible for technology de-



Friedman, Batya; Hendry, David (2019):

Value Sensitive Design.

Shaping technology with moral imagination.
Cambridge, MA: The MIT Press.

256 pp.,

40.00 USD,

ISBN 9780262039536

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.78>
Received: Aug. 09, 2021; revised version accepted: Sept. 09, 2021;
published online: Dec. 20, 2021 (editorial peer review)

sign should always remain vigilant towards them (p. 36). The theoretical construct “progress, not perfection” is the trademark element of VSD that most clearly highlights the pragmatic character of the approach. Due to the often complex and ambiguous situation surrounding values and their implications for the technology to be designed, it is often unclear what an ethically correct behaviour of those responsible for technology design would have to look like. The authors therefore argue, with reference to the procedures for evaluating technical systems, that enabling action should also be prioritized in technology design and thereby achieve progress “through practice” (p. 51).

Variety of methods

The authors offer no less than seventeen methods that have been developed under the umbrella term VSD so far. These have names such as “value dams and flows”, “value scenario” or “multi-lifespan co-design” and all of them are explained and categorized according to their intended purpose. Value dams and flows, for example, fall into the category of value analysis and are intended to help resolve value tensions of design alternatives in a project. Value dams are determined first and represent de-

ment (CTA), which is described as an “approach [...] with aims similar to VSD” (p. 128). Readers familiar with CTA will recognize these similarities, but also the differences from VSD throughout the book. What is in sum insightful about these practical examples is to learn which values have been identified as “key values” and how value conflicts have been dealt with in these projects.

Conclusion

With their book, Friedman and Hendry offer a pragmatic proposal for bridging theory and practice of ethically guided technology design. The book functions as an introductory work, explaining the state of the art of the approach and at the same time offering various ways to delve deeper into the subject. VSD methods provide a certain flexibility and adaptability often needed in practice. This can however also lead to some uncertainty as to when which method may be suitable. The authors seem to have had this impression as well, as they provide the reader with a ten-step guide to assist in selecting appropriate methods. Still, reading further literature on the methods (which the book points to) may be necessary. The authors reflect crit-

Value sensitive design enables design practitioners to achieve progress despite unresolved ethical, legal or social questions.

sign alternatives that are rejected by stakeholders, even if those who reject represent only a small percentage of the stakeholders. The stakeholders then evaluate the remaining design alternatives again. Those design alternatives that receive the strongest approval are then called value flows and receive a higher priority in design activities going forward (p. 73). Value scenarios, as another example, are used for value representation and/or value elicitation. These are narratives, created by stakeholders or practitioners to identify “(1) implications for direct and indirect stakeholders, (2) key values, (3) widespread use, (4) indirect impacts, (5) longer-term use, and (6) systemic effects” (p. 64). It is also very important to both authors that VSD does not claim to be a standalone approach, but that VSD methods can be used “in concert” with other existing (VSD) methods (p. 85).

Application examples

The chapter on application examples greatly expands the reader’s understanding of VSD. While reading the theory chapter, I wondered how well VSD can be applied when designing technical artefacts that contain no or only a few IT components. This chapter answers that it can indeed be done. The contributions presented in this chapter span technologies such as pacemakers, humanoid robots, or wastewater treatment plants, and thus tackle topics that also offer interesting insights for TA practitioners. In one example, reference is even made to TA and specifically to the approach of Constructive Technology Assess-

icisms of their approach, but stay true to their core belief – the need for pragmatism. The successes and new insights from the application of VSD described show that VSD can offer viable ways for ethically guided technology design. That said, the book does not answer clearly how well VSD methods can handle tight time and resource schedules that are often encountered in practice. Overall, VSD and TA can learn a lot from each other. VSD can draw on TAs experience and knowledge in dealing with complexity and TA can benefit from VSDs refreshing attitude and creative methods.

MEETING REPORT

Changing Values, Changing Technologies

Pascal Vetter, *Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT), Karlstr. 11, 76133 Karlsruhe, DE*
(pascal.vetter@kit.edu)  0000-0002-1588-0130

80

The conference Changing Values, Changing Technologies took place on 12 and 13 October 2021 at the Delft University of Technology. Around 40 scholars from the fields of philosophy, social science, law, and anthropology came together to listen to and discuss presentations addressing the interactions of values and technologies. The majority of presentations highlighted different facets of the interrelations between value change and technological change, put value change into context with technical and moral progress, or discussed different methods for studying and anticipating value change.

Keynotes

Tsjalling Swierstra started with a presentation on technologically induced moral change (TiMC). He proposed to approach and study moral change as a force field of changing relationships, which he memorably compared to the functional interrelationships of keys and levers on a DJ's mixing console. In her keynote, Helen Nissenbaum discussed whether there is currently a change in values regarding privacy. What I found particularly remarkable were her experiences with the *privacy paradox*, i. e., the phenomenon that people often make contradictory statements on the importance they (allegedly) attribute to privacy and their actual behavior. The conference concluded with Webb Keane's keynote. He examined social media, robots, and algorithms and showed how ethical questions about life with technical devices expand not only the boundaries of ethics but also the boundaries of being human.

Interrelations between value change and technological change

The dynamics and interrelations between value change and technological change were a prominent topic of the conference, dis-

cussed with reference to various modern technologies and related research questions. Lonneke Poort and Sanne Taekema looked at the relationship between ethics, technology, and law. As a case study, they used the EU regulatory framework on genetically modified crops, the regulatory suitability of which is currently being questioned by some EU member states in light of the developments around CRISPR-Cas9. Based on this case study, Poort and Taekema admitted that the law can hardly keep up with the pace of technical developments. However, they argued that law should not be seen as rigid and inflexible as it is sometimes perceived. Rather, in their view, law can enable communication about moral and social values and technology. In their presentation, they also addressed the double role of legal regulation of modern technologies, which can take the form of a facilitator of innovation or a gatekeeper protecting the moral limits of technology – a complex (because occasionally contradictory) task, as is also well known in technology assessment (TA).

Joshua Schulz presented his research on the use of AI systems in medicine. He showed that due to the increased use of AI systems, more and more digital models are applied in medical practice, which encourages medicine to focus on disease prediction, risk management, and behavior optimization. From this development, he derived that Western clinical ethics and medical practice will face a paradigm crisis in the future. He justified this assessment by arguing that the shift to lifelong medical management will shift the identification of appropriate life goals to medicine. In Western societies, however, the identification of appropriate life goals usually falls into the domain of ethics, religion, or society itself, which is why medicine may not be suitable for the task. Schulz also pointed out that predictive medicine may perceive patients increasingly as 'bundles of risk' – a statement that has a special flavor considering the COVID-19 pandemic. Overall, his research represents another example of direct and indirect consequences of technologies and may foreshadow upcoming fields of work for TA.

Katharina Bauer and Julia Hermann spoke about moral education in the light of techno-moral change. Using human-robot interaction in elderly care, they showed that coming into contact with new technologies can require individuals to refine their existing moral skills and sensibilities, even to the point of developing completely new moral skills. With reference to robots in elderly care, they illustrated that the use of robots may not only change work processes but also the self-image of the nursing staff. From the standpoint of moral education, Bauer and Hermann's main thesis is that learning 'moral resilience' should be fostered. To support this position, they explore mechanisms of moral learning and shed some light on psychological aspects of those mechanisms. The goal of their research is to learn how people can best be trained to reinterpret existing moral norms and values in the light of new technological developments.

Value change, technical progress, and moral progress

Five presentations dealt with the relationship between value change, technical progress, and moral progress. I attended the

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.80>
Received: Oct. 06, 2021; revised version accepted: Oct. 07, 2021;
published online: Dec. 20, 2021 (editorial peer review)

talk by Matthew Dennis and Steven Umbrello, who addressed the COVID-19 pandemic and the associated value conflicts and value disruptions. One of their research objects were the so-called ‘immunity passports’ that are currently being discussed in several countries. They pointed out that science is often critical of these immunity passports and criticized that governments often downplay these scientific concerns. This in turn results in individuals often not having a choice to use such immunity tools, as most of them are mandatory. Regarding the role of technology in the pandemic, they pointed out that many people were and still are forced to rethink how they work, study, shop, and entertain themselves, sometimes even in radically new ways. To accomplish this, digital technologies have often been promoted

Fabio Tollon’s presentation focused on the relationship between affordances and artifacts in particular. He put forward the following thesis: When a value changes, the intended design of a technical artifact also changes. Building on Michael Klenk’s affordance account, he aimed to show that a more holistic approach and interpretation of affordances can help explain why certain actions might be made more likely than others. For this purpose, it is not sufficient to merely consider the designed properties of artifacts but that the psychology of those who will use these artifacts must also be included: human beings. Interesting parallels of his work to TA are the questions of how to deal with the unintended use of technical artifacts and how to account for value change in the design and development of technology.

AI leads predictive medicine to perceive patients increasingly as ‘bundles of risk.’

to compensate for direct and indirect limitations in the social sphere. They argued that such ‘technofixes,’ however, often lack important components compared to a sound, socially negotiated solution. Overall, they see the danger of a “digitally divided world” on the horizon, which will particularly affect people who do not have a stable and secure home environment. This is another link to TA, since it sheds light on the people who are among the losers of technologically induced change.

Methods for studying and anticipating value change

Another subject of the conference was methods that can be used to study and anticipate value change. From the point of view of empirical philosophy, Marianne Boenink and Olya Kudina showed that the four methods – living labs, socio-technical systems modeling, techno-moral scenarios, and participatory design – could be useful not only for studying value change and the interaction of values and technology but also for integrating them into the design and development of technology. The speakers’ opinion is that identifying and facilitating value change is possible through sensitivity to dynamics and enabling anticipation, but admitted that the presented methods also have the weak point of being normatively inconclusive. How these methods can be applied in practice has been shown through a case study of the Dutch ‘CoronaMelder’ app. In the development of this app, ethicists were involved in the design process, where they helped identify and prioritize the values of privacy and solidarity as most important. Boenink and Kudina concluded that a pragmatic and constructivist approach to values is necessary, but that this view always comes with the risk of falling into the trap of contemporary value bias. Empirical philosophy and TA can learn a lot from each other here, since both fields work with these methods and existing experiences and best practices promise fruitful discussions.

Tristan de Wildt and Ibo van de Poel presented a pragmatic view on values and value change with their agent-based model. Their overall object of investigation is value change and how it can result from the interaction between technology and society. Their model simulated negotiation processes that are set in motion when a society is confronted with issues or conflicts that arise in relation to the use of a technology. Within the model, they were able to modify the four variables of needs, technologies, values, and moral problems and demonstrated how different characteristics and combinations of these variables affected the result of a simulation. In addition, they showed how they were able to simulate different types of society in terms of adoption of values and preferences for innovation. As with most simulations, the question remains of what insights can be derived from such simulations for practice. Nevertheless, their presentation was a refreshingly practical and illustrative stimulus for the conference.

Further information

The conference was part of the ERC funded research project ‘Design for Value Change.’ For more information, see <https://www.valuechange.eu/>.

MEETING REPORT

International Transdisciplinarity Conference 2021

Creating spaces and cultivating mindsets
for learning and experimentation

Christina Benighaus, *Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) & KIT-Zentrum Mensch und Technik (MuT),
Karlsruher Institut für Technologie (KIT), Karlstr. 11, 76133 Karlsruhe, DE
(benighaus@kit.edu) ☎ 0000-0002-5537-5593*
Richard Beecroft, *KIT-Zentrum Mensch und Technik (MuT),
Karlsruher Institut für Technologie (KIT), 76131 Karlsruhe, DE
(richard.beecroft@kit.edu) ☎ 0000-0002-3170-5201*

Die International Transdisciplinarity Online-Conference 2021 (ITD 21), organisiert vom Network for Transdisciplinary Research und dem Transdisciplinarity Lab der ETH Zürich, kam mit über 500 Teilnehmenden wuchtig daher. Entsprechend hoch waren unsere Erwartungen an die mehr als 200 Real-time-Sessions und Posterbeiträge. Auch inhaltlich hatte sich die Tagung den hohen Anspruch gesetzt, eine Plattform zu bieten „für Engagement und Diskussionen, die transdisziplinäre Forschung, (Un-)Lernen und Praxis miteinander verbindet“.

Das Thema war herausfordernd für eine Tagung, die ausschließlich online stattfinden konnte: ‚Creating spaces and cultivating mindsets for learning and experimentation‘ (Räume schaffen und Denkweisen für Lernen und Experimentieren kultivieren). Wie wird eine Online-Tagung dem gerecht? Mit knappen, vorbereitenden Inputs schöpfte die Konferenz die digitalen Möglichkeiten für Texte, Bilder, Filme oder Narrative voll aus. Ebenso lohnte sich der erste Tag zur Vernetzung der transdisziplinären Nachwuchsforscher:innen, für die mehrtägige internationale Reisen auch ohne Pandemie oft kaum organisierbar sind. Das richtige ‚Tagungsgefühl‘ wollte aber trotz aller inspirierenden Elemente im Programm nicht so recht aufkommen. Vielleicht war das Angebot zu breit gestreut oder können Online-Veranstaltungen dieses Tagungsgefühl einfach nicht bieten?

Was die Online-Tagung ohne Reisekosten wiederum ermöglichte: Die Community ist international geworden. Dominierten in den Anfangszeiten des Netzwerks noch der deutschsprachige und nordamerikanische Diskurs, sind inzwischen Anzahl und kulturelle Hintergründe der Teilnehmenden stark gewachsen. Auch die außergewöhnlich sozial gestaffelten Konferenzgebühren ermöglichten die Teilnahme für viele, die wegen zu hoher Kosten außen vor geblieben wären. Insbesondere für die Diskussion von Nachhaltigkeitsinitiativen lohnte sich dieser internationale und interkulturelle Blick. Die Themen der Tagung waren so vielfältig wie die internationale Community der transdisziplinären (TD) Forschung. In vielen Sessions wurde vor allem um eine allgemeingültige Definition der TD-Arbeit gerungen, die kontextuellen Unterschiede und Gemeinsamkeiten der inter- und transdisziplinären Forschung präsentiert. Es scheint weiterhin ein langer, herausfordernder Weg zu sein, das Fachwissen aus den verschiedensten Wissensgebieten durch TD-Konzepte zu integrieren und einen Beitrag zur Koproduktion von Wissen zu leisten.

Die Vielfalt der präsentierten Fallstudien und Beispiele sowie das Engagement der Community waren begeisternd und mitreißend, allerdings blieben viele Diskussionen sehr nah an der Praxis. Theoriearbeit, also was denn dieses am Lernen orientierte Selbstverständnis der transdisziplinären Community bedeutet, blieb eher im Hintergrund. Ebenso hätten Bezüge auf einschlägige Theoriegebäude, die sich mit Lernprozessen beschäftigen, wie zum Beispiel Psychologie, Bildungstheorie oder Innovationstheorien, stärker hervorgehoben werden können. Mit der Öffnung hin zu Bildungsthemen werden auch Anschlüsse an Diskurse zu inneren Dimensionen der Nachhaltigkeit – z. B. Werte- und Bewusstseinswandel – einfacher, die aus einem klassischen Wissenschaftsverständnis heraus schwer zugänglich

*Wie können wir Wissen als valide
qualifizieren, das in trans-
disziplinären Settings erzeugt wird?*

sind. Aber es ging auch um die wissenschaftliche Glaubwürdigkeit integrierten Wissens, das in einer solchen TD-Forschung produziert wird. Zentrale Fragen lauteten beispielsweise: „Wie stellen wir die wissenschaftliche Nachvollziehbarkeit von Methoden zur Wissensintegration und der Ergebnisse solcher Methoden sicher?“ Oder, „Wie können wir Wissen als valide qualifizieren, das in transdisziplinären Settings erzeugt wird?“

Und eine gute Nachricht: Bildung ist im Diskurs zur Transdisziplinarität angekommen! Dies nicht mehr nur in dem Sinne, dass Forschungsprozesse durch Lehrinitiativen flankiert werden, sondern dass die transdisziplinäre Forschung auch grundlegend als Lernprozess zwischen unterschiedlichen Akteur:innen, Perspektiven, Systemen oder Institutionen verstanden wird.

LETTER

Gegendarstellung

Uwe Schimanks Fehldeutungen
der Echtzeitgesellschaft

Johannes Weyer, TU Dortmund, Fakultät Sozialwissenschaften,
Fachgebiet Techniksoziologie, Otto-Hahn-Str. 4, 44227 Dortmund, DE
(johannes.eyer@tu-dortmund.de)  0000-0002-0181-8723

83

Dabei betonte das Tagungsthema nicht, wie es sonst leicht passiert, die Lernerfolge bzw. deren Messung, sondern die genuin bildungswissenschaftlichen Fragestellungen: also welche Rahmen für Lernprozesse geschaffen werden sollten und wie sich die Geisteshaltung, das *Mindset*, der Beteiligten auf transdisziplinäre Prozesse einstimmen lässt.

Experimentieren ist zu einem systematischen Ankerpunkt in der transdisziplinären Methodologie geworden. Das ist keineswegs selbstverständlich für einen Diskurs, der in der Anfangszeit gerade durch das Beharren auf das systematische Erzeugen von Wissen erfolgreich wurde und vielfach auf der Ebene von Wissensintegration und Aushandlung, aber nicht primär auf der Ebene von Intervention und Experiment agierte. Mit den Bezügen auf die transformative Forschung und der Entstehung von längerfristigen Kooperationsstrukturen, die praktische Transformationen erleichtern (*living labs*, *transformative innovation labs*, Reallabore und verwandte Ansätze) wurden bei dieser Tagung wesentliche Ankerpunkte gelegt.

Neben diesen neuen Schwerpunkten blieben aber auch eine Reihe Themen sichtbar, von denen wir für die TD-Forschung hoffen können, dass sich in den nächsten Jahren Lösungen abzeichnen:

- Wie evaluiert man die Wirksamkeit und die wissenschaftliche Qualität transdisziplinärer Prozesse?
- Wie gelingt transdisziplinäre Koproduktion von Wissen statt eines immer noch währenden multidisziplinären Dialoges?
- Wie können digitale Methoden und Werkzeuge die TD-Arbeit sinnvoll unterstützen?
- Wie muss sich die Rolle der TD-Forschung um eine nachhaltige Entwicklung in der Gesellschaft in den nächsten Jahren verändern?
- Wie gelingt die Förderung wissenschaftlicher Karrieren in der transdisziplinären Forschung, solange Promotionsordnungen etc. disziplinär organisiert sind?
- Wie gelingt die Finanzierung von Forschungsprojekten zwischen den Domänen großer Forschungsförderer?

Gemeinsam mit der intensiven Diskussion von methodologischen Fragen, die diesem Diskurs eigen ist, entsteht Selbstreflexion. Resümee: Eine Tagung allein kann die TD-Landschaft nicht weiterentwickeln. In den nachfolgenden Konferenzen wäre eine stärkere Orientierung auf die Inhalte und Wirkungen transformativer Forschung gewinnbringend und es wird noch viele Diskurse dazu geben müssen. Was mit der ITD 21 auf jeden Fall gelungen ist, ist weltweit denen eine geistige Heimat anzubieten, die bislang eher als Einzelkämpfer:innen transdisziplinär unterwegs waren und sich jetzt in einer großen Gemeinschaft der transdisziplinären Arbeit wiederfinden.

In Jg. 30, Heft 1 (2021) der Zeitschrift TATuP ist eine Rezension meines Buchs „Die Echtzeitgesellschaft“ erschienen, die voller Fehler und Falschbehauptungen ist. Uwe Schimank behandelt Themen, die in dem Buch nicht vorkommen, etwa „Ambivalenz“, „Transintentionalität“, „Reflexionsentlastung“, „großtechnische Systeme zweiter Ordnung“ und vieles anderes mehr. Andere Themen deutet er in einer Weise um, die den in meinem Buch dargestellten Forschungsergebnissen diametral widersprechen. Zudem versäumt er es, die von mir entwickelten zentralen Gedanken und Argumente wie auch die Ergebnisse der von mir durchgeführten empirischen Studien zu erwähnen beziehungsweise kritisch zu würdigen.

Unerwähntes, Gegenteiliges und Übergangenes

Im Detail: Ausgiebig und an mehreren Stellen seiner Rezension breitet Schimank die Trias „sachlich, sozial, zeitlich“ als analytischen Rahmen einer Befassung mit smarter Technik aus. In meiner „Echtzeitgesellschaft“ findet diese Trias nur ein einziges Mal und eher en passant Erwähnung (S. 17). Mein analytischer Rahmen ist die Unterscheidung der drei Ebenen Mikro, Meso und Makro wie auch das Zusammenspiel dieser drei Ebenen in Mikro-Meso-Makro-Modellen.

Jeweils ein Kapitel des Buchs befasst sich mit der Mensch-Maschine-Interaktion (Mikro), dem Risikomanagement in Organisationen (Meso) sowie der Steuerung komplexer Systeme (Makro). Davon erfährt die Leser:in in Schimanks Rezension nichts. Auch dass ich zu allen drei Themenbereichen Ergebnisse umfangreicher empirischer Studien vorgelegt habe, bleibt unerwähnt.

Ein zweites Argument, das Schimanks Rezension prägt, ist das des Kontrollverlusts als Folge des Umgangs mit smarter Technik. In der Tat spielt dieses Thema in meinem Buch eine wichtige Rolle. Allerdings zeigen die empirischen Studien z. B.

This is an article distributed under the terms of the Creative Commons Attribution License CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.14512/tatup.30.3.83>
Received: Jun. 28, 2021; revised version accepted: Aug. 03, 2021;
published online: Dec. 20, 2021 (editorial peer review)

zur Pilotenarbeit bzw. zum smarten Auto, von denen ich in der „Echtzeitgesellschaft“ berichte, das genaue Gegenteil von dem, was Schimank behauptet. Ich zitiere:

- „Allerdings führt das Zusammenspiel von Mensch und autonomer Technik nicht zwangsläufig zu einem Kontrollverlust aufseiten des menschlichen Bedieners.“ (S. 67)
- „Von einem Gefühl der Ohnmacht und des Kontrollverlusts angesichts zunehmender Digitalisierung kann bislang nicht die Rede sein.“ (S. 159, ähnlich S. 80 – in Abgrenzung von Hartmut Rosa)

Auch bei der Analyse des Risikomanagements in Organisation geht es mir darum, Strategien des Umgangs mit Unsicherheit aufzuzeigen, also darzulegen, wie Organisationen, die Echtzeitsysteme managen, die Kontrolle bewahren können – nicht (bzw. nicht nur), wie sie damit gescheitert sind.

Mal – allerdings in abschätziger Weise, die meinen Versuch diskreditieren soll, die Soziologie mithilfe dieser neuen Methode zu bereichern und dazu beizutragen, bislang ungeklärte Fragen zu beantworten.

Abstrakte Vermutungen vs. empirische Überprüfung

Schimank deutet an mehreren Stellen an, dass er noch weitere Argumente in der Hinterhand hätte, mit denen er meiner Kritik am Zustand der deutschen Soziologie entgegentreten könnte. Diese Debatte muss an einem anderen Ort geführt werden (u. a. Weyer 2020). Nur so viel: Ich halte nicht viel davon, ständig neue Begriffe zu erfinden und immer wieder abstrakte Vermutungen über dies oder jenes zu generieren, wie es große Teile der deutschen Soziologie und vor allem ihre führenden Vertreter:innen mit Leidenschaft tun.

Eine Soziologie, die von den anderen Disziplinen ernst genommen werden will, muss Modelle von Gesellschaft entwi-

Auch in der Soziologie müssen Modelle, Annahmen und vermutete Zusammenhänge empirisch überprüfbar sein.

Modellierung versus Sozialtheorie

In der „Echtzeitgesellschaft“ geht es nicht um die Entwicklung einer neuen Sozialtheorie. Im Zentrum steht vielmehr mein – von Schimank ebenfalls übergangenes – Plädoyer, komplexe Sozialsysteme zu modellieren, um auf diese Weise zu einem Verständnis ihrer Funktionsweise zu gelangen und, darauf basierend, soziale Dynamiken wissenschaftlich fundiert zu erklären. Also im Grunde nichts anderes, als Colemans Boot weiterzuentwickeln. Diesen Ansatz präsentiere ich als ein methodisches Instrument, das die Soziologie in die Lage versetzen könnte, abstrakte, theoriebasierte Vermutungen empirisch zu überprüfen. So, wie es andere, evidenzbasierte Wissenschaften tun.

Was Schimank ebenfalls nicht erwähnt, ist mein Anspruch, mit Hilfe dieses modellbasierten Ansatzes insbesondere steuerungstheoretische Fragen zu bearbeiten, darunter die immer wieder diskutierte Frage nach der Wirksamkeit unterschiedlicher Steuerungsmodi. In der „Echtzeitgesellschaft“ lege ich dar, dass man mit Hilfe von Simulationsexperimenten überprüfen kann, wie die Steuerung komplexer Sozialsysteme funktioniert und welche Governance-Modi vielversprechend sind (z. B. in puncto Verkehrswende).

Dabei identifiziere ich einen neuen Modus der zentralen Steuerung dezentraler Systeme, für den ich den Begriff „Echtzeitsteuerung“ verwende und der für mich ein wesentliches Merkmal der Echtzeitgesellschaft ist. Dieser von mir häufig verwendete Begriff taucht in Schimanks Rezension nicht ein einziges Mal auf, ebenso wie der Begriff „Modell“. Zudem erwähnt er den zentralen Begriff „Simulation“ lediglich ein einziges

Mal, die die vermuteten Zusammenhänge (z. B. von Handeln und Struktur) abbilden und es ermöglichen, die eigenen Annahmen empirisch zu überprüfen. Das und nichts anderes ist mein Anliegen. Und wenn ich dabei Wege aufzeige, wie man das – bislang ungelöste – Aggregationsproblem angehen könnte, dann bin ich offenbar nicht völlig auf dem Irrweg.

Ich zähle Uwe Schimank neben Renate Mayntz, Helmut Willke, Hartmut Esser und anderen zu meinen geistigen Kronzeugen, von denen ich viel gelernt und deren Ideen ich in der „Echtzeitgesellschaft“ verarbeitet habe. Warum Schimank dies vollkommen verdreht und verfälscht darstellt, ist mir unerklärlich.

Literatur

Weyer, Johannes (2019): Die Echtzeitgesellschaft. Wie smarte Technik unser Leben steuert. Frankfurt am Main: Campus.

Weyer, Johannes (2020): Muster ohne Wert. Nassehis folgenloses Rasonieren über die digitale Gesellschaft. In: TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 29 (3), S. 76–77. <https://doi.org/10.14512/tatup.29.3.76>

Information

Die ursprüngliche Rezension durch Uwe Schimank erschien unter dem Titel „Sorge ohne Fatalismus. Ambivalenzen immer tiefer gehender Digitalisierung“ in TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 30 (1), S. 83–84. <https://doi.org/10.14512/tatup.30.1.83>

REPLY

Replik auf J. Weyers Gegendarstellung

Uwe Schimank, SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik,
Universität Bremen, Mary-Somerville-Str. 9, 28359 Bremen, DE
(uwe.schimank@uni-bremen.de)

Ich habe Johannes Weyers Buch „Die Echtzeitgesellschaft“ in Heft 1 (2021) dieser Zeitschrift als ein wichtiges und innovatives Buch besprochen. Der Autor findet indes, dass meine Besprechung „voller Fehler und Falschbehauptungen“ ist. Habe ich also gelobt, was gar nicht lobenswert ist? Ich vermute eher: Ich habe in seinen Augen das Falsche oder zu wenig gelobt.

Das Falsche gelobt: Der Autor schreibt mir als Rezensenten offenbar die Rolle zu, die er als Autor allemal besser kann: in seinen Begrifflichkeiten und seinem Bezugsrahmen eine knappe, aber erschöpfende Inhaltsangabe des Buches zu liefern. Was ich, da ich – die erste Aufgabe eines Rezensenten – fast

nichts zu kritisieren hatte, stattdessen getan habe, ist, das Buch als Beitrag in größere Horizonte der Debatten über ‚Technik und Gesellschaft‘ einzuordnen. Diese Einordnungen findet der Autor offensichtlich unangemessen oder bemerkt sie gar nicht. Das mögen die LeserInnen beurteilen.

Zu wenig gelobt: Als Rahmenbedingung sei der vorgegebene Umfang der Besprechung angeführt. Ich wäre beispielsweise sehr gern auf die Möglichkeiten und Grenzen agentenbasierter Modellierung sozialer Dynamiken näher eingegangen. Ich stimme mit dem Autor überein, dass hier große, vom soziologischen Mainstream übersehene Chancen der Theoriebildung bestehen; und der Autor hat dazu sehr spannende Analysen vorgelegt. Aber ein Allheilmittel, um die chronischen Schwächen soziologischer Theorie zu kurieren, ist Simulation nicht. Vor allem widmet sie sich nur der „Logik der Aggregation“ – was gut ist, weil die Soziologie sich um diese viel zu wenig kümmert, was aber eben nicht alles ist, was die Soziologie zu analysieren hat. Allein diese Stärken/Schwächen-Bilanz auszuführen und die Leistung des Autors differenziert zu würdigen, hätte aber schon die Grenzen der Rezension gesprengt.

Wissenschaftlicher Erkenntnisfortschritt lebt davon, Kritik ebenso wie Anregungen abzuwägen. Wer dies nicht tut, endet im schlimmsten Falle selbstgewiss im Autismus. Dieses Schicksal sehe ich für Johannes Weyer nicht. Doch er sollte „inkongruente Perspektiven“ (Luhmann) auf seine Erkenntnisse auf sich wirken lassen, anstatt sie pauschal abzuwimmeln.

Information

Diese Replik bezieht sich auf Johannes Weyers Gegendarstellung zu Uwe Schimanks Rezension „Sorge ohne Fatalismus. Ambivalenzen immer tiefer gehender Digitalisierung“ in TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 30 (3), S. 83–84. <https://doi.org/10.14512/tatup.30.3.83>

This is an article distributed under the terms of the Creative Commons Attribution License CCBY 4.0 (<https://creativecommons.org/licenses/by/4.0/>) <https://doi.org/10.14512/tatup.30.3.85>
Received: Aug. 13, 2021; revised version accepted: Aug. 30, 2021;
published online: Dec. 20, 2021 (editorial peer review)



U wie Update

Durch die Überhitzung des Planeten droht die Unbewohnbarkeit ganzer Regionen. Neben dieser schlechten Nachricht gibt es eine gute: Die Klimawende ist noch möglich – mit den richtigen Technologien und einem gestärkten demokratischen System. Wolfgang Oels zeigt, wie wir demokratische Systeme stärken und dabei die Gewaltenteilung verbessern können.

W. Oels

Democracy For Future

Das demokratische Update zur Klimawende

112 Seiten, Broschur, 12 Euro

ISBN 978-3-96238-333-6

Bestellbar im Buchhandel und unter www.oekom.de.
Auch als E-Book erhältlich.

Die guten Seiten der Zukunft



PRACTICE

EPTA und die Pandemie

Michael Nentwich, Institut für Technikfolgen-Abschätzung, Apostelgasse 23,
1030 Wien, AT (mnent@oeaw.ac.at) ☎ 0000-0003-2269-0076



86

Während für das Netzwerk European Parliamentary Technology Assessment (EPTA) das Jahr 2020 coronabedingt ohne bemerkenswerte Netzwerkaktivitäten verlaufen ist, hat die EPTA-Präsidenschaft 2021, das Rathenau Institut (Niederlande), wieder starke Lebenszeichen gesetzt – und setzt dieses Programm trotz des unerwarteten, tragischen Ablebens seiner Direktorin Melanie Peters im Sommer fort. Wegen der Pandemie konnte die EPTA-Konferenz nicht in Den Haag stattfinden, dafür widmeten sich parlamentarische Abgeordnete und TA-Praktiker*innen aber am 9. November 2021 in einem hybriden Format der Rolle der TA in der Pandemie. Traditionell vorbereitet wurde die Konferenz mit 18 Länderberichten, darunter auch von TA-Swiss, TAB und ITA, sowie einer ausführlichen Synthese, verfasst vom Rathenau Institut und dem Norwegian Board of Technology. Der Bericht verfolgt zwei Ziele: zum einen zu dokumentieren, wie die verschiedenen Regierungen auf die Krise reagiert haben; zum anderen aus der Pandemie für die Rolle der TA und die Entscheidungsfindung bei wissenschaftlicher Unsicherheit zu lernen.

Das Synthesekapitel des Berichts (EPTA 2021) wirft einerseits einen Blick zurück und reflektiert über Herausforderungen und Chancen parlamentarischer TA, andererseits identifiziert es zukünftige Themen parlamentarischer TA in einer Post-COVID-Welt. Der ländervergleichende Rückblick stellt drei Hauptaufgaben heraus: in Zeiten der Krise eine Balance zwischen Reflexion und Aktion zu finden, Transparenz und öffentliches Vertrauen zu fördern sowie auf unbekannte Zukünfte vorzubereiten, also ‚antizipative Governance‘ zu unterstützen. Wenngleich der TA alle drei Herausforderungen aus anderen Praxiskontexten bekannt sind, werden sie im Zusammenhang mit der Pandemie besonders deutlich. Erkennbar wurden auch die höchst

unterschiedlichen Rollen der verschiedenen parlamentarischen TA-Einrichtungen: Während manche von der Politik aktiv einbezogen wurden und gleichsam im Zentrum wissenschaftsbasierter Entscheidungsfindung agierten, leisteten andere vorwiegend Beiträge aus eigener Initiative. Für fast alle EPTA-Einrichtungen hieß es, das ursprünglich geplante Arbeitsprogramm teilweise hintenanzustellen. International zu beobachten war, dass wissenschaftliche Einrichtungen in der Pandemie allgemein sichtbarer wurden, darunter teilweise auch TA-Einrichtungen.

Die Autor*innen der Synthese blicken auf Basis der gesammelten EPTA-Erfahrungen in die Zukunft und empfehlen der TA-Community mit Nachdruck, den Fokus auf einige kritische Themen zu legen: Angesichts der wichtigen Rolle, die Digitalisierung in verschiedensten Varianten in der Pandemie gespielt hat, gehe es nun darum, eine inklusive, faire, nachhaltige und damit resiliente Digitalisierung voranzutreiben. Speziell Homeoffice – als TA-Thema seit den 1990er-Jahren als ‚Telearbeit‘ bekannt – hat eine Renaissance erlebt; jetzt gelte es, diese Entwicklung als voraussichtlich wesentliches Merkmal des modernen Lebens durch TA zu begleiten, damit Telearbeit integrativ, verantwortungsbewusst und ganzheitlich integriert werde. Für die Zukunft stehe der Aufbau gesellschaftlicher Widerstandsfähigkeit gegen Krisen ganz oben auf der Agenda; Parlamente und EPTA sollten die Diskussion über diese transnationalen politischen Fragen gemeinsam anstoßen. EPTA-Mitglieder sollten darüber hinaus verstärkt die Anfälligkeit komplexer sozio-technischer Infrastrukturen in ihren künftigen Analysen und Überlegungen einbeziehen und fragen, was in welcher Form mit Hilfe der Technologie politisch machbar ist. Weiter sehen die Autor*innen eine entscheidende Rolle für parlamentarische TA bei der Entwicklung von Dateninfrastrukturen, die die länderübergreifende Analyse von Gesundheitsdaten zu Zwecken der öffentlichen Gesundheit ermöglichen und gleichzeitig die Rechte der EU-Bürger*innen wahren. Schließlich hätten TA-Einrichtungen einiges zur Frage beizutragen, wie wir langfristig mit COVID-19 leben wollen, etwa wie sich die neuen technologischen und therapeutischen Ansätze auf die Gesellschaft und ihre Bürger*innen auswirken und wie diese mit Blick auf Gemeinwohl und Nachhaltigkeit weiterentwickelt werden können.

Literatur

European Parliamentary Technology Assessment (EPTA) (2021): Technology assessment and decision-making under scientific uncertainty. Lessons from the COVID-19 pandemic. EPTA Report 2021. Online verfügbar unter http://epub.oeaw.ac.at/ita/ita-papers/EPTA_report_2021.pdf, zuletzt geprüft am 08. 11. 2021.

VERSCHENKEN SIE EIN JAHR die Zukunft des Wirtschaftens

mit einem Abo von ÖKOLOGISCHES WIRTSCHAFTEN

Vier Ausgaben für 74,40 Euro (zzgl. Versand)

Bestellung an: abo@oekom.de oder +49/(089) 54 41 84-225
oder unter: www.oekologisches-wirtschaften.net

4 AUSGABEN
PLUS GUTSCHEIN ZUM
DOWNLOAD



PASSENDES GESCHENK GESUCHT !

VERSCHENKEN SIE EIN JAHR
FORSCHUNG OHNE FACHGRENZEN
MIT EINEM ABO DER GAIA

Vier Ausgaben für 118,- Euro (zzgl. Versand)

Bestellung an abo@oekom.de oder +49/(0)89/54 41 84-225
oder unter www.gaia-online.net



GAIA
ÖKOLOGISCHE PERSPEKTIVEN FÜR WISSENSCHAFT UND GESELLSCHAFT
ist eine transdisziplinäre Zeitschrift für Wissenschaftler
und Wissenschaftsinteressierte, die sich mit Hintergründen, Analysen
und Lösungen von Umwelt- und Nachhaltigkeitsproblemen befassen.
MIT PEER REVIEW

TATuP Dates

TATuP

TATuP 1/2022 erscheint im
März 2022 mit dem Special Topic

„Technikfolgenab- schätzung im Kontext von (Hochschul-) Bildung: Theorien, Praxisansätze und Konzepte“

Praxis und Methoden der Technikfolgenab-
schätzung (TA) werden in vielfältigen Bil-
dungszusammenhängen angewendet und
können dazu beitragen, über disziplinäre Gren-
zen hinweg gesellschaftlich relevante Frage-
stellungen aus verschiedenen Perspektiven
zu bearbeiten. Wie soll oder kann sich TA als
transdisziplinäres Forschungsfeld weiterent-
wickeln?

Das TATuP-Thema in Heft 1/2022 nimmt
beide Perspektiven ein: Die Praxis in diversen
trans- und interdisziplinären Bildungskontex-
ten sowie die eigene (Aus-)Bildungspraxis der
TA.

Herausgegeben durch Sabrina Eimler
(Hochschule Ruhr West – Forschungsinstitut
Positive Computing) und Elke Hemminger
(Evangelische Hochschule RWL Bochum –
Technik, Medien, Empirische Sozialforschung).

30 JAHRGÄNGE: TATuP FEIERT JUBILÄUM

Anlässlich des 30-jährigen Jubiläums haben wir die Herausgeberinnen und Herausgeber sowie
die Mitglieder des wissenschaftlichen Beirats gefragt, welche Beiträge sie empfehlen, was sie
an der Zeitschrift besonders schätzen und ihr für die Zukunft wünschen. Alle Antworten unter:
www.tatup.de/TATuP30

NEUE WISSENSCHAFTLICHE REDAKTEURIN

Leonie Seng ist seit Oktober 2021 wissenschaftliche Redakteurin bei TATuP. Ihre Forschungs-
schwerpunkte liegen in den Bereichen Technikphilosophie und Ethik künstlicher Intelligenz.

TATuP DANKT HERZLICH ALLEN GUTACHTERINNEN UND GUTACHTERN

für Begutachtungen der zur Publikation im Jahr 2021 vorgesehenen Manuskripte!

P. Alfaro D'Alencon, S. Bauriedl, U. Bechtold, S. Böschen, S. Bühner, A. Burchart, J. Cas, Y. Egorova,
L. Eltrop, H. Fangerau, J. Feuchter, T. Fleischer, A. Förster, G. Grieser, M. Herrgen, L. Hilty, D. Hofmann,
W. Houkes, K. Hufendiek, A. Källén, M. Knecht, J. Kopfmüller, K. Kornwachs, J. Lachmund, R. Lind-
ner, V. Lipphardt, S. Maasen, D. Mahr, P. McCann, R. Meyer, M. Nentwich, J. Niewöhner, A. Nordmann,
R. Otte, F. Pallas, J. H. Passoth, K. Pfeffer, S. Rapp, W. Raskob, R. Roschy, A. Rossnagel, A. Sauter,
T. Sinozic, M. Sommer, M. Sonnberger, A. Strüver, B. Thull, W. Viehöfer, M. Weber, P. Wehling, J. Weyer,
D. Zupan

BEITRÄGE EINREICHEN UND ONLINE LESEN



www.tatup.de

AUF DEM LAUFENDEN BLEIBEN



www.oekom.de/newsletter

KOMMENTIEREN, TEILEN, LIKEN



www.facebook.com/TAjournal



www.twitter.com/TAjournal



www.tatup.de/youtube

Überlebenswichtig – und bald Mangelware?

Grundlage allen Lebens und zugleich verheerender Teil von Brandbomben und Pestiziden: Phosphor ist ein ambivalenter Stoff. Noch bedenklicher allerdings ist, dass abbaubares Phosphat langsam rar wird. Ein Buch über Vergangenheit und Zukunft eines essenziellen Elements.

Stefan Emeis, Kerstin Schlögl-Flierl (Hrsg.)

Phosphor Fluch und Segen eines Elements



224 Seiten, Hardcover, 25 Euro
ISBN: 978-3-96238-282-7
Erscheinungstermin: 14.09.2021
Auch als E-Book erhältlich



oekom.de

DIE GUTEN SEITEN DER ZUKUNFT



Weißes Wunder in Gefahr

Hitzewellen an den Polen und in Sibirien lassen immer größere Eismassen dahinschmelzen. Mit ihnen schwinden Süßwasservorräte und Lebensräume. »Das Eisbuch« stellt die gefährdeten Landschaften in außergewöhnlichen Infografiken vor und erkundet Möglichkeiten ihrer Rettung.

Esther Gonstalla

Das Eisbuch Alles, was man wissen muss, in 50 Grafiken



112 Seiten, Hardcover, zweifarbig, 24 Euro
ISBN: 978-3-96238-287-2
Erscheinungstermin: 05.10.2021
Auch als E-Book erhältlich



oekom.de

DIE GUTEN SEITEN DER ZUKUNFT



TATuP

ist die weltweit einzige interdisziplinäre Zeitschrift für Technikfolgenabschätzung. Wir fördern Debatten über technologische Innovation, Politik, Gesellschaft und Nachhaltigkeit. In TATuP finden Sie...

- Antworten auf drängende Fragen unserer Zeit;
- höchste wissenschaftliche Qualität durch faire Begutachtung und Transparenz;
- 100 % Open-Access: freier Zugang zu allen Artikeln und kostenfreie Publikation.

Herausgegeben vom Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)
am Karlsruher Institut für Technologie (KIT)

itas Institute for
Technology Assessment
and Systems Analysis

TATuP und oekom – passt natürlich!

Mit dieser Zeitschrift halten Sie ein echtes Stück Nachhaltigkeit in den Händen. Sie unterstützen eine Produktion mit hohen ökologischen Ansprüchen. Der oekom verlag...

- verwendet 100 % Recyclingpapier;
- verzichtet auf Plastikfolie;
- kompensiert alle klimaschädigenden Emissionen;
- druckt in Deutschland – und sorgt damit für kurze Transportwege.

Weitere Informationen finden Sie unter www.natürlich-oekom.de
und #natürlich_oekom.

 **oekom**