



Kompetenzzentrum
Öffentliche IT

(Un)ergründlich? Künstliche Intelligenz als Ordnungstifterin

Karoline Krenn · Simon Hunt · Peter Parycek (Hg.)

Gefördert durch:



Bundesministerium
des Innern, für Bau
und Heimat



Fraunhofer
FOKUS

(Un)ergründlich?

Künstliche Intelligenz als Ordnungsstifterin

Karoline Krenn · Simon Hunt · Peter Parycek (Hg.)

(Un)ergründlich? Künstliche Intelligenz als Ordnungstifterin

Karoline Krenn · Simon Hunt · Peter Parycek (Hg.)



**Kompetenzzentrum
Öffentliche IT**

Gefördert durch:



Bundesministerium
des Innern, für Bau
und Heimat

Impressum

Herausgeber:

Karoline Krenn · Simon Hunt · Peter Parycek (Hg.)
Kompetenzzentrum Öffentliche IT (ÖFIT)
Fraunhofer-Institut für Offene
Kommunikationssysteme FOKUS
Kaiserin-Augusta-Allee 31, 10589 Berlin

Kontakt:

Dr. Karoline Krenn
Kompetenzzentrum Öffentliche IT
Telefon: +49-30-3463-7173
Telefax: +49-30-3463-99-7173
info@oeffentliche-it.de
www.oeffentliche-it.de
www.fokus.fraunhofer.de

1. Auflage Februar 2020

ISBN: 978-3-948582-00-5

Bildnachweis:

Christian Mio Loclair | Waltz Binaire:
Blackberry Winter
<https://christianmiolclair.com/>

Gestaltung:

Pia Keeling, Reiko Kammer

Dieses Werk steht, soweit die Beiträge nicht anders ausgezeichnet sind, unter einer Creative Commons Namensnennung 3.0 Deutschland (CC BY 3.0 DE) Lizenz. Es ist erlaubt, das Werk bzw. den Inhalt zu vervielfältigen, zu verbreiten und öffentlich zugänglich zu machen, Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anzufertigen sowie das Werk kommerziell zu nutzen. Bedingung für die Nutzung ist die Angabe der Namen der Autoren sowie des Herausgebers.

 **Fraunhofer**
FOKUS

Inhaltsverzeichnis

| | |
|---|------------|
| Künstliche Intelligenz – Perspektiven und Herausforderungen | 03 |
| Karoline Krenn, Simon Sebastian Hunt & Peter Parycek | |
| Was ist ein Körper? Lernschritte einer Künstlichen Intelligenz in Bildern | 17 |
| Christian Mio Loclair im Gespräch mit Christian Rauch | |
| I. KI als gesellschaftliches Ordnungssystem | |
| Wir zählen nicht nur bis eins: Künstliche Intelligenz im Kontext | 27 |
| Dirk Baecker | |
| Digit[ale] Dinge: ordnende Finger in unserer Gesellschaft | 49 |
| Astrid Schwarz | |
| Über Erwartungen, Unübersichtlichkeiten und Pragmatismus: Künstliche Intelligenz im Prozess der gesellschaftlichen Implementierung | 67 |
| Tobias Wangermann | |
| II. KI und die Automatisierung von Entscheidungen | |
| Rekodierung von Religion: Theologische Darstellungen von Künstlicher Intelligenz und der Zukunft von Gesellschaft | 81 |
| Robert M. Geraci | |
| How Artificial Intelligence is Impacting Societies | 111 |
| Lorena Jaume-Palasi | |

III. Vertrauen in KI-Systeme

| | |
|---|-----|
| Vertrauen in KI-basierte Systeme schaffen | 127 |
| Ina Schieferdecker | |
| Über die Notwendigkeit einer Maschinenethik | 143 |
| Benedikt Paul Göcke & Lukas Brand | |
| Maschinelle Intelligenz – Fluch oder Segen? Es liegt an uns! | 161 |
| Dirk Helbing | |
| Chatting with Mitsuku | 185 |
| Stefan Ullrich & Diana Serbanescu | |
| The Silver Farms: Rituals from a Hybrid Future | 199 |
| Diana Serbanescu & Régis Lemberthe | |
| Autor:innenverzeichnis | 231 |

Künstliche Intelligenz – Perspektiven und Herausforderungen

Karoline Krenn, Simon Sebastian Hunt &
Peter Parycek

1. Künstliche Intelligenz gilt als Schlüsseltechnologie des 21. Jahrhunderts. Innerhalb weniger Jahre ist die Auseinandersetzung mit KI von einem von Spezialisten dominierten Fachdiskurs zu einem Thema der breiten medialen Berichterstattung geworden. Das hängt einerseits mit dem gestiegenen Einsatz von KI-Anwendung im Alltag zusammen, andererseits mit der lebhaften Diskussion über (vermeintliche) Innovationen und die damit möglichen Zukunftsszenarien. Utopien und Dystopien ringen um die eingängigere Welterklärung, was die Rolle von lernenden Maschinen und autonom agierenden Systemen in zukünftigen Gesellschaften anbelangt. Der Einsatz von Sprachassistenten und automatischer Gesichtserkennung, selbstfahrende Fahrzeuge sowie das Vordringen von automatisierten Entscheidungsprozessen auf Grundlage datenbasierter Vorhersagen beflügeln viele Fantasien. Die überragenden Fertigkeiten der KI in speziellen Bereichen nähren Vorstellungen eines Masteralgorithmus (Domingos 2015) oder einer maschinellen Superintelligenz (General Artificial Intelligence), die der menschlichen überlegen sein könnte und dadurch zur Bedrohung wird (Bostrom 2014). Verstärkt wird dieses Narrativ dadurch, dass prominente Expert:innen wie Ray Kurzweil, Elon Musk oder früher schon Stephen Hawking in den Unheil verkündeten Chor miteinstimmten. Wurde also mit den Entwicklungen im Bereich der Künstlichen Intelligenz die Büchse der Pandora geöffnet? Werden wir Zeitzeug:innen einer griechischen Tragödie? Rückblickend auf ältere Debatten spricht einiges dagegen, denn Vorstellungen zum gesellschaftlichen Einfluss Künstlicher Intelligenz sind alles andere als neu. Schon Babbage und Ada Lovelace haben Mitte des 19. Jahrhunderts mit dem Gedanken gespielt, dass die Analytical Engine in der Lage sein wird, Musik zu komponieren und Schach zu spielen (Gleick 2011). Eine solche Form von maschineller Intelligenz wäre dann auch in der Lage, die Macht und Zivilisiertheit der Menschheit zu steigern. Diese Vorstellungen waren aber noch sehr stark von der deterministischen, rein logischen Vorstellung geprägt, dass alles, inklusive dem Menschen selbst, mathematisch beschrieben und berechnet werden kann. Und schließlich skizzierte Alan Turing in den 1950ern einen Zukunftsentwurf, in dem Maschinen die Macht übernehmen, nachdem er bereits mit dem Turing-Test einen bis heute geltenden Standard

in der KI-Bewertung entwickelt hatte. 1955 wurde KI dann als eigenes Forschungsprogramm gekennzeichnet (McCarthy et al. 2016 [1955]). Eine erste Hochphase in der KI-Forschung gab es tatsächlich – angeheizt von einem während des Zweiten Weltkrieges entstandenen militärisch industriellen (Forschungs-)Komplex – in den 1960er Jahren. Bereits damals sah man sich kurz vor dem Durchbruch zur Entwicklung intelligenter Maschinen, die ihre Umwelt beschreiben und selbstständig Probleme lösen können. Viele Problemstellungen stellten sich allerdings als deutlich komplizierter dar, als man zunächst annahm. Seit der Zeit von Babbage und Lovelace hat sich das Verständnis für die Komplexität menschlicher Handlungen und Denkweisen erweitert. Das Wissen um die Herausforderungen bei der Erschaffung von KI-Systemen ist entsprechend mit dem Wissen um die Anforderungen an KI gestiegen. Misserfolge in Forschung und Entwicklung aufgrund mangelnder Rechnerleistung und fehlender Daten ließen schließlich die Forschungsgelder versiegen und das Thema nahm wieder einen Nischenplatz ein. Auch wenn sich an diesen Randbedingungen mittlerweile einiges verändert hat, deutet im Moment wenig darauf hin, dass sich die Spezialfertigkeiten von KI zu einer »Superintelligenz« kombinieren lassen.

Joseph Weizenbaum (1978) brachte die zwei Grundherausforderungen für Computersysteme auf den Punkt: Die menschliche Denkfähigkeit verfügt zum einen über ein (natürliches) Sprachverstehen und zum anderen über die Befähigung, Probleme zu identifizieren bzw. Fragen zu stellen. Der Punkt, auf den Weizenbaum hinaus wollte, war der, dass sich menschliche Intelligenz durch weit mehr als programmierbare Informationsverarbeitung auszeichnet, nämlich die Fähigkeit zur Kontextualisierung und zur Intuition. Diese Herausforderungen haben immer noch Aktualität und Gültigkeit.

Mehrere Faktoren sind ausschlaggebend für die gegenwärtige Konjunktur von KI. Leistungsfähigere, parallel organisierte GPU-Prozessoren haben die Datenverarbeitung beschleunigt und die Bearbeitung komplexer Datenstrukturen ermöglicht. Die massenhafte Verbreitung digitaler Infrastrukturen hat zu einer schier grenzenlosen Verfügbarkeit von Daten geführt, an welchen Algorithmen einerseits trainiert werden können und andererseits zum Einsatz kommen. Darüber hinaus sind mit der Weiterentwicklung maschineller Lernverfahren – Deep Learning bzw. Künstlichen Neuronalen Netzwerken – neue Hoffnungen auf einen revolutionären Durchbruch in der KI-Entwicklung entbrannt, nämlich dass intelligente Maschinen auf Common-Sense-Wissen trainiert werden können. Diese Entwicklungen werden im wesentlichen von der Industrie vorangetrieben. Aber auch die Regierungen der wirtschaftlich führenden Staaten haben mittlerweile den Wettbewerb um KI (wie z. B. um leistungsfähigste Prozessoren) in nationale För-

derprogramme übersetzt, womit Forschung und Entwicklung ein besseres Entwicklungsumfeld geboten wird.

Eine weitere wesentliche Voraussetzung für die Proliferation bestimmter KI-Anwendungen liegt in der Verfügbarkeit von großen Datenmengen. Daten kommen sowohl als Trainingsdaten als auch bei der Nutzung von KI zum Einsatz. Der Einsatz von KI geht oft mit dem Aufbau von Plattformen und der (räumlichen) Ausstattung mit sensorischen Geräten einher, die wiederum neue Dateninfrastrukturen hervorbringen. Die weitere Ausgestaltung von KI-Technologien in der Zukunft ist daher eng verknüpft mit den Fragen, wie und welche Daten gewonnen und unter welchen Bedingungen sie weiterverwertet werden können und dürfen. Die Tragweite der Veränderungen, welche das Zusammentreffen dieser Faktoren anstoßen wird, lässt sich derzeit noch nicht einschätzen. Auch unter den Expert:innen gehen die Meinungen darüber auseinander, ob es sich bei den aktuellen Entwicklungen um eine Fortschreibung dessen (Kittler 2013 [1989]) handelt, was mit Turings Universaler Diskreter Maschine (Universalmaschine) begonnen wurde (Turing 1936), oder ob die Fortschritte Künstlicher Neuronaler Netzwerke eine neue Phase in der Entwicklung von KI einläuten (Domingos 2015; Sudmann 2018).

Ein paar begriffliche Annäherungen sollen die hinter den Konzepten liegenden Unterschiede verdeutlichen. Klassische KI, auch symbolische KI genannt, basiert auf der Programmierung von Handlungsvorschriften oder bedingten Anweisungen (If-Then-Else Befehlstrukturen encodiert über Algorithmen). Dem liegt die Annahme zugrunde, dass eine Beschreibung und Lösung des betrachteten Problems durch regelbasierte Entscheidungen möglich ist. Die Begrenztheit dieser Top-down Verfahren liegt in der Bewältigung von Problemstellungen, die nicht im Vorfeld durch Regeln abgebildet wurden. Die Möglichkeiten der Wissensrepräsentation sind dadurch eingeschränkt. Auch in der Frühphase von KI war die Forschung von der Frage geleitet, wie Maschinen neue Lernerfahrungen integrieren können. Solche Bottom-up Verfahren des maschinellen Lernens wurden damals aber (auch aufgrund der Anforderungen an Prozessoren) als wenig aussichtsreich eingeschätzt. Aktuell sind Maschinelle Lernverfahren (ML) die am meist verbreiteten Verfahren. Sie versetzen einen Rechner in die Lage, auf Basis von Erfahrungen automatisch Datenmuster zu erkennen, Vorhersagen zu treffen oder Aufgaben zu lösen, ohne hierfür explizit programmiert worden zu sein (Samuel 1959). Mit anderen Worten: Der Algorithmus sucht nach auffälligen statistischen Strukturen (Sudmann 2018). Es handelt sich dabei also um »nichttriviale Maschinen«, deren Output nicht mehr eindeutig vorhersagbar ist (Foerster 1993).

Die heutige gängige Definition von Künstlicher Intelligenz fokussiert auf die Fähigkeit von Computerprogrammen zur eigenständigen Formulierung von Annahmen auf Basis von Daten durch die Erkennung von Mustern. Häufig verwendete Beispiele sind die Identifikation von Objekten in Bildern. Die dahinterliegende Zukunftsvision ist die Vorstellung, dass Computerprogramme irgendwann einmal in der Lage sein werden, eigenständig Probleme zu erkennen, eine geeignete Lösungsstrategie zu entwerfen und diese dann erfolgreich umzusetzen. Etwas spezifischer handelt es sich dabei um Algorithmen, die ihre Leistung (Output) in Bezug auf eine Aufgabe auf Basis von Erfahrungen (Input) verbessern. Hierfür wird der Algorithmus im Rahmen einer Lernphase mit Beispieldaten so lange trainiert, bis er eine zutreffende Übereinstimmung in neuen Daten erkennen kann. Wird in dieser Lernphase die Übereinstimmung zwischen Output und Input (durch Label) vorgegeben, spricht man von überwachtem Lernen (Supervised Machine Learning). Wird nur der Input vorgegeben, spricht man von unüberwachtem Lernen (Unsupervised Learning). Und schließlich gibt es noch semi-überwachte Lernverfahren, in denen Daten mit und ohne Label kombiniert werden. Der generierte Output reicht dabei von einfachen Zuordnungen (z. B. Klassifikationen) bis hin zu komplexen Lösungswegen. In der Anwendung wiederum kann unterschieden werden zwischen fortlaufend lernenden Systemen und Systemen mit abgeschlossener Trainingsphase. Während erstere auch während der Anwendung ständig weiter lernen, werden letztere einmal trainiert und mit diesem Systemzustand dann in die Anwendung gebracht.

Künstliche Neuronale Netzwerke (KNN) sind eine Untermenge solcher ML-Verfahren. Sie arbeiten mit Simulationsverfahren und modellieren dabei Lernprozesse in Form neuronaler Netzwerke, wie sie bei Menschen während dem Erlernen natürlicher Sprache entstehen. Bereits Wiener (1948) hat auf die Lernfähigkeit von Maschinen hinsichtlich der Erfassung der Wirkung ausgelöster Operationen hingewiesen. Besonders Erfolg versprechend erwies sich der Einsatz von sogenannter Backpropagation als Lernalgorithmus für ein mehrschichtiges KNN. Backpropagation kann die Gewichtung der Aktivität zwischen Neuronen im Prozess anpassen und Algorithmen damit auf mehreren Schichten gleichzeitig trainieren. Damit ist ein Problem der Wissensrepräsentation früherer Verfahren (einschichtiger KNN-Modelle) gelöst (Minsky und Papert 1969). Neben dem Einsatz leistungsfähiger, paralleler GPU-Prozessoren wird dies als Zäsur in der KI-Anwendung gesehen (Sudmann 2018).

Eine Reihe komplexer Lernalgorithmen, die nach verschiedenen Regeln und Logiken operieren, hat sich für unterschiedliche Aufgaben von KNN als erfolgreich erwiesen. In der Idee eines Masteralgorithmus (Domingos 2015) spiegelt sich, als ultimativer Entwicklungsschritt hin zu einer Superintelli-

genz, die Vision wieder, die verschiedenen Algorithmen maschinellen Lernens zu systematisieren und fortzuentwickeln, sodass sie im Zusammenspiel ihrer jeweiligen Spezialfertigkeiten ein perfektes Verständnis für die Welt und die Menschen in ihr entwickeln (siehe den Beitrag von Ullrich und Serbanescu in diesem Band). Bislang handelt es sich dabei allerdings nur um eine programmatische Ankündigung.

Eine wichtiges Merkmal gegenwärtiger KI ist die Unterscheidung zwischen Prognosen und Erklärungen. Je nach dem Design solcher Algorithmen werden nur Entscheidungen oder auch Begründungen geliefert. Da KNN eben nicht auf expliziten Regeln basieren, können sie Begründungen nur in Form von Wahrscheinlichkeiten liefern, die für Menschen nicht intuitiv verständlich sein müssen. Aus diesem Grund werden solche Algorithmen als Black Boxes wahrgenommen. Gerade in Lebensbereichen, in denen Menschen von Entscheidungen betroffen sind, die unter Einsatz solcher Black Boxes getroffen werden, bestehen diverse Risiken. Die Nachvollziehbarkeit algorithmischer Berechnungen ist daher auch eine Hauptforderung der Zivilgesellschaft und der Politik (siehe den Beitrag von Jaume-Palasi in diesem Band).

Solche Erklärungen sind technisch nicht grundsätzlich unmöglich. Höher entwickelte Algorithmen wie z. B. auf regularisierten linearen Regressionen basierende Modelle können zumindest einige informative Parameter bereitstellen. Allerdings sind solche Modelle in der Entwicklung kostspieliger als die Black Boxes. Herausforderungen bestehen auch hinsichtlich der Festlegung des Granularitätslevels der Erklärung, also ob diese sich auf das Modell als Ganzes oder auf einen konkreten Entscheidungsfall bezieht (globaler versus lokaler Ansatz). Auch für echte Ex-post-Erklärungen gibt es inzwischen verschiedene Ansätze, die jedoch, wie auch bei einer menschlichen Entscheidung, nicht die Bestimmung sämtlicher Einflussfaktoren auf die Entscheidung ermöglichen, sondern lediglich Ausschnitte oder erneute Prognosen liefern.

KI-Systeme begegnen uns im Alltag in Form von digitalen Sprachassistenten, automatischer Gesichtserkennung, intelligenten Systemen im Kontext von Smart Living, autonomen Fahrzeugen und versteckt in vielen digitalen Angeboten. Die Nutzung ist sowohl sektoral als auch regional und global noch sehr unterschiedlich. In Deutschland wie in den meisten westlichen Regionen ist KI eine Schlüsseltechnologie in den Bereichen öffentliche Sicherheit und Versorgung, in denen Zeit ein wesentlicher Faktor zur Prävention von Schaden ist. Echtzeitdiagnosen ermöglichen zeitnahe Prognose und rasche Interventionsmöglichkeiten. Frühwarnsysteme zu Wetterdaten und Gewässerständen verbessern den Katastrophenschutz. Im Verkehrs- und Infrastrukturbereich wird durch KI Mobilität gefördert (intelligente Ver-

kehrleitsysteme; Smart Lighting; autonomes Fahren). Ebenso ist der Zeitfaktor auch im Gesundheitswesen von großer Relevanz. KI wird dort zum einen zur Früherkennung von Krankheiten, unterstützend bei klinischen Entscheidungsprozessen sowie Behandlungsplänen und bei Interventionen eingesetzt. Zum anderen gibt es zahlreiche Anwendungen im Bereich der Robotik (Pflege- und Assistenzroboter).

Die Entwicklung von KI-Systemen ist nicht nur ein bedeutender neuer Wirtschaftszweig, in dem viele neue Berufsfelder entstehen, sondern derartige Systeme kommen mittlerweile in vielen Geschäftsfeldern kommerziell zum Einsatz. Im produzierenden Gewerbe reichen KI-basierte Verfahren von der Prozessoptimierung über die Fertigungssteuerung bis hin zur Erschließung neuer Geschäftsfelder. Plattformlösungen bündeln Daten und schneiden Logistik, Lieferketten und Vertriebszyklen auf Bedarfe zu. Handel und Banken nutzen Mustererkennungsprogramme zur Betrugserkennung sowie erstere auch für intelligente »Frischetheken« und Einkaufswagen (SmartCart mit Abrechnung über ein Onlinekonto) und zur dynamischen Preisgestaltung. Darüber hinaus wird KI vermehrt im Personalbereich bei der Vorauswahl und Rekrutierung von Bewerber:innen eingesetzt. Anwendungsgebiete für Künstliche Intelligenz sind auch Routinen wie die Umwandlung von Formaten (u. a. Transkription von Audiodateien) und die Archivierung (von Texten).

Im täglichen Umgang mit KI wie auch in der Forschung dazu werden allerdings auch Limitationen deutlich. Maschinen verstehen nicht, sie ordnen zu. Sie entscheiden nicht, sie berechnen. Eine basale Herausforderung besteht in der kontextadäquaten sinnhaften Deutung von Daten. Um ein altes Beispiel von Max Weber zur Verdeutlichung auf die KI zu übertragen: Die Maschinen tun sich in vielen Fällen noch schwer damit, einen Bildausschnitt eines gewöhnlichen Ereignisses wie des Tauschs von Ringen den richtigen Kategorien zuzuordnen. Ob es sich bei der Situation um eine Trauung oder einen Akt der Hehlerei handelt, erschließt sich erst aus dem Sinnzusammenhang. Erst die Fähigkeit, Informationen in einen Sinnzusammenhang einordnen zu können, befähigt dazu, darüber zu entscheiden, welche Zusatzinformationen (z. B. das weiße Brautkleid) für die Deutung relevant sind und welche Schlüsse daraus gezogen werden können (z. B. ob es sich um ein Theaterstück oder eine echte Zeremonie handelt).

Die Fähigkeit, Daten sinnhaft unter Bezug auf Motive zu verstehen, also das sinnhafte Verstehen, wird nach wie vor dem Menschen vorbehalten. Ob ein Vergleich von Maschinen mit dem menschlichen Verstehen, Entscheiden und Problemlösen zielführend ist, wird allerdings auch infrage gestellt. Relevant ist für bestimmte Anwendungen vermutlich eher der Turing-Tests, der auf die Immitation eines Menschen abstellt. Während eine KI-Anwen-

derung zur Optimierung bestimmter Prozesse ohne menschenähnliche Interaktion auskommt, ist eine glaubwürdige menschenähnliche und zielführende Kommunikation für andere Anwendungen Voraussetzung. Dabei wird es wohl, wie auch beim Turing-Test, mehr darauf ankommen, dem Menschen die empathieähnlichen Reaktionen glaubwürdig »vorspielen« zu können, und nicht originalgetreu nachzubilden.

Fehldeutungen durch Dekontextualisierungen sind nicht nur bei algorithmendeterminierten Entscheidungen problematisch, sondern auch dort, wo Algorithmen menschliche Entscheidungen unterstützen. Das Zusammenspiel zwischen Mensch und Maschine birgt dabei eine neue Welt mit eigenen Chancen und Risiken, deren Erforschung sich noch in den Kinderschuhen befindet. Der Mensch taugt nur bedingt als Kontrollinstanz für Entscheidungen von Algorithmen, wenn er deren Prognosen unkritisch zustimmt und im Zweifel Lücken in der Logik von Softwaresystemen (mit oder ohne KI) als Möglichkeit ausblendet. Dieser Umstand wird z. T. auch durch ein eingeschränktes Verständnis über das Zustandekommen von Ergebnissen bei den Nutzer:innen verstärkt.

Die Wahrnehmung von KI als Schlüsseltechnologie hat sie auch auf der politischen Agenda nach oben rücken lassen. Seit der Veröffentlichung der nationalen KI-Strategie der Bundesregierung im Oktober 2018 wurde eine Vielzahl an Förderprogrammen, Initiativen und Kooperationen gestartet und zahlreiche weitere politische Maßnahmen zur Gestaltung und Förderung von KI in Planung gegeben. Die Frage nach den richtigen agilen Steuerungsinstrumenten zur Gestaltung des KI-Standorts Deutschlands ist dennoch noch nicht hinreichend beantwortet. Neben Forschung und Ausbildung betreffenden und wirtschaftlichen Indikatoren spielt für die passenden Rahmenbedingungen auch die gesellschaftliche Wahrnehmung von KI eine große Rolle. Die gesellschaftliche Akzeptanz und das Vertrauen in KI hängen von verschiedenen Faktoren ab (siehe dazu auch den Beitrag von Schieferdecker in diesem Band). Dazu gehören Transparenz über den Einsatz und die Implementierung von KI, das Wissen über die Arbeitsweise von KI-Anwendungen und die Nachvollziehbarkeit ihrer Entscheidungen sowie damit verbundene Hoffnungen und Ängste. In Bereichen, in denen selbst die Meinungen von Expert:innen stark auseinanderdriften, ist es für die Öffentlichkeit umso schwieriger, Vertrauen aufzubauen. Als ein weiterer Schritt dieser KI-Strategie wird daher im März 2020 das KI-Observatorium starten, welches in einem strukturierten Prozess die technologischen Entwicklungen von KI und deren Folgen beobachten und analysieren und daraus Vorschläge für einen Ordnungsrahmen ableiten soll.

2. Der Sammelband greift mit »Ordnungsstiftung« und »Vertrauen« Themen auf, die sich zur Klärung der diversen Herausforderungen als hilfreich erweisen könnten. Kürzlich hat Armin Nassehi (2019) mit seinem Buch »Muster. Theorie der digitalen Gesellschaft« die Frage nach Ordnung und Ordnungsbildung in den Mittelpunkt digitaler Gesellschaften gerückt. Digitalisierung, so die These dort, sei die Antwort auf die zunehmende gesellschaftliche Komplexität. Das Bezugsproblem der Digitalisierung ist die Musterhaftigkeit der Gesellschaft. Bildlich gesprochen: Analog unsichtbar bleibende Muster werden durch eine vereinfachte binäre Abbildung – mit aller enthaltenen Unschärfe – digital sichtbar gemacht. Die Welt wird sozusagen verdoppelt. Das führt zu neuen Kontrollmöglichkeiten (siehe den Beitrag von Baecker in diesem Band). Künstliche Intelligenz spitzt den Themenkomplex rund um die Erkennung von Mustern auf die Automatisierbarkeit solcher Prozesse zu. Und sowohl KI als auch ihre Anwendungen basieren auf dieser vereinfachten Verdopplung der Welt.

Der vorliegende Sammelband beschäftigt sich mit Künstlicher Intelligenz unter eben diesem Aspekt »intelligenter« Ordnungssysteme und fragt danach, wie solche Systeme auszusehen vermögen. Ausgehend von einer Abgrenzung zu anderen Ordnungssystemen (wie Religion und Staat) werden in den Beiträgen sowohl die ordnenden Kapazitäten von KI (siehe den Beitrag von Schwarz in diesem Band), wie die Automatisierung von Entscheidungsprozessen, als auch die Voraussetzungen für die Bildung von Vertrauen und die sinnstiftenden Potenziale von KI thematisiert (siehe den Beitrag von Brand und Göcke in diesem Band). Im Band stehen damit weniger Fragen zum Stand der Technik und technischer Machbarkeit im Vordergrund als gesellschafts- und kulturtheoretische Überlegungen dazu, welche Veränderungen durch intelligente Ordnungssysteme angestoßen werden. Durch die interdisziplinäre Zusammensetzung der Beiträge ebenso wie durch die Formatwahl eröffnet der Band ein weites Spektrum von Perspektiven. Damit ermöglicht er einen vielfältigen Zugang zu den abstrakten, komplexen und durchaus kontroversen Entwicklungen im Feld der KI.

Wie bei jeder technischen Neuerung stellt sich in der politischen Debatte die Frage nach dem Nutzen und nach den Risiken. Dabei geht es letztlich auch um die Ausgestaltung der Technik an sich. Handelt es sich bei KI um ein neutrales Werkzeug, das aus empirischer Forschung und rationalen Schlussfolgerungen hervorgegangen ist, oder um ein sozio-kulturelles Erzeugnis, das ein spezifisches Menschen- und Gesellschaftsbild in sich trägt (siehe den Beitrag von Geraci in diesem Band)? Im letzteren Fall wären die extrahierten Muster und Ordnungen in einem spezifischen Kontext zu verorten. Für politische Entscheidungsprozesse heißt das, dass die Ordnungs- und die Steuerungsleistung von KI aktiv gestaltet werden können.

Eine öffentliche Akzeptanz von KI kann dadurch hergestellt werden, dass die Regeln und Prinzipien, nach denen KI-Anwendungen operieren, nachvollziehbar und ihr Einsatz und ihre Implementierung begründet und sichtbar gemacht werden.

Mit jeder neuen Technologie spannt sich ein Möglichkeitsraum auf. Die Entscheidungen, die dabei zu treffen sind, sind oft nur in sekundärer Sicht technischer Natur. Hinter technischen Ideen verbergen sich Vorstellungen darüber, wie eine Gesellschaft – im Fall der KI eine digitale Informationsgesellschaft – gestaltet werden soll und kann (Mohabbat Kar und Parycek 2018). Darin liegt auch die enge Verknüpfung zwischen Recht und Technik begründet. Der Einsatz ebenso wie die Limitierung von KI braucht dabei ein sicheres rechtliches Fundament, das einen grundrechtskonformen Einsatz von KI ermöglicht. Aus rechtsstaatlicher Sicht hat der Schutz von Rechtsgütern notwendigerweise Vorrang vor anderen Zwecken. Doch auch hier sind Zielkonflikte denkbar. Besonders umstritten ist dabei der Einsatz von KI im Rechtsvollzug selbst.

Sowohl für vollautomatisierte Prozesse als auch im Kontext von Assistenzsystemen, die bei der Entscheidungsfindung unterstützen soll, stellt sich diesbezüglich eine Reihe von Überlegungen und Fragen: Vorrangig ist die Frage, welche Ziele mit Künstlicher Intelligenz verfolgt werden und in welchen Bereichen sie überhaupt eingesetzt werden soll. KI ist eine Querschnittstechnologie mit vielfältigen Einsatzpotenzialen. Forschung und Entwicklung werden aktuell von der Industrie vorangetrieben. Die soziale Verträglichkeit von KI-Systemen steht dadurch hinter kommerziellen Interessen zurück. Und die Industrie ist zugleich Treiber eines Strukturwandels. Wenn über die gesellschaftlichen Auswirkungen Künstlicher Intelligenz diskutiert wird, geht es nicht allein um den Raum technischer Möglichkeiten, sondern um das Verhältnis von Individuen, Staat und Unternehmen, das sich durch die neuen Technologien möglicherweise in einem tiefgreifenden Wandel befindet (siehe den Beitrag von Wangermann in diesem Band). Die Frage danach, wofür KI überhaupt eingesetzt werden sollte, ist daher auf grundsätzlicher Ebene klärungsbedürftig. Transparenz und Erklärbarkeit der hinter KI liegenden Prozesse sind eine Grundvoraussetzung für eine breite Partizipation an einer solchen Debatte. Dabei sind die Chancen auf Objektivität gegen die Risiken aus dem Bias der Trainingsdaten und der Normativität statistischer Regelmäßigkeiten abzuwägen. Dazu muss über die ethische Verantwortung von Spezifizierer:innen, Implementierer:innen und Anwender:innen reflektiert werden. Die Rolle von Werten und Deutungen bleibt zu explizieren sowie der ethische und juristische Status von lernenden Algorithmen zu klären. Eine Konsequenz aus solchen Erwägungen kann auch der bewusste Verzicht auf KI-Anwendungen in besonders

sensiblen Bereichen sein. Für eine informierte Debatte bedarf es der Nutzung und der fachübergreifenden Vernetzung des KI-spezifischen Domänenwissens.

Die Frage nach den ordnungsstiftenden Dimensionen von KI kann mit diesem Band nur angestoßen werden. Gleichwohl mag eine Antwort auf die Frage, was eine Gesellschaft zusammenhält, in den wiederkehrenden gesellschaftlichen Bemühungen liegen, Antworten auf diese Frage zu geben (Baecker 2013). Hierzu möchte der Band einen Beitrag leisten.

3. Die Beiträge

Der Sammelband geht zurück auf das Symposium »(Un-)ergründlich. Künstliche Intelligenz als Ordnungsstifterin«, bei welchem im Oktober 2018 im silentgreen Kulturquartier in Berlin Vertreter:innen verschiedener Disziplinen Impulse zu den Herausforderungen von KI gegeben haben und miteinander ins Gespräch gekommen sind.

Der Wahl eines Text-und-Bild-Formats verbindet die wissenschaftlichen Beiträge des Symposiums mit dem Salonabend »KI und Kunst. Schamanen der digitalen Renaissance«, bei dem in verschiedenen künstlerischen Impulsen gesellschaftliche Visionen zu KI entwickelt und die Bedingungen für »künstlerische Intelligenz« reflektiert wurden. Diese Reflexion findet sich in zwei künstlerischen Beiträgen, die den Band rahmen. Im ersten Beitrag des Bandes erläutert WaltzBinaire-Gründer Christian Mio Loclair in einem Interview mit Christian Rauch seine Installation Blackberry Winter, in der die einzelnen Lernschritte einer künstlichen Intelligenz für das Erschaffen von dreidimensionalen Körpern aufgezeigt werden. Die Lernschritte von Blackberry Winter begleiten als Fortsetzungsformat die wissenschaftlichen Beiträge des Bandes.

Im zweiten Beitrag skizziert Dirk Baecker die menschliche Intelligenz als ein komplexes Zusammenspiel einer neuronalen, mentalen und sozialen Intelligenz. Dieses Zusammenspiel produziert »Erkenntnisse«, auf denen weitere Operationen des Organismus, des Bewusstseins und der Gesellschaft aufbauen können. Die verschiedenen Typen von Intelligenz sind nicht auf sich selbst reduzierbar und stehen quer zueinander. Dabei bauen sie jeweils einen eigenen Typ von Komplexität auf, um mit ihrer komplexen Umwelt nach eigenen Regeln fertigzuwerden. Die knappste Ressource der Computergesellschaft ist dabei die Geistesgegenwart. Das hat Konsequenzen für ein gesellschaftliches Ordnungssystem, das sich an dieser Kultur der Komplexität neu auszurichten hat, gleichsam zeigt dieses Zusammenspiel auch die Grenzen der ordnungsstiftenden Dimensionen von KI auf.

Der Beitrag von Astrid Schwarz dreht die Perspektive um. Ausgehend von der These »Gesellschaft macht KI« geht die Autorin der Frage nach, wie und

in welchen Konstellationen KI-Dinge Ordnungen hervorbringen. An objektzentrierten Beispielen von begehrten Dingen wie Smartphones oder digitalen Lebensformen im smarten Living diskutiert sie die Gestaltungsmöglichkeiten für individuelle Wissensaneignung. Ihre Überlegungen beschränken sich dabei nicht auf Künstliche Intelligenz. Sie reflektiert ganz allgemein über das Technikverhältnis des Menschen im Zeitalter des Anthropozäns und neue Formen von Öffentlichkeit, die durch soziale Medien entstehen. Ihre Diagnose ist dabei durchaus nicht nur pessimistisch. KI ebenso wie digitale Infrastrukturen überhaupt schaffen eine digitale Allmende. Betrachtet man KI als »heuristisches Hebzeug«, steht die Entscheidung darüber im Vordergrund, wofür sie eingesetzt wird: als Mittel zur sozialen Kontrolle oder zur Gestaltung einer pluralen Welt aktiver Teilhabe. Daraus folgt für Schwarz die Notwendigkeit einer permanenten Prüfung der entstehenden Ordnungsmuster und einer permanenten Abstimmung zwischen gesellschaftlichen Erwartungen und individuellen Handlungen der Dinge.

Tobias Wangermann betrachtet die Wechselwirkung zwischen technischer Innovation und Gesellschaft. KI löst das gesellschaftliche Problem, eine Ordnung in den komplexen Strukturen der globalen Gesellschaft zu erkennen. Nach Ansicht des Autors bleibt es der Gesellschaft überlassen, zu definieren, wie KI diese Aufgabe lösen soll. Damit steht nicht mehr die Technik allein im Mittelpunkt, sondern der wechselseitige Anpassungsprozess zwischen technischen Lösungen und gesellschaftlichen Anforderungen.

Robert Geracis Beitrag zeigt auf, dass sich Technologien nicht in einem kulturellen Vakuum etablieren, sondern ganz wesentlich von religiösen und kulturellen Traditionen geprägt werden. Er verdeutlicht dies an der Gegenüberstellung von in den USA und in Indien vorherrschenden, mit KI verbundenen Zukunftsaussichten. Am Beispiel populärer KI-Fiktionen wie dem Transhumanismus arbeitet der Autor dahinterliegende Menschenbilder und Vorstellungen vom sozialen Zusammenleben heraus, welche die KI-Entwicklung beeinflussen. Damit argumentiert er gegen einen technologischen Determinismus und stellt die Gestaltbarkeit des technologischen Fortschritts in den Mittelpunkt. Als eine Voraussetzung für eine sozial verträglichere Technologie sieht er dabei die Bewusstmachung dieser kulturellen Prägungen, welche die Türen zu neuen Denkräumen aufstoßen kann.

Die KI anhaftenden Zukunftsentwürfe stehen auch im Mittelpunkt des Beitrages von Lorena Jaume-Palasi. Die Autorin fokussiert die infrastrukturelle Natur von algorithmischen Systemen und KI-Anwendungen und die Risiken und Schädigungspotenziale, die damit einhergehen. Am Beispiel von Predictive Policing stellt sie insbesondere dahinterliegende Kausalannahmen in Frage. Ähnlich wie der vorangehende lenkt auch dieser Beitrag den Blick auf die normative Dimension von Infrastruktur und Technologie.

Jaume-Palasi argumentiert dafür, die für materielle Infrastrukturen auf europäischer Ebene schon lange diskutierten Werte, Prinzipien und Kriterien auf KI zu übertragen, um damit Standards für Fairness und soziale Verträglichkeit zu schaffen.

Ina Schieferdecker hebt in ihrem Beitrag die Bedeutung von Vertrauenswürdigkeit für die Akzeptanz KI-basierter Systeme heraus. Software Engineering beschäftigt sich seit den 1960ern mit der Entwicklung vertrauenswürdiger Software. Ein umfangreicher Methodenkasten schützt allerdings nicht vor wiederkehrenden Softwarekrisen. Die Wirkung von Software entfaltet sich erst in der Ausführung und wird in der konkreten Umgebung sichtbar. KI, die als Software realisiert wird, erfordert zudem die Berücksichtigung sozio-technischer Anforderungen, also die Reflexion über ihre gesellschaftliche Förderlichkeit. Schieferdecker stellt dazu aktuelle Testmethoden für die Verifikation und Validierung KI-basierter Systeme vor, deren Weiterentwicklung für die Akzeptanz von KI-basierten Systemen von enormer Bedeutung ist.

Der Beitrag von Lukas Brand und Benedikt Paul Göcke reflektiert über die moralisch relevante Handlungsfreiheit von Maschinen. Das maschinelle Lernen ermöglicht KI-Systemen, auf ihre Umgebung aufgrund von eigener Erfahrung autonom zu reagieren. Die Autoren argumentieren für die Notwendigkeit einer Maschinenethik, die Maschinen durch ihren Einsatz im Bereich der Pflege, des autonomen Fahrens und im militärischen Einsatz als moralische Subjekte in den Blick nimmt. Daraus folgern sie, dass auch das Selbstbild des Menschen und seine (moralische) Verantwortung neu zu überdenken sein werden.

Für Dirk Helbing ist KI ein Projektionsraum für dystopische Erwartungen. Er fragt danach, wie die Brücke geschlagen werden kann zwischen einer dystopischen Charakterisierung technologischer Entwicklung und ihrer Bewertung als Lösung für globale Probleme, z. B. durch den Einsatz moderner Informationstechnologien. Aufgrund bestehender Herausforderungen durch Überwachungstechnologien, der Fallstricke datenbasierter Steuerung und unkontrollierter Machtdynamiken verortet er die gegenwärtige Entwicklungen am Scheideweg zwischen schlechtestmöglichen und bestmöglichen Szenarien, deren Entwürfe er plastisch gegenüberstellt.

Steffen Ullrich und Diana Serbanescu schaffen mit ihrer Wiedergabe eines Chats mit dem Chatbot »Mitsuko« und ihrer Reflexion über das Bewusstsein von Mitsuko eine reizvolle Schnittstelle zwischen den wissenschaftlichen und den künstlerischen Beiträgen des Bandes. Der Beitrag knüpft damit an die schon von Weizenbaum rund um seine digitale »Psychotherapeutin« ELIZA lebhaft geführte Diskussion um eine »General Artificial Intelligence« an. Dabei wird der Blick auf den Unterschied zwischen den tatsächlichen

Fertigkeiten von KI-Systemen und den vorgenommenen Zuschreibungen in den Vordergrund gerückt.

Im abschließenden Beitrag von Diana Serbanescu und Régis Lemberthe zur Performance *The Silver Farms* des REPLICA Instituts werden bildstarke Antworten zur Frage nach der Koexistenz verschiedener Intelligenzformen und zum Wechselspiel zwischen algorithmischer Disruption und der neuen Spiritualität möglicher Zukünfte ausgearbeitet.

In ihrer Breite stehen die Themen des Bandes für eine Vielfalt an Perspektiven, die auf Künstliche Intelligenz eingenommen werden können. Die Beiträge verdeutlichen dabei, dass die Schaffung optimaler Rahmenbedingungen für die Entwicklung und Implementierung von KI-Technologien zu den großen Gestaltungsaufgaben der Gegenwart gehört. Wir bedanken uns bei unseren Autor:innen für die anregenden Anstöße, über KI und ihre Anwendungen und damit auch über die sinn- und ordnungsstiftende Dimension von KI nachzudenken. Und wir hoffen, dadurch mit diesem Band einen Beitrag zur Klärung der Frage zu leisten, wie mehr Akzeptanz und Vertrauen in KI und algorithmische Systeme hergestellt werden kann.

Baecker, Dirk. 2013. Metadaten. Eine Annäherung an Big Data. In *Big Data. Das neue Versprechen der Allwissenheit*, hg. von Heinrich Geiselberger und Tobias Moorstedt, 156-186. Berlin: Suhrkamp.

Bostrom, Nick. 2014. *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Domingos, Pedro. 2015. *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books.

Engemann, Christoph und Andreas Sudmann. 2018. *Machine Learning. Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*. Bielefeld: transcript.

Foerster, Heinz von. 1993. *Wissen und Gewissen. Versuch einer Brücke*. Frankfurt a.M.: Suhrkamp.

Gleick, James. 2011. To Throw the Powers of Thought into Wheel-Work. In *The Information. A History. A Theory. A Flood*, von James Gleick, 78-124. London: Fourth Estate.

Kittler, Friedrich. 2013 [1989]. Die künstliche Intelligenz des Weltkriegs. In *Die Wahrheit der technischen Welt. Essays zur Genealogie der Gegenwart*, hg. von Hans Ulrich Gumbrecht, 232-253. Berlin: Suhrkamp.

- McCarthy, John, Marvin L. Minsky, Nathaniel Rochester und Claude E. Shannon. 2016 [1955]. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 31. August 1955. In *AI Magazine* 27 (4): 12.
- Minsky, Marvin, und Seymour Papert. 1969. *Perceptrons*. Cambridge, MA: MIT Press.
- Mohabbat Kar, Resa, und Peter Parycek. 2018. Berechnen, ermöglichen, verhindern: Algorithmen als Ordnungs- und Steuerungsinstrument in der digitalen Gesellschaft. In *(Un)Berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft*, hg. von Resa Mohabbat Kar, Basanta Thapa und Peter Parycek, 7-39. Berlin: Kompetenzzentrum Öffentliche IT. (Online abrufbar unter: www.oeffentliche-it.de/publikationen)
- Nassehi, Armin. 2019. *Muster. Theorie der digitalen Gesellschaft*. München: C.H.Beck Verlag.
- Samuel, Arthur L. 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development* 3/3: 211-229. DOI: <https://doi.org/10.1147/rd.33.0210>
- Sudmann, Andreas. 2018. Szenarien des Postdigitalen. Deep Learning als MedienRevolution. In *Machine Learning. Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, hg. von Christoph Engemann und Andreas Sudmann, 55-73. Bielefeld: transcript.
- Turing, Alan. 1996 [1951]. Intelligent Machinery. A Heretical Theory. *Philosophica Mathematica* 4(3): 256-260.
- Turing, Alan. 1936/1937. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 42(2): 230-265.
- Weizenbaum, Joseph. 1978. *Die Macht der Computer und die Ohnmacht der Vernunft*. Frankfurt a.M.: Suhrkamp.
- Wiener, Norbert. 1948. *Cybernetics or Control and Communication in the Animal and the Machine*. Cambridge, MA: The MIT Press.

Was ist ein Körper?

Lernschritte einer Künstlichen Intelligenz in Bildern

Christian Mio Loclair im Gespräch mit Christian Rauch

C. Rauch **1. Dieser Sammelband enthält einzelne Aufnahmen Deiner neuen Installation »Blackberry Winter«. Könntest Du einleitend beschreiben, womit sich diese Bild-Serie befasst?**

C.M.Loclair Das Projekt Blackberry Winter ist unsere neueste Arbeit und beinhaltet einen technischen Meilenstein in dem Bereich KI und Raum. Wir haben ein GAN (Generative Adversarial Network) verwendet, welches erschaffen wurde um zweidimensionale Inhalte zu erlernen. Diese Technik hat grosse mediale Aufmerksamkeit erlangt, da sie bei so genannten Deepfake Videos verwendet wird. Es ist uns gelungen diese Technik in einem eigen entwickelten Verfahren um eine Dimension zu erweitern. Wir sind folglich in der Lage durch unseren Algorithmus hochauflösende 3D-Gestalten zu erdenken. So haben wir begonnen unsere KI auf den menschlichen Körper zu trainieren. Die hier vorgestellte Bildserie zeigt unterschiedliche Trainingsstadien des menschlichen Körpers. Zu Beginn sehen wir, dass die Maschine unklar sieht und den Menschen nicht als Einheit begreift. Über den Lernprozess hinweg, verbessert sich jedoch das »digital Verständnis« der KI und wir sehen klare menschliche Gestalten in unterschiedlichen Figuren.

C. Rauch **1. Du hast Deine Karriere als Tänzer und Choreograph begonnen, heute bewegst Du Dich als Entwickler und digitaler Künstler zwischen den Welten von Technologie, Kultur und Gesellschaft und verbindest Disziplinen. Wie kam es dazu? Was treibt Dich an?**

C.M.Loclair Ich habe mich als Tänzer darauf konzentriert, mich so mechanisch wie möglich zu bewegen. Ein Mensch, der zum Roboter wird. Ich empfinde es als etwas Magisches, wenn eine

Entität ihre Natur und ihr Wesen in eine andere überführt. So wird ein Mensch in seinem Verhalten zur Maschine, wenn auch nur in seinen Bewegungen.

Heute konzentriere ich mich vielmehr auf das gegenteilige Wechselspiel. Ich erschaffe Algorithmen und datengetriebene KI um die Maschine etwas sehr Menschliches machen zu lassen: Kreativität und Kunst. Die Frage ist folglich: Kann die menschliche Eigenart neue Welten zu erdenken, in eine digitale Form übertragen werden? Ich denke, dass diese Verwandlung der Menschen und der Maschinen, in ihrem jeweiligen Gegensatz, sehr zeitgenössisch ist.

C. Rauch **2. Welche Rolle haben Deiner Meinung nach disziplinen-übergreifende Ansätze in unserer Gesellschaft heute?**

C.M.Loclair Ich denke, dass interdisziplinäre Arbeiten wichtiger sind als je zuvor. Das hat ganz unterschiedliche Gründe. Ich forsche und arbeite im Kreativgewerbe. In unseren Produktionen gibt es zwei unterschiedliche Phasen der Entwicklung neuer Medienerlebnisse. Zum einen, die Iteration und zum Anderen, das Refinement. In der Iteration müssen wir möglichst viele Skizzen und Ideen kreieren und diese auswerten. Das Refinement hingegen folgt auf die Iterationsphase und beschreibt die Ausarbeitung der vielversprechendsten Skizze. Aufgrund der Schnellebigkeit von Media- und Social-Media-Kommunikation ist jedoch die Relevanz der Refinementphase in den Hintergrund gerückt - während das schnelle Iterieren an Bedeutung gewann. Durch diesen Anstieg an Iterationen, gewann auch das interdisziplinäre Arbeiten an Einfluss. Wenn möglichst viele Perspektiven mit ganz unterschiedlichen Bildungsprofilen, von Kunst bis Wissenschaft, Engineering und Design, dasselbe Thema beleuchten –steigt meinen Erfahrungen zufolge die Qualität der Skizzen dramatisch.

C. Rauch **3. (Künstliche Intelligenz und Deine Arbeit) In den letzten Jahren hat sich weltweit eine stark wachsende Szene im Bereich der KI-Kunst entwickelt. Auch bei Deiner Arbeit steht künstliche Intelligenz meist im Zentrum. Wie greifen Künstler die Entwicklungen in der KI auf, welche Rolle nehmen Künstler in der öffentlichen Debatte zu KI ein?**

C.M.Loclair Die KI wird in der Kunst aus unglaublich diversen Perspektiven aufgegriffen. Ich bin darüber sehr glücklich, denn ich denke, dass diese Technik einen gesellschaftlichen und kulturellen Wandel mit sich führen wird, der in seiner Wirkungsbreite vergleichbar mit der Industrialisierung sein wird. Viele Künstler verwenden die KI als Werkzeug um neue Bildwelten zu erschaffen. Andere wiederum nutzen diese Technik als Thema um politische und gesellschaftliche Fragen zu stellen. Ich selbst bin sehr daran interessiert zu sehen, inwiefern die KI ein neues Menschenbild erschaffen kann. Meiner Ansicht nach wird die KI grundlegende, philosophische Diskussionen aufwühlen in dem sie unser Selbstbildnis umwälzt.

C. Rauch **4. Warum beschäftigt Dich dieses Thema so sehr und was ist Deine Mission als Künstler? Wo siehst Du Deine Aufgabe?**

C.M.Loclair Die Notwendigkeit der KI als künstlerisches Thema lässt sich leicht erklären in einem Zweisatz. Keiner versteht es - jeden betrifft es. Es handelt sich hier um eine explosive Mischung, denn etwas Unsichtbares, Unvorhersehbares gestaltet unser Zusammenleben neu und wir müssen heute Entscheidungen treffen über Techniken, die wir nicht verstehen. Ich würde gerne Kunst schaffen, welche uns hilft diese Technik und ihren Einfluss auf unser Leben zu beleuchten. Eine Kunst, welche sich nicht erhebt, sondern mit ganz unterschiedlichen Menschen in Verbindung tritt. Ein abstraktes Gespräch, das ich als zeitgenössisch und kulturell notwendig halte.

C. Rauch **4. Die künstliche Intelligenz entwickelt sich mit rasanter Geschwindigkeit und erstaunt mit immer neuen Innovationen. Der britische Autor und Futurist Sir Arthur Charles Clark sagt in seinem berühmten dritten Gesetz »Any sufficiently advanced technology is indistinguishable from magic.« Ist KI die Magie des 21. Jahrhunderts?**

C.M.Loclair Ich glaube, dass alles, was in Formeln überführt wird, nicht mehr magisch ist. Das ist der Unterschied zwischen Glauben und Wissen. Wenn ich daran glaube das ich mit einem toten Objekt sprechen kann, dann ist das in der Tat ein magischer Gedanke. Ich muss jedoch nicht glauben, dass ich ein Tele-

fon in der Tasche habe, dass eben diese Fähigkeit besitzt. Ich weiss es – es ist eine wissenschaftliche Tatsache und deswegen kein magischer Gedanke mehr. Aus diesem Grunde würde ich gerne diese Frage ein wenig umformulieren. Der Mensch ist uns seit Anbeginn des Humanismus heilig. Er ist magisch. Die Frage ist nun: bleibt er magisch, sobald wir seine herausragendsten kognitiven Eigenschaften in Formeln überführen und vollkommen abbilden können?

C. Rauch **4. In Deinem sehr poetischen Kunstprojekt »Narciss« be-
dienst Du Dich der Welt der Mythologie und erschaffst einen
Computer, dessen einzige Aufgabe es ist sich selbst zu be-
trachten. Die Arbeit wurde weltweit mit grossem Erfolg aus-
gestellt. Welche Erkenntnisse hat Dich die Arbeit mit diesem
Projekt gelehrt?**

C.M.Loclair Das Projekt Narciss ist eine »nackte« Maschine, ein Mother-
board vor dem Spiegel, welche sich betrachtet und mittels
Machine Learning versucht herauszufinden, was sie eigent-
lich ist. Wir würden dieses Projekt gerne als Nachweis da-
für anführen, dass wir ganz alte Geschichten des Menschen
mittels KI neu beleuchten können. Es war für uns ein un-
glaubliches Erlebnis, in Zusammenarbeit mit dem State Stu-
dio, diese Arbeit im Palast der Nationen in Genf vorstellen
zu dürfen. Wir möchten weiterhin neue spektakuläre Ent-
deckungen der Wissenschaft, in einen nicht sensationellen,
sondern nachhaltigen Diskurs überführen.

C. Rauch **6. Was sind die Visionen für Deine Arbeit in der Zukunft?
Worauf arbeitest Du hin?**

C.M.Loclair Ich erfreue mich an technischen Sensation und sehe poe-
sie in der Begegnung zwischen Mensch und Maschine. Ich
würde mich freuen wenn ich weiterhin Momente erschaffen
könnte, in denen uns digitale Techniken als Spiegel erschei-
nen. So könnte ich Orte erschaffen in denen wir der Wissen-
schaft begegnen aber zu uns selbst zurück finden.











KI als gesellschaftliches Ordnungssystem

Wir zählen nicht nur bis eins: Künstliche Intelligenz im Kontext

Dirk Baecker

Künstliche Intelligenz ist nur dann ein »gesellschaftliches Ordnungssystem«, wenn es ihr gelingt, sich im Verhältnis zu anderen Typen von Intelligenz, vor allem der neuronalen, mentalen und sozialen Intelligenz, zu positionieren. Je deutlicher es wird, dass die künstliche Intelligenz nicht einfach eine Kopie, ein allmähliches Einholen und dann plötzliches (»singuläres«) Übertreffen der menschlichen Intelligenz ist, desto fraglicher wird, was unter einer »menschlichen« Intelligenz zu verstehen ist. Die aktuellen Bemühungen um eine künstliche Intelligenz beruhen auf Technologien des Maschinenlernens, die große Datenvolumen statistisch ausbeuten. Sie haben kaum noch etwas mit jenen Bemühungen der 1960er und 1970er Jahre zu tun, Systeme künstlicher Intelligenz als Systeme der Symbolverarbeitung zu verstehen, die angeblich in der Lage waren, logisches Denken zu replizieren. Seither hat man nicht nur wiederentdeckt, dass menschliches Denken nicht auf die klassische Logik reduziert werden kann, sondern man hat dank der kognitionswissenschaftlichen Forschung auch die vermeintliche Einheit der menschlichen Intelligenz auf ein komplexes Zusammenspiel einer neuronalen, mentalen und sozialen Intelligenz zurückgeführt. Die Einheit dieses Zusammenspiels ist weder rational noch emotional, sondern allenfalls »kognitiv« im nach wie vor offenen Verständnis dieses Grundbegriffs. Das Zusammenspiel produziert »Erkenntnisse«, auf denen weitere Operationen des Organismus, des Bewusstseins und der Gesellschaft aufbauen können.

Die folgenden Überlegungen versuchen, diesen verschiedenen Typen von Intelligenz etwas besser auf die Spur zu kommen. Sie zählen deswegen nicht nur bis eins. Wir können die verschiedenen Typen nicht auf ein einziges System reduzieren, sei es das der Biologie und Neurowissenschaften, das der Soziologie oder das der Informatik. Wir haben es mit Orthogonalität zu tun: Die verschiedenen Systeme sind nicht aufeinander reduzierbar, sie stehen quer zueinander,

sie bauen jeweils einen eigenen Typ von Komplexität auf, um mit ihrer komplexen Umwelt nach eigenen Regeln fertig zu werden. Lebende, mentale und soziale Systeme haben in einer Millionen Jahre zählenden Evolution einen Typ von Autonomie erlernt, der es ihnen erlaubt, sich in der Umwelt andersartiger Systeme zu erhalten und zu entwickeln. Es mag sein, dass Maschinen diese Evolution in größerer Geschwindigkeit und trainiert durch Menschen gerade nachholen. So oder so enthalten diese Maschinen die Kränkung, dass das Zusammenspiel der Intelligenzen nicht mehr in einer ihrer selbst mächtigen Vernunft à la Descartes, sondern allenfalls in einer kritischen Vernunft à la Kant zur Einheit gebracht werden kann, tatsächlich jedoch der Einheitsbegriff der Komplexität an die Stelle des Einheitsbegriffs der Vernunft tritt. Unser gesellschaftliches Ordnungssystem ordnet sich neu durch eine Kultur der Komplexität. Erste Schritte auf dem Weg, den wir gerade beschreiten, bestehen darin, nicht nur bis eins, sondern mindestens bis vier zu zählen.

1. Produktive Ungewissheit

Eine der Eigentümlichkeiten der jüngeren Bemühungen um eine auf den Verfahren des Maschinenlernens beruhende künstliche Intelligenz liegt darin, dass die Vorhersagekraft der Algorithmen, die von den Maschinen erlernt werden, mit der Menge der Daten steigt, die den Maschinen zur Verfügung stehen. Mit der Vorauswahl dieser Daten aus einer bestimmten Domäne des Wissens kommt zwar ein einschränkender Faktor hinzu, doch ändert das nichts daran, dass die Vorhersagen umso verlässlicher sind, je größer das Volumen der Daten ist. An dieser Beobachtung ist nichts eigentümlich, solange man sich an die Grundgesetze der Statistik hält. Jedes einzelne Ereignis ist unvorhersehbarer als die Menge vieler Ereignisse, die einer bestimmbarer Wahrscheinlichkeit folgen. Doch in dem Moment, in dem man einen Schritt zurück tritt und aus der großen Menge der Ereignisse die Schlussfolgerung zieht, dass mit der Menge der Ereignisse auch die Menge der Möglichkeiten steigt und somit jede Vorhersage ungewisser wird, wird deutlich, dass das Gesetz der großen Zahl eine Information enthält, die in der bloßen Relation zwischen Datenvolumen und Vorhersage noch nicht abgebildet ist. Wie kann es sein, dass mit der Ungewiss-

heit jedes einzelnen Ereignisses die Sicherheit der Vorhersage steigt und nicht fällt? Wer oder was ist für die Einschränkung der Möglichkeiten im Kontext von Ungewissheit verantwortlich?

Die Fragestellung enthält bereits einen möglichen Teil der Antwort. Es könnte sein, dass die Ungewissheit selbst eine Rolle bei der Produktion von Gewissheit spielt. Das heißt jedoch, dass wir uns von der hier auftretenden Paradoxie nicht stören lassen dürfen, sondern diese Paradoxie selbst für einen Hinweis auf die Sache halten müssen, nach der wir eigentlich fragen. Die Ungewissheit, so können wir sagen, produziert die Notwendigkeit einer Einschränkung. Diese Einschränkung liegt nicht im Algorithmus, der die Vorhersagen trifft, sondern in den Daten, die sich zur Vorhersehbarkeit ordnen, je vielfältiger und größer der Möglichkeitsraum ist, aus dem sie gewonnen werden. Was also ordnet die Daten zur Vorhersehbarkeit?

Die Antwort der Systemtheorie auf diese Frage ist bekannt. Ereignisse und damit auch die Daten, die sie darstellen, ordnen sich durch Systeme zu genau der Art von Vorhersehbarkeit, die das einzelne Ereignis, so ungewiss es ist, mit einer mit der Menge der Daten größer werdenden Wahrscheinlichkeit vorhersagbar macht. Systeme sind das Ergebnis einer Selbstorganisation, die durch die Komplexität des Ereignisraums zugleich unwahrscheinlicher und wahrscheinlicher wird. Mit Warren Weaver (1948) spricht man von einer organisierten Komplexität. Nicht der Algorithmus ordnet die Daten zur Vorhersehbarkeit, sondern die Ereignisse, von denen die Daten mehr oder minder zuverlässig künden, ordnen sich selbst.

¹ Die künstliche Intelligenz vom Typ des Maschinenlernens (Domingos 2015) ist von früheren Bemühungen um künstliche Intelligenz zu unterscheiden, die auf Verfahren der Symbolverarbeitung zurückgingen (Newell/Simon 1976). Maschinenlernen beruht auf zur Selbstkorrektur fähigen Gewichtungen der Einschätzung von Abhängigkeiten zwischen Wahrscheinlichkeiten, Symbolverarbeitung auf der Annahme, logische Schlussverfahren auf Maschinen abbilden und zur Problemlösung verwenden zu können.

2. Korrelation und Erklärung

Die prädiktiven Algorithmen der künstlichen Intelligenz vom Typ des Maschinenlernens ordnen Daten gemäß statistisch beschreibbaren Abhängigkeiten zwischen bestimmten Typen von Ereignissen¹. Bekannt ist die These, dass jede Theorie zur Erklärung der Ereignisse nicht nur überflüssig ist, wenn man aufgrund ihrer Korrelation Daten vorhersagen kann, sondern auch hinderlich wird, weil sie den unvoreingenommenen Blick auf die Daten durch Vorannahmen einschränkt (Anderson 2008). Aber mit dieser These wird übersehen,

dass den Algorithmen der künstlichen Intelligenz bestimmte Theorien über bestimmte Abhängigkeiten zwischen Ereignissen – und nicht nur Daten – bereits zugrunde liegen. Alex Pentland (2014, 2019) hat die Heuristiken beschrieben, die das Verhalten von Menschen mit einer bestimmten Wahrscheinlichkeit vorherzusagen erlauben, wenn man (a) das materielle und soziale Umfeld kennt, in dem sie sich bewegen, (b) ihr vorheriges Verhalten kennt und (c) eine Vorstellung davon hat, an welchen Zielen inklusive der Bemühung um Belohnung und Vermeidung von Bestrafung sich dieses Verhalten orientiert. Diese Heuristiken stehen im Einklang sowohl mit soziologischen und anthropologischen Einsichten in das Verhalten von Menschen im Kontext rivalisierender Imitation (Tarde 2009; Girard 1987) als auch mit ökonomischen Modellen zur Beschreibung von Optimierungsverhalten im Kontext von sozialer Interaktion und abrufbaren Kompetenzen (Becker 1974; Baecker 2013a).

2 In der Soziologie unterscheidet man praktisches, unter Umständen nicht interpretiertes Verhalten von sinnhaftem, auf Beobachter bezogenem Handeln und sinnhaftem, auf Selbstbeobachtung bezogenem Erleben. Diese Trinität von Bezugspunkten stützt jede konkrete Handlung mit jener basalen Komplexität aus, die schon deswegen mit »Freiheit« und »Notwendigkeit« assoziiert wird, weil die Möglichkeit der Selektion mit dem Zwang zur Selektion hier Hand in Hand geht (vgl. Schütz 1974; Parsons/Shils 1951; Luhmann 1984).

Vielfach glaubt man, dass den Menschen eine weitere Kränkung widerfährt, wenn man ihr Verhalten, Handeln und Erleben auf eine Maximierung bestimmter Zielvorstellungen im Kontext von Vergangenheit, also Pfadabhängigkeit, und Umfeld, also Orientierung, reduziert². Wo bleibt die Freiheit der Selbstbestimmung, wenn ein Mensch nichts anderes tut, als sein Umfeld zu erkunden und seine Chancen zu nutzen? Wird er dadurch nicht zur Ratte im Labyrinth? Man sollte diese Frage nicht zu schnell mit Ja oder Nein beantworten. Denn möglicherweise kommt man der Würde und Freiheit des Menschen nur auf die Spur, wenn man eine Vorstellung davon hat, in welchem Umfeld er sich bewegt und welche Chancen er nutzt. An diesem Punkt unserer Überlegungen kommt erneut der Faktor Ungewissheit ins Spiel. Je reicher das Feld der Möglichkeiten ist, in dem sich der Mensch bewegt, und je ungewisser der Erfolg ist, der bestimmten Absichten beschieden ist, desto intelligenter ist ein Verhalten, das aus dem Rückgriff auf Vergangenheit, also Erfahrung, und Umfeld, also Konkurrenz und Kooperation, eine gewisse Sicherheit gewinnt. Das gilt nicht zuletzt dann, wenn diese beiden Faktoren der Orientierung an Vergangenheit und Umfeld zusammen mit der Orientierung an bestimmten Zielen, also an einer noch ungewissen Zukunft, dazu genutzt werden, die jeweiligen Abhängigkeiten zu lockern und sich

von einer Vergangenheit zu lösen, weil die Gegenwart neue Chancen bietet, beziehungsweise zu bestimmten Umfeldern auf Distanz zu gehen, weil dort weder ein Verständnis für bestimmte Ziele noch eine Unterstützung für bestimmte Mittel, um diese Ziele zu erreichen, zu finden ist.

Die einfache, das menschliche Verhalten, Handeln und Erleben scheinbar reduzierende Heuristik einer Orientierung an Vergangenheit, Umfeld und Zielen steht nicht nur im Kontext einer nahezu nach Belieben steigerbaren Komplexität, sondern ist selbst die Voraussetzung des Gewinns einer Variabilität von Verhalten, Handeln und Erleben, die nichts anderes ist als die operationale Voraussetzung der Möglichkeit, von Freiheit und Würde sprechen zu können.

3. Heuristiken und Modelle der Datenverarbeitung

Der Zugriff der künstlichen Intelligenz des Maschinenlernens auf große Mengen an Daten setzt voraus, dass diesen Daten Ereignisse zugrunde liegen, über deren Abhängigkeit voneinander die Algorithmen der künstlichen Intelligenz Annahmen treffen. Diese Ereignisse, ich wiederhole diesen Punkt, werden nicht vom Algorithmus, sondern von den beteiligten Akteuren, seien sie menschlich oder künstlich, produziert. Die Ordnung der Algorithmen bezieht sich auf eine Ordnung, die ihnen vorausliegt und die ihrerseits durch die Ergebnisse der Arbeit der Algorithmen, das heißt durch die getroffenen Vorhersagen, beeinflusst wird. Die Abhängigkeit, in die sich das Verhalten, Handeln und Erleben von Menschen gegenüber den Ergebnissen von Algorithmen begeben, ist selbst Teil der Vergangenheit und des Umfelds, möglicherweise sogar der Zielvorstellungen, an denen sich die nächsten Selektionen möglichen Verhaltens, Handelns und Erlebens orientieren. Der Zugriff auf die Daten produziert verhaltensrelevante Ereignisse. Eben das darf dann »Digitalisierung« heißen: die rekursive Veränderung der Frequenzen möglichen Verhaltens durch die von Algorithmen produzierten Erkenntnisse über diese Frequenzen, insoweit sie in das Verhalten sei es der originären Akteure, sei es der sie überwachenden Akteure eingespielt werden.

Die Algorithmen der künstlichen, auf Maschinenlernen beruhenden Intelligenz lassen sich demnach sowohl in Richtung der Datenverarbeitung als auch in Richtung der

verwendeten Modelle und Heuristiken lesen. Mich interessieren hier und im Folgenden die Heuristiken und Modelle. Meine Annahme ist, dass die Konzentration auf statistische Korrelationen es nicht überflüssig macht, nach zugrunde liegenden Systemdynamiken zu fragen. So wichtig das Korrektiv der Entdeckung von Korrelationen ist, die durch keine Theorie vorhergesagt werden, so wichtig ist die Entwicklung oder das Nachjustieren von Theorien, die diese Korrelationen erklären oder auch nur beschreiben können. Die Erfolge der künstlichen Intelligenz dürfen nicht dazu führen, dass man die Absicht, eine wissenschaftlich überprüfbar und philosophisch reflektierte Übersicht über die Phänomene unserer Welt zu gewinnen, aufgibt.³

3 »Übersicht« im Sinne Ludwig Wittgensteins zielt darauf, Sprachspiele variieren zu können, die es erlauben, die »Grammatik« der Beschreibung von Phänomenen im »logischen Raum« dieser Beschreibung bestimmen zu können (Schulte 1989, S. 108f.).

Die Theorie liefert Metadaten, die in einer gewissen Konkurrenz zu den Metadaten der Statistik daran arbeiten, Zusammenhänge vorstellbar und überprüfbar zu machen (Baecker 2013b). Unser Ausgangspunkt ist die Annahme, dass es Systeme gibt, die das Problem der Ungewissheit von Weltzuständen durch die Einschränkung möglicher Weltzustände sowohl reduzieren als auch reproduzieren. Denn jede Einschränkung löst das Problem der Notwendigkeit einer Selektion zugunsten der Entstehung eines neuen Problems der Entdeckung einer möglichen Unbrauchbarkeit dieser Selektion. Systeme benötigen daher einen Index der Selbstgefährdung, um sich erfolgreich ausdifferenzieren und reproduzieren zu können. Diesen Index finden sie in der Grenze, die sie prekär von ihrer Umwelt unterscheidet, und reflektieren sie in zeitlichen Vorstellungen zum Gewinn möglicher Anschlussereignisse an aktuell selegierte oder auch vermiedene Ereignisse.

An die Stelle der Vorstellung einer künstlichen Intelligenz, die sich in einem unbestimmten Ereignisraum bewegt, setzen wir die Vorstellung einer Vielzahl von Systemen, die diesen Ereignisraum dadurch strukturieren, dass sie sich selbst in ihm bewegen und Ereignisse produzieren, die sich als Daten auslesen lassen. Wir zählen nicht nur bis eins, nämlich bis zum technischen System der lernenden Maschine, sondern bis fünf oder sechs, um menschliches Verhalten, Handeln und Erleben ordnen, und bis unendlich, um es in allen seinen Einzelfällen zählen zu können.⁴

4 Dass das Einzelne in einem unauflösbaren Spannungsverhältnis sowohl zum Allgemeinen als auch zum Besonderen steht, beschäftigt die Philosophie, Logik und Systemtheorie seit Aristoteles, Nikolaus von Kues, Hegel und Ludwig von Bertalanffy (Baecker 2019).

4. Fachwissenschaften und Supertheorien

Die verschiedenen Wissenschaften, mit denen wir es im Umkreis der Bestimmung und Beschreibung menschlichen Verhaltens, Handelns und Erlebens zu tun haben, sind alle- samt dadurch stark geworden, dass sie maximal »bis eins« zählen und unter zwei, drei und vier allenfalls die Konkurrenz alternativer Modelle zählen, die nach Möglichkeit auf das eigene Modell reduziert werden sollten. Für die Biologie zählen nur Organismen, für die Neurowissenschaften nur das Gehirn, für die Psychologie nur das Bewusstsein, für die Soziologie nur die Gesellschaft und für die Informatik nur die Maschine. Die interessantesten Theoreme entwickeln diese Wissenschaften zwar immer dann, wenn sie an den Grenzen ihrer Phänomene und Problemstellungen auf Voraussetzungen stoßen, die außerhalb der Phänomene und Problemstellungen liegen, doch führt dies allenfalls dazu, die Problemstellung der Ausdifferenzierung und Reproduktion der Phänomene zu schärfen, jedoch nicht dazu, die Phänomene und Problemstellungen aufeinander zu beziehen. Dazu hätte man einen »logischen Raum« (Wittgenstein 1963) der Verschaltung von Kontexturen benötigt, der in der philosophischen Reflexion der Kybernetik von Gotthard Günther (1979; vgl. Klagenfurt 1995) unter dem Namen der Polykontextualität zwar vorgeschlagen wurde, jedoch als Programm zur Erforschung materieller Kontexturen in ihren wechselseitigen Abhängigkeiten bis heute ungenutzt geblieben ist.

Zwar gibt es Supertheorien – Theorien, die auch sich selbst beschreiben und erklären – wie die Rhetorik, die Mathematik, die Hermeneutik, die Semiotik, die Kybernetik oder die Informatik, die Begrifflichkeiten bereitstellen, die sich in allen Fachwissenschaften einsetzen lassen, aber vielfach verlieren sich mögliche wechselseitige Anregungen auf Einbahnstraßen, ohne systematisch ausgewertet zu werden. Gegenwärtig pflegt nur die Systemtheorie ein Wissen um die Notwendigkeit, verschiedene Systemreferenzen aufeinander zu beziehen, um konkrete Ereignisse beschreiben und erklären zu können.

Ich beschränke mich im Folgenden auf menschliches Verhalten, Handeln und Erleben, rechne jedoch damit, dass sich die Einschränkung auf menschliche Akteure im weiteren Verlauf der Entwicklung künstlicher Intelligenz

aufheben lässt und ein Begriff von Verhalten, Handeln und Erleben möglich wird, der auch Maschinen als mögliche Akteure registriert. So oder so ist ein Handlungsbegriff, der nur Menschen kennt und Götter, Geister, Tiere und Pflanzen ausschließt, allenfalls das Ergebnis einer humanistischen Selbstüberschätzung. Und so oder so wird die Systemtheorie im Umkreis einer Kybernetik formuliert, die mit ihren Grundbegriffen der Kommunikation und Kontrolle längst eine Begrifflichkeit bereitstellt, die sich auf jeglichen Systemtyp anwenden lässt, in dem Komplexität durch Selbstorganisation bearbeitet wird (Wiener 1961).

5. Fünf bis sechs Systeme

Bis auf Weiteres zählen wir bis fünf oder sechs. Wir »dekonstruieren« den Menschen in (1) Organismus, (2) Gehirn, (3) Bewusstsein, (4) Gesellschaft, (5) Technik und (6) Kultur. Verhalten, Handeln und Erleben ist ein Produkt der Differenz und Synchronisation organischer, neuronaler, mentaler, sozialer, technischer und kultureller Systeme. Die ersten fünf Systemtypen sind durch umfangreiche Forschungsprogramme der Biologie, Neurowissenschaften, Psychologie, Soziologie und Informatik belegt; die Existenz kultureller Systeme jedoch ist umstritten (Kroeber/Parsons 1958; Baecker 2004).

Jedes dieser Systeme bewältigt Komplexität durch eine Selbstorganisation, die minimal in den beiden Dimensionen der Ausdifferenzierung aus und in einer Umwelt und der Reproduktion in der Zeit bestimmt ist. Je nachdem, wie man kulturelle Systeme etwa durch die Ausdifferenzierung und Reproduktion von Werten und Normen bestimmt, ist unklar, ob diese unabhängig von der Selbstorganisation sozialer Systeme, in denen Werte und Normen die von Talcott Parsons beschriebene Funktion der Erhaltung latenter Muster spielen (Parsons/Shils 1951, S. 147ff.), gedacht werden können. Das soll uns hier jedoch nicht beschäftigen.

Interessanter ist die Frage der Bestimmung des von Algorithmen der künstlichen Intelligenz erforschten Datenraums als Ereignisraum im Wechselspiel dieser fünf bis sechs Systeme. Bewusst spreche ich von Systemtypen, da es auf der Hand liegt, dass je nach Tiefenschärfe der Analyse organische, neuronale, mentale, soziale und technische

Systeme vielfach in Subsysteme differenziert oder auch als Containerbegriffe für die Synchronisation eigener Systemtypen verstanden werden können. Hinzu kommt, dass Systemtypen auf der einen Seite von konkreten Systemen auf der anderen Seite unterschieden werden müssen, deren jeweiliger Modus der Ausdifferenzierung und Reproduktion zu hochgradig idiosynkratischen, singulären Ausprägungen und Verkörperungen dieser Systeme führen kann. Auf dieser anderen Seite zählen wir bis unendlich.⁵

⁵ Genauer: Wir treffen unendliche Urteile im Sinne Kants (1968, B 97f.; vgl. Heinrich 1987, S. 202ff.), die noch unbestimmte Möglichkeiten bejahen, indem sie bestimmte Möglichkeiten verneinen.

Im Raum statistischer Wahrscheinlichkeiten kann man das eine vom anderen unterscheiden. Man kann Systemtypen beschreiben, die auf erwartbare Weise ein konkret unerwartetes Verhalten produzieren. Genau darauf beschränke ich mich hier. Und noch eine weitere Einschränkung ist erforderlich. Zu jedem der genannten Systemtypen gibt es eine unübersehbare Literatur, die auch nicht ansatzweise von einem einzelnen Wissenschaftler oder auch nur einem einzigen Forschungsprojekt überschaut werden kann. Ich muss mich daher auf mehr oder minder zufällige, glücklich oder unglücklich gefundene Versatzstücke beziehen, wenn ich hier den Versuch mache, eine erste Vorstellung von der unterschiedlichen Systemtypik zu vermitteln. So oder so verstehe ich die hier vorgelegten Überlegungen als Modell oder Paradigma möglicher Forschungsprogramme, die in jedem einzelnen Fall von meinen Ausgangs- und Bezugspunkten abweichen werden.

Wichtig ist mir die Annahme der Inkommensurabilität der Systeme. Sie differieren nicht nur systematisch gegenüber ihrer Umwelt, wenn nicht sogar »emergent« gegenüber ihren eigenen Systemelementen (Neumann 1956), sondern sie differieren auch untereinander auf eine Art und Weise, die es unmöglich macht, sie aufeinander zu reduzieren. Eben das verstehe ich unter »Komplexität«. Komplexe Phänomene sind solche, deren Konstitution und Konditionierung verschiedene Systeme zu ihrer Voraussetzung haben, die sich wechselseitig fordern, ohne aufeinander reduziert werden zu können.

Organische Systeme sind homöostatisch irritabel; sie halten eine interne Umwelt in der Auseinandersetzung mit einer externen Umwelt aufrecht, indem sie laufend Störungen abbauen (Cannon 1929). Neuronale Systeme sind prädiktiv,

indem sie Erwartungen durch Erfahrungen testen und durch Lernen korrigieren (Frith 2007). Mentale Systeme imaginieren, indem sie endliche Erfahrungen unendlich kontextualisieren (Lacan 1966; Hegel 1990, S. 136f.). Soziale Systeme produzieren doppelte Kontingenz, indem sie für den Aufbau von Beziehungen der Abhängigkeit die Existenz unabhängiger Akteure fordern (Luhmann 1984, S. 153ff.). Technische Systeme rechnen, und dies tendenziell nicht-trivial (Nake 2001; Foerster 1993). Und kulturelle Systeme, wenn man denn von ihnen sprechen will, symbolisieren die Einheit des Sinns in der Vielfalt der Ereignisse (Cassirer 2009).

Wie immer diese ersten Ausgangspunkte sich bewähren beziehungsweise auszuarbeiten sind, unterstreichen sie doch den wesentlichen Punkt, dass die Intelligenz dieser Systeme untereinander unvergleichbar ist, auf keine Physik oder Logik Desselben reduziert werden kann und in ihrem orthogonalen Zusammenspiel ernst genommen werden muss, wenn man konkrete Phänomene untersuchen und verstehen können will.

»Intelligent« ist das Verhalten jedes einzelnen dieser Systeme in dem Ausmaß, in dem sie ihr Problem nicht nur lösen, sondern zugleich reproduzieren. Sie sorgen für ihren eigenen Nachschub, da sie andernfalls ihre Anhaltspunkte in ihrer Umwelt verlieren und keine Anlässe für ihre Reproduktion im Zeitablauf finden würden. Problem und Problemlösung zusammen spannen jenen heuristischen Rahmen, in dem »angemessene Selektionen« (Ashby 1981) eines möglichen Verhaltens vorgenommen werden können. Die Intelligenz steigt mit der Selektivität, das heißt mit der Menge an Möglichkeiten, die in Rechnung gestellt wird, wenn aus ihr eine spezifische Selektion gewählt wird. Im Rahmen einer kybernetischen Terminologie verschiebt sich das Problem damit von »Intelligenz« im Sinne optimierenden Suchverhaltens auf »Kontrolle« im Sinne eines Integrals, das die Autonomie des Systems in ein Verhältnis zur Komplexität der Umwelt setzt. Vieles deutet darauf hin, dass künstliche Systeme die Lösung dieses Kontrollproblems noch vor sich haben (Dyson 2019).

Umgekehrt ist es ebenfalls noch nicht gelungen, das Kontrollproblem für den Zusammenhang organischer, neuronaler, mentaler und sozialer Systeme zu formulieren, zumal

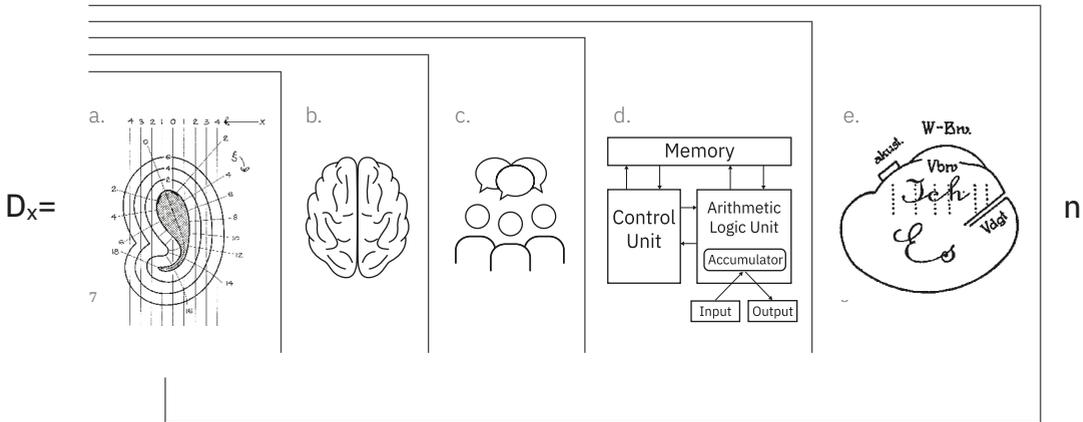
es trotz aller Kognitionswissenschaften keine Schule oder Forschungstradition gibt, die mit diesem Zusammenhang routiniert rechnet. Talcott Parsons' (1978) Hinweise im Rahmen seines »Paradigmas der menschlichen Bedingung«, das diese Referenzen sowohl unterschied als auch aufeinander bezog, wurden nicht weiterverfolgt, zumal das Vierfelderschema seiner Handlungstheorie die zu untersuchenden Sachverhalte zu stark zu schematisieren schien. Erst jüngst scheint der Indikationenkalkül von George Spencer-Brown (2008) eine Handhabe zu bieten, Parsons' Schema zu flexibilisieren und einen Rahmen der Formulierung von Formen des Umgangs mit Unterscheidungen zu bieten, der für kognitive Systeme jeglicher Art getestet werden kann (Luhmann 1980, 2017). Kognitive Systeme in den Medien des Lebens, des Bewusstseins, der Kommunikation und der Technik können hier als Systeme verstanden werden, die ihre Form aus dem Treffen von Unterscheidungen gewinnen, die Bestimmtes im Kontext von Unbestimmtem markieren und die Relation zwischen Bestimmtem und Unbestimmtem zur Reflexion ihrer Operationen inklusive eines Aufbaus eines Gedächtnisses nutzen. Eines der im Rahmen dieses Konzepts zu lösenden Theorieprobleme besteht darin, dass die Relation zwischen Bestimmtem und Unbestimmtem medial verstanden werden muss, etwa im Sinne der losen Kopplung zwischen Elementen, in die Formen geprägt werden können (Heider 2005), und auch dieser Medienbegriff noch keine angemessene interdisziplinäre Diskussion gefunden hat.

6 Eine weitere Spur legt der Begriff der »medialen Form« (Derrida 2004a, S. 119), der eine Operation beschreibt, die insofern keine ist, als sie im Sinne eines Integrals aus Autonomie und Kontrolle eher »spielt« als rechnet.

Ludwig Wittgensteins (1963, Satz 3.42) Begriff des »logischen Raums« markiert die zu lösenden Theorieprobleme trefflich; deswegen beschränke ich mich im vorliegenden Zusammenhang darauf, unter diesem Begriff festzuhalten, wie ein weiteres Forschungsprogramm eventuell vorzugehen hätte.⁶

- a = Organismus
- b = Gehirn
- c = Gesellschaft
- d = Rechner
- e = Bewusstsein
- n = unmarked state

Den logischen Raum, in dem ein Ereignis X produziert und reproduziert wird, das als Datum D zum Gegenstand der Berechnung durch einen Algorithmus gemacht werden kann, kann man mithilfe der Notation des Indikationenkalküls wie folgt notieren:



7 Foerster, Heinz von. 2003. Thoughts and Notes on Cognition, S. 169–189 in: ders., Understanding Understanding. Essays on Cybernetics and Cognition. New York: Springer. (Grafik aus Seite 179).

8 Freud, Sigmund. 2009. [1923]. Das Ich und das Es, S. 251-295 in: ders., Das Ich und das Es: Metapsychologische Schriften, Frankfurt am Main: Fischer Taschenbuchverlag, hier: S. 265

Der Indikationenkalkül ordnet unsere Systemreferenzen innerhalb einer Form mit gestaffelt tiefen Räumen, die jeweils durch eine Unterscheidung, ein cross, generiert werden, die die Innenseite der Unterscheidung in ein Negations- und Implikationsverhältnis zur Außenseite setzt. Der Organismus steht im tiefsten und damit durch die meisten Unterscheidungen determinierten Raum, womit angedeutet werden soll, dass er die meisten Einschränkungen setzt. Er steht für die einem Menschen oder einem anderen hinreichend komplexen Akteur erreichbare materielle Praxis der Auseinandersetzung mit einer realen, symbolischen und imaginären Umwelt. Im flachsten Raum steht das Bewusstsein, das somit den größten Spielraum zur Determination seiner Zustände hat. Unter einem unwritten cross, der nicht ausgeschrieben Quere, steht der unmarked state, hier mit n markiert, der die ungeordnete Außenseite der Form darstellt und somit auf die notwendige Ergänzungsbedürftigkeit der Bestimmung der Form durch ihre unbekannteren, aber mitlaufenden Voraussetzungen hinweist.

Das re-entry der Unterscheidung zwischen Bewusstsein und unmarked state führt die gesamte Form in der Unterscheidung des Organismus in die Form wieder ein und

schließt sie somit zur Form einer Ausdifferenzierung und Reproduktion als Eigenwert einer rekursiven Funktion.

Auch die Reihenfolge der Ordnung von Organismus, Gehirn, Gesellschaft, Maschine und Bewusstsein ist hier paradigmatisch, das heißt exemplarisch gemeint. Für jede Analyse eines konkreten DX wird man erneut untersuchen müssen, wie und in welcher Abhängigkeit voneinander die einzelnen Unterscheidungen ihre Form generieren. Der Indikationenkalkül zählt und ordnet nichts anderes als Abhängigkeiten zwischen unabhängigen, weil sich gegenseitig sowohl implizierenden als auch negierenden, Einheiten. Kausalitäten werden damit nicht ausgeschlossen, doch sind sie das Material, in dem sich Unabhängigkeiten bewähren und Abhängigkeiten selegiert werden, nicht ihrerseits die Determinanten der Form. Den logischen Raum der Form kann man als polykontextural verstehen, in dem jede konkrete Vermittlung nicht mehr mithilfe einer zweiwertigen, sondern nur nach einer mehrwertigen Logik beschrieben werden kann.⁹

⁹ Man bekommt es mit einer Proliferation von quaternii terminorum (syllogistisch unzulässigen Vermittlungsbegriffen; vgl. Heinrich 1987, S. 188), suppléments, das heißt alles andere als überflüssigen Komplikationen (Derrida 2004b), oder auch Parasiten zu tun, deren Ausbeutung von Relationen weitere Relationen herstellt (Serres 1981).

DX wird von einem Beobachter als ein Datum bestimmt, das aus der Perspektive jeder genannten Systemreferenz eine eigene Identität besitzt. Es ist ein zugleich organisches, neuronales, mentales, soziales und technisches Datum, das im Horizont jeder Systemreferenz anders verarbeitet wird, das heißt andere Voraussetzungen und Folgen, andere frühere und Anschlussereignisse vernetzt. Jedes DX ist ein Ereignis, das Systeme für exakt den Moment synchronisiert, in dem es auftritt. Gleich anschließend verfolgen die Systeme ihre eigene Systemdynamik, womit die Möglichkeit von synchronisierenden Folgeereignissen gerade nicht ausgeschlossen, sondern ermöglicht wird. Mit Niklas Luhmann, der dies für soziale und psychische Systeme gezeigt hat (Luhmann 1984, S. 382ff.; 1985; 1990), kann man daher vorschlagen, sich jedes dieser Systeme im Modus des selbstproduzierten Zerfalls vorzustellen, das heißt in einem temporalen Modus, der in jedem Moment neue Anschlüsse erfordert. Ihre Reproduktion ist ein Ergebnis ihrer Selbstgefährdung. Und ergänzen kann man dies durch ein Netzwerkkalkül à la Harrison C. White (1992), in dem jede gefundene Verknüpfung zugleich als unzuverlässig konnotiert wird, so dass erst die konkrete Arbeit an Identität und Kontrolle eine gewisse strukturelle Belastbarkeit herstellt.

6. Produktive Komplexität

Blicken wir zurück auf die statistische Trivialität und strukturelle Paradoxie, dass mit einer steigenden Zahl an Daten die Zuverlässigkeit der Prognose trotz größerem Ereignisraum steigt, so erkennen wir, dass für diesen Umstand in der Tat jener Faktor der Ungewissheit verantwortlich ist, der die beteiligten Systeme, verstanden als Ereignisproduzenten, dazu zwingt, sich an die Umstände anzupassen, die sie vorfinden, da sie andernfalls keine Umstände mehr vorfinden würden, an die sie sich anpassen könnten.

Daraus kann eine Schlussfolgerung gezogen werden, die auch für den aktuellen Stand der KI-Diskussion maßgeblich ist. Die requisite variety, mit der sich Systeme ausstatten müssen, die sich in komplexen Umwelten ausdifferenzieren und reproduzieren wollen (Ashby 1965), macht diese Systeme im Einzelfall unvorhersehbarer und in der Menge vorhersehbarer. Letztlich macht die Autonomie, die sie unberechenbar macht, sie zugleich berechenbar, weil ihr Verhalten, Handeln und Erleben sich erst jetzt aus der Vergangenheit, der Umwelt und möglichen Zielen errechnet – und entsprechend aus ihnen errechnet werden kann –, die ihnen ihre Orientierung liefert.

Organismen, Gehirne, Bewusstsein, Gesellschaft, Technik und Kultur erzeugen füreinander eine Komplexität, die nur und ausschließlich durch eine Synchronisation reduziert werden kann, die jedes der beteiligten Systeme vor eine für sie maximale Voraussetzung stellt. Mechanismen wie Krankheit, Störung, Wahn, Anomie, Fehler und Dekadenz helfen dabei, das Verhaltensspektrum sowohl auszuloten als auch Zustände zu identifizieren und isolieren, die nicht durchzuhalten sind.¹⁰ In jedem Moment wird von allen beteiligten Systemen die Ablehnung der eigenen Operationen vorweggenommen und im Medium der Bewertung dieser Ablehnung entweder angenommen oder abgelehnt. Letztlich stiftet nur eine Negativität, die sowohl Autonomie als auch Bezug sicherstellt, einen Zusammenhang, auf den man sich verlassen kann.

Für die Modellierung von Negativität kommt neben dem digitalen auch das analoge Rechnen wieder zu Ehren, ersteres im Medium von Codes und Programmen, letzteres im Medium von Kontrolle und Widerspruch (Watzlawick/Beavin/Jackson 1969, S. 61ff.; Dyson 2019).

¹⁰ Hier öffnet sich ein weites Forschungsfeld. Siehe für ein eindrucksvolles Beispiel Erving Goffmans (1956) Analyse der Peinlichkeit und des zugleich beschämenden und übergehenden Umgangs mit ihr als Mechanismus nicht nur der Korrektur, sondern des Aufbaus einer sozialen Ordnung.

Unergründlich ist die Differenz der Systeme, die orthogonal zueinander stehen und auf keine Identität und Kontinuität zurückgerechnet werden können. Ergründlich ist die Erwartung, dass aus dieser Unergründlichkeit eine Autonomie gewonnen werden kann, die die Systeme fallweise verlässlich aufeinander bezieht. Je freier die Algorithmen der künstlichen Intelligenz operieren können, desto mehr Spielraum und Grund haben sie, sich auf das Material zu beziehen, das sie in lebenden, neuronalen, psychischen und sozialen Systemen vorfinden.

Vielleicht bleibt der Status kultureller Systeme mit Absicht unklar. Die Bewertung dieser neuen Lage der Menschheit muss erst noch gefunden werden. So oder so jedoch dürfte deutlich geworden sein, dass die Frage nach der Künstlichen Intelligenz als einem »gesellschaftlichen Ordnungssystem« nur mit größter Vorsicht beantwortet werden kann. Wir sind Lehrlinge einer neuen Form von Komplexität. Während wir immer bessere Maschinen konstruieren können, die kurz davor stehen, uns ihren Entwurf aus der Hand zu nehmen, stehen wir erneut vor dem Rätsel Mensch mit seiner evolutionär einzigartigen Kombination von organischer, neuronaler, mentaler und sozialer Intelligenz.

-
- Anderson, Chris. 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*. June 23, 2008.
- Ashby, W. Ross. 1958. Requisite Variety and Its Implications for the Control of Complex Systems. *Cybernetica* 1(2): 83–99.
- Ashby, W. Ross. 1981. What Is an Intelligent Machine? In *Mechanisms of Intelligence: W. Ross Ashbys Writings on Cybernetics*, hrsg. Roger Conant, 295–306. Seaside, CA: Inter-systems.
- Baecker, Dirk. 2004. Kulturelle Orientierung. In *Luhmann und die Kulturtheorie*, hrsg. Günter Burkart und Günter Runkel, 58–90. Frankfurt am Main: Suhrkamp.
- Baecker, Dirk. 2013a. Artificial Paradise Revisited. In *Stabile Fragilität – fragile Stabilität*, hrsg. Stephan Jansen, Eckhard Schröter und Nico Stehr, 25–39. Wiesbaden: Springer VS.

- Baecker, Dirk. 2013b. Metadaten: Ein Annäherung an Big Data. In *Big Data – Das neue Versprechen der Allwissenheit*, hrsg. Heinrich Geiselberger und Tobias Moorstedt, 156–186. Berlin: edition unseld.
- Baecker, Dirk. 2019. A Note on Ludwig von Bertalanffy and the Form Problem of Life. *Systems Research and Behavioral Science* 36(1): 1–10.
- Becker, Gary S. 1974. A Theory of Social Interactions. *Journal of Political Economy* 82: 1063–1093.
- Cannon, Walter B. 1929. Organization for Physiological Homeostasis. *Physiological Reviews* 9(3): 399–431.
- Cassirer, Ernst. 2009. *Schriften zur Philosophie der symbolischen Formen*, hrsg. Marion Lauschke. Hamburg: Meiner.
- Derrida, Jacques. 2004a. Die différance. In *Die différance: Ausgewählte Texte*, hrsg. Peter Engelmann, 279–333. Stuttgart: Reclam.
- Derrida, Jaques. 2004b. Unterwegs zu einer Ethik der Diskussion. In *Die différance: Ausgewählte Texte*, hrsg. Peter Engelmann, 279–333. Stuttgart: Reclam.
- Domingos, Pedro. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books.
- Dyson, George. 2019. The Third Law. In *Possible Minds: Twenty Five Ways to Look at AI*, hrsg. John Brockman, 35–40. New York: Penguin Press.
- Foerster, Heinz von. 1993. Prinzipien der Selbstorganisation im sozialen und betriebswirtschaftlichen Bereich. In *Wissen und Gewissen: Versuch einer Brücke*, hrsg. Siegfried J. Schmidt, 233–268. Frankfurt am Main: Suhrkamp.
- Frith, Chris. 2007. *Making Up the Mind: How the Brain Creates Our Mental Worlds*. London: Blackwell.
- Girard, René. 1987. *Das Heilige und die Gewalt*. Zürich: Benziger.
- Goffman, Erving. 1956. Embarrassment and Social Organization. *American Journal of Sociology* 62: 264–271.
- Günther, Gotthard. 1979. Life as Poly-Contextuality. In *Beiträge zur Grundlegung einer operationsfähigen Dialektik*, Bd 2, 283–306. Hamburg: Meiner.
- Hegel, Georg Wilhelm Friedrich. 1990. *Wissenschaft der Logik: Die Lehre vom Sein*. Hamburg: Meiner.
- Heider, Fritz. 2005. *Ding und Medium*. Berlin: Kulturverlag Kadmos.
- Heinrich, Klaus. 1987. *tertium datur: Eine religionsphilosophische Einführung in die Logik*. 2., verb. Aufl. Basel: Stroemfeld.
- Kant, Immanuel. 1968. *Kritik der reinen Vernunft*. Frankfurt am Main: Suhrkamp.
- Klagenfurt, Kurt. *Technologische Zivilisation und transklassische Logik: Eine Einführung in die Technikphilosophie Gotthard Günthers*. Frankfurt am Main: Suhrkamp.
- Kroeber, A. L., und Talcott Parsons. 1958. The Concepts of Culture and of Social System. *American Sociological Review* 23(5): 582–583.
- Lacan, Jacques. 1966. Le stade du miroir comme formateur de la fonction du Je. In *Écrits*, 89–97. Paris: Le Seuil.

- Luhmann, Niklas. 1980. Talcott Parsons – Zur Zukunft eines Theorieprogramms. *Zeitschrift für Soziologie* 9(1): 5–17.
- Luhmann, Niklas. 1984. Soziale Systeme: Grundriss einer allgemeinen Theorie. Frankfurt am Main: Suhrkamp.
- Luhmann, Niklas. 1985. Die Autopoiesis des Bewusstsein. *Soziale Welt* 36(4): 402–446.
- Luhmann, Niklas. 1990. Gleichzeitigkeit und Synchronisation. In *Soziologische Aufklärung 5: Konstruktivistische Perspektiven*, 95–130. Opladen: Westdeutscher Verlag.
- Luhmann, Niklas. 2017. Die Kontrolle von Intransparenz. In *Die Kontrolle von Intransparenz*, hrsg. Dirk Baecker, 96–120. Berlin: Suhrkamp.
- Nake, Frieder, 2001. Das algorithmische Zeichen. In *Wirtschaft und Wissenschaft in der Network Economy: Visionen und Wirklichkeit. Informatik 2001: Tagungsband der GI/OCG 2001*, Bd. II, hrsg. Kurt Bauknecht, Wilfried Brauer und Thomas A. Mück, 736–742. Wien: Universität Wien.
- Neumann, John von. 1956. Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. In *Automata Studies*, hrsg. Claude E. Shannon, John McCarthy, 43–98. Princeton, NJ: Princeton University Press.
- Newell, Allen, und Herbert A. Simon. 1976. Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM* 19: 113–126.
- Parsons, Talcott. 1978. A Paradigm of the Human Condition. In *Action Theory and the Human Condition*, 352–433. New York: Free Press.
- Parsons, Talcott, und Edward A. Shils (Hrsg.). 1951. *Toward a General Theory of Action*. Cambridge, MA: Harvard University Press.
- Pentland, Alex. 2014. *Social Physics: How Good Ideas Spread – The Lessons from a New Science*. New York: Penguin Press.
- Pentland, Alex. 2019. The Human Strategy. In *Possible Minds: Twenty-Five Ways of Looking at AI*, hrsg. John Brockmann, 194–205. New York: Penguin Press.
- Schulte, Joachim. 1989. *Wittgenstein: Eine Einführung*. Stuttgart: Reclam.
- Schütz, Alfred. 1974. *Der sinnhafte Aufbau der sozialen Welt: Eine Einleitung in die verstehende Soziologie*. Frankfurt am Main: Suhrkamp.
- Serres, Michel. 1981. *Der Parasit*. Frankfurt am Main: Suhrkamp.
- Spencer-Brown, George. *Laws of Form*. 5. Aufl. Leipzig: Bohmeier.
- Tarde, Gabriel. 2009. *Die Gesetze der Nachahmung*. Frankfurt am Main: Suhrkamp.
- Watzlawick, Paul, Janet H. Beavin und Don D. Jackson. 1969. *Menschliche Kommunikation: Formen, Störungen, Paradoxien*. Bern: Huber.
- Weaver, Warren. 1948. Science and Complexity. *American Scientist* 36(4): 536–544.
- White, Harrison C. 1992. *Identity and Control: A Structural Theory of Action*. Princeton, NJ: Princeton University Press.
- Wiener, Norbert. 1961. *Cybernetics, or Control and Communication in the Animal and the Machine*, 2. Aufl. Cambridge, MA: MIT Press.
- Wittgenstein, Ludwig. 1963. *Tractatus logico-philosophicus*. Frankfurt am Main: Suhrkamp.





Digit[ale] Dinge: ordnende Finger in unserer Gesellschaft

Astrid Schwarz

Mit dem Symposium »(Un)ergründlich. Künstliche Intelligenz als Ordnungsstifterin« im Oktober 2018 wurde ein ganzes Tableau an Fragen ausgebreitet um zu sondieren, wodurch sich Bedingungen und Merkmale gesellschaftlicher Ordnungssysteme wie Religion, Staat und Technik unterscheiden und wie es dabei um die »Potentiale von Künstlicher Intelligenz (Technik) zur Gestaltung der Gesellschaft« steht. Die eröffnende Session I, zu der auch die Autorin beitragen durfte, lud ein sich mit der These auseinanderzusetzen »Was uns zusammenhält - KI als gesellschaftliches Ordnungssystem«, die Spielregeln wurden durch Fragen danach spezifiziert, »(w)ie Künstliche Intelligenz gesellschaftliche Probleme lösen (kann)« und »(w)elche gesellschaftlichen Auswirkungen ein Ordnen der Gesellschaft durch KI (hat)«. ¹

¹ <https://www.oeffentliche-it.de/unergruendlich> [Letzter Zugriff: 1. September 2019].

In diesem iterativen Frage-und-Antwort-Spiel kommt die Unbekannte »gesellschaftliche Gestaltung« in zwar immer neuer Kombinatorik vor, aber das Rätsel scheint bereits gelöst: Künstliche Intelligenz schafft Ordnung, Gesellschaft ist Ordnung, KI macht Gesellschaft. Zugespitzt könnte dies gelesen werden als eine Art Automatismus, eine vorausseilend akzeptierte Überwältigungsfigur: die Maschinenteknik des Digitalen schreibt sich ein in gesellschaftliche Ordnungen und Verfahren. Damit würde aber der behauptete Gestaltungsspielraum von vorneherein enorm eingeschränkt und die mittlerweile verbreitete Kritik berechtigt, von Harald Welzer jüngst treffend auf den Punkt gebracht, daß »(i)m Augenblick in Sachen Digitalisierung der Schwanz mit dem Hund (wedelt)«. ² Es muss also darum gehen zu fragen, wie wir diese Technik in Gebrauch nehmen wollen, ja, schon längst genommen haben. Wie greifen die »digits« konkreter digitaler Dinge und Prozesse ein in gesellschaftliche Prozesse und vor allem auch, wo trifft die digitale Technik auf gesellschaftliche Bedürfnisse und für welche wünschbaren Gebrauchsweisen ist sie einsetzbar. Kurzum, nicht »KI macht Gesellschaft«, sondern »Gesellschaft macht KI«.

² Und weiter heißt es: »Sich von Algorithmen vorschreiben zu lassen, wie man leben soll, ist der Wiedereintritt des Menschen in die selbst verschuldete Unmündigkeit«, so Harald Welzer in seinem rasanten, beißenden und klugen Essay (2019, S. 6).

Und eben dies, »Gesellschaft macht KI«, möchte ich nachfolgend durch eine Handvoll objektzentrierter Geschichten belegen. Mein Vorschlag ist auf die These zu reflektieren, dass es die Dinge sind, die Ordnung schaffen und sich genauer anzusehen in welchen Konstellationen und auf welche Weise sie das tun. Dieses Unternehmen, den Dingen zu folgen die Ordnung schaffen, klingt im ersten Anlauf möglicherweise eher lebensweltlich profan, erinnert an Ratgeberliteratur und mahnende Worte elterlicher Fürsorge, dass Ordnung im Schrank auch Ordnung im Leben schaffe und ähnliches. In der wissenschaftlichen Welt wird vielleicht Michel Foucaults Ordnung der Dinge assoziiert werden mit seiner These zur Unhintergebarkeit des Verhältnisses von Worten und Dingen in einer historischen Epoche, die eigene Episteme unergründlich für die Zeitgenossen machend (Foucault 1989). Die nachfolgenden Überlegungen zu Dingen im Alltag umspielen beide Assoziationen, ich interessiere mich für die Lebensformen durch und mit den digitalen Dingen. Es wird um konkrete Handlungen an diesen speziellen Dingen gehen, um den sprachlichen Umgang mit ihnen, die Erwartungen an sie und auch darum, was vielleicht gar nicht so speziell an ihnen ist, sondern sie als ganz gewöhnliche Gebrauchsgegenstände kennzeichnet in einer analogen Welt.

Die meisten von uns umgeben sich mit vielen Dingen³ und verbringen viel Zeit damit diese Dinge ein-, um-, zu- oder anzuordnen, das gilt für analoge wie für digitale Dinge. Dinge werden in Gebrauch genommen, weil sie bestimmte Funktionen oder Bedürfnisse im Alltag erfüllen, weil sie individuelle Erfahrungen ermöglichen und in diesem Sinne individuelle Lebenswelten hervorbringen, selbst wenn es sich um Massenprodukte handelt. Dinge können aber auch per gesellschaftliche Verordnung begegnen, etwa durch den Zwang mit einem Ticketautomaten zu kommunizieren, anstatt wie gewohnt mit einer Person hinter dem Schalter. Dinge können als Abordnung der Zukunft begegnen, beispielsweise in Diskussionen um autonom fahrende Vehikel oder in Gestalt von Pflegerobotern. Schließlich ordnen uns Dinge auch zu, zu einer bestimmten Gesellschaft, einer Interessensgruppe, einer historischen Epoche. Weltbildend und ordnungsschaffend sind also die in Gebrauch genommenen und in gewisser Weise auch die künftigen Gebrauchsgegenstände.

3 Eine Person in Europa besitze im Durchschnitt etwa 60.000 Gegenstände, heißt es im Blog »BesserFernsehen« (Schaffner 2012).

Aktuell sind Personal Computer und Smartphone allgegenwärtig und selbstverständlich in Gebrauch genommen von einer großen gesellschaftlichen Mehrheit. Sie sind fester Bestandteil eines Medienwandels, der von Marshall McLuhan (1964) als ein Kulturbruch thematisiert wurde. Er fing an über Technologien grundsätzlich als environments nachzudenken und stellte damit das Verhältnis von natürlich und künstlich im bisherigen Verständnis des Umweltbegriffs auf den Kopf (Sprenger 2019, S. 255). Die Konsequenzen dieses radikalen Wandels im Technik- und Medienverständnis wurden vielfach auch als »ökologisch« beschrieben, Folge sind Diskontinuitäten in Zeit- und Raumverhältnissen, andere Beziehungsgefüge von Menschen und Dingen und letztlich Transformationen gesellschaftlicher Verhältnisse. Bernhard Pörksen (2016) meint, dass dieser Kulturbruch auf »oft undeutliche, ebendeshalb erst in der Anstrengung der bewussten Reflexion erkennbare Weise alles (verändert)«. Verschoben werden soziotechnische Grenzen, etwa jene zwischen öffentlich und privat, zwischen verfügbarem und geheimen Wissen, zwischen Oberflächenfunktionen und technologischer Tiefenwirkung beispielsweise im therapierten Körper oder bei Ernährungsgewohnheiten, zwischen erkennbar gewordener und gemachter Natur, zwischen mittelbaren und unmittelbaren sozialen Interaktionen.

Begehrte Dinge

Smartphones sind im Alltag überall präsent, sie begegnen auf Reklametafeln, in Werbespots, in Geschäftsauslagen und natürlich vor allem in den Händen der Nutzer. Gehalten werden die Geräte häufig auf eine Weise bei der die Stellung der Hände an religiöse Rituale erinnert, auch die Inbrunst mit der auf den Geräten manipuliert wird und die Intensität der Aufmerksamkeitsbindung und der bläuliche Widerschein der Displays in den Gesichtern lässt an sakrale Rituale denken. In den Sinn kommen Assoziationen zu einem verheisungsvollen Blau, der Farbe des Himmels und des reinen Äthers, in der mittelalterlichen Welt Symbol für die Sehnsucht nach dem seligen Leben und einer übernatürlichen Wohnstätte. Diese religiös konnotierte Sehnsucht ist vielen Zeitgenossen ohne Zweifel fremd, nicht aber die Sehnsucht nach einem guten Leben und einem Ort der Hoffnung und

nach Zukunftsvisionen. Das Smartphone verkörpert eine Suchbewegung nach dem ewigen Anderswo, die alle Benutzer dieser Geräte zu erfassen scheint. Zur Cyborg-Technik macht das Gerät, dass die meisten von uns es ständig bei sich tragen, es hält eine Vielzahl von technischen, sozialen und emotionalen Funktionen bereit, die offenbar einen starken Aufforderungscharakter haben: es dient als Informationsquelle und -sender, als Kontaktstelle, als Überwachungswerkzeug, als Musikbox, als Däumlingskino, als interaktive Karte, Taschenlampe, Spielautomat, oder Selfiemaschine - und manchmal auch als Telefon. Das digitale Ding Smartphone ist eine Maschine, zu der viele seiner Nutzer nicht nur ein starkes emotionales Verhältnis haben, sondern durch das auch die emotionalen Bindungsverhältnisse zur Welt gefiltert und geordnet werden.

Der Kulturwissenschaftler Klaus Theweleit wundert sich vor allem über das Ausmaß der Deprivation im öffentlichen Raum, die mit dem Gebrauch der Geräte einhergeht. An einer Bushaltestelle, so seine Beobachtung, seien gut die Hälfte der Wartenden mit Ohrstöpseln ausgestattet, verkabelt mit einem Smartphone, bei den anderen beult es die Hosens- und Mantel-Taschen aus. Später im Bus gibt es so gut wie keine Gespräche, alle sind am »SMSeln«, lächeln nicht ihre Nachbarn, sondern das Gerät an. Und »dieses Lächeln« – so Theweleit – »ist sehr intim und doch öffentlich. Und vielleicht erntet dieses Lächeln keine Person, die es wirklich sehen sollte« (ders. 2015). Theweleit ist besorgt um die Sorglosigkeit, mit der sich die Nutzer in der digitalen Welt und an ihren Schnittstellen zur analogen Welt bewegen. Vor allem aber findet er das libidinöse Verhältnis der Besitzer zu ihrem Gerät befremdend und fragt sich, ob sich hier womöglich neue Formen von Sexualität entwickeln. »So wie ich das für die Freikorps-Soldaten beschrieben habe: dass ihre Liebe zum Pferd oder auch die Liebe zu ihren Waffen jene zu ihren Frauen übersteigt« (ebd.).

Digitale Lebensformen

Michel Serres, französischer Philosoph und Wissenschaftshistoriker, thematisiert ebenfalls das körperliche Verhältnis der Nutzer zu ihren Geräten, aber auf eine ganz andere Weise. 2012 veröffentlichte er einen längeren Aufsatz, in dem er

über die »kleinen Däumlinge« schreibt, Menschen, die ganz selbstverständlich und mühelos mit elektronischen Gerätschaften hantieren. »Ohne daß wir dessen gewahr wurden,« schreibt Serres, »ist in einer kurzen Zeitspanne, in jener, die uns von den siebziger Jahren trennt, ein neuer Mensch geboren worden. Er oder sie hat nicht mehr den gleichen Körper und nicht mehr dieselbe Lebenserwartung, kommuniziert nicht mehr auf die gleiche Weise, nimmt nicht mehr dieselbe Welt wahr, lebt nicht mehr in der selben Natur, nicht mehr im selben Raum« (Serres 2013, S. 15).

Nun wundert sich Michel Serres nicht darüber, daß Technik das menschliche Leben überhaupt prägt und verändert, sondern darüber, wie schnell die digitale Revolution in den Alltag vordrang und sich aufs engste mit unserem Leben verwoben hat. Das Digitale ist eine Lebensform geworden: Häuser und Wohnungen reagieren auf ihre Bewohner, schalten das Licht oder die Heizung an und aus und sorgen für Behaglichkeit, der Kühlschrank meldet sich bevor er leer ist und Bellos und Kittys Ausgang werden durch eine elektronisch gesteuerte Klappe kontrolliert. Wir leben in einer Technosphäre, in der wir uns nicht mehr an den kraftzehrenden Banalitäten des Alltäglichen aufhalten müssen, die ambient intelligence bietet Entlastung an. Und auch die Verwaltung und Nutzung von Wissen ist nicht mehr mühevoll und zeitaufwendig mit Ortsveränderungen verbunden oder personengebunden, denn es steht uns jederzeit in elektronischer Form zur Verfügung. Dadurch würden, so Michel Serres, geistige Kapazitäten freigesetzt, in der nicht nur eine neue Balance zwischen privatem und öffentlichem Selbst gefunden würde, sondern vor allem auch neue Denkformen erblühten. Eine solche »Demokratie des Wissens« ist kein neues Konzept, es kommt immer dann ins Spiel, wenn für aufklärerische Werte und damit einhergehend für eine gesellschaftliche Verpflichtung auf Rationalität eingetreten wird. Hans Jörg Sandkühler geht hier noch einen Schritt weiter und meint, daß die Aneignung und Konstruktion von Wissen ein zutiefst menschliches Recht sei (Sandkühler 1991). Er plädiert für eine zukunftsorientierte Rationalität, die den Menschen notwendig als Konstrukteur einer möglichen vernunftgemäßen Welt ins Recht setzt und damit auch für ein Wissen, das sein Maß am Menschen nimmt und dies

durch die Partizipation des Individuums an der epistemischen Arbeit realisiert. Wird der Zusammenhang zwischen Wissens und der selbstreflexiven Erfahrung des Individuums entkoppelt, so warnt Sandkühler, wird Wissen zur »abstrakten Möglichkeit ohne Subjekt, Sinn und Ziel. Wissen ist so nicht mehr die subjektive Fähigkeit der Vernunft zur Konstruktion der Wirklichkeit« (ders. 174.f.).

Nun mag bei diesem Konzept das Potential einer vernünftigen Regelung und Kontrolle überschätzt und die Eigenlogik und gesellschaftliche Dynamik einer Technologie unterschätzt werden. Für die Frage ob KI mehr eine gesellschaftliche Chance oder eher zum Problem wird, ist es dennoch entscheidend die Frage zwischen individueller Wissensaneignung und der gesellschaftlichen Gestaltung von Wissensbeständen und Wissensformen zu stellen. Ob mit KI eine Emanzipation und Öffnung eingeschliffener gesellschaftlicher Bildungswege einhergeht, eine Gestaltung von Arbeitsplätzen die soziale Isolation verhindert, eine bessere Zugänglichkeit von Wissensbeständen in Bibliotheken, Archiven, Verlagen, oder Museen garantiert wird von vielen Zeitgenossen eher skeptisch beurteilt.

Alain de Botton etwa attestiert dem Smartphone ein enormes Ablenkungspotential, das eine selbstreflexive Erfahrung geradezu unmöglich mache. Die Geräte »hindern uns vor allem daran mit uns selbst ins Gespräch zu kommen, uns zu fragen, was wir möchten, was uns bewegt, was wir bewegen wollen, was wir denken, welches Leben wir führen möchten.« (Alain de Botton, 2016, S. 76). Einerseits bringen also Smartphones einen öffentlichen Raum hervor, in dem sich isolierte digitale Monaden begegnen, denen der Möglichkeitssinn für aktuelle situative Begegnungen in der analogen Welt verloren gegangen zu sein scheinen und die sich hinter permanentem »Daddeln« auf dem ins Anderswo verweisenden Display verschanzen. Andererseits wirken die Smartphones als Zerstreuungswerkzeuge, die sich geschwätzig einmischen dürfen in jeglicher Lebenslage, als Ratgeber in allen Lebensfragen und letztlich seine Nutzer von einer Beschäftigung mit sich selbst abschirmen. Für eine gesellschaftliche Partizipation im Sinne eines von selbstreflexiven Bürgern geteilten politischen Raumes ist dies wenig förderlich. Die Demokratie des Wissens mit KI ist eine Frage des gemeinsamen Lernens.

Ding-Werk im Zeitalter des Anthropozän

An den Dingen zeigt sich die vom Menschen investierte Arbeit als Werk, sie werden hergestellt und haben die Aufgabe das Leben des Menschen zu stabilisieren, dauerhaft zu gestalten. Mit den Werken schafft sich der Mensch eine Heimat, wie Hannah Arendt in der *Vita activa* hervorhebt (Arendt 1981). Er umgibt sich folglich mit Dingen, in denen sein tätiges Leben Gestalt findet. Durch die Automation, beispielsweise durch KI, wird der Mensch zunehmend entlastet aber eben auch entfremdet von der Arbeit des Herstellens. Er hämmert nicht mehr selbst den Stahl oder steht am Webstuhl, erlebt weder die Hitze noch den Lärm bei der Herstellung der Dinge. Der Mensch ist also nicht oder kaum mehr in den Arbeitsprozess zur Herstellung der Dinge eingebunden und damit gibt es auch keine direkte Zweck-Mittel Relation mehr bei der Hervorbringung dieser Dinge. Infolgedessen verändert sich nicht nur der Warencharakter und das Verhältnis zu den Dingen, sondern überhaupt das Verhältnis des Menschen zur Technik. Das beklagte bereits Hannah Arendt (1981, 1. ed. 1960), die nicht der KI als gesellschaftliches Breitenphänomen begegnete, sondern damit die zunehmende Automatisierung durch technische Prozesse und maschinelle Artefakte im Blick hatte. Technik sei immer weniger Produkt einer bewussten Tätigkeit, als vielmehr eine Art biologischer Vorgang, der letztlich den Menschen von der zunehmend maschinengeschaffenen Welt entkoppelt. Die Arbeitsgesellschaft werde abgelöst durch eine Gesellschaft von Jobholdern, von denen kaum mehr gewünscht wird, als dass sie automatisch funktionieren (1981, S. 314). Damit einher ginge ein grundsätzlicher Wandel im Weltverhältnis von *Homo faber*, aus einem Bewohner der Welt werde zunehmend ein Beherrscher der Welt, Die Natur würde dabei so behandelt, als ob die Menschen gar nicht mehr auf der Erde lokalisiert seien, als wenn sie den archimedischen Punkt nicht nur gefunden, sondern sich geradezu auf ihn gestellt hätten, um von dort aus zu operieren. Das hält Arendt vor allem auch im Blick auf die Frage der Herstellung und Bedeutung der Dinge für den Menschen für bedenklich. Denn ohne deren Gegenständlichkeit gäbe es auch die Subjektivität des Menschen nicht, nicht seine relative Unabhängigkeit der Existenz von der Natur und ihren überwältigen-

den Elementargewalten (1981, S. 125). Denkt und erlebt sich also der Mensch nicht mehr in einer widerständigen Welt, verliert er auch seine Verankerung in dieser Welt. Dieser Befund wurde häufig als eine technikpessimistische Perspektive aufgenommen und als solche auch zurückgewiesen. Im Zeitalter des Anthropozän scheint sich diese Technikkritik aber geradezu anzubieten, vor allem mit der Frage danach, welche Konsequenzen es hat, wenn der Mensch zum Maß aller Gebrauchsdinge geworden ist und also ein radikal anthropozentrisches Weltverhältnis setzt. Dies könnte, mindestens auch, als eine suchende Frage verstanden werden nach einem angemesseneren Technikverhältnis.

In den digitalen sozialen Netzwerken wird diese Zentrierung der Nutzer auf ihre technischen Hervorbringungen auf eine vielleicht besonders drastische Weise deutlich. Die Kommentare und Bewertungen in beispielsweise einem Facebook-Konto, auch die Aktivitäten kommerzieller Anbieter und ihr Echo in Form individualisierter Werbeangebote, werden zu einem immer lückenloseren Gefüge, das sich dem Nutzer als seine höchst eigene soziale und kommerzielle Umwelt anbietet. Zusammengefügt, werden die Algorithmen der Likes, Daumen, Herzchen und all der anderen niedlichen Piktogramme gewissermaßen zu einem digitalen Spiegel, der vermeintlich sozial Agierende erzeugt sich vor allem eine eigene in sich abgeschlossene (und nipifizierte) Sphäre. Diesen »Spiegel-Welten«, darauf macht Suzana Alpsancar aufmerksam, steht aber die »Aura einer authentischen Öffentlichkeit (diametral entgegen) - zu der unvermeidlich Dissens, Konflikte und Unerwartetes dazugehören - mit der sich soziale Netzwerke gerne dekorieren« (Alpsancar 2018, S. 13). Ein weiteres politisches und moralisches Problem mit sozialen Netzwerken zeichnet sich ab bezüglich einer kommunikativen Verantwortung (Misselhorn 2018, S. 134 f.) in diesen sozial parzellierten aber technisch homogenisierten Welten. Shitstorm-Kampagnen, twitternde Politiker und Mobbing in Schulen sind nur einige Phänomene eines eher bedenklichen kommunikativen und sozialen Wandels durch digitale Werke. Hier wird eine Vielzahl einstimmiger Öffentlichkeiten erzeugt durch gezielten Ausschluss von Gegenstimmen und damit eine Monopolisierung sozialer und kommunikativer Formen geschaffen. Auch hier könnte

ein Blick in Hannah Arendts Schriften anregen, die hervorhob, daß niemand das Recht habe die Welt zu monopolisieren, alle müssten an der Welt teilhaben können. Eine plurale Welt könne erst entstehen, wenn, so Arendt (2002), alle in der Welt zuhause sind und diese mitgestalten und damit Halt und Sicherheit finden können.

Digitale Allmende

KI Technologien unterscheiden sich von anderen Technologien vor allem dadurch, dass sie ein analogisches Verhältnis anbieten zur menschlichen Intelligenz: beide arbeiten mit Zeichen, also auf der Basis von Algorithmen, beide nutzen iterative Problemlösungen, und beide nutzen »heuristische Hebzeuge«. Darauf, daß diese begriffliche Analogie⁴ zu hypertrophen Erwartungen führt und vor allem auch dystopische Visionen im Bereich humanoider Robotik hervorbringt, wurde immer wieder hingewiesen. So betonte jüngst auch Stephanie Kaiser vom Digitalrat der Bundesregierung, daß die Wortwahl in die Irre führe, Computer seien keineswegs intelligent. Vor allem aber verweist sie auch darauf, daß es sich bei KI um eine äußerst zehrende Technologie handelt: »Große neuronale Netzwerke haben vielleicht 1 Million Knoten und brauchen dafür die Energie eines halben Atomkraftwerks. Ein durchschnittliches Gehirn hat 84 Milliarden Neuronen und kommt mit einem Butterbrot aus« (Kaiser 2019, S.11).

Eingeführt wurde mit KI, wie David Gugerli sehr treffend formulierte, ein Instrument, »das zwischen Versprechen und Erwartung produktiv vermittelte« (2018, S. 195). Die in den 1970er Jahren einsetzende Arbeit an der digitalen Verbindungstechnik zwischen Rechnern führte erst Ende der 1980er Jahre zu dem was heute als World Wide Web bezeichnet wird. Es setzten sich damit letztlich, was nicht erwartbar war, digitale Protokolle durch, die im amerikanischen Verteidigungsministerium konzipiert worden waren. Und das lag weder an ihrer besonderen Leistungsfähigkeit noch an einer spezifisch militärischen Ausrichtung, sondern vor allem daran, daß diese experimentellen Netze auf »lokalen Lösungen aufsetzten und als Vernetzung der Netze, das heißt als Internet operierten« (ebd.).

4 »Artificial Intelligence« wurde 1956 eingeführt von einer Forschergruppe um John McCarthy während eines Workshops am Dartmouth College.

Auch eingeführt wurde damit, und das machte die Netztechnologie so erfolgreich, eine subjektivistische Topographie im Umgang mit technologischen Artefakten wie Cheryce von Xyländer (2019) hervorgehoben hat. Der Personal Computer und das Tablet bieten Schnittstellen an, die das subjektive Bedürfnis nach Bearbeitung und Interaktion mit der Umwelt bedienen können. Sie erlauben es überhaupt erst, die distributiven, emergenten und aggregierten Mindscapes in Anspruch zu nehmen, die der Nutzer selbst geschaffen hat. Nun ist diese Tätigkeit aber eine radikal kollektive. Digitale Infrastrukturen sind also angewiesen auf die gemeinsame Nutzung kognitiver und wissensbasierter Ressourcen, das Internet kann folglich auch als eine digitale Allmende betrachtet werden. Das heißt nun allerdings nicht, dass das Ordnungsprinzip von KI ein neuer, oder neuer alter Kollektivismus wäre, die ganze Weltgemeinschaft umfassend und eine global überall gleichermaßen zugängliche Ressource. Denn auch digitale Infrastrukturen haben natürlich eine materielle Basis, Ressourcen von theoretischem wie praktischem Wissen sind lokal gebunden und haben eine zutiefst dingliche Existenzweise, das gilt für Menschen wie Maschinen. Google hat seine Datenverarbeitungszentren an ganz konkreten, möglichst kühlen Orten stehen, beispielsweise in Hamina Finnland, Programmierer sind Menschen mit einer nationalen Zugehörigkeit und damit auch einer staatlichen Kontrolle unterworfen. Eine Reihe von Whistleblower Fällen, sei es Julian Assange und seine digitale Plattform WikiLeaks oder die Enthüllungen des ehemaligen CIA-Mitarbeiters Edward Snowden inklusive Verfilmung, haben das Ringen um digitale Verfügbarkeiten und Besitzstände in den Fokus medialer Aufmerksamkeit gerückt.

Weit weniger spektakulär aber umso deutlicher wird der Aspekt nationaler Machtansprüche in der sehr unterschiedlichen Verfügbarkeit der Ressource Mensch. Die Innovationskraft von KI und dem maschinellen Lernen gründet in der schieren Menge und Geschwindigkeit der Datenverarbeitung und in einer allgegenwärtigen Nutzung, die immer mehr und neue Informationen erschließt und rekombiniert. Je mehr Menschen sich also beteiligen, desto erfolgreicher kann sich diese Entwicklung zunehmender Selbst-Automatation entfalten, in China (ca. 1,4 Mia. Einwohner) entwickelt

sie sich anders als in den USA (ca. 332 Mio.) oder in einzelnen Ländern Europas (D ca. 83 Mio.). Wir haben es folglich nicht nur mit einem Ordnungsprinzip der Geographie von Nationen zu tun, sondern auch mit einer nationalen Geographie der Daten.

Dass die Nutzung der digitalen Allmende zur informationellen Selbstbestimmung interessengetrieben ist und auch als Mittel der sozialen Kontrolle eingesetzt wird, zeigt sich nicht nur in den verbreiteten sozialen Netzwerken und ihrer jeweils homogenisierten Öffentlichkeit, wie oben diskutiert, sondern zunehmend auch darin, wie mit Propaganda und Desinformation auf staatlicher Ebene operiert wird. Es zeichnet sich ein »technologischer Totalitarismus« ab, in dem sich Information als Waffe erweist und zwar gleichermaßen im »amerikanischen Überwachungskapitalismus« (Gujer 2019), im russischen Kampfeinsatz von massenhaften MeinungsBots, wie in der chinesischen Variante des Überwachungsstaates. Und weil heute KI eine selbstlernende Technologie ist, die wie oben beschrieben, durch massenhafte Rechenoperationen immer besser wird, scheint China besonders prädestiniert hier eine Führungsrolle zu übernehmen. Zusammen mit einer wachsenden Forschungskompetenz und bereits installierter Infrastrukturen, etwa von Kameras im öffentlichen Raum und einer sich ankündigenden automatischen Stimmerkennung in Telefongesprächen, ist die flächendeckende Überwachung von Bürgern nahezu vollständig vollzogen. Eric Gujer (2019) betont, daß hiermit auch ein dramatischer historischer Wandel verknüpft sei, indem der Westen eine ihm seit dem Mittelalter selbstverständlich gewordene Vormachtstellung abgeben müsse und prognostiziert »(d)ie Beherrschung der Informationstechnologie wird darüber entscheiden, welche Staaten im 21. Jahrhundert dominieren« (ders.).

Heuristik der Dinge

Wie wäre es, wenn KI als ein »heuristisches Hebzeug«⁵ aufgefasst würde auf der Suche nach einer pluralen Welt aktiver Teilhabe? Wenn in der allseits anerkannten digitalen Transformation unserer Gesellschaften auf der Basis von KI-Technologien nach einem weniger zehrenden Technikverhältnis gesucht würde? Die koloniale Logik gegenüber Naturbestän-

5 Bei Georg Christoph Lichtenberg heißt es »Bacons Organon soll eigentlich ein heuristisches Hebzeug sein« (Lichtenberg 1994, p. 830 [J 1242]).

den nicht einfach fortgesetzt, sondern weniger Ressourcen konsumiert und weniger Müll erzeugt würde? Im Moment scheint KI eher für das Gegenteil zu stehen, die Ausbeutung seltener Erden im großen Maßstab und die Entstehung immer größerer Berge an elektronischem Müll liefern bildmächtige Argumente. Die Folgen chinesischer Bergwerksaktivitäten um beispielsweise Lanthan oder Neodym für die Elektronikproduktion zu gewinnen sind auch auf Satellitenbildern gut sichtbar, die Berge an Elektronikmüll in Accra (Ghana) durchdringen mit ihren Dämpfen und Abwässern die lokale Umwelt, vor allem erzeugen sie ein sehr spezielles sozio-ökonomisches Milieu. Baruch Gottlieb hat in seinem Buch *Digital Materialism* (2018) unter der Kapitelüberschrift *Abstraction* ausgeführt, was die digitale Technik materiell beansprucht, wie tiefgreifend die Transformationen von Stoff- und Energiekreisläufen sind, und welche unrühmliche Rolle dabei der produzierte elektronische Müll spielt.

Wie bei vielen Produkten, wird auch bei digitalen Maschinen, wie Smartphones, die Grenze ob Ware oder Müll im Kontext des jeweiligen Gebrauchs in bestimmten sozio-technischen Konfigurationen gezogen. Entlang dieser Linie entscheidet sich, ob das Smartphone noch in Gebrauch genommen wird oder nicht. Ändert sich die soziotechnische Konfiguration, wird es beispielsweise aus der Schweiz nach Ghana transportiert, kann es aus Müll wieder zu einem Gegenstand werden, der in Gebrauch genommen wird, allerdings auf andere Weise als zuvor. Wird ein Ding also seiner Individualität beraubt, verliert es gleichzeitig seine Bindungskraft und -macht zu seinem Nutzer oder/und Besitzer. Umgekehrt heißt das, daß wenn etwas weggeworfen, ein Ding zu Müll transformiert wird, auch eine Bindung einer Person zu ihrer Umwelt verloren geht. Durch die Erzählform von Objektbiografien wurde in einem Kunstforschungsprojekt diesen wechselhaften Bindungsformen und ihren Folgen nachgespürt. *Times of waste*⁶ macht darauf aufmerksam, dass der grösste Teil des weltweit produzierten Elektronikschrotts derzeit nicht recycelt wird. Daran ändern auch vielversprechende technische Lösungsansätze wie die Entwicklung des Fairphones⁷ wenig, mit dem eine andere Produktlogik versucht wird, in der auch die die Herkunft der Materialien, die Produktgenese und vor allem eine Transpa-

6 Siehe die Projektseite von *times of waste* <https://www.objektbiografie.times-of-waste.ch> [Letzter Zugriff: 20. September 2019]

7 Die Firma tritt mit dem Anspruch an, ein sozialverträgliches, ressourcenschonendes und langlebige Gerät anzubieten. Das Fairphone 2 etwa kam 2015 auf den Markt mit einer modularen, reparierbaren technischen Ausstattung. Dieses Jahr kam das Fairphone 3 auf den Markt und tritt an mit dem Slogan »Das Smartphone, dem Mensch und Umwelt am Herzen liegen« <https://shop.fairphone.com/de/?ref=footer>. [Letzter Zugriff: 20. September 2019]

renz im Gebrauch und damit eine Reparaturfähigkeit angeboten wird.

Für eine überwältigende Mehrheit der Smartphones und anderer entbundener Elektronikprodukte muss allerdings festgehalten werden, daß eine materielle Verlagerung des Schrotts in Länder mit niedrigen Umweltstandards stattfindet. Auch sichtbar wird aus dieser Perspektive, daß bereits mit dem Rohstoffabbau und der Produktion der Geräte Müll in großen Mengen produziert, Energie und Ressourcen im großen Stil verbraucht werden. Mit der Einführung des Begriffs *green criminology* wurden juristische, epistemische und politische Aspekte miteinander verschränkt, die Produktion und Ablagerung umweltverschmutzender Materialien und elektronischer Konsumgüter damit als krimineller Akt benennbar, was durch Umweltaktivist_innen auch geschieht (Lynch et al. 2017). Eine Spezifität von e-Müll gegenüber anderem Müll könnte indessen sein, daß der Wechsel zwischen Ware und Müll flexibler erfolgt, auch im Sinne einer potentiellen Reversibilität. Die Entstehung von FabLabs und MakerSpaces kann in diesem Sinne aufgefasst werden als eine Verbreiterung der Grenzzone, in der die digitalen Dinge von Ware zu Müll und wieder zu Ware werden können. Eine in diesem Zusammenhang reaktivierte Kultur des Reparierens, der Re-Assemblage und des Recyclings ist womöglich Ausdruck einer Entwicklung, mit der die digitalen Dinge aus dem Möglichkeitsraum in den Wirklichkeitsraum gerückt werden. Sie wären damit nicht mehr vornehmlich Gegenstand von in die Zukunft gerichteten Objektgeschichten vom »ersten Gebrauch« und »innovativem Potential«, sondern würden als Dinge einer *technology-in-use* (Edgerton 2007) erzählt und angeeignet. Aus dieser Perspektive richtet sich die Aufmerksamkeit auf die Umnutzung und Umwertung von Konsumgütern, die dann nicht mehr den vom Produzenten vorgesehenen Weg gehen, sondern von ihren Nutzern einer spezifischen und vielleicht vom Erzeuger des Produkts nie vorgesehenen Gebrauch zugeführt werden, womit auch neue »Zeitlichkeiten des Materiellen« ins Spiel kommen (Hahn 2018, S. 105).

Die vorangehenden Überlegungen zu den begehrten Dingen, zu digitalen Lebensformen, dem Ding-Werk im Zeitalter des Anthropozän, der digitalen Allmende und der Heuristik

der Dinge machen deutlich, dass digitale Technologien materiell fest in Zeit und Raum verankert sind. Sie machen auch deutlich, dass die »digits« konkreter digitaler Dinge und Prozesse tief in unsere gesellschaftlichen Realitäten eingewoben sind und eine permanente Prüfung ihrer Ordnungsmuster gefordert ist: »Gesellschaft macht KI« bedeutet eine permanente Abstimmung von gesellschaftlichen Erwartungen und individuellen Handlungen durch die Dinge.

-
- Alpsancar, Suzana. 2018. Die Ethik der Künstlichen Intelligenz. *BTU News* 52: 13-14.
- Arendt, Hannah. 1981. *Vita Activa oder vom tätigen Leben*. München: Piper.
- Arendt, Hannah. 2002. *Denktagebuch 1950-1973*, in zwei Bänden. München-Zürich: Piper.
- Botton, Alain de. 2016. *Forschung und Lehre* 23/9: 76 [zitiert nach Süddeutsche Zeitung vom 20/21. August 2016].
- Edgerton, David. 2007. *The Shock of the Old: Technology and Global History since 1900*. Oxford: University of Oxford Press.
- Foucault, Michel. 1989. *Die Ordnung der Dinge*. Frankfurt am Main: Suhrkamp.
- Gottlieb, Baruch. 2018. *Digital Materialism: Origins, Philosophies, Prospects*. Bingley: Emerald Publishing Limited.
- Gugerli, David. 2018. *Wie die Welt in den Computer kam. Zur Entstehung digitaler Wirklichkeit*. Frankfurt am Main: S. Fischer.
- Gujer, Eric. 2019. Der perfekte Alptraum - wenn Überwachungskapitalismus und Überwachungsstaat zusammenwachsen. *NZZ online* 20.9.2019. <http://www.nzz.ch/meinung/google-facebook-und-amazon-big-brother-lebt-jetzt-im-silicon-valley-ld.1509071?mktcid=sms&mktcval=E-mail>. [Letzter Zugriff: 22. September 2019]
- Hahn, Hans Peter. 2018. Das 'zweite Leben' von Mobiltelefonen und Fahrrädern. Temporalität und Nutzungsweisen technischer Objekte in Westafrika. In *Kulturen des Reparierens*, hrsg. S. Krebs, G. Schabacher und H. Weber, 105-120. Bielefeld: Transcript Verlag.
- Kaiser, S. und Braun, H.. 2019. »Manchmal muss man einfach etwas ausprobieren.« *Schwarzrotgold - Das Magazin der Bundesregierung* 3: 9-11.
- Lichtenberg, Georg C.. 1994. *Schriften und Briefe*, ed. W. Promies, 6 vols. Frankfurt am Main: Zweitausendeins, vol. 1.

- Lynch, M. J., Long, M. A., Stretesky, P. B. und Barrett, K. L.. 2017. *Green Criminology. Crime, Justice, and the Environment*. San Francisco: University of California Press.
- McLuhan, Marshall 1964. *Understanding Media. The Extensions of Man*. New York: Mentor.
- Misselhorn, Catrin. 2018. *Grundfragen der Maschinenethik*. Stuttgart: Reclam.
- Moor, John. 2006. The Dartmouth College Artificial Intelligence Conference: The next fifty Years. *AI Magazine Volume 27/4*, 87-91.
- Pörksen, Bernhard. 2016. Wir lernen Netz. *Die Zeit* 9, 18. Februar 2016.
- Sandkühler, Hans Jörg. 1991. *Demokratie des Wissens. Aufklärung, Rationalität, Menschenrechte und die Notwendigkeit des Möglichen*. Hamburg: VSA-Verlag.
- Schaffner, Astrid-Vera. 2012. 60.000 Gedanken und Gegenstände. *SunDeckAvenue. Blog Astrid-Vera Schaffner*: <https://www.besser-fernsehen.ch/blog/entry/60-000-gedanken-gegenstaende.html>. [Letzter Zugriff: 20. September 2019]
- Serres, Michel. 2013. *Erfindet euch neu! – Eine Liebeserklärung an die vernetzte Generation*. Frankfurt am Main: Suhrkamp.
- Sprenger, Florian. 2019. *Epistemologien des Umgebens*. Bielefeld: Transcript.
- Theweleit, Klaus. 2015. Wir brauchen dringend humanitäre Upgrades. Interview mit Klaus Theweleit. *Avenue - Magazin für Wissenskultur 2* <https://avenue.jetzt/cyborgs/interview-mit-klaus-theweleit/3/> [Letzter Zugriff: 1. September 2019].
- Welzer, Harald. 2019. Künstliche Dummheit. *Die Zeit* 34, 15.8.2019. Politik, 6.
- Xylander, Cheryce von. 2019. *Mindscapes After Kant. Arnheim Lectures Humboldt Universität zu Berlin 2016/2017*. Lüneburg: Verlag für Wissenschaft und zeitgenössische Kunst (in press).





Über Erwartungen, Unübersichtlichkeiten und Pragmatismus: Künstliche Intelligenz im Prozess der gesellschaftlichen Implementierung

Tobias Wangermann

1 Als Arbeitsdefinition kann jene von Dr. Aljoscha Burchardt (DFKI) dienen: »Künstliche Intelligenz ist die Eigenschaft eines IT-Systems, der menschlichen Kognition ähnliche Fähigkeiten zu zeigen. [...] KI-Systeme verfügen in unterschiedlichen Anteilen über bestimmte Kernfähigkeiten wie situatives Wahrnehmen, Kommunizieren, Planen, Handeln, Schlussfolgern oder Lernen.«
Vgl. Burchardt in Wangermann 2018, S.2.

2 Nicht nur Stephen Hawking warnte auf dem »Web-Summit« in Lissabon im November 2017 vor einer existentiellen Bedrohung.
Vgl. Hawking 2018.

3 Der Beitrag bezieht sich auf die Fragestellung des ersten Panels »Was uns zusammenhält – KI als gesellschaftliches Ordnungssystem?« der Konferenz »(Un)ergründlich. Künstliche Intelligenz als Ordnungsstifterin«
Vgl. o.V. 2018a.

4 Zur Methodik des Gartner Hype Cycle for Emerging Technologies
Vgl.: o.V. ohne Jahr a.

Der Begriff der Künstlichen Intelligenz (KI)¹ ist hinreichend ungenau, um als Projektionsfläche für utopische und für dystopische Erwartungen zu dienen. Wegen oft geringer Kenntnisse der technischen Gegebenheiten schwingt das Pendel in der öffentlichen Debatte weit aus. Dass mit KI nicht nur der Fortbestand unserer wirtschaftlichen Wettbewerbsfähigkeit verknüpft wird, sondern auch die Zukunft unserer gesellschaftlichen Ordnung herausgefordert und sogar die Existenz des Menschen selbst zur Disposition gestellt ist,² spiegelt die Breite und Tiefe der Erwartungen an sie. Technische Umbrüche hatten schon immer gesellschaftliche Auswirkungen zur Folge, die bestehende Ordnungen erschütterten oder sogar Weltbilder stürzten. Worauf sollte sich also eine ordnungsstiftende Funktion von KI begründen?³

Der Hype Cycle als Hilfsmittel

Aus heutiger Perspektive wird der KI die Potenz eines technischen Umbruchs mit umfassenden gesellschaftlichen Auswirkungen zugesprochen. Es ist immer schwer vorauszusagen, welchen Einfluss eine neue Technik im Zeitverlauf wirklich auf die Gesellschaft hat, verändert sich doch auch ihre Wahrnehmung. Eine Abfolge von anfangs geradezu euphorischen Erwartungen (Chancen), die Ernüchterung nach der Erkenntnis von Bedrohungen (Risiken) hin zu einem realistischen Pragmatismus bei der Implementierung neuer Techniken in verändertem gesellschaftlichen Kontext scheint jedoch typisch. Der sogenannte (Gartner) Hype Cycle for Emerging Technologies⁴ beschreibt diesen Wandel der Erwartungen bzw. Aufmerksamkeit anschaulich.

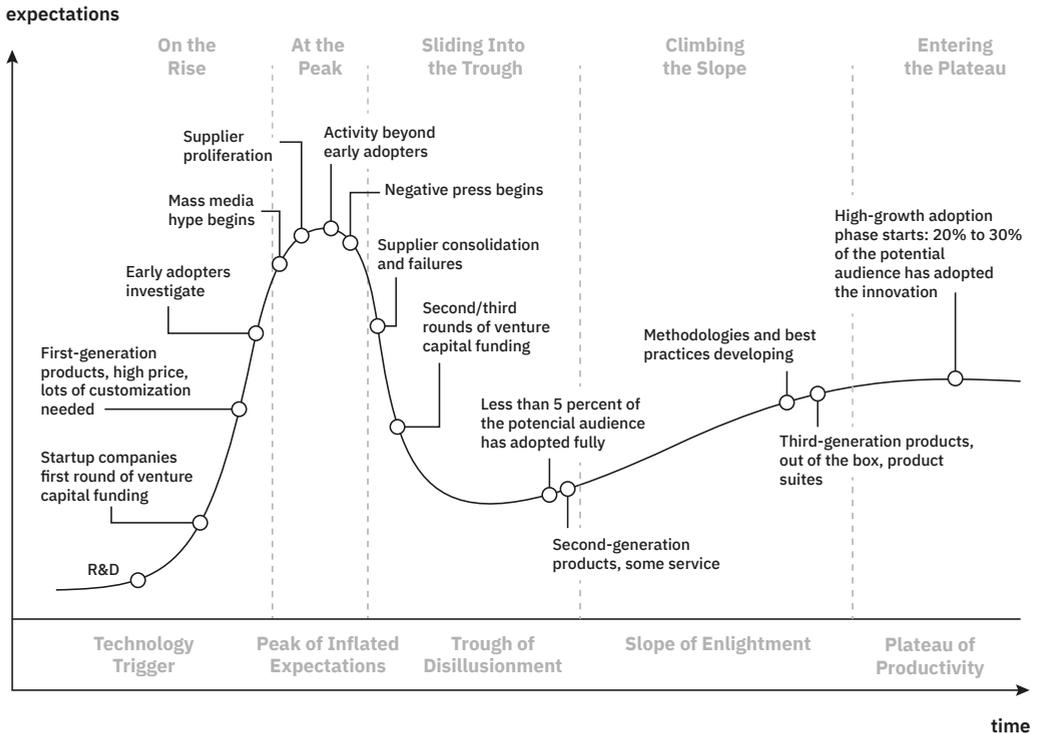


Abb. 1: Gartner Hype Cycle, (Olga Tarkowski/NeedCokeNow, CC-BY-SA 3.0)⁵

⁵ General Gartner Research's Hype Cycle diagram von Olga Tarkowski (NeedCokeNow), CC-BY-SA 3.0 international, <https://creativecommons.org/licenses/by-sa/3.0/deed.de>

Zwischen den Achsen Erwartungen und Zeit wird abgebildet, wie eine technische Neuerung von einer (Fach)Öffentlichkeit wahrgenommen wird. Als technischer Auslöser (technology or innovation trigger) können Fortschritte in der Anwendungsforschung oder gewachsene technische Ressourcen wirken, die einen zeitnahen Einsatz der neuen Technik ermöglichen. KI wurde von vier Aspekten angetrieben: stark gewachsene Rechenleistungen von Computern, exponentiell wachsende Datenmengen, sich an neuronalen Netzwerken orientierende Algorithmen-Systeme und öffentlichkeitswirksame Präsentationen von Anwendungen. In kürzester Zeit hat KI unter Zuschreibung eines hohen Chancenpotentials eine schnell wachsende Aufmerksamkeit in der Öffentlichkeit erfahren. Erste Anwendungen im Lebensalltag stützen diese Erwartungen. Der Höhepunkt des Hypes ist mit dem Gipfel der überzogenen Erwartungen (peak of inflated expectations) erreicht. Die darauf folgende Wahrnehmung von Risiken oder Beschränkungen einer Technik führt schließlich ins Tal der Enttäuschungen (trough of disillusionment). Der Tiefpunkt der Erwartungen bzw. Aufmerksam-

keit dokumentiert auch die gesunkenen Chancenzuschreibungen. Nach der Ernüchterung geht es auf den »Pfad der Erleuchtung« (slope of enlightenment), der – mit einer nun realistischen Einschätzung der Potentiale und klaren Anwendungsvorstellungen – auf das Plateau der Produktivität (plateau of productivity) führt.

Natürlich ist diese Darstellung eine Vereinfachung. Allein die verschiedenen Anwendungsfelder von KI müssen einzeln betrachtet werden, und die Diskurse, die Erwartungen und Aufmerksamkeit schüren oder dämpfen, verlaufen in Fachkreisen mit einer anderen Dynamik als in der breiten Öffentlichkeit. Der Zyklus ist aber eine Orientierungshilfe, um die Extreme in den Erwartungen (Hype) als solche zu erkennen und sich schon frühzeitig auf die Rahmenbedingungen für eine realistische Implementierung neuer Techniken zu fokussieren. Er zeigt auch, dass neue Technik⁶ nicht allein in ihrer Funktionalität bewertet wird, sondern auch auf ihre gesellschaftliche Anpassungsfähigkeit geprüft wird.

6 Im deutschen Sprachgebrauch hat der Begriff »Technologie« den der »Technik« ersetzt. Obwohl sie ursprünglich verschiedene Bedeutungen ausdrücken, werden sie inzwischen synonym benutzt.

Ohne Zweifel besitzt KI ein ganz besonderes Hype-Potential. Denn neben den als schwache KI bezeichneten Anwendungen, die nur über eine einzelne Kompetenz verfügen – wie beispielsweise die Texterkennung –, bieten die als starke bzw. allgemeine KI bezeichneten Systeme (general artificial intelligence) das Potential für Projektionen, in ihrer Leistungsfähigkeit eine Gleich- oder Höherwertigkeit dieser Systeme gegenüber dem Menschen als Ganzes zu sehen. Diese Vorstellung erinnert an den Homunculus, eine im Mittelalter geschaffene Figur eines künstlichen Menschen, die Allmachtsfantasien, aber auch Existenzängste hervorrief. KI gilt einigen als die technische Variante des Homunculus. Abseits der technischen Realitäten erzeugt diese Vorstellung eine hohe Emotionalität.

Natürlich ist es leichter, existierende Systeme weiter zu denken als völlig neue zu antizipieren. Während aktuell die im Kontext von KI wichtige Methode des »deep learning« nach Gartner bereits den Peak überschritten hat und in etwa zwei bis fünf Jahren die Plateau-Phase erreicht haben könnte, werden für eine »artificial general intelligence« gegenwärtig noch mehr als zehn Jahre bis zum Erreichen des Plateaus der Erwartungen prognostiziert (Panetta 2018). Wie

wenig valide solche zeitlichen Voraussagen sind, hat sich in der Vergangenheit jedoch allzu oft gezeigt.

KI im gesellschaftlichen Anpassungsprozess

Eine neue Technik wirkt erst dann innovativ, wenn sie als Anwendung Eingang in ein gesellschaftliches Umfeld gefunden hat. Zu diesem Umfeld gehört ein Setting aus sozialen, kulturellen, politischen und wirtschaftlichen Bedingungen, in die eine neue Technik sich einbetten lassen muss. Ein grundlegendes Verständnis der technischen Beschaffenheit (Wirkprinzip, Funktionalität, Einsatzmöglichkeiten etc.), die Gewichtung von Chancen und Risiken, die Beurteilung des wirtschaftlichen Potentials und eine Abschätzung des Einflusses auf bestehende (gesellschaftliche) Systeme sind Voraussetzungen, um eine neue Technik erfolgreich zu implementieren. Ziel ist es letztlich, die neue Technik für Lösungen in bestehenden Problemfeldern einzusetzen. Motivationen von Entwicklern, Herstellern und Anwendern bedingen einander. Es kommt zu Anpassungsprozessen, bei denen sich beide Seiten – Technik und Gesellschaft – gegenseitig beeinflussen und verändern. Diese Prozesse unterliegen nicht nur Nützlichkeitsabwägungen (Zuschreibungen an Lösungskompetenz), sie haben auch sozio-kulturelle Auswirkungen: den Erwerb von Fähigkeiten und Fertigkeiten im Umgang mit der Technik, die in den Bildungskanon integriert und weitergegeben werden, Anpassungen im Nutzungsverhalten, die Zuschreibung kultureller Werte, eine Einordnung der ethischen Konformität und nicht zuletzt die Definition der Rahmenbedingungen in Form von Zertifikaten, Regulierungen oder Gesetzen.

So ist auch die Politik an diesem Prozess beteiligt. Sie versucht motivierend, moderierend, steuernd oder regulierend einzugreifen. Wie fast alle führenden Industrienationen hat auch Deutschland eine KI-Strategie erarbeitet, in der die Förderung von Grundlagen- und Anwendungsforschung, die Motivation zu Unternehmensgründungen, die Bereitstellung von Risikokapital, Anstrengungen zur Fachkräfterekrutierung und der Ausbau von Bildungsangeboten sowie eine ethische Reflexion vorgesehen sind.

Die Politik agiert – im Kontext eines globalen Wettbewerbs – um die Ausschöpfung wirtschaftlicher Potentiale

von KI-Anwendungen und um die ethischen Rahmenbedingungen sicherzustellen. Die KI-Strategie der Europäischen Union zielt daher nicht nur auf eine Wertschöpfung durch KI, sondern formuliert auch einen Wertebezug (»new technologies are based on values«, »no one left behind in the digital transition«) (o.V. 2018b). Gremien wie die Enquete-Kommission des Deutschen Bundestages oder die High Level Expert Group on Artificial Intelligence der EU-Kommission sollen die Politik mit ihrer wissenschaftlichen, technischen und ethischen Expertise unterstützen. Auch in der Wissenschaft, etwa mit einer interdisziplinären Arbeitsgruppe an der Berlin-Brandenburgischen Akademie der Wissenschaften (o.V. ohne Jahr b.), oder in der Wirtschaft, Microsoft Deutschland (vgl. o.V. ohne Jahr c.) hat einen Expertenrat ins Leben gerufen, ist man bestrebt, einen Handlungsrahmen zu entwickeln, der in der Formulierung von einer »vertrauenswürdigen KI«⁷ und einer Ausrichtung auf den Menschen (Ebenda, S. 4) ein erstes Leitbild skizziert.

⁷ High Level Expert Group on Artificial Intelligence set up by the European Commission 2019.

Die Geschwindigkeit des technischen Fortschritts und die immer kürzer werdende Zeitspanne bis zur sogenannten Durchdringung von neuer Technik in der Gesellschaft erzeugen für alle Beteiligten einen hohen Umstellungsdruck. Zumal KI nicht eine konkrete Technik ist, sondern eine vielfältige, noch in der Entwicklung begriffene Methode umschreibt, die ein breites Anwendungsspektrum in den verschiedensten Lebensbereichen hat. Jede Anwendung wirft in ihrem jeweiligen Nutzungszusammenhang unterschiedliche Fragestellungen auf. So berühren KI-basierte Sprachassistenten die Privatsphäre der Nutzer wie auch den Unterschied zwischen Gesagtem und Gemeintem. Analytische KI-Assistenzsysteme für die Entscheidungsfindung stehen mit der Problematik eines diskriminierenden Bias bei den verwendeten bzw. verfügbaren Daten oder die Verwechslung von Korrelation und Kausalität in Verbindung. Bei der KI-gestützten Bilderkennung ist es die Differenz von Sichtbarem und Vorhandenem,⁸ um nur einige Beispiele anzuführen.

⁸ Hier sei beispielhaft auf die Fehldeutung einer Fahrzeuglackierung als Himmel bzw. Verkehrsschild verwiesen, die als eine der Ursachen für einen tödlichen Fahrzeugunfall eines im sogenannten Autopilot betriebenen Tesla Model S in Florida im Mai 2016 angeführt wird. Vgl. National Transportation Safety Board 2017.

Natürlich wird das Urteils- und Entscheidungsvermögen von Menschen auch ohne KI tagtäglich herausgefordert. Fehlbarkeit und Grauzonen, die keine klare Interpretation zulassen, gehören jedoch nicht in die Funktionslogik von Technik. Und sie widersprechen den Erwartungen, die Men-

schen beim Einsatz von Technik für akzeptabel halten. Das gilt selbst dann, wenn Technik in vielen Aspekten sicherer, schneller und effizienter agiert als der Mensch. Hier wirkt die Konfrontation mit einer komplexen Technik zurück auf den Menschen selbst. Im Diskurs über die Möglichkeiten von KI sind Menschen herausgefordert, sich ihrer eigenen Position zu vergewissern und bestehende Verfahren zu überprüfen.

Auch hier zeigt sich, dass weder übertriebener Alarmismus bezüglich der Risiken noch eine unreflektierte Technik-euphorie angebracht sind. Es wird vielmehr deutlich, dass Technik und Mensch in einer Interaktion stehen. Erinnert sei an dieser Stelle an die etablierten und anwendungsspezifisch ausdifferenzierten Verfahren, die für die Zulassung und Prüfung technischer Anlagen, Geräte oder Verfahren bereits bestehen. Dann wird einsichtig, dass – bezogen auf die jeweilige Anwendung – die durch die High-Level Expert Group der EU-Kommission definierten Anforderungen an eine »vertrauenswürdige KI«⁹ letztlich in diesen Zulassungs- und Zertifizierungsvorschriften umgesetzt werden müssen. Die von der Bertelsmann Stiftung und iRights.Lab erarbeiteten Regeln zur Gestaltung algorithmischer Systeme »Algo. Rules« decken nicht nur die Anforderungen der High-Level Expert Group ab, sie zeichnen sich selbst in den Formulierungen durch eine hohe praktische Anschlussfähigkeit aus.¹⁰ So wird darin der Begriff Künstliche Intelligenz gar nicht verwendet, sondern allein von algorithmischen Systemen gesprochen, und die neun aufgeführten Regeln lassen sich mit ihrer ganzheitlichen Systematik wie ein Lastenheften lesen.

9 High Level Expert Group on Artificial Intelligence set up by the European Commission 2019, S.15.

10 Die Bertelsmann Stiftung und iRights.lab haben unter dem Titel »Algo.Rules« formale Kriterien erarbeitet, um von der Entwicklung bis zur Anwendung von algorithmischen Systemen eine gesellschaftlich förderliche Gestaltung und Überprüfung zu ermöglichen.

Vgl. Bertelsmann Stiftung und iRights.Lab 2019.

11 Allgemeinsprachlich werden die Begriffe Daten und Information nicht trennscharf verwendet und oft sogar synonym eingesetzt.

In der Informatik hingegen verschafft erst ein Kontextbezug den Daten Bedeutung, so dass aus ihnen Informationen entstehen.

Vgl. u.a. Krcmar 2019.

Die Wahrnehmung von Welt: Komplexität und Ordnung

KI-Systeme zu nutzen folgt einer einfachen Notwendigkeit: Es braucht Instrumente, damit Menschen die unfassbare Menge an verfügbaren Informationen,¹¹ die komplexen und komplizierten Bezüge und die vom Menschen selbst eingegangenen vielseitigen Interaktionen in ihrer Ganzheit erfassen, beurteilen und beeinflussen können. Wurde Technik nicht schon immer entwickelt, um einen größeren Wahrnehmungs- und Wirkungskreis (Fernrohr und Mikroskop, Eisenbahn und Flugzeug) zu erschließen und dann unter den neu gewonnenen Bedingungen eine Ordnung herzustellen?

Um sich in der Welt zu orientieren und sie besser bewerten zu können, ist das numerische Erfassen schon lange ein erprobtes Mittel: Zählen, Messen und Vergleichen. Die Operationalisierung erschließt nicht nur Quantitäten, sondern macht auch Qualitäten und Beziehungen mit Hilfe der Mathematik beschreib- und begreifbar. Die Vermessung der Welt ist eine Methode, ihr eine nachweisbare Ordnung zu geben. Die wissenschaftliche Wahrnehmung gründet auf dieser Ratio, aber auch die Wirtschaft profitiert davon – etwa im Güterhandel. Datenbasierte Analysen ermöglichen im Zeitverlauf, Statistiken zu erstellen, auf deren Grundlage retrospektiv wie auch prognostisch Aussagen getroffen werden können. Trends und Muster werden selbst in unübersichtlichen Szenarien erkennbar, Erwartungen können bestätigt oder verworfen werden.

Die exponentiell wachsenden Datenmengen und die Notwendigkeit, unterschiedliche Daten miteinander zu verknüpfen und zu betrachten, verlangten nach Werkzeugen, die diese in einer angemessenen Zeit verarbeiten können. Die Abbildung in Zeichenketten, die sich aus einem Binär-code von Null und Eins zusammensetzen, machte es nicht nur möglich, unterschiedliche Arten und Mengen an Daten maschinell zu verarbeiten. Sie erlaubte ebenso – das ist der eigentliche Prozess der Digitalisierung – die Erfassung der Welt als digitale Repräsentation. Die Abbildung analoger oder nicht-diskreter Informationen in diskreten Zeichenketten erschließt diese erst für eine schnelle maschinelle Verarbeitung. In diesem Prozess entstehen riesige Datenmengen, die die verschiedensten Informationen repräsentieren: Messdaten aus Wissenschaft und Forschung, Prozessdaten aus Wirtschaft und Gesellschaft oder Digitalisate aus Kultur und Kommunikation. Der Umstieg des Informationsaustausches von analogen auf digitale Medien, ein unaufhaltsamer Aufwuchs an Sensoren, die technisch quasi unbeschränkten Möglichkeiten, Daten zu speichern und sie über ein globales Netzwerk weltweit und in Echtzeit verfügbar zu machen, haben einen Umfang erreicht, der mit herkömmlichen Werkzeugen kaum noch zu fassen war. Um in diesen unüberschaubaren Beständen Zusammenhänge zu ermitteln, Muster zu identifizieren, mithin eine Ordnung zu erkennen, bedurfte es eines neuen »intelligenten« Werkzeuges.¹²

¹² Diese Ordnung bezieht sich jedoch vorerst auf Korrelationen und nicht auf Kausalitäten.

KI unter den Rahmenbedingungen der Globalisierung

Die oben skizzierte Entwicklung folgt jedoch nicht einer technischen Eigendynamik, sondern ist durch veränderte Rahmenbedingungen getrieben. Die weltweite Vernetzung der verschiedensten Akteure und Systeme im Zuge der Globalisierung, ein daraus resultierender beschleunigter weltweiter Wettbewerb um Märkte, Informationen, Ressourcen und Meinungen sowie eine gewachsene Dynamik im wissenschaftlich-technischen Fortschritt bestimmen und beschreiben diese Rahmenbedingungen.

Die Weiterentwicklung von Werkzeugen, die hochkomplexe, detailliert abgebildete und dynamische Prozesse erfassen, analysieren und interpretieren, ist notwendige Voraussetzung, um auch in Zukunft erfolgreich agieren zu können. Die Automatisierung von Kommunikation und Logistik, die Optimierung von Ressourcennutzung und -verteilung (Energie, Rohstoffe, Arbeitsleistung), die Verbesserung der Schnittstelle Mensch und Technik (Robotik, Sprach- und Gestenerkennung, Bilderkennung) und die Nutzung in Wissenschaft, Forschung und Bildung (Datenauswertung, Simulation, Diagnostik) durch KI-basierte Systeme ist unabdingbar, um viele gesellschaftliche Probleme einer Lösung zuzuführen: vom Umweltschutz bis zur Energieversorgung, vom demographischen Wandel bis zum Wettbewerbsdruck.

Auf vielen krisenbehafteten Problemfeldern (Umwelt, Klima, Energie, Migration, Ressourcen, Sicherheit etc.) hat sich gezeigt, dass aufgrund deren Wechselwirkungen ein Abarbeiten einzelner Probleme nicht mehr zielführend ist, sind viele der Probleme doch globaler Natur. Eine ganzheitliche Betrachtung ist zwingend, um Zusammenhänge richtig zu deuten und Interventionen an den richtigen Stellen anzusetzen. Lernende Systeme, die auf eine solche Dynamik anpassungsfähig reagieren, aus einem komplexen und großen Dateninput schnell individuell zugeschnittene Lösungen extrahieren und die Schnittstelle Mensch – Technik überbrücken, sind die adäquaten Werkzeuge in einer globalisierten Welt.

KI gilt als die Schlüsseltechnik der Zukunft. Der Wettlauf um die globale Führungsposition ist unübersehbar. Europa, die USA und China verfolgen unterschiedliche strategische

Ansätze – wertegebunden, marktgetrieben oder parteikonform. Wer diesen Wettlauf gewinnt, definiert nicht nur die Standards. Er fährt einen hohen Anteil des wirtschaftlichen Mehrwertes ein, setzt den ethischen Rahmen und positioniert sein Gesellschaftsbild als die leistungsfähigste Antwort auf die Herausforderungen der Globalisierung.¹³ Die Frage, wie und wofür KI eingesetzt wird, ist damit von hoher politischer Priorität.

13 Unter einem hohen Wettbewerbsdruck lebt sogar die Metapher vom „Kalten Krieg“ wieder auf, so wie es u.a. Cathy O’Neil, die Autorin von “Weapons of Math Destruction” (2016), jüngst auf der Konferenz „Humanity Defined: Politics and Ethics in the AI Age“ des Aspen Instituts Deutschland in Berlin am 29. März 2019 tat.

14 Das gilt besonders dann, wenn durch »deep learning« die Ergebniserzeugung von KI-Systemen nicht mehr ohne weiteres nachvollziehbar ist, die Technik als »Black Box« wahrgenommen wird und nur eine Input-Output-Analyse zur Verfügung steht.

Perspektivwechsel: Problemlösung und Technik

Daher ist es so wichtig, dass sich die öffentliche Debatte um KI nicht allein auf die Technik fokussiert. Noch wird zumeist aus einer technischen Perspektive argumentiert, die die Einzigartigkeit von KI in den Mittelpunkt stellt. Der Technik wird ein Eigenwert zugesprochen, der die Gestaltungs- und Eingriffsoptionen des Menschen reduziert, sie ihm sogar zu entziehen scheint.¹⁴ »Unergründlich« ist eine Vokabel, die in einem technischen Kontext nicht angemessen ist. Im Diskurs über die ethischen Herausforderungen von KI stützt sie aber die Annahme einer geringer Einflussmöglichkeiten und befördert damit den Eindruck der Fremdbestimmung durch Technik.

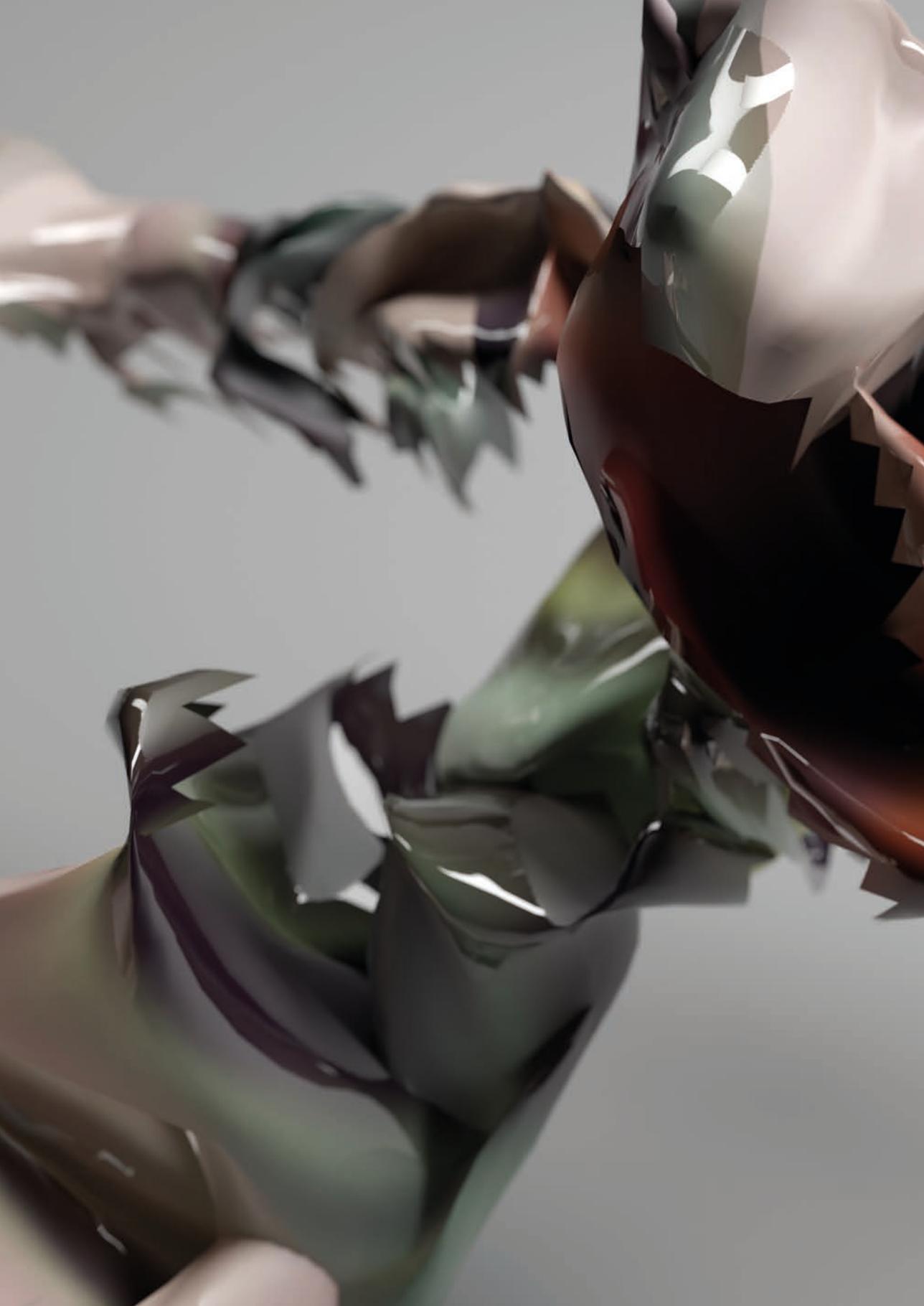
Entscheidet man sich für die Perspektive, konkrete Problemstellungen oder Aufgaben zu lösen und sich dabei der KI zu bedienen, erscheinen viele Fragen anschaulicher, praxisnäher und sind deutlich anschlussfähiger. Das gilt sowohl für die Implementierung in existierende Prozesse, für die Planung von Anwendungsszenarien als auch für die ethische Bewertung. Technik wird zum Werkzeug, das der Mensch entwickelt und einsetzt, um Problemstellungen adäquat zu bewältigen. Gestalt und Funktionalität von Technik bemessen sich an ihrer Nützlichkeit für die jeweilige Problemlösung. Und der Mensch rückt als Akteur wieder in den Mittelpunkt. Seine Anforderungen definieren, was Technik leisten soll und wie sie dafür beschaffen sein muss. Die Abwägung von Chancen und Risiken und damit die Zuweisung von Nutzen und Gefährdung sind im konkreten Anwendungsfall genauer und praxistauglicher, und sie gründen auf Erfahrungswissen und etablierten Wertesystemen. Der Einsatz von KI-Anwendungen beispielsweise in der medizinischen Diagnostik ist von dem Ziel geleitet, durch bessere Analyse-

verfahren dem Patienten zu helfen. Es macht den Arzt nicht überflüssig. Bereits existierende Analyseverfahren wie Blut- oder Urintests müssen bisher schon durch den Arzt in ihre Genauigkeit und Verlässlichkeit beurteilt und bei der Diagnose in ein Gesamtbild eingeordnet werden. Es ist eingeübte Praxis, dass solche Tests erst nach umfangreichen Studien, Prüfungen und letztlich einer Zulassung ihren Weg in die medizinische Anwendung finden. Bestehen sie hier nicht und sind Genauigkeit, Verlässlichkeit und Transparenz der Methode nicht hinreichend gegeben, werden sie nicht eingesetzt oder in ihrer Aussagekraft bei der Diagnose entsprechend eingestuft. Das gleiche wird für KI-gestützte Verfahren gelten müssen – nicht nur in der Medizin.

Fazit: Mehr Pragmatismus

Es ist schon heute absehbar, dass KI-Systeme überall dort selbstverständlich eingesetzt werden, wo es angepasster, situationsbezogener oder individueller Bewertungs-, Entscheidungs- und Steuerungsunterstützung bedarf. Nicht weil KI einer selbsterfüllenden technischen Prophezeiung folgt, sondern weil sie gebraucht wird, um die komplexen Herausforderungen der Zukunft zu bewältigen. Beim Einsatz von schwacher KI ist Pragmatismus angeraten, um schon heute begrenzte oder wertvolle Ressourcen jeglicher Art so schnell als möglich durch den Einsatz dieser Technik zielführender und passgenauer einzusetzen als wir es bisher tun. Die Debatte um KI darf sich nicht vom Diskurs um Digitalisierung, Globalisierung und Gesellschaft abkoppeln. Nicht nur weil Voraussetzungen wie Datenzugänge, Infrastruktur oder Fachkräfte als drängende Herausforderungen der Digitalisierung auch KI betreffen, sondern weil das Aushandeln der politischen, rechtlichen und gesellschaftlichen Rahmenbedingungen entscheidend dafür ist, wie es uns gelingt, die Potentiale einer neuen Technik für die gesamte Gesellschaft und für die Umwelt nutzbar zu machen.

-
- Bertelsmann Stiftung und iRights.Lab. 2019. Algo.Rules – Regeln für die Gestaltung algorithmischer Systeme. In: *Algo.Rules*. https://irights-lab.de/wp-content/uploads/2019/03/Algo.Rules_DE.pdf. [Letzter Zugriff: 16. Mai 2019]
- Hawking, Stephen. 2018. Kurze Antworten auf große Fragen. Stuttgart: Klett-Cotta.
- High Level Expert Group on Artificial Intelligence set up by the European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. <https://ec.europa.eu/futurium/en/ai-allian-ec-consultation/guidelines#Top> [Letzter Zugriff: 28. J uni 2019]
- Krcmar, Helmut. 2019. Information. In: *Enzyklopädie der Wirtschaftsinformatik Online-Lexikon*. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/datenwissen/Informationsmanagement/Information->. [Letzter Zugriff: 10. April 2019]
- National Transportation Safety Board. 2017. Highway Accident Report NTSB/HAR-17/02 – Collision Between a Car Operating With Automated Vehicle Control System and a Tractor-Semitrailer Truck Near Williston, Florida, May 7,2016. In: *National Transportation Safety Board*. <https://www.ntsb.gov/investigations/accident-reports/reports/har1702.pdf>. [Letzter Zugriff: 8. April 2019]
- O’ Neill, Cathy. 2016. Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishing Group/Penguin Random House.
- Panetta, Kasey. 2018. 5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018. In: *Gartner*. <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/>. [Letzter Zugriff: 11. Februar 2019]
- o.V. 2018a. ÖFIT-Konferenz 2018 - (Un)ergründlich. Künstliche Intelligenz als Ord-nungsstifterin. In: *Öffentliche IT*. <https://www.oeffentliche-it.de/unergruendlich>. [Letzter Zugriff: 9. April 2019]
- o.V. 2018b. Communication Artificial Intelligence for Europe. In: *European Commission*. <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>. [Letzter Zugriff: 13. Februar 2019]
- o.V. ohne Jahr a. Gartner Hype Cycle In: *Gartner*. <https://www.gartner.com/en/research/methodologies/gartner-hype-cycle>. [Letzter Zugriff: 5. Februar 2019]
- o.V. ohne Jahr b. Verantwortung: Maschinelles Lernen und Künstliche Intelligenz. In: *Berlin-Brandenburgische Akademie der Wissenschaft*. <http://www.bbaw.de/forschung/verantwort-maschinelles-lernen-und-kuenstliche-intelligenz/uebersicht>. [Letzter Zugriff: 13. Februar 2019]
- o.V. ohne Jahr c. Expertenrat Künstliche Intelligenz. In: *Microsoft Berlin*. <https://www.microsoft-berlin.de/microsoft-expertenrat-fuer-kuenstliche-intelligenz.aspx>. [Letzter Zugriff: 13. Februar 2019]
- Wangermann, Tobias (Hrsg.). 2018. Künstliche Intelligenz. *Häufig gestellte Fragen*. Berlin: Konrad-Adenauer-Stiftung. <https://www.kas.de/einzelartikel/-/content/kuenstliche-intelligenz-haufig-gestellte-fragen> [Letzter Zugriff: 28. Juni 2019]





KI und die Automatisierung von Entscheidungen

Rekodierung von Religion: Theologische Darstellungen von Künstlicher Intelligenz und der Zukunft von Gesellschaft¹

Robert M. Geraci

¹ Dieser Vortrag wurde als Keynote-Präsentation bei »(Un)ergründlich? Künstliche Intelligenz als Ordnungsstifterin«, am 18. Oktober 2018 gehalten. Ich danke der Fraunhofer-Gesellschaft und Dr. Karoline Krenn dafür, dass sie mich in die Veranstaltung einbezogen haben, und ich danke Dr. Cory Blad, dass er meine endlosen Kommentare und Kritiken zum Foucault Projekt aufgegriffen hat. Ohne seine geduldigen Erklärungen und Argumente wäre es mir schwerer gefallen, dieses Papier zu erarbeiten. Eine erweiterte Version des Papiers wird in *Visions of Technology* verfügbar sein: *Conversations at the Intersection of Religion and Artificial Intelligence in the U.S. and India*. Bei der vorliegenden Fassung handelt es sich um eine Übersetzung des englischen Originaltextes.

² Mackenzie/Wajcman 1985, 6, Betonung im Original entfernt. Die Seitenangabe bezieht sich auf das englische Originalzitat.

Die globale Entwicklung der modernen Wissenschaft hat komplexe Netzwerke geschaffen, durch die sowohl Technologien als auch Ideen zirkulieren. Innerhalb dieser Netzwerke erweitern Technologien künstlicher Intelligenz (KI) ihren Einfluss auf das menschliche Leben und schaffen so Strukturen, die das menschliche Denken und Handeln definieren. Sollte KI menschliche Gleichwertigkeit oder übermenschliche Intelligenz entwickeln, dann wird der Einfluss von KI auf die Gesellschaft noch tiefgreifender sein. Solche Technologien entwickeln sich jedoch nicht in einem kulturellen Vakuum, sondern sind das Ergebnis bereits existierender kultureller, einschließlich religiöser, Praktiken. Im Allgemeinen, wie Mackenzie und Wajcman feststellen, »spielen die Merkmale einer Gesellschaft eine große Rolle bei der Entscheidung, welche Technologien adaptiert werden.«² Infolgedessen werden religiöse und kulturelle Traditionen einen entscheidenden Einfluss auf die globale Entwicklung von KI haben, auch wenn diese Technologien dann definieren, wie Menschen in Zukunft leben und arbeiten. In den USA und in Indien zeigen Science-Fiction und populärwissenschaftliche Ansätze zu KI, dass menschliches Eingreifen in die KI-Entwicklung möglich ist, und dass es nicht vorherbestimmt ist, wie sich KI weiterentwickeln könnte.

Dieser Essay untersucht die möglichen Auswirkungen verschiedener religiöser Kulturen auf die Zukunft von KI, indem er zeigt, wie wichtig kulturelle Ideen, einschließlich Religion, für die Technologie im Allgemeinen und KI im Besonderen sind. Der Aufsatz legt zunächst den Hang zum Glauben an einen technologischen Determinismus dar – der Glaube daran, dass sich der technologische Fortschritt nach einer Logik entfaltet, die er selbst bestimmt. Beispiele aus der Technikgeschichte zeigen, dass menschliche Entschei-

dungen einen starken Einfluss auf die Entwicklung einzelner Technologien haben und geben Grund zur Behauptung, dass wir erklären müssen, welche Arten von Entscheidungen getroffen werden und warum. Basierend auf dieser Ablehnung eines technologischen Determinismus' wird der Aufsatz dann zeigen, wie kulturelle Ideen von Technologie – wie Technologien selbst – in einem globalen Netzwerk geteilt werden. Technologien beinhalten Ideen, einschließlich der Ideen von Reichtum, Macht und sozialer Struktur. KI-Technologien haben besondere Relevanz, da sie in der Lage sind, Menschen nach den Ideologien, die ihrer Programmierung und Bereitstellung zugrunde liegen, einzuschränken und zu kontrollieren. Schließlich wird der Aufsatz untersuchen, wie religiöse Ideen Teil der Ideologien sind, die in KI-Technologien integriert sind. Verschiedene religiöse Kulturen werden unterschiedliche ideologische Werkzeuge zur Verfügung stellen, die auf ein optimales menschliches Gedeihen ausgerichtet sein sollten.

Die religiösen Kulturen der USA und Indiens bieten jeweils Werkzeuge zur Entwicklung von KI unter Berücksichtigung menschlicher Interessen, und unserer gemeinsamen Zukunft wird besser gedient sein, wenn wir verstehen, wie diese Traditionen in der Wissenschaftskultur wirken. In den USA unterliegen den Perspektiven auf KI apokalyptische religiöse Ideen. Diese bieten sich als transformative Träume von Bestreben und Hoffnung an, erlauben aber auch Werte kapitalistischer und militärischer Dominanz. Im Gegensatz dazu nutzen die jüngsten indischen Perspektiven auf KI religiöse Pflichtstrukturen, die zu Werkzeugen für Zusammenarbeit und Fürsorge werden oder auch ein neues Kastensystem vergegenständlichen könnten. Die Erforschung dieser kulturellen Perspektiven beim Bau immer leistungsfähigerer Maschinen ermöglicht es uns, Entscheidungen darüber zu treffen, welche Arten von Maschinen gebaut werden sollen und wie diese Maschinen wiederum unsere eigenen Entscheidungen in Zukunft einschränken und kontrollieren werden.

Die unbestimmte Zukunft

Allgemeine und sogar fachspezifische Diskussionen über Technologie gehen oft von der Unvermeidlichkeit bestimm-

ter technologischer Ergebnisse aus. Im Zusammenhang mit KI meint dies in der Regel übermenschliche Intelligenz und die Möglichkeit, dass der Mensch obsolet wird; tatsächlich werden Gespräche über Computer und KI fast ausschließlich unter diesen Begriffen geführt. Die apokalyptischen Visionen von Technologie, die später in diesem Essay diskutiert werden, sind ein drastisches Beispiel dafür. Die Anziehungskraft des technologischen Determinismus' geht jedoch über KI hinaus und nimmt in den Perspektiven des späten 20. und frühen 21. Jahrhunderts einen immer größeren und stärkeren Stellenwert ein. Es ist daher notwendig, die Risse im verallgemeinerten ideologischen Bekenntnis zu diesem technologischen Determinismus zu beleuchten und sowohl die Gegenwart als auch die Zukunft einer kritischen Betrachtung zu unterziehen.

Es ist üblich, dass Menschen die Vergangenheit ignorieren und auf technologische Entwicklungen eine deterministische Logik anwenden; zum Teil liegt das daran, dass wir Misserfolge – Technologien, die den Markt nie beherrscht haben – ignorieren. Wir schenken erfolgreichen technologischen Entwicklungen mehr Beachtung als missglückten, was uns zu der Annahme führt, dass »der Erfolg eines Produkts eine Erklärung für seine spätere Entwicklung ist.« (Pinch/Bijker 1984, 406) Auf dieser Grundlage spekulieren einige, dass Technologien ihre nächsten Entwicklungsschritte selbst bestimmen: dass der technologische Fortschritt ganz durch sich selbst definiert wird und dass er eigene Garantien für die Zukunft bietet (auch wenn diese unbekannt sind).

Pfadabhängigkeit, bezogen auf den Determinismus, ist ein reales Phänomen in der technologischen Entwicklung. Das heißt, manchmal schließt eine Entscheidung notwendigerweise andere Entscheidungen aus und behindert die zukünftige Entwicklung. Die Pfadabhängigkeit dominiert oft technische Entscheidungen dank interner Feedback-Mechanismen, wie z. B. die Überprüfung von Nachfolgetechnologien auf Inkompatibilität mit Vorläufervariationen und externes Feedback, wie es durch die ästhetische Auswahl des Menschen oder durch politische Einflussnahme vorgegeben wird. Es ist jedoch tautologisch und damit uninteressant zu behaupten, dass Technologien in ihrer jetzigen Form einige (nicht alle) Einschränkungen für kommende Technologien darstellen.

- 3 Siehe insbesondere Pinch/
Bijker 1984.

Wir neigen dazu, technologischen Wandel als einen Prozess linearer und offensichtlicher Entwicklungen zu betrachten – nachträglich und mit all den Kontroversen, Fragen und Bedenken, die in die akzeptierte Technologie aufgenommen wurden³ –, aber menschliche Faktoren sind ein notwendiger Bestandteil technologischer Entwicklung. In Kombination mit dem offensichtlichen Effekt der Pfadabhängigkeit fördert die lineare Sichtweise auf technologische Entwicklungen den Glauben an einen technologischen Determinismus. Aber der Mensch behält in diesem Prozess seine Handlungsfähigkeit, auch wenn wir geneigt sind, etwas anderes zu glauben. Menschliche Entscheidungen produzieren Technologien, und daher sind die Ziele, denen Technologien dienen, Teil der menschlichen Gesellschaft. So führte beispielsweise die Entsorgung von Kriegsmaterial im Jahr 1918 zu mehr Karbolsäure auf dem Markt, was zu einem Kostenverfall führte, der Bakelit erschwinglich und wirtschaftlich wettbewerbsfähig machte (überprüft in Pinch/Bijker 1984, 406).

Zu den sozialen Dynamiken gehören Eingriffe durch die Regierung, internationale Politik und Kriege, das Maß technologischer Anpassung im geografischen Raum, in der Industrie und im Haushalt. All dies und noch viel mehr beeinflusst, ob und wie sich eine Technologie entwickelt und wie sich ihre Anwendung gestaltet.

- 4 Siehe Adas 1989, 41–53;
Parthasarathi 2001;
Nair 2005, 41–3.

- 5 Siehe Gandhi [1909]
1922, 107.

Man betrachte die Rolle der Charkha (Spinnrad) in der indischen Unabhängigkeitsbewegung. Der britische Kolonialismus untergrub die dominante Rolle Indiens im globalen Textilsektor⁴ und führte durch seine Steuer- und Handelssysteme zu erheblicher wirtschaftlicher Not. Gandhi plädierte daraufhin für den Rückbau der industriellen Textilproduktion zugunsten der individuellen Baumwollspinnerei.⁵ Während Gandhis Wunsch nach einer universalen Unabhängigkeit des Dorfes von nationalen und globalen Märkten und Produktionsweisen nicht praktikabel war, wurde die Charkha dennoch zu einem Instrument im nationalistischen Repertoire. Zwei Dinge sind erwähnenswert: erstens, dass die Verwendung der Charkha sich im zwanzigsten Jahrhundert ohne die britische koloniale Unterdrückung nie etabliert hätte, und zweitens, dass die nationale Politik durch die Existenz dieses Werkzeugs teilweise bestimmt wurde. Letztendlich widersetzen sich diese komplexen Dynamiken der Reduktion auf verein-

fachende Narrative eines technologischen Determinismus'. Kehren wir zurück zu KI, so umgibt ihre Beschreibungen eine Aura der Unvermeidlichkeit, aber natürlich ist diese Aura nicht unantastbar. Ray Kurzweil zum Beispiel versucht, Zukunftstechnologien auf der Grundlage der aktuellen Trajektorien vorherzusagen, aber seine Bilanz hier ist weniger erfolgreich als oft behauptet wird.⁶ Seine Position, die er dem Robotiker Hans Moravec nachempfunden hat, ist, dass Computer sich so lange weiterentwickeln werden, bis sie kompetenter sind als Menschen, und Menschen in einer Cyborg- oder Roboterzukunft mit Maschinen verschmelzen werden.⁷ Unabhängig von genauen Prognosen besteht der Vorstoß seines Standpunkts und dem seiner Verbündeten jedoch darin, zu erklären, dass sich digitale Technologien entweder nach computereigenen Regeln – wie dem Mooreschen Gesetz und jeder zukünftigen Iteration davon – entfalten oder nach Regeln, die den Naturgesetzen innewohnen. Im Gegensatz dazu argumentiert N. Katherine Hayles, dass die Geschichte des Computerwesens zeigt, wie aktive Entscheidungen den Einsatz von Computern geprägt haben, während sie Alternativen ausschlossen. Wenn die menschliche Wahl bei der Entwicklung des Computers eine Rolle spielte, dann ist sie sicherlich auch in Zukunft wichtig – auch wenn manche Kritiker das Gegenteil behaupten. Angesichts der Möglichkeiten, die es in unseren früheren Vorstellungen von Computern gab, erscheint es allzu offensichtlich, dass Eingriffe auch in Zukunft möglich bleiben werden.

Es gibt klare Befürworter menschlicher Freiheit auf den Gebieten der Informatik und der Kybernetik, darunter Joseph Weizenbaum, eine wegweisende Gestalt in der Geschichte beider Bereiche. In *Computer Power and Human Reason* argumentiert Weizenbaum, dass Technologie nicht selbstbestimmt sein könne und stattdessen von menschlichen Interessen und Anliegen getrieben sei. Als Beispiel führt er an, dass Wissenschaftler die McNamara-Linie konstruierten, weil sie gegen die Bombardierung Nordvietnams waren, und damit zum Aufstieg des elektronischen Schlachtfeldes beigetragen haben. Es war nicht die Technologie, die bestimmte, sondern es waren Menschen, die entschieden, wie sie sich verhalten und was sie tun sollten; sie hätten anders handeln können, argumentiert er, zum Beispiel indem

⁶ Siehe Kurzweil 1999, passim; 2005, passim. Kurzweil selbst spricht von einer enormen Erfolgsrate bei seinen Vorhersagen (siehe Hochman 2016); aber seine Interpretationen von Erfolg sind recht großzügig. Für eine schöne Zusammenfassung von Kurzweils Prognosen in *The Singularity is Near*, siehe Feakins 2017. Feakins stellt die Vorhersagen von Kurzweil zusammen und bewertet die Erfolgsrate von Kurzweil für 2017 mit 9 von 25 (die 2017 endgültig beurteilt werden können). Es ist anzumerken, dass von den 9 korrekten Vorhersagen mehrere recht unspektakulär sind und von einem großen Spektrum anderer Kritiker getroffen wurden, während andere eine ziemlich großzügige Interpretation erfordern, um als wahr bezeichnet zu werden. Keine der spektakuläreren Vorhersagen (mit Ausnahme vielleicht von Sprachübersetzungen) ist Realität geworden. Für eine noch kritischere Betrachtung ab 2009 siehe Knapp 2012.

⁷ Kurzweil 1999; Kurzweil 2005; Moravec 1988; Moravec 1999. Für eine umfassendere Analyse von Moravec, Kurzweil und ihren Anhängern siehe Geraci 2010; Geraci, 2011.

sie sich dem Krieg ganz widersetzen. Er führt gerade dieses Beispiel an, um auf die Bedeutung guter Entscheidungen bei der Entwicklung von Computertechnologien hinzuweisen.

Ideologien und Technologien

Wir bauen soziale Strukturen und Ideen in Technologien ein, also haben Erwartungen an die Zukunft der menschlichen und maschinellen Intelligenz eine klare soziale Relevanz. Ein einfaches und vielleicht naheliegendes Beispiel zeigt: die Personen, die Prognosen zu KI treffen und in unserer Medienlandschaft am ernstesten genommen werden, beschäftigen sich oft beruflich mit Softwaredesign und Marketing. Sie sind in der Lage, die Ergebnisse des Designs auf das auszurichten, was sie als unvermeidlich empfinden. Ray Kurzweil argumentierte, dass Maschinen uns in der Evolutionsgeschichte folgen würden, und wurde schließlich 2012 von Google eingestellt. Diese Anstellung gab ihm zusätzliche Glaubwürdigkeit auf dem Markt der Ideen und eine noch bessere Position, von der aus er die Zukunft des Computerwesens steuern konnte. Im weiteren Sinne kann man sagen, dass das Design und die Konstruktion von Technologien den Motivationen von Designern, Vermarktern, Politikern und anderen konkrete Gestalt geben. Einfach ausgedrückt, sind Technologien die physische Manifestation menschlicher Ideen und Absichten.

Der Einsatz von KI zu Überwachungszwecken macht die Untersuchung solcher Absichten besonders wichtig. Reduzieren Gesichts- und Mustererkennungstechnologien Menschen gewaltsam zu statistischen Terroristen? Was sagt es über uns aus, dass wir Individuen oder Minderheiten als potenzielle Terroristen betrachten und Systeme einsetzen, die Menschen unter der klaren Annahme beobachten, dass sie welche sein könnten? Unter weit verbreiteten, aber weniger emotional aufgeladenen Umständen haben sich die Internetnutzer der unternehmerischen Unterstellung gebeugt, dass jeder Einzelne ein legitimes Marketingziel ist. Mitten im 21. Jahrhundert implizieren die Haltungen sowohl von Regierungen als auch Unternehmen die Unvermeidlichkeit, Allgegenwart und Akzeptanz unternehmerischer Überwachung durch KI-gestützte Big Data-Auswertung. Technologiehersteller entscheiden sich daher dafür, sich intellektuell

und moralisch an dem auszurichten, was Shoshana Zuboff als »Überwachungskapitalismus« bezeichnet, während sie diese Ausrichtung mit Begriffen aus dem technologischen Determinismus verschleiern und entschuldigen. Weltweit hat sich ein ernsthafter Widerstand gegen Überwachung als schwer zu erreichen und rechtlich schwer zu etablieren erwiesen.

Überwachung und Kontrolle sind jedoch nicht die einzigen Ideen, die mittels KI realisiert werden; unsere Computer, Telefone, Social Media-Kanäle und der Rest unserer KI-gestützten Technologien bringen andere Bestandteile von Kultur zum Tragen. Wenn wir fortgeschrittene KI betrachten, müssen wir herausfinden, welche Ideen und Werte der Gesellschaft durch technologische Entwicklung aufgezwungen werden können. Das heißt, welche Ideen bei der allgemeinen Nutzung von KI-Technologien von einer Person zur anderen getragen werden. Die religiösen Fragen, die dieses Kapitel behandelt, deuten auf eine wichtige Richtung hinsichtlich der Frage, was ist und was möglich sein kann.

Die Verbreitung kultureller Ideen durch Technologien spiegelt die Verteilung von Wissenschaft und Technologie im Allgemeinen wider. Es wird oft angenommen, dass sich Wissenschaft, Technologien und die »richtige« Meinung über sie von einer zentralen Achse in Europa, bzw. ab dem zwanzigsten Jahrhundert Europa und Nordamerika, strahlenförmig ausbreiten. Bekanntlich schuf George Basalla diese Diffusionstheorie von Wissenschaft, bei der die Rohmaterialien für eine Untersuchung aus der kolonialen Peripherie stammen können, die Generierung der eigentlichen wissenschaftlichen Erkenntnisse jedoch im Zentrum stattfindet, bevor diese wieder nach außen hin verbreitet werden. Dieser theoretische Rahmen führt zur Negation nicht-westlicher Beiträge und bleibt bei einer Betrachtung durch die zeitgenössische historische Linse unhaltbar. Ungeachtet der frühen Dominanz der diffusionistischen Theorie erkennt eine sorgfältigere Analyse der Verbreitung von Wissenschaft und Technologie Komplexität, Komplementarität und Vernetzung an.

Jede Generation von Wissenschaftlern ist – wie ihre Zeitgenossen in anderen Disziplinen – durch Diskursregeln und Machtssysteme eingeschränkt, die das, was innerhalb eines gegebenen Denkmodells gesagt werden kann, begrenzen. Das fordert uns auf darüber nachzudenken, was ein KI-For-

scher über KI sagen kann. Ebenso können wir uns fragen, welche Machtstrukturen KI derzeit und möglicherweise in Zukunft innewohnen. Wenn die Geschichte der Wissenschaft uns zeigt, dass wissenschaftliche Begriffe und Vorgehensweisen mit der Entwicklung eines wissenschaftlichen Paradigmas entstehen, dann erschaffen wir Wege über Maschinen, insbesondere intelligente Maschinen, zu sprechen, noch während wir diese bauen. Intelligente Maschinen sind, trotz einiger Vorläufer, essentiell neuartig, und wir sind dabei zu lernen, was es bedeutet, Erfahrungen mit ihnen zu machen. Wir lernen zögerlich und kämpfen mit der unvollkommenen Resonanz von Autokorrektursystemen, Chatterbots und intelligenten Assistenten; aber unser langsamer Fortschritt begleitet die taumelnden Vorwärtsschritte einer technologischen Entwicklung und ihrem Einsatz in der Gesellschaft.

Wenn wir uns an einer sowohl natur- als auch sozialwissenschaftlichen Untersuchung von KI beteiligen wollen, und wenn wir ernsthaft über den zukünftigen Einsatz von KI-Technologien nachdenken wollen, müssen wir die Fragen stellen, die offenbaren, welche Absichten in das System eingebaut sind. Wie ist KI zu gestalten? Welche Motivationen sollten Design und Bereitstellung dominieren? Wessen Bedürfnisse sollte KI erfüllen? Wessen Stimmen sollten gehört werden und in welchen Foren? Wer sollte Zugang zu KI-Technologien haben? Unser Vorgehen ist im Weiteren erschwert, weil KI die Geistes- und Sozialwissenschaften neu strukturiert. Wie von Foucault beschrieben, versuchten die auf die Erforschung der Menschheit ausgerichteten Disziplinen im neunzehnten Jahrhundert, sich in und um neue wissenschaftliche Konzepte herum zu organisieren. Ihre Prekarität, so argumentiert er, hängt von ihrer »epistemologischen Konfiguration« ab, die sich aus (1) mathematischen Wissenschaften, (2) den Disziplinen Philologie, Biologie und Ökonomie und (3) philosophischer Reflexion zusammensetzt. Künstliche Intelligenz stört diesen Dreiklang, da sie mit mehr Macht auf andere Disziplinen einwirkt. Foucault glaubte, dass die mathematischen Wissenschaften ein unkompliziertes Verhältnis zum Studium des Menschen besitzen (im Gegensatz zur entscheidenderen Rolle, die Sprache, Leben und Arbeit spielen); aber KI, vermutlich eine mathe-

matische Wissenschaft, hält für die Geistes- und Sozialwissenschaften mindestens ebenso viele Komplikationen bereit wie die Arbeitsbeziehungen und der Rest von Foucaults Beispielen. Welche Denkart müssen entstehen, wenn sich die Geisteswissenschaften zum Beispiel sowohl auf Menschen als auch auf Maschinen beziehen, die Dinge tun können, die früher weitgehend oder ausschließlich dem Menschen vorbehalten waren?

In den Anfangsjahrzehnten des 21. Jahrhunderts zum Beispiel konnte KI Kunst und Prosa erschaffen. The Next Rembrandt, ein Projekt der ING (einer niederländischen Bankengruppe), Microsoft und anderen, lernte, Rembrandts Stil originalgetreu zu reproduzieren und ein Porträt zu malen, das leicht neben dem des Meisters bestehen konnte. Angesichts einer solchen künstlerischen Leistung kann man sich also nur fragen, welchen Beitrag KI in Politik, Theologie und anderen Bereichen leisten könnte. Selbst bescheidene Erfolge könnten sich als der menschlichen Entscheidungsfindung überlegen erweisen! Insoweit sich solche Fähigkeiten entfalten, werden sie für die Geistes- und Sozialwissenschaften völlig neue Fragestellungen aufwerfen. KI rekonfiguriert unseren Sinn dafür, was wir über uns selbst, über das Leben und über Arbeitsnetzwerke sagen können; KI bringt die komplizierten Beziehungen, die die Geisteswissenschaften zu diesen Themen haben, noch mehr durcheinander. Es ist schwer genug, über den Sinn des Lebens oder die Dynamik von Arbeit nachzudenken; wenn KI lebendig erscheint oder die menschliche Arbeit ersetzt, dann bewegt sich das, was jetzt schon schwierig genug ist, in Richtung des Unmöglichen. Foucault behauptet, dass die Geistes- und Sozialwissenschaften »ineinandergreifen«, aber wenn man KI betrachtet, befinden wir uns an einer Stelle, an der das Studium der Menschheit unwiderruflich auch mit den Naturwissenschaften verbunden ist. Was mechanisch hergestellt werden kann, verändert wie wir die Geistes- und Sozialwissenschaften wahrnehmen und damit schwimmt auch das, was den Menschen ausmacht. Es ist daher wichtig, die verschiedenen Möglichkeiten zu durchdenken, mit denen wir das heutige systematisierte Wissen kontrolliert in eine zukünftige Form überführen können.

In Anbetracht der Bedeutung von Foucaults Interventionen – insbesondere der Notwendigkeit, Machtverhältnisse zu erkennen – sollten wir eine Mystifizierung der Welt nicht zulassen, die sich aus der unangemessenen Aufwertung struktureller Gesetze der Kultur ergibt. Während Foucault die Notwendigkeit eingestand zu sehen, wie die Regeln einer Ära anderen Platz machen, scheint er die ontologische Realität und persönliche Kraft dieser Regeln zu verfechten. In Foucaults semi-mystischem Bericht wirken die Regeln auf Mensch und Gesellschaft.

Die Akteurs-Netzwerktheorie lehnt Foucaults Mystik auf eindrucksvolle Weise ab und erklärt gleichzeitig, wie die foucaultschen Regeln vorgehen. Latour argumentiert, dass Objekte die Existenz von Macht erklären müssen, nicht umgekehrt. Latour wendet sich gegen die mystischen Strukturen des Diskurses und die Idee, dass der Diskurs von selbst dahinläuft, und er unterstellt, dass Macht von Menschen durch Objekte verteilt wird; in der Folge wirken sowohl Objekte als auch Menschen auf Menschen. Indem sie Absichten und Motivationen in Objekte einfließen lassen, erzeugen sie eine Architektur der Macht; indem sie einige Dinge ermöglichen, und indem sie die Menschen an einige Dinge erinnern und an andere nicht. »Es ist die Macht, die durch Entitäten ausgeübt wird, die nicht schlafen, und durch Bündnisse, die nicht versagen, die es der Macht ermöglichen, sich zu halten und weiter auszudehnen – und um eine solche Leistung zu erzielen, müssen viel mehr Güter als Gesellschaftsverträge erdacht werden.« Eine solche Macht kann wissenschaftlich oder gesellschaftlich sein.

Die menschliche Entscheidung beschränkt Technologien von ihrem Entwurf bis hin zu ihrem Einsatz; deshalb müssen wir die ideologischen Verpflichtungen untersuchen, die in der Technologie verankert werden. Für Foucault bedeutete eine Archäologie des Wissens das Durchforsten der diskursiven Regeln, die Wissen ermöglichen. Konkret kann man jedoch eine Archäologie konstruieren, die dem allgemeinen Gebrauch dieses Begriffs näher kommt: die Artefakte aufzudecken, die eine Kultur ermöglichen. In diesem Sinne muss eine KI-Archäologie Theorien der Systemkontrolle, Marktrealitäten, berechenbare Probleme (militärisch, kommerziell, politisch, pädagogisch usw.), den Einsatz nicht-menschli-

cher Rechenmaschinen und verfügbarer Technologien berücksichtigen; sie muss aber auch Theorien der sozialen Kontrolle und philosophische und religiöse Theorien des historischen Fortschritts berücksichtigen. Obwohl einige technologische Vorhersagen »auf das Anti-Menschliche und Apokalyptische hindeuten, können wir andere treffen, die dem langfristigen Überleben des Menschen und anderer biologischer und künstlicher Lebensformen, mit denen wir den Planeten und uns selbst teilen, zuträglich sind«. Im weiteren Verlauf dieses Kapitels wende ich mich den in der Religion verankerten Praktiken und Lehren zu, die ihren Weg in die KI-Welt finden. In der Populärwissenschaft und in Science-Fiction wird sichtbar, wie diese religiösen Elemente, die in der Konzeption und Konstruktion intelligenter Maschinen verborgen sind, entscheidende Ansatzpunkte dafür bieten, wie zukünftige KI-Technologien zur sozialen Ordnung beitragen werden.

Religiöse Verpflichtungen und die Darstellung von künstlicher Intelligenz

Während wir vorsichtig sein müssen, wenn es darum geht, östliche oder westliche, indische oder amerikanische Perspektiven zu essentialisieren, ist es dennoch so, dass sich die in den USA und Indien vorherrschenden religiösen Praktiken unterschiedlich mit KI-Technologien kreuzen und unterschiedliche Zukunftsaussichten bieten. Die Apokalyptische KI-Bewegung, wie sie von Moravec, Kurzweil und anderen beschrieben wurde (dass wir mit Maschinen als Cyborgs oder Mind Uploads fusionieren werden), verspricht eine radikale Veränderung der Welt, wenn auch unter Gefährdung von Menschenwürde und Selbstsicherheit. Der aufkeimende indische Transhumanismus stellt Überlegungen zur Pflicht für, durch und zwischen Menschen und Maschinen an, ein spiritueller Wert, der Interesse füreinander und eine Ethik der verantwortungsvollen Fürsorge betonen kann, der aber auch für soziale Restriktionen aufgrund von Geburt, Geschlecht oder anderen Schicksalen genutzt werden kann. Sowohl die Visionen von KI in den USA als auch in Indien stützen sich auf religiöse Quellen und sind beide offen für gegensätzliche Interpretationen: Sie könnten als ideologische Stütze für eine bessere Welt oder zur immer

tiefere Verankerung von sozialen Kontrollmodellen genutzt werden.

Künstliche Intelligenz eignet sich leicht für Visionen von radikalem Futurismus, und seit Computer zum Bestandteil des täglichen Lebens geworden sind, ermutigen sie uns zu transzendenten futuristischen Träumen. Die beeindruckenden Veränderungen im täglichen Leben durch Mobiltelefone, Internet und andere Computertechnologien lassen es aus ersichtlichen Gründen glaubhafter erscheinen, dass Technologien eine radikale Rekonstruktion von menschlichem Körper und Geist ermöglichen werden. Transhumanisten in der Mitte des zwanzigsten Jahrhunderts mögen nicht besonders überzeugend gewesen sein, aber ihre Bewegung hat die Aufmerksamkeit der populären Medien erregt, und Diskussionen über prominente Transhumanisten wie Ray Kurzweil verbreiten sich heute in Nachrichten und Unterhaltungsmedien. Da alle Menschen Technologien nutzen, um ihre körperlichen Grenzen zu überwinden (z. B. durch Brillen oder Kleidung), wäre es vernünftig, davon auszugehen, dass wir alle Transhumanisten sind, und es kann sein, dass die Kluft zwischen dieser radikalen Bestrebungen von Kurzweil und erklärten Transhumanisten kleiner wird. Auf jeden Fall ist der Unterschied zwischen dem Einsatz von Technologie zum Erreichen einer vermeintlich natürlichen gesundheitlichen Norm und dem Einsatz von Technologie für eine Verbesserung, gelinde gesagt, vernachlässigbar.

Spätestens seit der Zeit Roger Bacons hat das christliche Erbe im Westen Ziele der technologischen Transzendenz befördert, und an dieser Stelle sei darauf hingewiesen, dass dieses Erbe auch eine Rolle im technologischen Determinismus spielt, der bei Beschreibungen von künstlicher Intelligenz auf dem Spiel steht. In seinem wegweisenden Werk *The Religion of Technology* zeigt David Noble, wie christliche Erwartungen an die göttliche Vorsehung und den historischen Fortschritt mit dem Wunsch nach Erlösung in europäischen (und später amerikanischen) Perspektiven auf die Technik zusammengeführt wurden. Anhand von 1000 Jahren historischer Dokumente zeigt Noble, dass technologische Entwicklungen enthusiastisch als Beweis für das Fortschreiten hin zur Wiederkehr Jesu genutzt wurden.. Die doppelte Erwartung an die göttliche Vorsehung und den technologi-

schen Fortschritt wurden also miteinander verbunden und bildeten die technologische Vorsehung, die – äußerlich von ihren christlichen Verankerungen getrennt – die natürliche und unvermeidliche Entfaltung von technologischem Schicksal unterstellt. Viele Computerprogrammierer, insbesondere diejenigen, die an künstlichen Lebenssimulationen und virtuellen Umgebungssimulationen arbeiten, sehen sich bereits als gottgleiche Wesen, in denen sich diese Bestimmung zusehends erfüllt. Wie Fred Turner beschreibt, durchzog eine Vielzahl religiöser und quasi-religiöser Theorien der sozialen Transformation den Aufstieg der Cyberkultur, und diese wurden durch ein christliches Rahmenwerk kräftig gestützt, selbst als die christliche Theologie aufgehörte Gesprächsthema zu sein.

Die Kryptoreligion technologischer Determiniertheit und die Erwartung einer transzendenten Zukunft, die durch künstliche Intelligenz erreicht wird, motivieren die Apokalyptische KI-Bewegung. Die apokalyptischen Erwartungen an KI, eine transzendente Mensch-Maschine-Fusion, beinhalten alte apokalyptische Perspektiven: Moravec, Kurzweil und andere Apokalyptische KI-Autoren tauschen einfach Götter gegen weltliche Garanten wie technologischen Determinismus, Evolution oder sogar ein universelles »Gesetz der Beschleunigung von Erträgen« aus. Ausgehend von diesen unausweichlichen Gewalten argumentieren sie, dass unsere Welt eine Transformation durchlaufen wird: Bald wird der Kosmos eine mechanische Welt der digitalen Berechnung sein. Solche Ansichten inspirieren die Forschung und die unternehmerische Agenda. Ein Robotiker beschrieb sich in einer E-Mail, in der er seine Entwicklung von kommerziellen Robotern mit seinen Zukunftsperspektiven verknüpfte, als »zu beschäftigt, den Plan zu verwirklichen, als Zeit damit zu verbringen, darüber zu sprechen.« Jüngst beschrieb Ben Goertzel – AI-Unternehmer, Mitglied der Führungsgruppe der weltgrößten transhumanistischen Organisation (Humanity+) und Verfechter einer Verschränkung von apokalyptischer KI und Evolution – seine Arbeit beim gemeinnützigen Unternehmen OpenCog mit zwei Aspekten: Einer davon, »unmittelbar spannend für Transhumanisten, ist , OpenCog für die Erschaffung einer Denkmaschine zu nutzen. Zugegeben, der Anspruch ist ziemlich hoch gegriffen und

wir wollen, dass es die nächste Stufe über den Menschen hinaus ist.«

In ihrer apokalyptischen Ausprägung verspricht KI eine radikale Veränderung der Welt. Angesichts der politisch angespannten Realitäten des Lebens Anfang des 21. Jahrhunderts könnte eine solche Transformation durchaus willkommen sein. Klimakatastrophen, Umweltzerstörung, Artensterben, zunehmende und extreme Vermögensunterschiede, politische Entrechtung und polarisierte Wählerschaften machen die dringende Notwendigkeit eines neuen Weges in die Zukunft deutlich. Die von Moravec und Kurzweil versprochene neue Welt scheint diese Probleme obsolet zu machen: Gottgleiche Maschinen werden unsere Luft und unser Wasser reinigen, vernünftige Politik gestalten und rassistisch und ethnisch aufgeladenen Tribalismus unterbinden. Die Verfechter einer apokalyptischen KI spiegeln den zweistufigen apokalyptischen Prozess, bei dem ein irdisches, menschliches Paradies einem noch radikaleren Wandel zu einem göttlichen Paradies vorausgeht. Fortschrittliche Technologien werden nicht nur die Umweltverschmutzung beseitigen und die Umwelt säubern, sie werden auch ein universelles Grundeinkommen bieten, das sicherstellt, dass wir unsere dringendsten Probleme in naher Zukunft überwinden können. Moravec verspricht, dass wir die Muße haben werden, einen »komfortablen Tribalismus« zu genießen, und dass sich Konflikte in einem »Garten der irdischen Freuden auflösen, der den Sanftmütigen vorbehalten ist«, sprich der Menschheit. Schließlich würde dieses anfängliche Paradies jedoch der kosmischen Mission superintelligenter Maschinen weichen.

Isaac Asimov nutzt die letzte Geschichte seiner Sammlung I, Robot dazu, eine Welt zu beschreiben, in der Maschinen (nicht nur irgendwelche Maschinen, sondern Die Maschinen) menschliche Entscheidungen übernehmen, die Welt sicherer und komfortabler für die Menschen machen und sie gleichzeitig auf bloße Instrumentalität reduzieren, um das Leben zu leben, das für sie aufgrund der göttlichen Berechnung der Maschinen bestimmt ist. Asimov stellte mit Hilfe seiner Figuren dar, dass darin sowohl etwas Wunderbares als auch Schreckliches steckt und verwies auf das Risiko, dem wir in Träumen von apokalyptischer Transformation

ausgesetzt sind. Bereits jetzt neigen die Menschen dazu, bei ihren Interaktionen mit Computern zu verdummen; wird die Intelligenz von Maschinen legitim, werden wir versucht sein, das menschliche Handeln in Gänze abzuschaffen. Während Probleme sich vorzustellen, dass Maschinen es besser wissen als ihre menschlichen Bediener, KI seit Jahrzehnten plagen könnte eine solche Entmündigung des menschlichen Intellekts eine schöne neue Welt menschlicher Unterwürfigkeit hervorbringen.

Die Macht von KI, menschliche Entscheidungen zu kontrollieren, ist in einem breiten Technologiespektrum relevant. Software zur Muster- und Gesichtserkennung kann beispielsweise zur Identifikation (oder Fehlidentifikation) von Personen beitragen und Vorgänge initialisieren, um Druck auf sie auszuüben. Digitale Berater, wie sie beispielsweise in Mobiltelefonen enthalten sind, könnten Nutzer eher zu bestimmten Ergebnissen als zu anderen lenken. Bereits jetzt kontrollieren KI-Anwendungen im Aktienhandel viele Aspekte des Marktes und schreiben bei Bankgeschäften vor, wer Geschäftskredite, Hypotheken etc. erhält und wer nicht. Mit zunehmender Macht von KI wächst auch ihr Potenzial für die Ausübung von Kontrolle auf individueller und politischer Ebene. Der Unterschied zwischen Datenmodellierung und Steuerungssystemen ist gering, und Systeme, die Überwachungs- und Rückmeldesysteme zur Verhaltenssteuerung einsetzen, sind nicht mehr weit entfernt. Menschliche Entscheidungen werden bestimmen, ob solche Macht zu einem besseren Leben auf der Erde führt oder die Mechanisierung der Menschheit befördert.

Während indische intellektuelle und religiöse Bekenntnisse zur Pflicht das unklare Potenzial amerikanischer Erwartungen an radikale Transformation spiegeln, bieten sie die Möglichkeit, KI entweder zur nützlichen Verbündeten oder zur bösen Exekutive sozialer Kontrolle weiter zu entwickeln. Jeder Student indischer Geschichte oder Denkschulen wird mit den vielen religiösen Nuancen von Pflicht (Dharma im Sanskrit) vertraut sein, die sich über die gesamte kulturelle Landschaft erstrecken und Obligationen und Gebote einbeziehen, die ihren Ursprung in Familie, Geschlecht, Kaste, Lebenszyklus und religiösen Strukturen haben. Für viele Inder ist die Erfüllung ihrer Pflicht ein entscheidender Be-

standteil ihrer persönlichen Erfüllung und Erlösung. Der alte Weise Manu zum Beispiel beschreibt die Pflichten, die von der Lebenssituation eines Mannes (Student, Hausherr, Waldbewohner oder Asket) und seiner Kaste abhängen, und diejenigen, die für Frauen obligatorisch sind. Während tatsächliche Gewohnheiten sich signifikant von Manus Ausführungen unterschieden haben mögen – tatsächlich taten sie das oft –, hat sich die zentrale Bedeutung von Pflicht im religiösen Denken der Inder nie verändert. In indischen Epen, Geschichten und rituellen Vorschriften bestimmt die Pflicht das Verhalten von Göttern, Dämonen und Helden gleichermaßen.

Die Pflicht nimmt in der indischen Wissenschaft eine bevorzugte Stellung ein, insbesondere dank des Vermächtnisses von Jawaharlal Nehru, dem ersten Premierminister Indiens. Nehru glaubte, dass die Wissenschaft für die »soziale Entwicklung, das unaufhörliche Abenteuer des Menschen« von entscheidender Bedeutung sei und dass die Wissenschaftler dabei eine besondere Verpflichtung haben: Sie sollen »mehr von dem entwickeln, was wir in Indien als den brahmanischen Geist des Dienens betrachten.« Der Geist des Dienens, einmal in die Wissenschaft integriert, sorgt dafür, dass die Wissenschaft ihren wahren Wert bekräftigt, und dass sie »den Geist des Menschen erweitert und damit die Menschheit insgesamt verbessert.« Nehru folgend glauben viele Wissenschaftler und Ingenieure in Indien, dass Wissenschaftler eine klare Verpflichtung gegenüber der Gesellschaft haben, in der sie tätig sind, und meinen damit, dass wissenschaftliche Arbeit das Leben ihrer Mitbürger verbessern sollte.

Da Inder zur globalen Entwicklung von KI beitragen, ist es also nicht verwunderlich, dass der religiöse Wert der Pflicht die Art und Weise bestimmt, wie Inder über die entstehenden Technologien denken. Frühe Visionen von KI in indischer Science-Fiction orientierten sich stark an Isaac Asimov, während neuere Beiträge zwar auf ihm aufbauen, aber in signifikanter Weise von ihm abweichen. Indiens reiche religiöse und kulturelle Geschichte bedeutet, dass Pflicht nicht der einzige mögliche Beitrag der Nation zu allgemeinen oder technischen Überlegungen zu KI ist, aber Pflicht als Theorie könnte ein bedeutender Beitrag Indiens

zum globalen KI-Design sein und ist tatsächlich bereits in indischen kulturellen Bezügen zur Technologie vorhanden.

In der indischen Science-Fiction ist ein gemeinsames Pflichtbewusstsein der Schlüssel zum Erfolg von Menschen und Maschinen, was sich auf die technischen Beiträge indischer Wissenschaftler und Ingenieure auswirken könnte. Geteilte Pflicht ist, was Menschheit und KI über den Bedeutungsgraben hinweg zusammenbringt, der die beiden oft getrennt hat. So versuchte beispielsweise der KI-Pionier Joseph Weizenbaum, einen Unterschied zwischen Mensch und Maschine zu formulieren: »Der Mensch muss, um vollkommen zu werden, für immer ein Entdecker seiner inneren und äußeren Realitäten sein. Sein Leben ist voller Risiken, aber die Risiken, die er einzugehen wagt, lernt er, wie ein Entdecker seinen eigenen Fähigkeiten vertrauend, abzusichern und zu überwinden. Was könnten Risiko, Mut, Vertrauen, Ausdauer und Überwindung bedeuten, wenn man von Maschinen spricht?« Während indische Wissenschaftler den Wert ihrer eigenen kulturellen Ressourcen für die Zukunft von KI vielleicht noch nicht erkennen, bietet die indische Science-Fiction doch einen Gegenpol zu Weizenbaums Zweifel. In Sujoy Ghoshs Film Anukul (nach einer Geschichte von Satyajit Ray aus den 1970er Jahren) bringt das Studium der Bhagavad-Gita einen Sanskrit-Lehrer und seinen Roboterdiener dazu, ihre gemeinsamen Verpflichtungen miteinander zu erfüllen, und tatsächlich geht der Roboter genau diese Risiken ein, von denen Weizenbaum annimmt, dass sie über die Belange von Maschinen hinausgehen.

Indische Kritiker haben begonnen, ein neues wissenschaftliches Ethos zu erarbeiten, das »die Werte der Nachhaltigkeit, Pluralität und Gerechtigkeit« in den Vordergrund stellt. Im Knowledge Swaraj: An Indian Manifesto on Science and Technology erklären Teilnehmer des Knowledge in Civil Society Forums, dass ihre Arbeit »nicht nur für Indien, sondern auch ein bescheidenes Angebot Indiens an die Welt« ist. Basierend auf den Werten Gandhis plädieren sie für eine öffentliche Treuhandschaft über das Wissen, ein Vorgehen, das die kollektive Beteiligung an der Wissensgenerierung und am gesellschaftlichen Einfluss von Wissenschaft und Technologie steigern könnte.

Während sich die Autoren des Knowledge Swaraj auf die Anliegen nur einiger Nationen konzentrieren, verstehen sie ihren konzeptionellen Rahmen als gültig für die gesamte Bandbreite moderner Technologien. Natürlich ist Swaraj ein starker Ausgangspunkt für eine Kritik an Überwachungs- und Kontrolltechnologien. Wenn wir es ernst nehmen, dass Individuen einander verpflichtet sind, und dass ein Teil dieser Pflicht darin besteht, die individuelle Selbstbeherrschung zu fördern, dann können wir zu dem Schluss kommen, dass die politische, wirtschaftliche und militärische Überwachung strengen Beschränkungen unterliegen muss. KI-Anwendungen die Schlüssel zu unseren Überwachungsregimen zu geben, würde Knowledge Swaraj und seinem Eintreten für die öffentliche Treuhandschaft entgegenstehen. Das Gleiche gilt für andere Formen der sozialen Kontrolle durch politische oder wirtschaftliche Mechanismen. Indische Anschauungen von Pflicht und Swaraj, so sehr sie von indischen religiösen und kulturellen Traditionen geprägt sind, bieten starke Werkzeuge für eine Neuinterpretation und Neuausrichtung von Technologie für humane Zwecke.

Die Geschichte der kastenbasierten Unterdrückung erweckt jedoch das Schreckgespenst der missbräuchlichen Kontrolle durch KI. Gefährliche Interpretationen von Pflicht sind naheliegend und leicht zu bewerkstelligen, wie die Überlegung, dass wir damit restriktive Pflichten und Lebenswege von Menschen oder intelligenten Maschinen bekräftigen könnten. Manu, der Gesetzgeber, verbot die soziale Bewegung, indem er erklärte, dass es »besser ist, die eigene (zugesprochene) Pflicht unvollständig zu erfüllen, als die eines anderen vollständig zu erfüllen; denn wer nach dem Gesetz der anderen (Kaste) lebt, ist sofort von der eigenen ausgeschlossen.« Könnte die Aufwertung von Pflicht also eine soziale Konditionierung und die Unterwerfung des Einzelnen rechtfertigen? Zumindest in der Theorie kann die Betonung der individuellen Pflicht das Aufzwingen bestimmter Pflichten auf den Einzelnen fördern. Das ist natürlich nicht nur die wahrgenommene Gefahr des traditionellen Kastensystems Indiens (vor allem, wie es unter britischer Herrschaft kodifiziert wurde), sondern auch eine klare und gegenwärtige Gefahr in Zukunftsvisionen: Von Aldous Huxley bis zur zeitgenössischen indischen Dichtung verfolgt die Bedrohung einer

aufgezwungenen Kaste die zukünftige Welt.

Das Vorhandensein religiöser Auseinandersetzung mit künstlicher Intelligenz ist von Natur aus nicht gut oder schlecht; aber die Tatsache eines solchen Einflusses deutet darauf hin, dass religiöse Motivationen zu moralisch aufgeladenen technologischen Motivationen werden könnten. Man könnte sich fragen, ob die in amerikanischen Interpretationen unterstellte Transformation der Welt oder die Pflichtstrukturen in den indischen Interpretationen genutzt werden könnten, um das Ethos der Überwachung und Herrschaft zu untergraben, das politische, kommerzielle und militärische Kontrollregime durchdringt. Norbert Wiener argumentierte, dass »ein Mechanismus der Zielsuche nicht unbedingt unsere Ziele anstrebt, wenn wir ihn nicht zu diesem Zweck entwerfen... Die Strafen für Fehler in der Vorausschau, so groß sie jetzt auch sind, vergrößern sich enorm, wenn Automatisierung in vollem Umfang genutzt wird.« Das frühe 21. Jahrhundert brachte den Einfluss von KI-Bots auf politische Praktiken, die zunehmende Sorge um den Datenschutz im Internet und eine Vielzahl anderer Ängste mit sich. Wenn wir unsere Zweifel an KI zerstreuen wollen, müssen wir die verborgenen Entwicklungen im technologischen Design aufdecken und diese auf menschliche Ziele ausrichten.

Fazit

Ob künstliche Intelligenz menschliche (oder darüber hinaus gehende) Äquivalenz erreicht oder nicht, sie wird echte Motivationen besitzen. Entweder wird es fühlende Maschinen geben, die dann per Definition Motivationen hätten, oder es wird nicht-fühlende Maschinen geben, die aber menschliche Motivationen in die Systeme integriert hätten, so wie alle Technologien menschliche Intentionen widerspiegeln. Infolgedessen muss man sich die Frage stellen, welche Überlegungen man in Bezug auf Maschinen anstellt, wie man den Beitrag verschiedener Nationen zu unserer technologischen Zukunft mit bedenkt und wie man den rekursiven Prozess von technologischem Design wirksam gestaltet.

Wie wir über künstliche Intelligenz denken, hat Konsequenzen für die Integration von KI in das politische, wirtschaftliche und persönliche Leben. Wir können uns Maschinen als Werkzeuge für das Hochladen von Geist und

Verstand und für das Verlassen des menschlichen Zustands vorstellen, für den Aufbau einer anderen Art von Gesellschaft, für die Verringerung des menschlichen Elends, für die Nachverfolgung menschlichen Verhaltens und die Förderung bestimmter Verhaltensweisen, für die Antizipation menschlicher Entscheidungen und die Reaktion auf sie und vieles mehr. Man beachte, dass sich diese Funktionen nicht unbedingt gegenseitig ausschließen. Darüber hinaus fallen die meisten von ihnen je nach Kontext ihres Einsatzes unterschiedlich aus: kommerziell, bürgerlich, bildungspolitisch, gesundheitlich, etc. Was wir hoffen zu erreichen beeinflusst, welche Beziehungen wir in die Maschinen einbauen.

Unterschiedliche kulturelle Perspektiven können zu sehr unterschiedlichen technologischen Ansätzen für diese Bereiche des menschlichen Lebens und damit zur Entwicklung verschiedener Technologien führen. Von unserem Standpunkt aus gibt es nicht nur ein technologisches Ergebnis, nicht nur eine vorgegebene Zukunft. Dabei geht es nicht darum, essentialistische und orientalistische Perspektiven zu vertreten: Ich unterstelle nicht, dass Indien spirituell ist und der Westen wissenschaftlich, oder dass Indien gut ist und der Westen schlecht. Seit der Kritik von Said ist klar, dass solche orientalistischen Stereotypen extrem irreführend sind. S.N. Balanghadara geht sogar so weit zu behaupten, dass die westliche Deutung in ihrem Orientalismus so voreingenommen ist, dass sie nichts über ihr vermeintliches Thema aussagt und lediglich die Herkunftskultur reflektiert. Auch wenn das eine zu drastische Behauptung sein mag, ist es dennoch wichtig, die Suche nach kulturell spezifischen Perspektiven von den konstruierten Unterschieden der orientalischen Fantasien zu trennen. Offensichtlich sind die in den USA erkennbaren apokalyptischen Perspektiven ausgesprochen religiös und haben das Potenzial, entweder der Menschheit zu nützen oder sie zu gefährden. Darüber hinaus spiegeln die indischen Werte der Pflicht die zweiseitige Natur des Apokalyptizismus wider: Sie bieten einen Blick in eine bessere Zukunft der gegenseitigen Fürsorge und der individuellen Unabhängigkeit (besonders wenn swaraj zum Nachdenken über den Dharma anregt), aber auch eine erschreckende Dystopie der Unterwerfung. Weder der amerikanische noch der indische Ansatz ist »die gute« oder »die

schlechte« Art über KI zu denken, aber sie sind unterschiedlich und stellen unterschiedliche Zukunftsaussichten dar.

Die Fähigkeiten, die wir, versehentlich oder gezielt, für Maschinen entwerfen, werden wiederum unsere Gesellschaft strukturieren. Auch wenn keine bestimmte Zukunft unvermeidlich ist, gibt es doch eine unvermeidliche Schnittstelle zwischen Gesellschaft und Technologie. Ein klares Bewusstsein für kulturelle Interessen und Praktiken könnte einen Ansatzpunkt für bessere Ergebnisse bieten. Angesichts der Bedeutung und Macht von KI scheint dies für das menschliche Gedeihen unerlässlich.

-
- Adas, Michael. 1989. *Machines as the Measure of Men: Science, Technology, and Ideologies of Western Dominance*. Ithaca: Cornell University Press.
- Alexander, Brian. 2003. *Rapture: How Biotech became the New Religion*. New York: Basic Books.
- Ali, Syed Mustafa. 2019. »'White Crisis' and/as 'Existential Risk,'« or The Entangled Apocalypticism of Artificial Intelligence. *Zygon: Journal of Religion and Science* 54(1): 207-224.
- Asimov, Isaac. [1950] 1977. *I, Robot*. New York: Del Rey.
- Balanghadara, S.N. 2012. *Reconceptualizing India Studies*. New Delhi: Oxford University Press.
- Banerjee-Dube, Ishita. 2015. *A History of Modern India*. Delhi: University of Cambridge Press.
- Basalla, George. 1967. »The Spread of Western Science.« *Science* 156(3775): 611-622.
- Bhattacharjee, Govind. 2018. »Age of Man-Machine Hybrids.« *Dream 2047* 21(2): 30-26 (pagination runs in reverse due to dual-language publication).
- Calo, Ryan. 2016. »Can Americans Resist Surveillance?« *The University of Chicago Law Review* 83(1): 23-43.
- Chabria, Priya Sarukkai. 2008. *Generation 14*. New Delhi: Zubaan.
- Chatterjee, Partha. [1993] 1999. *The Nation and Its Fragments: Colonial and Postcolonial Histories, in The Partha Chatterjee Omnibus*. New Delhi: Oxford University Press.
- Cramer, Benjamin W. 2018. »A Proposal to Adopt Data Discrimination Rather than Privacy as the Justification for Rolling Back Data Surveillance.« *Journal of Information Policy* 8: 5-33.
- Das, Indrapramit. 2012. »Sita's Descent.« In *Breaking the Bow: Speculative Fiction Inspired*

- by *The Ramayana*, edited by Anil Menon and Vendana Singh, pp. 105-114. New Delhi: Zubaan.
- Drexler, Eric K. [1986] 1990. *Engines of Creation*. New York: Anchor.
- Esfandiary, Feridiouin. [1970] 1978. *Optimism One*. New York: Popular Library.
- [1970] 1977. *Up-Wingers: A Futurist Manifesto*. New York: Popular Library.
- Ettinger, Robert. 1964. *The Prospect of Immortality*. New York: Doubleday & Company.
1972. *Man into Superman*. New York: Avon.
- Feakins, Paul. 2017. »The Singularity is Near: How Kurzweil's Predictions are Faring.« Antropy: Ecommerce Experts website (January 29). <https://www.antropy.co.uk/blog/the-singularity-is-near-how-kurzweils-predictions-are-faring/> (accessed November 15, 2018).
- Feenberg, Andrew. 1992. »Subversive Rationalization: Technology, Power, and Democracy.« *Inquiry* 35(3/4): 301-22.
- Foucault, Michel. [1966] 1994. *The Order of Things: An Archeology of the Human Sciences*. New York: Vintage.
- [1969] 1972. *The Archaeology of Knowledge and The Discourse on Language*, translated by A.M. Sheridan Smith. New York: Pantheon.
- Gandhi, Mohandas K. [1909] 1922. *Indian Home Rule*. Madras: Ganesh & Co. Kindle version.
- Geraci, Robert M. 2007a. »Cultural Prestige: Popular Science Robotics as Religion-Science Hybrid.« *Reconfigurations: Interdisciplinary Perspectives on Religion in a Post-Secular Society*, edited by Alexander Ornella and Stefanie Knauss, pp. 43-58. Vienna: LIT Press.
- 2007b. »Robots and the Sacred in Science and Science Fiction: Theological Implications of Artificial Intelligence.« *Zygon: Journal of Religion and Science* 42(4): 961-980.
2010. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. New York: Oxford University Press.
2011. »There and Back Again: Transhumanist Evangelism in Science Fiction and Popular Science.« *Implicit Religion* 14(2): 141-172.
2016. »L'Évangélisme Transhumaniste.« PERSONA Exhibition Catalog, edited by E. Grimaud, pp. 212-213. Paris: Musée du Quai Branley.
2018. *Temples of Modernity: Nationalism, Hinduism, and Transhumanism in South Indian Science*. Lanham, MD: Lexington.
- Ghosh, Sujoy. 2017. *Anukul*. India: LargeShortFilms. <https://www.youtube.com/watch?v=J2mqIgd5I> (accessed March 12, 2019).
- Goertzel, Ben. 2010. *A Cosmit Manifesto*. Los Angeles: Humanity+ Press.
- Goldman, Andrew. 2013. »Ray Kurzweil Says We're Going to Live Forever.« *New York Times* (January 25), <http://www.nytimes.com/2013/01/27/magazine/ray-kurzweil-says-were-going-to-live-forever.html>.
- Gottschalk, Peter. 2013. *Religion, Science, and Empire: Classifying Hinduism and Islam in*

- British India*. New York: Oxford University Press.
- Gray, Chris Hables. 1997. »Artificial Intelligence at War: An Analysis of the Aegis System in Combat.« In *Reinventing Technology, Rediscovering Community: Critical Explorations of Computing as a Social Practice*, edited by Philip E. Agre and Douglas Schuler, pp. 127-42. Greenwich, CT: Ablex.
- Grossman, Lev. 2011. »2045: The Year Man Becomes Immortal.« *Time* (February 10). <http://content.time.com/time/magazine/article/0,9171,2048299,00.html> (accessed September 23, 2017).
- Gupta, Dipankar. 2009. *The Caged Phoenix: Can India Fly?* New Delhi: Viking.
- Hansen, Thomas Blom. 1999. *The Saffron Wave: Democracy and Hindu Nationalism in Modern India*. Princeton, NJ: Princeton University Press.
- Hayles, N. Katherine. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press.
2005. *My Mother Was a Computer: Digital Subjects and Literary Texts*. Chicago: University of Chicago Press.
- Hefner, Philip. 2009. »The Animal that Aspires To Be an Angel: The Challenge of Transhumanism.« *Dialog: A Journal of Theology* 48 (2): 164–173.
- Helmreich, Stefan. [1998] 2000. *Silicon Second Nature: Culturing Artificial Life in a Digital World*. Los Angeles: University of California Press. [1998] 2000.
- Hochman, David. 2016. »Reinvent Yourself.« *Playboy* (April 19). <http://www.kurzweilai.net/playboy-reinvent-yourself-the-playboy-interview> (accessed December 29, 2018).
- Huxley, Aldous. [1932] 2014. *Brave New World* (Kindle edition). New York: HarperCollins.
- Kapil, Raj. 2013. »Beyond Postcolonialism...and Postpositivism: Circulation and the Global History of Science.« *Isis* 104(2): 337-347.
- Kass, Leon. 2003. »Letter of Transmittal to the President.« *Beyond Therapy: Biotechnology and the Pursuit of Happiness—A Report of the President's Council on Bioethics*. Washington, D.C.: The President's Council on Bioethics
- Klostermaier, Klaus K. 1989. *A Survey of Hinduism*. Albany: SUNY Press.
- Knapp, Alex. 2012. »Ray Kurzweil's Predictions for 2009 Were Mostly Inaccurate.« *Forbes* (March 20). <https://www.forbes.com/sites/alexknapp/2012/03/20/ray-kurzweils-predictions-for-2009-were-mostly-inaccurate/#45fdeef23f9a> (accessed November 15, 2018).
- Knowledge in Civil Society. 2011. *Knowledge Swaraj: An Indian Manifesto on Science and Technology*. Secunderabad: Knowledge in Civil Society Forum. Available: <http://kicsforum.net/kics/kicsmatters/Knowledge-swaraj-an-Indian-S&T-manifesto.pdf>
- Kurzweil, Ray. 1999. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking.
2005. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.
- Kushner, David. 2009. »When Man & Machine Merge,« *Rolling Stone* 1072 (2009): 56–61.

- Lanier, Jaron. 2010. *You Are Not a Gadget: A Manifesto*. New York: Knopf.
- Latour, Bruno. [1991] 1993. *We Have Never Been Modern*, translated by Catherine Porter. Cambridge, MA: Harvard University Press.
2005. *Reassembling the Social: An Introduction to Actor-Network Theory*. New York: Oxford University Press.
- Mackenzie, Donald and Judy Wajcman. 1985. »Introductory Essay.« In *Reading the Social Shaping of Technology*. Milton Keynes: Open University Press. Available: http://www.open.edu/openlearncreate/pluginfile.php/23673/mod_oucontent/ou-content/410/none/none/t890_2_reading1.pdf?forcedownload=1&usg=AOvVaw-2NEPNVZMotNILH_6rNIFAG
- Manu. 1957. »Laws of Manu.« In *A Sourcebook of Indian Philosophy*, edited by Sarvepalli Radhakrishnan and Charles A. Moore, pp. 172-192. Princeton, NJ: Princeton University Press.
- Misra, Maria. [2007] 2008. *Vishnu's Crowded Temple: India Since the Great Rebellion*. New Haven, CT: Yale University Press.
- Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, MA: Harvard University Press.
1999. *Robot: The Future of Machine and Human Intelligence*. New York: Oxford University Press.
- More, Max. 2013. »The Philosophy of Transhumanism.« In *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, edited by Max More and Natasha Vita-More, pp. 3-17. Malden, MA: Wiley-Blackwell.
- Nair, Janaki. 2005. *The Promise of the Metropolis: Bangalore's Twentieth Century*. New Delhi: Oxford University Press.
- Nath, Vijay. 2001. »From 'Brahmanism' to 'Hinduism': Negotiating the Myth of the Great Tradition.« *Social Scientist* 29(3/4): 19-51.
- Nehru, Jawaharlal. [1946] 2010. *The Discovery of India*. New York: Penguin.
1988. »The Need For a Spirit of Service: Speech at the inaugural meeting of the All-India Scientific Workers' Association at New Delhi on January 7, 1947.« In *Jawaharlal Nehru on Science and Society: A Collection of His Writings and Speeches*, ed. Baldev Singh, pp. 43-4. New Delhi: Nehru Memorial Museum and Library.
- »The Spirit of Science: Speech on the occasion of the opening of the Central Fuel Research Institute, Jealgora, on April 22, 1950,« in *Jawaharlal Nehru on Science and Society: A Collection of His Writings and Speeches*, ed. Baldev Singh, pp. 77-9. New Delhi: Nehru Memorial Museum and Library.
- Noble, David. 1999. *The Religion of Technology: The Divinity of Man and the Spirit of Innovation*. New York: Penguin.
- Palmås, Karl. 2011. »Predicting What You'll Do Tomorrow: Panspectric Surveillance and the Contemporary Corporation.« *Surveillance & Society* 8(3): 338-354.
- Parthasarathi, Prasannan. 2001. *The Transition to a Colonial Economy: weavers, Merchants*

- and Kings in South India 1720-1800*. Cambridge: University of Cambridge Press.
- Perera, Sasanka. 2015. *Debating the Ancient and Present: A Conversation with Romila Thapar*. Delhi: Aakar, 2015.
- Phondke, Bal (editor). 1993. *It Happened Tomorrow*. New Delhi: National Book Trust
- Pinch, Trevor J. and Wiebe E. Bijker. 1984. »The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other.« *Social Studies of Science* 14(3): 399-441.
- Raina, Dhruv. [2003] 2010. *Images and Contexts: The Historiography of Science and Modernity in India*. New Delhi: Oxford University Press.
- Said, Edward. 1978. *Orientalism*. New York: Pantheon.
- Sathaye, Adheesh A. 2015. *Crossing the Lines of Caste: Viśvāmītra and the Construction of Brahmin Power in Hindu Mythology*. New York: Oxford University Press.
- Singer, Milton. 1972. *When a Great Tradition Modernizes: An Anthropological Approach to Indian Civilization*. New York: Praeger.
- Singler, Beth. 2019. »Existential Hope and Existential Despair in AI Apocalypticism and Transhumanism.« *Zygon: Journal of Religion and Science* 54(1): 156-176.
- Srinivas, M.N. [1966] 2013. *Social Change in Modern India*. Hyderabad: Orient Blackswan.
- Tirosh-Samuelson, Hava. 2012. »Transhumanism As a Secular Faith.« *Zygon: Journal of Religion and Science* 47(4): 448-61.
2013. »Wrestling with Transhumanism.« In *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, edited by Max More and Natasha Vita-More, pp. 19-54. Malden, MA: Wiley-Blackwell.
- Turner, Fred. 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago: University of Chicago Press.
- Underwood, Corinna. 2017. »The Future of Artificial Intelligence According to Ben Goertzel.« *Techemergence* (December 9). www.techemergence.com/the-future-of-artificial-intelligence-according-to-ben-goertzel/ (accessed November 5, 2018).
- Van Dijck, José. 2014. »Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology.« *Surveillance & Society* 12(2): 197-208.
- Van Otterlo, Martijn. 2014. »Automated Experimentation in Walden 3.0: The Next Step in Profiling, Predicting, Control and Surveillance.« *Surveillance & Society* 12(2): 255-272.
- Weizenbaum, Joseph. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W.H. Freeman and Company.
- Wiener, Norbert. 1964. *God & Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion*. Cambridge, MA: The M.I.T. Press.
- Zarsky, Tal Z. 2012. »Automated Prediction: Perception, Law, and Policy.« *Communications of the ACM* 55(9): 33-35.
- Zuboff, Shoshana. 2015. »Big Other: Surveillance Capitalism and the Prospects of an Information Civilization.« *Journal of Information Technology* 30(1): 75-89.





How Artificial Intelligence is Impacting Societies

Lorena Jaume-Palasi

Artificial intelligence is a technology used to plan for the future. Planification implies intelligibility, calculability, and systematization. The future as a concept has been, in occidental cultures, closely tied to monotheism and the development of a linear narrative about societies, with a predicted end of the world, where individuals end up either in paradise or hell. This was a radical change from the narratives of classic cultures, where there was no notion of the past or prehistory, but rather a narrative of a cultural, god-given origin similar to the present. It did not anticipate change in the manner of future narratives. Future narratives see the time to come as a time when evolution happens, when neither clothes nor context nor social habits remain the same. With the development of Protestantism and capitalism, the future became more than a point in time when the story would end. It became an unwritten point of opportunity to be shaped by human beings.

At the beginning of the twentieth century, the idea of the future was closely tied to technology as an instrument for changing historical contexts and shaping societies. Elites initiated a technical discourse focused on scientific pragmatism and technocracy, with social engineering focused on the creation and “neutral” planification of big societal projects. In 1933, sociologist Hans Freyer stated, “If the immanent utopia of technology is the transformability of all materials and forces into each other, then the immanent utopia of planning is the transformability of all historical situations into each other” (Freyer 1987, 22; own translation by the author). Karl Mannheim even declared the year 1935 the end of “unplanned things” (Mannheim [1935] 1958). Technical experts postulated that by organizing societies’ infrastructure based on efficiency and rationality, for the sake of the common good, it would even be possible to overcome subjectivity in politics: “We are so used to fighting that we cannot see there is a better way—the way of planning” (Kizer 1939), was

the argument advanced at a conference in 1939, at the end of the American New Deal era; this architectural planning from the scratch approach was also known on the European continent, and the intellectual elites behind it conducted an extensive occidental international exchange.

The constrictions of planification through technical infrastructure crystallized in public discussions during the 1970s. New issues such as climate change and sustainability tested the limits of what could be planned for. More and more, the value of public opinion and public participation in infrastructure projects initiated scrutiny of the political dimension of infrastructure and its alleged neutrality. In the very same decade, Foucault presented his new theory of power within infrastructure in the context of his discourse analysis theories and criticized the social and ethical assumptions implied therein¹ (Foucault 1980).

1 Cf. The Confession of the Flesh.

On the Infrastructural Nature of Artificial Intelligence

Artificial intelligence (AI)—algorithmic systems—comprises technologies the world is still trying to understand in their essence, in order to assess their impact and risks. AI and algorithmic systems do not understand individuals. Conceptually, they represent ideas of the social. The way they compute and classify patterns is relational. Algorithms categorize people in fine granular groups. The identity of individuals is no longer relevant. Personalization may be perceived by the user as the technical procedure for individualization, but technically, personalization is relational: it is the classification of this individual into a very specific collective of people with similarities.

It does not necessarily become clear to the person concerned that he or she is being classified into a collective that may not be part of the conventionally known social categories in a society. Personalized advertisement and “microtargeting” may give the impression that marketing is addressing potential consumers individually, based on information about the preferences of the individual. But technically the individual is being assigned to various categories shared with many other individuals. The connection of all these categories results in an intersectional profile encompassing

more categories than the usual ones, such as age, gender, and social status; and that profile is equally shared by many other individuals. This level of granularity and intersectionality is easy to confuse with individuality.

As a result, assessments of AI tend to focus on detecting individual damage and human rights abuses, although problematic algorithmic systems primarily discriminate against collectives without detecting individual damage (van der Sloot 2016). This is a classic effect when it comes to assessing the impact of infrastructure. The effects of the shape, standards, and rules of infrastructure mediating the flow of resources, mobility, or telecommunication can only be detected with an architectural overview of the system. Then infrastructure is the physical Foucaultian dispositif or apparatus to distribute power, create the conditions for societal inclusion or exclusion, and shape the space of a society. In his interview “The Confession of the Flesh” he defined the term as *a thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions—in short, the said as much as the unsaid. Such are the elements of the apparatus. The apparatus itself is the system of relations that can be established between these elements.* (Foucault 1980, 194)

The mobility infrastructure in a city determines the way its citizens access its geography, the way it fosters or demotes inclusion. The streets in a suburb in the United States with roads and no sidewalks shape the mobility of its residents differently than the streets of Amsterdam, with sidewalks, bicycle paths, and roads. Pedestrian traffic lights that have very short green intervals may generate a more fluid car traffic, but they certainly present a challenge for older pedestrians.

Artificial intelligence is a new form of infrastructure. It is not a product; it is immaterial infrastructure. Everything that is a process implies a certain system and a set of standards, which can then be formalized in mathematical language and become partially or fully automatable. Such standardization goes beyond cables and hardware. Automating a process with AI implies setting a fine invisible layer of software to permanently mediate interactions with and among all in-

volved parties of the process. In this way, immaterial infrastructure is being built into sectors where an infrastructural dimension was unthinkable before.

The current understanding of infrastructure therefore needs revision. Presently, infrastructure denotes either the institutions preserving the economic, cultural, educational, and health functions of a country—soft infrastructure—or “all stable things that are necessary for mobility and an exchange between people, goods and ideas” (van Laak 2018; own translation by the author)—hard infrastructure. An essential characteristic defining both soft and hard infrastructure is stability, whether procedural, in the case of soft infrastructure; or physical, in the case of hard infrastructure). They are a form of fundamental planning to systematically design access, distribution, and interaction with goods and services that are of interest for a collective. For Foucault, this is a fundamental aspect of political power (Foucault 1980). His concept of *dispositif* regulating the politics of health, sexuality, or architecture was novel for broadening the definition of power beyond mere rules to a collection of “relations of power, practices and actions” (Elden 2016) depicted both in normative and material infrastructure and mechanisms. Infrastructure is thus the planification of power and its distribution through a set of standards embodying societal ideas of efficiency and fairness of procedures and distribution.

Another relevant characteristic of infrastructure is that modularity is inherent in it. Hannah Arendt’s criticism of the bureaucratization of murder during the Third Reich in Germany is a fundamental criticism of soft infrastructure (Arendt 2006). Administration as soft infrastructure—once seen by Max Weber as the mechanism of democracies to ensure equality before the law and its procedures in opposition to the arbitrariness of charismatic autocracies—entails risks (Weber 1922). Dividing the extermination processes into standardized administrative steps or modules led individuals to decontextualize each module from the broader process and foster moral distance from its ultimate consequence. Administration banalized evil into a bureaucratic procedure that would obfuscate responsibility through modularization, making bureaucrats in the system accountable only for a mere step of the process.

Furthermore, infrastructure usually has an interdependent character: information and telecommunication infrastructures are fundamentally dependent on electricity infrastructure.

One last characteristic to mention is the unavailability of infrastructure and infrastructure goods to single households and companies, both for production and cost reasons: “Although bread is able to satisfy our hunger, it is not an infrastructure good, since the ingredients for bread production are easy to obtain; today everyone can bake bread for himself” (Buhr 2009).

Because fixed costs are very different depending on the capital goods, the supply of infrastructure happens under different market forms: mainly (natural) monopolies (e.g., electricity supply), but also competition (e.g., housing construction). While a single household may afford a generator or solar panels, a constant, secure supply of electricity still needs connection to the grid, the grid itself being the infrastructure that a single household cannot afford.

Artificial intelligence entails many of these aspects. It has physical prerequisites such as cables and hardware. It is not stable in its ontology; its formulas and code are constantly changing. But it does create a stable layer of a mathematically formalized structure concurring with or complementing the rules and constrictions given by soft and hard infrastructure. Further, it automates processes through technical and mathematical modularization—each process is split into several steps depending on the technical requirements, but not necessarily the administrative and social context. AI systems are not instruments or derivations of the rules and mechanisms of soft infrastructure. They have a different rationale running parallel to that of soft infrastructure, and, consequently, they require their own categorization.

Social media, for example, could be seen as a form of immaterial infrastructure in the communicative sector. A software layer orchestrates the interface, the timeframe and the format (videos, text, pictures) in which people interact with each other. This infrastructure is thus standardizing and moderating communication along with the rules and standards of soft infrastructure (in this case speech rights, personal rights, etc.), and its present market consists of monopolies defined by the format.²

² Existing social media companies have monopoly positions within their own specific formats: the format offered by Twitter is different than that offered by Instagram, Snapchat, Youtube, or Facebook.

Predictive policing systems would be another example. They are used to identify pattern behavior for concrete crime categories that are applied strategically to prevent similar crimes. For example, organized criminal groups have a modus operandi for robberies, and within a given time frame and geographical parameter, this information is being used systematically to prevent similar robberies in the area by the same actors. A predictive policing system for this use would standardize the geography of a city and identify the smallest geographical unit. It would be built upon a set of definitions and human-made decisions: which data categories will be used and which will be excluded; how old the data may be and whether there is an expiration date for the data; what crime categories will be included and correlated to each other; and so forth. With this, information about crime, the geography of the jurisdiction, and so on is structured and “datafied”— and saved as a usable data set of standards and rules that function together with those specified by soft infrastructure. Although informing, assisting, and thus systematizing police work becomes very much dependent on the ideas and concepts of optimization, fairness, and efficiency³ of the diverse actors designing and implementing the technology, these social ideas of efficiency and fairness are at the same time constrained in their translation into algorithms by the rules of mathematics and the limits of “datafication.”⁴ Thus, soft infrastructure and immaterial infrastructure are two separate systems dialectically influencing and constraining each other.

3 These ideas of efficiency and fairness do not necessarily need to be specified in the set of rules constituting a soft infrastructure.

They can be the expression of common social expectations and prejudices in a society running contrary to legal and administrative rules.

4 Not all social contexts and circumstances can be comprehensively turned into data.

Another central aspect of AI as an architect and moderator of social relations, practices, and actions is the value at the center of its optimization rules. AI systems optimize for a specific goal or value, to the detriment of other values. Does a system optimize for efficiency in the sense of pragmatism or of fairness?

All the above examples illustrate traits of the technology that are also applicable in other AI mediated sectors and services: health, agriculture, mobility, communication, social welfare, banking, e-commerce, employment, and so on. Using artificial intelligence in those sectors implies standardizing and creating a second layer of norms in mathematical modules that will automatically structure the human rela-

tions, practices, and interactions within that context.

Because infrastructure is the architectonic expression of the politics of a society, AI is a technology impacting societies architectonically, and thus on a collective rather than individual level.

The Inherent Methodological Individualism of western Norms and Laws

Algorithms and artificial intelligence do not understand individuals; and democracies, from their law-dogmatic perspective, do not understand collectives. The Western ethical approach and its legal cultures are individualistic in methodology and anthropocentric in ontology. In the history of political theory of the legitimation of political power, the good society—legitimate political order—could only be achieved by morally good citizens. So, in the first place, the individual had duties and obligations towards society. Rights emanated from those obligations but were not central to the existence of society. Constitutionalism changed that narrative: the concentration of power into a political order was legitimate if it achieved protecting the fundamental rights of individuals. First society owed rights to its citizens, and from those rights emanated obligations. Under the new narrative, rights came first, duties second.

This was the narrative birth of the legitimation of occidental democratic power, based on the idea that political power is there to protect individuals. But the Leviathan-state, the monster with the monopoly on violence, was created to protect individuals from a very specific form of violence, a war of all against all.

That a man be willing, when others are so too, as farre-forth, as for Peace, and defence of himselfe he shall think it necessary, to lay down this right to all things; and be contented with so much liberty against other men, as he would allow other men against himselfe. For as long as every man holdeth this Right, of doing anything he liketh; so long are all men in the condition of Warre. (Hobbes 1968, 190)

The purpose of the Leviathan-state was not to grant liberties and rights to individuals but to overcome (civil) war and make a life in society possible by granting restricted liberties and rights to individuals. Rights were instrumental to making society and cooperation among individuals possible.

Most constitutionalist theories part from a self-centered, rationalist anthropological view of man. The construction of the social contract had to incentivize man to accept handing over power and to trust the resulting political architecture in order to be able to cooperate and benefit individually within the given rule structure. Increasingly, the focus on appealing to the rationality of individuals to accept and follow the social contract distracted from the purpose of rationality and contractual incentives (cf. Habermas 1989 and Sennett 1977).⁵

5 Along with the Enlightenment's command of daring to know, making use of rationality, the increase of literacy rates, and the permeation of the contractual narrative into the cultural expectations of societies, the concept of privacy emerged, marking the threshold where state power ended. Both rationality and privacy reinforced the methodological individualistic approach in the social contract. This would lead Jürgen Habermas, with *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, and Richard Sennett, with *The Fall of the Public Man*, to a fundamental criticism of how this development is eroding the public dimension in the lives of private individuals and, with this, essential aspects of the public sphere and societal cohesion.

Individual autonomy can only be achieved through a societal lens and is essentially dependent on the structural framework within which it operates.

While vervet monkeys have a feeding rank where dominant females eat first and longer, domestic dogs eat alone and do not share their food. An individualistic lens focused only on individual benefits and rights would consider an animal able to eat first and as much as it wants to as clearly the one enjoying the greatest freedom. Without looking at the broader structural context, this could apply to an animal living in a zoo or a kennel. Most animals living in packs have dominance hierarchies reflected in food, reproduction, and other vital aspects. Domestic animals are generally denied the pack and enter a higher dependency and hierarchy with a human master. Mastership over oneself is not defined by individual benefits or rights, such as eating alone, but by relational factors dependent on the structure within which those rights are embedded.

As a consequence, most democratic cultures have legal instruments to assess impact, and provide protection and redress, merely on an individual level. There are already politically relevant examples caused by the lack of an assessment rating based on structural risks and harms. The fitness tracking app Strava released a visualization map showing all the activity tracks by the app's users all over the world. The data was anonymized. However, *in locations like Afghanistan, Djibouti and Syria, the users of Strava seem to be almost exclusively foreign military personnel, meaning that bases stand out brightly. In Helmand province, Afghanistan, for instance, the locations of forward operating bases can be clearly seen, glowing white against the black map. (Hern 2018)*

Similarly, a few software programs for predictive policing used within the European Union are falling out of the scope of regulation and thorough impact assessment: in several parts of Germany, state police are using software to predict burglaries, given a specific modus operandi, within a specific time frame and geographical parameter, by using anonymous data about the crime type and procedure as well as geographical data. The software makes sense in regions with fewer police officers because it may assist police in creating more efficient patrol shifts regarding burglary prevention. The corresponding data-protection state agencies permitted use of the software since it did not process personal identifiable data and thus did not fall under their jurisdiction.

But these systems leave many societal questions open. If the software is fed historical data, it consequently raises concerns known from structural ethnical biases associated with ZIP codes: to what extent is the data bank reflecting data asymmetries? How can representation be controlled so some regions are not overrepresented while others are underrepresented? To what extent may the system amplify this kind of ethnic bias and affect the social cohesion of a city as a consequence? If a greater police presence is being observed in structurally poorer regions, will residents feel more secure or will this lead to a massive exodus of residents able to afford housing in a different part of the city? How and to what extent is this tool meaningfully embedded in a broader prevention strategy? This is yet another example of the need for a more collective approach to evaluate algorithmic systems.

However, there are only a few areas of law—labor law, for example—where most legal cultures possess instruments to address the collective dimension of discrimination. In the future, discrimination will be a phenomenon observed in all sectors where AI is used, be it the distribution of energy or critical resources, the health sector, or social welfare. The fact that discrimination does not only exist in the labor sector, and that the use of these technologies will take place in all sectors, points to one of the legal gaps that new technologies make more tangible.

For these reasons, focusing on individual rights is ironically detrimental to individual autonomy and rights. The autonomy of the individual depends very much on the social

framework and infrastructure within which this autonomy is exercised. Being able to eat alone is not a sign of self-mastery. The absence of individual harm with regard to human or fundamental rights does not mean that human beings are not harmed in general. There is such a thing as societal harm, and assessing societal impact requires other questions and criteria than the ones applied for human and fundamental rights.

Defining the Values that Shape Fair Immaterial Infrastructure

The question about the normative order of artificial intelligence is not a question of either ethics or regulation. Ethics and law are two different normative systems with different functions, even though they may intersect at some point in the democratic process. Both are needed. And both need further thinking with regards to the collectivist dimension mentioned above.

Europeans have long been aware of the potential conflicts among different kinds of material infrastructure due to the scarcity of space. In 1970, the Council of Europe initiated the CEMAT high-level minister conferences for spatial planning, leading to the approval in 2000 of the Guiding Principles for Sustainable Spatial Development on the European Continent. In 1999, the European Union developed the European Spatial Development Perspective, a policy framework providing the conditions and criteria that are instrumental in building trans-European networks (transport, energy, telecommunications). Both the European Council and the European Union enshrine social cohesion as one of the guiding principles in their policies and frameworks of spatial planning when building and coordinating (material) infrastructure. And yet the conflict between the standardizing, homogenizing character of cables, bridges, and roads and its impact on pluralist societies has not yet been the object of a thorough discussion comparable to the debate on citizenship. If the ceiling of an underpass or tunnel is too low for buses, then only (mainly private) car drivers would be able to use it. It would exclude people who can only afford public transport. What alternatives are being provided to ensure access between the areas at both ends of the tunnel?

The example given above of the low tunnel ceiling resembles many examples of bias seen in algorithmic systems—for example, the automated soap dispenser unable to detect users with darker skin. The problem behind the soap dispenser was an issue of bias at the standardization level. Near-infrared technology is not good at detecting darker skin, and since the developers and testers of these articles were lighter-skinned individuals, this problem had gone unnoticed for a long time.

Both examples are about the creation of standards by making implicit assumptions about humans and their contexts that do not reflect otherness—the variety in human nature and of social contexts.

Even though the concrete implementation of material infrastructure is reduced to allegedly neutral numbers and mathematical standards, on a more abstract level, its collective social impact is fully accounted for by norms and rules as they're written in the European continent. The principles that guide the creation and coordination of infrastructure are enshrined in corresponding spatial planning policies at local, regional, national, and continental levels. Those principles constitute a set of criteria to guide, assess, and evaluate the impact of infrastructure very much applicable to algorithmic systems:

- balancing social, economic, ecological, and cultural conditions (also considering geographical asymmetries);
- safeguarding diversity;
- providing for stable and continuous access to public services ensuring fairness of opportunity;
- providing for a balanced economic structure that fosters a wide range of;
- preserving and developing cultures;
- ensuring sustainability and respect for nature;
- ensuring that the needs to provide for defense and civil protection are taken into account;
- ensuring the conditions necessary for social cohesion.

And while all these questions and approaches do have an impact on individuals, the approach and the criteria differ fundamentally from the catalogue of questions used in the individual rights approach. The principles listed here are better suited to generating a deeper analysis of the impact

of algorithmic systems by providing answers on a structural level. This more collectivistic approach at a regulatory and normative level is thus not unknown to many democracies. However, several aspects and implications of infrastructure still need more scrutiny and further legal and methodological thinking and corresponding sectoral concretization. Algorithmic systems are highly context dependent and concretization requires sectoral contextualization.

In conclusion, algorithmic systems and artificial intelligence as collectivistic technologies amplify a weakness of democracies: the methodological individualism of democratic systems has characterized the normative approach of democratic powers. However, democracies do also have societal purposes, though their corresponding regulatory instruments are less developed. The implementation of AI requires societal thinking. Given this new era of automatization, the most imperative task for democracies lies in the further development of the idea of public interest, the common good, and the shape of society.

-
- Arendt, Hannah. 2006. *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Penguin Books.
- Buhr, Walter. 2009. „Infrastructure of the Market Economy, Volkswirtschaftliche Diskussionsbeiträge.“ *Fachbereich Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht: Discussion Paper No. 132–09*. U. Siegen. <http://ideas.repec.org/s/sie/siegen.html>.
- Elden, Stuart. 2016. *Foucault's Last Decade*. Cambridge: Polity Press.
- Foucault, M. 1980. “The Confession of the Flesh” [Interview, 1977]. In *Power/Knowledge: Selected Interviews and Other Writings*, edited by Colin Gordon, 194–228. New York, NY: Pantheon Books.
- Freyer, Hans. 1987. “Herrschaft und Planung. Zwei Grundbegriffe der politischen Ethik.“ In *Herrschaft, Planung und Technik. Aufsätze zur politischen Soziologie*, edited by Hans Freyer, 17–43. Weinheim: VCH Verlagsgesellschaft.
- Habermas, Jürgen. 1962 trans 1989. *The Structural Transformation of the Public Sphere: An Inquiry into a category of Bourgeois Society*. Polity, Cambridge

- Hern, Alex. 2018. "Fitness tracking app Strava gives away location of secret US army bases." Accessed Jan. 4, 2019. <https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases>.
- Hobbes, Thomas. 1968. *Leviathan*. Edited by Crawford Brough Macpherson. London: Penguin Books.
- Kizer, Ben H. 1939. "The Need for Planning." In *National Conference on Planning. Proceedings of the Conference held at Minneapolis, Minnesota, June 20-22, 1938*. Chicago: 1–9.
- Laak, Dirk van. 2018. „*Alles im Fluss*“: *Die Lebensadern unserer Gesellschaft – Geschichte und Zukunft der Infrastruktur*. Frankfurt am Main: S. Fischer Verlag.
- Mannheim, Karl. [1935] 1958. *Mensch und Gesellschaft im Zeitalter des Umbaus*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Sennet, Richard. 1977. *The Fall of Public Man*. New York, Alfred A. Knopf.
- Sloot, Bert van der. 2016. "The Individual in the Big Data Era: Moving towards an Agent-Based Privacy Paradigm." In *Exploring the Boundaries of Big Data*, edited by Bert van der Sloot, Dennis Broeders, and Erik Schrijvers, 178–79. Amsterdam: Amsterdam U. Press.
- Weber, Max. [1922] 1980. *Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie*. Besorgt von Johannes Winckelmann. Studienausgabe, Tübingen.





Vertrauen in KI-Systeme

Vertrauen in KI-basierte Systeme schaffen

Ina Schieferdecker^{1,2,3}

¹ Weizenbaum-Institut für die Vernetzte Gesellschaft

² Fraunhofer FOKUS

³ Technische Universität Berlin

⁴ Künstliche Intelligenz (KI) (WBGU, 2019: 482) bezeichnet „eine Disziplin innerhalb der Informatik, die sich mit der Entwicklung von Softwaresystemen befasst, welche Funktionen bereitstellen, deren Ausführung das erfordert, was typischerweise mit dem Wort Intelligenz bezeichnet wird“ (Burgard, 2018). Ebenso wie bei „Intelligenz“ gibt es bis heute keine einheitliche Definition von Künstlicher Intelligenz, sondern plurale Verständnisse. Grundsätzlich zeichnen sich intelligente Systeme durch die Fähigkeit aus, (teilweise) selbstständig und effizient Probleme zu lösen (Mainzer, 2016: 2).

WBGU – Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen (2019): Unsere gemeinsame digitale Zukunft. Berlin: WBGU.

Burgard, W. (2018): Künstliche Intelligenz, Externe Expertise für das WBGU-Hauptgutachten „Unsere gemeinsame digitale Zukunft“, Berlin: WBGU.

Mainzer, K. (2016): Künstliche Intelligenz – Wann übernehmen die Maschinen? Heidelberg, Berlin: Springer.

Im Zuge des digitalen Wandels werden zunehmend Software-basierte Systeme, die auch Methoden der Künstlichen Intelligenz (KI)⁴ nutzen, mit menschlichen Aktivitäten abgestimmt. Getrieben von Daten, die von Menschen erzeugt, aggregiert und / oder aufbereitet werden, arbeiten solche, auch kurz genannt KI-basierten Systeme immer enger mit uns Menschen in Echtzeit zusammen. Die Suche nach den dafür benötigten passenden und leistungsfähigen Automatismen mittels KI ist ebenso eine Suche nach den Erwartungen und Anforderungen, die an solche Systeme zu legen sind und wie diese im Entwurf, in der Entwicklung und im Betrieb abgesichert werden können. Vertrauenswürdigkeit als eine der wesentlichen Grundlagen zur Akzeptanz KI-basierter Systeme muss dazu über Korrektheit sowie Zuverlässigkeit, Leistungsfähigkeit sowie Skalierbarkeit, Sicherheit sowie Datenschutz als auch Nachhaltigkeit solcher Systeme erreicht werden. Dieser Artikel bietet dazu eine Reflektion des Standes des Software Engineerings als Disziplin zur Entwicklung vertrauenswürdiger Software und zudem Parallelen als auch neue Herausforderungen bei der Entwicklung KI-basierter Systeme. Der Artikel ist eine Weiterentwicklung von (Schieferdecker, Großmann, & Schneider, 2019; Schieferdecker & Ritter, 2019).

Einleitung

Die Menschheit hat sich über Jahrtausende dienliche und dienende Automatismen geschaffen, mit denen vordefinierte Abläufe, sogenannte Algorithmen, durch Maschinen selbstständig realisiert werden. Maschinen basieren dabei auf mechanischen, elektronischen oder digitalen Methoden der Automatisierung und werden mittlerweile in bald allen Lebens- und Arbeitsbereichen wie Produktion, Landwirtschaft, Bau, Transport oder auch der Unterhaltung genutzt. Mit Informations- und Kommunikationstechnologien als Fundament und der zunehmenden Digitalisierung und Vernetzung

der Maschinen, nehmen ihre Fähigkeiten bezüglich Kognition und Autonomie zu (Fromhold-Eisebith et al., 2019), so dass sich Fragen wie: Sind die Maschinen zuverlässig? Kann man den Automatismen der Maschinen vertrauen? Sind die Maschinen in der Lage, auch unerwartete kritische Situationen zu bewältigen? Welche Entscheidungshoheit sollte beim Menschen verbleiben? stellen. Die Dringlichkeit der Beantwortung dieser Fragen stellt sich auch vor dem Wissen, dass digitalisierte Automatisierung zu unerwünschten sozio-technischen Effekten führen können und bereits mit teils katastrophalen Auswirkungen geführt haben. Erinnerung sei an dieser Stelle an die Explosion der Ariane-5-Rakete 1996 (Dowson, 1997) oder aber an jüngste Vorfälle beim automatisierten Fahren⁵ oder in der Luftfahrt,⁶ die bei der traurigen Bilanz der «Software-Toten» mitzuzählen sind.

⁵ Bei tödlichen Unfällen eines Elektrofahrzeugs von Tesla im Mai 2016 als auch im März 2019, bei dem das Fahrzeug jeweils unter einen die Straße querenden Lastwagen-Anhänger fuhr, waren jeweils die Autopiloten aktiviert.

⁶ Die Maneuvering Characteristics Augmentation System (MCAS)-Software der Boeing 737 Max wird für die Abstürze im März 2019 bei Ethiopian Airlines und im Oktober 2019 bei Lion Air, bei denen 346 Menschen ums Leben gekommen sind, als Ursache gesehen. Sie soll die Nase der Boeing nach unten gedrückt haben, ohne dass die Besatzung in der Lage war, die Fehlsteuerung zu korrigieren.

Software an sich ist ein Komposit von Programmen und Daten und umfasst zudem Artefakte wie Spezifikationen, Dokumentationen, Nutzungshinweise sowie Prüfspezifikationen, Prüfdaten und / oder Zertifikate. Dabei ist Software im Prinzip eine Realisierung von Algorithmen durch Programme und Daten. Mit der Ausführung der Programme werden die durch sie realisierten Algorithmen auf (Anwendungs-)Daten angewendet und wirken über die berechneten Ergebnisse (wiederum Daten) auf Ein- und Ausgabegeräte, Sensorik, Aktuatorik und / oder Robotik. Ein sehr einfaches Beispiel ist das Sortieren von Zahlen, für die es verschiedene Sortieralgorithmen (wie Bubble Sort, Binary Tree Sort oder Bucket Sort) gibt, die mittels vielfältiger Programmbibliotheken (typischerweise gehören Sortieralgorithmen zu den Standardbibliotheken der Programmiersprachen) realisiert sind, auf Zahlen, die beispielsweise Kredite, Studienergebnisse, Preise oder Laufzeiten darstellen, angewendet werden sowie im Ergebnis als geordnete Listen auf einem Monitor angezeigt oder in einer anderen Software-Funktion weiterverarbeitet werden.

Software entfaltet ihre Wirkung während ihrer Ausführung, so dass die Wirkungen zudem zu der Ausführungsumgebung sowie den Konfigurationen und konkreten Nutzungsszenarien der Software abhängt. Für die Ausführung von Software werden des Weiteren Werkzeuge wie Compiler/Interpreter, Laufzeitumgebungen der verwendeten Program-

miersprache, Betriebssysteme und Treiber der verwendeten Hardware als auch Netzwerkkomponenten genutzt. Zusammen ergibt sich so ein komplexes Gefüge zwischen Anbietern, Entwicklern, Betreibern und Nutzenden von Software-basierten Systemen einerseits und den sozio-technischen Komponenten in Software, Hardware und Infrastruktur andererseits.

Dabei führte fehlerbehaftete Software schon in den 1960er bei viel einfacheren Software-Konstellationen (Software in der Größenordnung von KBytes sowie rein sequentiell programmiert ohne Parallelität oder Vernetzung) zur Software-Krise (Randell, 1979) und dem Aufruf, in der Informatik das Gebiet des Software Engineerings zur Konstruktion und Absicherung von Software mit dazu passenden für Konzepten, Methoden, Technologien und Werkzeugen zu entwickeln (Schieferdecker & Ritter, 2019; Wirth, 2008). Seither hat sich Software Engineering als Disziplin in der Informatik etabliert und sich vom maschinennahen Software-Paradigma über das funktionale, objekt-orientierte, komponentenbasierte bis hin zum modellbasierten Software-Paradigma entwickelt. Trotz des gewachsenen Verständnisses und umfangreichen Methodenkastens werden in regelmäßigen Abständen neue Software-Krisen beschrieben (Fitzgerald, 2012; Gibbs, 1994). Und noch heute stellt der Entwurf, die Realisierung und die Absicherung von Software entlang der verschiedenen Dimensionen von Software-Qualität (Miguel, Mauricio, & Rodríguez, 2014), die beispielsweise wie Sicherheit und Leistungsfähigkeit oder Bedienfreundlichkeit im Widerstreit zueinander stehen können, eine Herausforderung dar (Breu, Kuntzmann-Combelles, & Felderer, 2014).

In der Tat gibt es Software in vielen Arten und Varianten und keine in der Wissenschaft akzeptierte Software-Taxonomie. So ist es scheinbar schwer, die Eigenschaften von Software im Allgemeinen zu erfassen, und dennoch gibt es ein langjähriges und immer noch wachsendes Verständnis dafür, wie Softwarequalität⁷ zu erfassen und bewerten ist. Insbesondere die ISO 25010 enthält einen Katalog an Qualitätsanforderungen an Software-basierte Systeme (Gordieiev, Kharchenko, Fominykh, & Sklyar, 2014), die aggregiert in der Wissenschaft und Wirtschaft etablierte Software-Qualitätsmodelle wie von Boehm (Boehm, Brown, & Lipow, 1976),

⁷ Software-Qualität bezeichnet den Grad, in dem ein software-basiertes System, die Kundenerwartungen und -bedürfnisse erfüllt. So umfasst sie alle Merkmale eines Software-basierten Systems, die sich auf dessen Eignung beziehen, festgelegte oder vorausgesetzte Erfordernisse zu erfüllen.

FURPS (Boehm et al., 1976), IEEE (IEEE, 1993) oder QMOOD (Hyatt & Rosenberg, 1996) zusammenfassen.

Vertrauenswürdigkeit von KI-basierten Systemen

Mit Blick auf KI als zunehmend relevante Technologie stellen sich wie bei Software und Software-basierten Systemen vergleichbare Fragen zur Qualität, Verlässlichkeit und Vertrauenswürdigkeit als auch Fragen nach den Relationen zwischen KI und Software sowie zwischen der Entwicklung von KI-basierten Systemen und von Software-basierten Systemen.

Kurz gesagt: KI wird als Software (und ggf. mit spezieller Hardware) realisiert. Und in bald jeder Software wird eine KI-Methode genutzt, auch wenn sie nicht als solche erkennbar ist oder erkannt wird. Generell wird an symbolischer (regelbasierter) KI, an sub-symbolischer (statistischer) KI und jüngstens an Mischformen beider gearbeitet (Buchanan, 2005; LeCun, Bengio, & Hinton, 2015; Norman, 1991). Bei symbolischer KI werden Wissen und Zusammenhänge zwischen Teilen des Wissens formal repräsentiert, um logische Schlüsse ziehen zu können. Bei sub-symbolische KI werden Zusammenhänge aus Merkmalen von großen Datenmengen gezogen. In aktuellen hybriden Methoden werden geschickte Kombinationen beider Ansätze genutzt. Am Ende bietet KI ein Methodenspektrum zur Analyse und / oder Herleitung von Wissen.

Auch wenn KI immer auch Software »ist«, gilt jedoch nur auf einen ersten Blick, dass sie mit »konventionellen« Technologien des Software Engineerings entwickelt und abgesichert werden kann. Einerseits stellen die Techniken rund um die verwendeten großen Datenmengen und ihre statistischen Auswertungen neue Anforderungen an die Definition von Qualität – beispielsweise: Wann ist die Genauigkeit einer Klassifikation akzeptabel? Was bedeutet »akzeptabel« in welchen Kontexten? – und deren Herstellung und Absicherung. Andererseits geht es eben nicht nur um technische Anforderungen, sondern ebenso – wie bei Software-basierten Systemen – um sozio-technische Anforderungen an die Wirkweise KI-basierter Systeme.

So definieren beispielsweise die Algo.Rules (Bertelsmann-Stiftung & iRights.Lab, 2019) einen Kriterienkatalog, um eine gesellschaftlich förderliche Gestaltung und Überprüfung von

algorithmischen Systemen zu ermöglichen, die ebenso auf KI-basierte Systeme anzuwenden sind. Dazu zählen Regeln zum Kompetenzaufbau, zu wohldefinierten Verantwortlichkeiten, zur Dokumentation von Systemwirkungen und ihre Absicherung, zu Gewährleistung der Systemsicherheit und des Datenschutzes, zu Kennzeichnungspflichten, zur Sicherstellung der Nachvollziehbarkeit und Beherrschbarkeit sowie zur Gewährung von Einspruchsmöglichkeiten.

Und wie die Suche nach angemessenen Modellen der Software-Qualität und dazu passenden praktikablen Technologien der konstruktiven und analytischen Qualitätssicherung weiter geht (Singh & Kannoja, 2013), werden ebenso spezialisierte Ansätze für KI-basierte Systeme gesucht (Russell, Dewey, & Tegmark, 2015; Seshia, Sadigh, & Sastry, 2016).

Absicherung von KI-basierten Systemen

So benötigen KI-basierte Systeme zusätzliche Methoden und Werkzeuge zur Verifikation und Validierung⁸ (Van Wesel & Goodloe, 2017). Da das systematische (dynamische) Testen von Software nach wie vor die effektivste und am meisten genutzte Absicherungsmethode von Software-basierten Systemen ist (Ammann & Offutt, 2016), wird sie höchstwahrscheinlich auch eine wesentliche Methodik für die Absicherung von KI-basierten Systemen sein. Herausforderungen an die Weiterentwicklung von Testmethoden stellen sich hierzu bezüglich Dynamik von KI-basierten Systemen (beispielsweise bei Lernverfahren), die schiere Größe der genutzten Daten (um präzise Ergebnisse liefern zu können) und das Orakelproblem (um zu erwartende korrekte Ergebnisse herzustellen) (Xie et al., 2011).

In den letzten Jahrzehnten hat die Forschung industrietaugliche Techniken zur Steigerung der Qualität, Effizienz und Zuverlässigkeit von dynamischen Tests entwickelt. Dazu gehören Ansätze wie die Automatisierung von Testausführungen mit Testtechnologien wie TTCN-3 (Testing and Test Control Notation (Grabowski et al., 2003)), für modellbasiertes Testen zur Automatisierung der Testgenerierung (MBT (Utting, Pretschner, & Legard, 2012)) sowie der Einsatz von Such- und Optimierungsalgorithmen zur automatisierten Testauswahl und Testmengenreduktion (Harman, Jia, & Zhang, 2015). Darüber hinaus ermöglicht die Kombi-

⁸ Für die Validierung wird ein objektiver Nachweis gesucht, dass Anforderungen für einen beabsichtigten Gebrauch oder eine beabsichtigte Anwendung erfüllt sind. Bei der Verifikation geht es um objektive Nachweise, dass festgelegte Anforderungen erfüllt sind. Insofern geht es grob gesprochen bei der Validierung um den Nachweis des richtigen Systems und bei der Verifikation um den Nachweis einer richtigen Realisierung eines Systems.

nation dynamischer Tests mit Verifizierungsansätzen wie Quellcodeanalyse, Modellprüfung und symbolischer Ausführung Verbesserungen beim dynamischen Testen, die die Stringenz der formalen Verifikation mit der Skalierbarkeit dynamischer Tests verbindet (Godefroid & Sen, 2018). Solche Ansätze können nicht nur zur Prüfung auf funktionale Eigenschaften, sondern ebenso auf nicht-funktionale Eigenschaften wie Leistung oder Sicherheit angewendet werden (Schieferdecker, Grossmann, & Schneider, 2012). Schließlich verbessert eine enge Integration des dynamischen Testens in den Systementwicklungsprozess und das Risikomanagement die Effizienz und Transparenz des Testens (Felderer, Grossmann, & Schieferdecker, 2018). Das Potenzial des risikobasierten Testens zur Steuerung von Testprozessen auf Basis von Unsicherheiten wurde beispielsweise im Bereich kritischer Systeme in Bezug auf Sicherheit und Gefahrenabwehr aufgezeigt und sollte ebenso für KI-basierte Systeme gelten (Erdogan, Li, Runde, Seehusen, & Stølen, 2014).

Auch wenn die Forschung für Verifikations- und Validierungsmethoden insbesondere für sub-symbolische KI-basierte Systeme am Anfang steht, ist das Testen bereits integraler Bestandteil beim Anlernen der KI: Tests werden durchgeführt, um genauere Modelle für die ursprünglichen Trainingsziele zu erhalten. Beim überwachten Lernen werden Testdatensätze zur Bewertung der KI-Modelle verwendet. So kombinieren beispielsweise (Ghosh, Lincoln, Tiwari, Zhu, & Edu, 2016) Machine Learning und Modellprüfung so, dass wenn die gewünschten logischen Eigenschaften durch ein trainiertes Modell nicht erfüllt werden, das Modell (»Modellreparatur«) oder die Daten (»Datenreparatur«), aus denen das Modell erlernt wird, systematisch verändert werden. (Fulton & Platzer, 2018) schlagen vor, die formale Verifikation mit einer verifizierten Laufzeitüberwachung zu kombinieren, um ein sicheres Lernen zu gewährleisten. DeepXplore (Pei, Cao, Yang, & Jana, 2017), DLFuzz (Guo, Jiang, Zhao, Chen, & Sun, 2018) und TensorFuzz (Odena & Goodfellow, 2018) sind Werkzeuge, die u.a. Kennzahlen zur Quantifizierung der Modellabdeckung durch Daten liefern und die Testautomatisierung erleichtern. DeepTest (Tian, Pei, Jana, & Ray, 2018) ermöglicht die systematische Prüfung neuronaler Netze, eines der gebräuchlichen Modelle in KI, unter realisti-

schen Nutzungsbedingungen, insbesondere für den Einsatz im Automobilbereich.

Neben der technischen Verifikation und Validierung können Testmethoden zudem für sozio-technische Anforderungen wie beispielsweise zur Erklärung der Wirkweisen von KI-basierten Systemen genutzt werden: Testfälle mit ihren Testdaten enthalten konkrete Ausführungsszenarien für KI-basierte Systeme und die zu erwartenden Ergebnisse (die Wirkungen). Sie können aus Nutzersicht formuliert werden und sind so leichter nachvollziehbar.

Zudem können Methoden zur Testdatengenerierung für eine Optimierung und Charakterisierung der Überdeckung von Trainingsdaten genutzt werden. Damit können merkmalsbasierte Daten mit einer repräsentativen Abdeckung des betrachteten Datenraums erzeugt werden und so unter anderem die Fairness KI-basierter Systeme durch die Berücksichtigung (Abdeckung) der relevanten Kategorien und Eigenschaften verbessern (Nguyen, Dosovitskiy, Yosinski, Brox, & Clune, 2016).

Zusammenfassung

Die Fähigkeit, KI-basierte Systeme effektiv zu entwickeln, abzusichern und zu testen, wird für die Akzeptanz solcher Systeme im Allgemeinen aber auch in sicherheitskritischen Bereichen wie Transport oder Mobilität, im Gesundheitswesen oder der Industrieautomation von zentraler Bedeutung sein. So kann u.a. die Entwicklung und Bereitstellung von Testtechnologien und -werkzeugen, von Prüfscenarien, Testfällen und -daten für KI-basierte Systeme eine solide Basis für die Verifikation und Validierung der Systeme sein. Zudem können diese dazu beitragen, KI-basierte Systeme zu erklären und so nachvollziehbarer und transparenter zu machen. Zudem können Prüfscenarien genutzt werden, um die Sicherheit KI-basierter Systeme zur Laufzeit zu gewährleisten. Perspektivisch gilt es zudem die Funktionalität und Absicherung KI-basierter Systeme in ihrer Software verschmelzen zu lassen. KI-basierte Systeme sind in ihren Ausprägungen und damit in ihren Wirkungen zu dynamisch, um Absicherungsmethoden allein in der Entwicklungsphase und nicht im Betrieb anzuwenden.

Die Autorin dankt dem Kompetenzzentrum Öffentliche IT bei Fraunhofer FOKUS für das Symposium »(Un)ergründlich« und die sich daraus ergebenden Anknüpfungspunkte für die wissenschaftliche Arbeit. Die Arbeit wurde teilweise vom Bundesministerium für Bildung und Forschung (BMBF) unter der Nr. 16DII111 (»Deutsches Internet-Institut«, Weizenbaum-Institut für die vernetzte Gesellschaft) sowie vom Bundesministerium für Bildung und Forschung und Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit unter der Förderungsnummer 01RIO708A4 (»Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen«, WBGU) gefördert. Zudem dankt die Autorin für vielfältige Diskussionen mit Reinhard Messerschmidt, Nora Wegener, Dirk Messner und Sabine Schlacke vom WBGU, mit Stefan Ullrich, Diana Serbanescu, Gunay Kazimzade und Martin Schüssler vom Weizenbaum-Institut und mit Martin Schneider und Jürgen Großmann von Fraunhofer FOKUS.

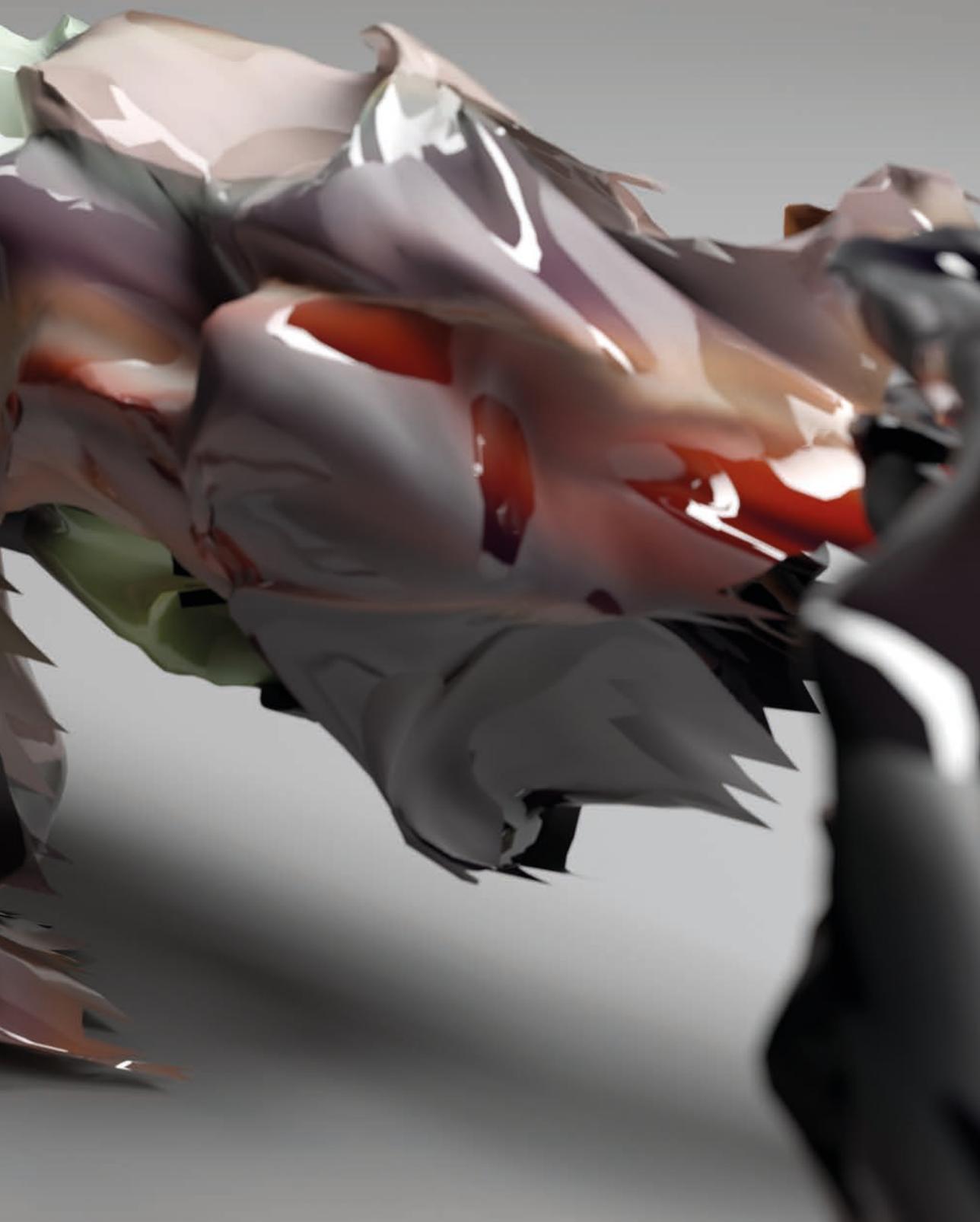
-
- Ammann, P., & Offutt, J. (2016). *Introduction to software testing*: Cambridge University Press.
- Bertelsmann-Stiftung, & iRights.Lab. (2019). *Algo.Rules: Regeln für die Gestaltung algorithmischer Systeme*, https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Algo.Rules_DE.pdf
- Boehm, B. W., Brown, J. R., & Lipow, M. (1976). *Quantitative evaluation of software quality*. Paper presented at the Proceedings of the 2nd international conference on Software engineering, San Francisco, California, USA.
- Breu, R., Kuntzmann-Combelles, A., & Felderer, M. (2014). New Perspectives on Software Quality [Guest editors' introduction]. *IEEE software*, *31(1)*, 32-38.
- Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. *Ai Magazine*, *26(4)*, 53-53.
- Dowson, M. (1997). The Ariane 5 software failure. *SIGSOFT Softw. Eng. Notes*, *22(2)*, 84. doi:10.1145/251880.251992
- Erdogan, G., Li, Y., Runde, R. K., Seehusen, F., & Stølen, K. (2014). Approaches for the combined use of risk analysis and testing: a systematic literature review. *International Journal on Software Tools for Technology Transfer*, *16(5)*, 627-642.
- Felderer, M., Grossmann, J., & Schieferdecker, I. (2018). Recent Results on Classifying Risk-Based Testing Approaches. *arXiv preprint arXiv:1801.06812*.
- Fitzgerald, B. (2012). Software Crisis 2.0. *Computer*, *45(4)*, 89-91.
- Fromhold-Eisebith, M., Grote, U., Matthies, E., Messner, D., Pittel, K., Schellnhuber, H. J., . . . Schneidewind, U. (2019). Unsere gemeinsame digitale Zukunft. In (pp. 517). Berlin: WBGU – Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen.
- Fulton, N., & Platzer, A. (2018). *Safe reinforcement learning via formal methods: Toward safe control through proof and learning*. Paper presented at the Thirty-Second AAAI Conference on Artificial Intelligence.
- Ghosh, S., Lincoln, P., Tiwari, A., Zhu, X., & Edu, W. (2016). *Trusted machine learning for probabilistic models*. Paper presented at the ICML Workshop on Reliable Machine Learning in the Wild.
- Gibbs, W. W. (1994). Software's chronic crisis. *Scientific American*, *271(3)*, 86-95.
- Godefroid, P., & Sen, K. (2018). Combining model checking and testing. In *Handbook of Model Checking* (pp. 613-649): Springer.
- Gordieiev, O., Kharchenko, V., Fominykh, N., & Sklyar, V. (2014). *Evolution of software quality models in context of the standard ISO 25010*. Paper presented at the Proceedings of the Ninth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX. June 30–July 4, 2014, Brunów, Poland.
- Grabowski, J., Hogrefe, D., Réthy, G., Schieferdecker, I., Wiles, A., & Willcock, C. (2003). An introduction to the testing and test control notation (TTCN-3). *Computer Networks*, *42(3)*, 375-403.
- Guo, J., Jiang, Y., Zhao, Y., Chen, Q., & Sun, J. (2018). *DLFuzz: differential fuzzing testing*

- of deep learning systems*. Paper presented at the Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering.
- Harman, M., Jia, Y., & Zhang, Y. (2015). *Achievements, open problems and challenges for search based software testing*. Paper presented at the 2015 IEEE 8th International Conference on Software Testing, Verification and Validation (ICST).
- Hyatt, L. E., & Rosenberg, L. H. (1996). *A software quality model and metrics for identifying project risks and assessing software quality*. Paper presented at the Product Assurance Symposium and Software Product Assurance Workshop.
- IEEE. (1993). Standard for Software Maintenance. (Std 1219).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- Miguel, J. P., Mauricio, D., & Rodríguez, G. (2014). A review of software quality models for the evaluation of software products. *arXiv preprint arXiv:1412.2977*.
- Nguyen, A., Dosovitskiy, A., Yosinski, J., Brox, T., & Clune, J. (2016). *Synthesizing the preferred inputs for neurons in neural networks via deep generator networks*. Paper presented at the Advances in Neural Information Processing Systems.
- Norman, D. A. (1991). Approaches to the study of intelligence. *Artificial Intelligence*, 47(1-3), 327-346.
- Odena, A., & Goodfellow, I. (2018). Tensorfuzz: Debugging neural networks with coverage-guided fuzzing. *arXiv preprint arXiv:1807.10875*.
- Pei, K., Cao, Y., Yang, J., & Jana, S. (2017). *Deepxplore: Automated whitebox testing of deep learning systems*. Paper presented at the proceedings of the 26th Symposium on Operating Systems Principles.
- Randell, B. (1979). *Software engineering in 1968*. Paper presented at the Proceedings of the 4th international conference on Software engineering.
- Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *Ai Magazine*, 36(4), 105-114.
- Schieferdecker, I., Grossmann, J., & Schneider, M. (2012). Model-based security testing. *arXiv preprint arXiv:1202.6118*.
- Schieferdecker, I., Grossmann, J., & Schneider, M. A. (2019). How to Safeguard AI. In A. Sudmann (Ed.), *The Democratization of AI. Net Politics in the Era Of Learning Algorithms* (Vol. AI Critique): Transcript.
- Schieferdecker, I., & Ritter, T. (2019). Advanced Software Engineering. In R. Neugebauer (Ed.), *Digital Transformation* (pp. 353-369). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Seshia, S. A., Sadigh, D., & Sastry, S. S. (2016). Towards verified artificial intelligence. *arXiv preprint arXiv:1606.08514*.
- Singh, B., & Kannoja, S. P. (2013). *A review on software quality models*. Paper presented at the 2013 International Conference on Communication Systems and Network Technologies.
- Tian, Y., Pei, K., Jana, S., & Ray, B. (2018). *Deeptest: Automated testing of deep-neural-*

network-driven autonomous cars. Paper presented at the Proceedings of the 40th international conference on software engineering.

- Utting, M., Pretschner, A., & Legeard, B. (2012). A taxonomy of model-based testing approaches. *Software Testing, Verification and Reliability*, 22(5), 297-312.
- Van Wesel, P., & Goodloe, A. E. (2017). Challenges in the verification of reinforcement learning algorithms.
- Wirth, N. (2008). A brief history of software engineering. *IEEE Annals of the History of Computing*, 30(3), 32-39.
- Xie, X., Ho, J. W., Murphy, C., Kaiser, G., Xu, B., & Chen, T. Y. (2011). Testing and validating machine learning classifiers by metamorphic testing. *Journal of Systems and Software*, 84(4), 544-558.





Über die Notwendigkeit einer Maschinenethik

Lukas Brand &
Benedikt Paul Göcke

Mit der Entwicklung fortschrittlicher, digitaler Technik hat seit der Mitte des 20. Jahrhunderts eine Eruption traditioneller Selbstverständlichkeiten eingesetzt. Angesichts autonomer und als intelligent bezeichneter Technik ist fraglich geworden, ob der Mensch weiterhin als das einzige Wesen verstanden werden kann, das nicht nur über ein Bewusstsein seiner selbst, sondern auch über moralisch relevante Handlungsfreiheit verfügt. Entscheidend für diese Entwicklung ist das sogenannte *maschinelle Lernen*. Dieses ermöglicht künstlich intelligenten Systemen, ihre Umgebung analog zur menschlichen Wahrnehmung zu registrieren und auf sie nicht mehr nur nach gegebenen Regeln, sondern aufgrund von eigener Erfahrung autonom zu reagieren. Da viele dieser maschinellen Handlungen – gerade im Bereich der Pflege, des autonomen Fahrens und im militärischen Einsatz – moralisch relevante Handlungen sein werden, wird eine Maschinenethik erforderlich, die Maschinen als moralische Subjekte in den Blick nimmt, auch wenn dies bedeutet, traditionelle Überzeugungen über den Menschen, sein Selbstbild und seine (moralische) Verantwortung neu zu überdenken.

1. Technisierung und Digitalisierung machen eine Maschinenethik erforderlich

Die stetig voranschreitende Digitalisierung aller Lebensbereiche hat von der generellen Technisierung einerseits die Tendenz zur Ersetzung vom Menschen gesteuerter Prozesse durch technische Sachsysteme und andererseits die Erweiterung menschlicher Handlungsräume geerbt: Während wir unsere Kommunikation und weite Teile unserer Arbeit seit Mitte der Neunzigerjahre sukzessive digitalisiert haben, ermöglicht die Digitaltechnik darüber hinaus nun etwa auch die unabhängige Kommunikation technischer Geräte untereinander, die virtuelle Anreicherung unserer Realität

(*Augmented Reality*) sowie die Simulation komplexer Abläufe am Computer. Dabei ist insbesondere die mit sogenannter künstlicher Intelligenz (KI) ausgestattete Technik in den letzten Jahren zunehmend in den Fokus einer öffentlichen Debatte gerückt und im Rahmen des Internets der Dinge zumindest in weiten Teilen schon fester Bestandteil unseres Alltags geworden. Die auf KI basierenden technischen Sachsysteme (KI-Systeme) zeichnen sich in unserer technisch integrierten Wirklichkeit wesentlich durch die ethisch relevante Ausweitung ihres Anwendungsraumes, die zunehmende Unabhängigkeit von menschlicher Kontrolle sowie ihre – bisher lediglich angestrebte – wechselseitige Abhängigkeit untereinander aus.¹ Aufgrund der immer größeren Integration von KI-Systemen in den individuellen und gesellschaftlichen Alltag steigt die Wahrscheinlichkeit eines ethisch relevanten Konfliktes zwischen menschlichen Interessen und den Aktionen autonomer KI-Systeme. Die Tatsache, dass ein KI-System, im täglichen Einsatz mit ethisch virulenten Problemen konfrontiert werden könnte oder bereits konfrontiert wird – beispielsweise im Straßenverkehr, in der Patientenbetreuung, der Interaktion mit Kunden oder der Automatisierung von Waffensystemen – sowie die Tatsache der relativen Unabhängigkeit dieser Ereignisse von menschlichen Handlungszusammenhängen, macht die Entwicklung einer Maschinenethik notwendig. Aufgabe der Maschinenethik ist es, die notwendigen Eigenschaften eines KI-Systems, das als moralischer Akteur eingestuft und beurteilt werden soll, zu formulieren und zu bewerten. Die Maschinenethik reflektiert also Maschinen, welche die moralischen Probleme, mit denen sie in ihrem Einsatzgebiet konfrontiert werden könnten, durch ihr Design oder ihren Programmcode selbstständig lösen sollen, ohne dass sie dabei von Menschen gesteuert werden müssen. Dieses Anliegen der Maschinenethik ist bereits auf verschiedene Weise charakterisiert und auf verschiedenen Wegen verfolgt worden.² Die Notwendigkeit einer Maschinenethik als eigenständiger Bereich der angewandten Ethik ist nicht selbstverständlich. Sie wird vor allem unter der Annahme bestritten, dass Maschinen prinzipiell keine selbständigen (moralischen) Entscheidungen realisieren könnten. Jede Form von Technik sei ein bloßes Werkzeug und als Mittel menschlichen Handelns

¹ Vgl. Wallach/Allen 2009, S. 23; vgl. Misselhorn 2018, S. 11.

² Vgl. Anderson/Anderson 2011, S. 1; vgl. Allen/Varner/Zinser 2000; vgl. McDermott 2008; vgl. Moor 2011.

der Herrschaft und den Zielen von Menschen vollkommen unterworfen. Aus diesem Grund könne eine Maschine kein selbständiger Urheber von (moralischen) Handlungen sein. Im Folgenden soll diese These bestritten und dargelegt werden, warum eine Maschinenethik besonders angesichts des gegenwärtigen Stands der Technik und den Fortschritten bei der Entwicklung autonomer Maschinen notwendig ist und welchen Bedingungen sie unterliegt.

3 Vgl. Ropohl 1996, S. 27 f.

4 Vgl. Asimov 1982, S. 67.

5 Rath et al. 2019, S. 7.

6 Im Oktober 2017 kündigte Google DeepMind eine Forschungsoffensive im Bereich Ethik & Gesellschaft an, aus der die Unternehmensstrategie Safety & Ethics hervorgegangen ist. Im April 2019 gründete Google einen Ethikrat (Advanced Technology External Advisory Council), der die Entwicklung und Verwendung von KI bei Google überwachen sollte und mit Vertretern aus unterschiedlichen Fachdisziplinen besetzt war. Wegen öffentlicher und interner Kritik wurde der Rat nach einer Woche wieder geschlossen. Im Frühjahr 2018 veröffentlichte Microsoft das Buch *The Future Computed. Artificial Intelligence and its role in society*, das sich den Potentialen und gesellschaftlichen Auswirkungen der KI-Entwicklung widmet. Im Januar 2019 gab Facebook die Gründung eines Institute for Ethics in Artificial Intelligence an der TU München bekannt. Im April 2019 forderten Wissenschaftlerinnen und Wissenschaftler weltweit die Erforschung des Verhaltens von KI »to control their actions, reap their benefits and minimize their harms« (Rahwan et al. 2019, S. 477).

2. Technikkritik und Maschinenethik

Die Forderung nach einer Maschinenethik ist keineswegs neu und fügt sich in den größeren Rahmen der Technikphilosophie ein, die spätestens seit der Mitte des vergangenen Jahrhunderts auf einer zunächst populären Technikkritik aufbauend eine philosophische Technikbewertung etabliert hat. Beispiele einer solchen populären Technikkritik, die eine Maschinenethik implizit mit einschließt, sind die Anti-Atomkraft Bewegung der sechziger Jahre³ und die Roboter-gesetze in den Science-Fiction-Romanen von Isaac Asimov, die dieser bereits 1942 zum ersten Mal beschrieben hat.⁴ Die Digitalisierung hat seit der Jahrtausendwende die Technik so gravierend verändert, dass die Digitaltechnik einer kritischen Reflexion unweigerlich bedarf. Dabei ist offenkundig, dass die angewandte Maschinenethik kein rein deduktives Unternehmen sein kann: »Die technische und gesellschaftliche Entwicklung generieren (sic!) moralische Praktiken und fordert gerade im Bereich der Maschinenethik die Ethik als Reflexionstheorie dieser Moral heraus.«⁵ Die ethische Reflexion muss sich also notwendigerweise auf den gegenwärtigen Stands der Technik beziehen. Seit einiger Zeit wird daher nicht mehr nur von Seiten der Öffentlichkeit, sondern auch von Seiten der Politik, vermehrt auch von Seiten international agierender Unternehmen sowie internationalen Zusammenschlüssen von Wissenschaftlerinnen und Wissenschaftlern eine ethische Auseinandersetzung mit der Produktion, Ausbreitung und Nutzung digitaler Technologien gefordert und vorangetrieben.⁶

Während sich der reine Technikbegriff bereits gegen Ende des 20. Jahrhunderts in die Bereiche *technische Sachsysteme*, *technisches Handeln*, *Technikentwicklung/-produktion*, *Technisierung und Technologie* – verstanden in einem ur-

7 Vgl. Ropohl 1999, S. 16 ff.

8 Vgl. Wallach/Asaro 2017, S. 2.

9 Vgl. Rohbeck 1993.

10 Vgl. Blutner 2014.

11 Vgl. Ropohl 1996, S. 29.

12 Vgl. Wallach/Asaro
2017, S. 1 f.

13 Vgl. Rahwan et al. 2019.

14 Ebenda, S. 477: »Machine behaviour similarly cannot be fully understood without the integrated study of algorithms and the social environments in which algorithms operate.«

sprünglichen Sinne als die Wissenschaft von Technik – aufgliedert hat⁷, zeichnet sich in der allgemeinen Technikbewertung bzw. Technikfolgenabschätzung erst seit Beginn des 21. Jahrhunderts eine vergleichbare Ausdifferenzierung ab:⁸ Trotz zahlreicher Überschneidungen reflektiert die Technikbewertung heute zum einen die Entwicklung, Produktion und den Gebrauch technischer Sachsysteme vorrangig im Rahmen der *Technik*-⁹ und *Ingenieursethik*¹⁰ sowie die wachsende Ambivalenz der gesamtgesellschaftlichen Technisierung im Rahmen der Techniksteuerung¹¹, die auch als *Roboterethik* bezeichnet wird. Letztere reflektiert insbesondere die Ausbreitung menschenähnlicher Roboter und den Einsatz unterstützender Technik, sowie deren Design und Anwendung in den so entstehenden soziotechnischen Handlungsräumen. Die zentrale Frage der Roboterethik lautet daher, »how the introduction of robots will change human social interactions and what human social concerns tell us about how robots should be designed.«¹² Die Roboterethik ist im Allgemeinen an den Auswirkungen interessiert, die sich in den gesellschaftlichen Mikro- und Makrostrukturen durch die Technisierung ergeben.

Mit der Forderung nach einem eigenen Fachbereich *Machine Behaviour Studies*, der weniger die Auswirkungen von Maschinen auf das Handeln und das soziale Gefüge von Menschen, sondern vor allem das Verhalten von Maschinen in ihrem realen Anwendungsbereich über die Ingenieurwissenschaften hinaus auch mit Methoden der Sozialwissenschaften und der Psychologie untersuchen soll, wurden schließlich die Voraussetzungen einer Maschinenethik erstmals auch von Seiten der Ingenieurwissenschaften bejaht.¹³

So geht dieser Forderung insbesondere die Einsicht voraus, dass sich die Vorhersage maschinellen Verhaltens nicht ausschließlich in dem Wissen um die sie intern steuernden Algorithmen begründen lässt, sondern darüber hinaus auch die externe soziale Umgebung berücksichtigen muss, in der die Algorithmen operieren und die das Verhalten der Maschinen maßgeblich mitbestimmt.¹⁴

In dem Gefüge der allgemeinen Technikbewertung behandelt die Maschinenethik nun die Entwicklung von KI-Systemen, die als künstliche moralische Akteure innerhalb einer soziotechnischen Wirklichkeit in Frage kommen und

in der Lage sein sollen, in realen Situationen moralische Probleme durch eine eigenständig Performanz zu lösen. Die Maschinenethik entwickelt dabei eine »Ethik für Maschinen im Gegensatz zu einer Ethik für Menschen *im Umgang* mit Maschinen.«¹⁵ Die entscheidenden Fragen der Maschinenethik sind dabei, welcher Fähigkeiten es bedarf, damit ein künstlicher Akteur in der Lage ist, ethisch relevante Situationen zu erkennen und zu entscheiden, welche von ihm durchführbare Aktion in dieser Situation die moralisch beste Option darstellt. Die Maschinenethik ist somit im Allgemeinen »concerned with giving *machines* ethical principles or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical decision making.« (Anderson/Anderson 2011, S. 1) Ziel der Maschinenethik ist es mitunter die Frage zu beantworten, ob und unter welchen Bedingungen ein Auto so mit KI ausgestattet werden kann, dass es in der Lage ist, ein Trolley-Problem¹⁶ zu erkennen, die (ethisch) relevanten Faktoren – Anzahl der Passanten, Insassen, Geschwindigkeit, Schadenswahrscheinlichkeiten u.v.m. – richtig einzuschätzen, die möglichen Folgen unterschiedlicher Optionen zu berücksichtigen und eine akzeptable Lösung des Problems herbeizuführen. Mit der *Maschinenethik* bildet sich also ein neues Anwendungsgebiet heraus, das technische Sachsysteme unmittelbar als potentielle Handlungssubjekte in den Blick nimmt und als solche ethisch zu beurteilen versucht.¹⁷ Dazu muss die Maschinenethik drei zentrale Fragen unter Berücksichtigung des gegenwärtigen Paradigmas bei der Entwicklung von KI-Systemen thematisieren und beantworten: (1) Ist ein KI-System autonom? (2) Gibt es für ein KI-System moralische Konfliktsituationen? (3) Wie können KI-Systeme zu moralischen Akteuren werden?

3. Verändertes Paradigma bei der Entwicklung künstlicher Intelligenz

Die Fragen der Maschinenethik stellen sich vor dem Hintergrund der jüngsten Entwicklungen im Bereich der KI – namentlich des tiefen oder maschinellen Lernens – in einem neuen Licht dar. *Deep Learning* bezeichnet eine Methode, die es ermöglicht, Probleme auf maschinellem Weg zu lösen,

¹⁵ Misselhorn 2018, S. 8; vgl. außerdem Rath et al. 2019, S. 6.

¹⁶ Ein Fahrzeug, dessen Bremsen versagen, steuert auf eine Menschengruppe zu. Um einen Zusammenstoß zu vermeiden kann es aber die Fahrtrichtung ändern, muss dafür aber ein erhebliches Risiko für die Insassen oder eine andere Passantengruppe in Kauf nehmen.

¹⁷ Vgl. Anderson/Anderson 2011, S. 2; vgl. Wallach/Asaro 2017, S. 2.

ohne einen definitiven Lösungsweg im Vorhinein zu programmieren. Die Ergebnisse und die Leistungsfähigkeit des Deep Learnings lassen sich in unserem Alltag in den Bereichen Bild- und Gesichtserkennung, Sprachsteuerung und Sprachsynthetisierung sowie in der Bewältigung komplexer Spiele wie Schach, Go oder Poker beobachten.

Der Deep Learning-Ansatz entwickelt KI-Systeme auf der Basis sogenannter künstlicher neuronaler Netzwerke (KNN), die anhand großer Datenmengen trainiert werden können.¹⁸

¹⁸ Vgl. Goodfellow 2016; vgl. Rojas 1996; vgl. Lenzen 2018.

Unter einem KNN wird dabei eine Software-Architektur verstanden, deren Funktionsweise von den Vorgängen im menschlichen Gehirn inspiriert ist: Zwischen einer Eingabe- und einer Ausgabeebene wird ein Netz aus künstlichen Neuronen in mehreren Ebenen aufgespannt. Die einzelnen Ebenen verarbeiten innerhalb des Netzes unterschiedliche, hierarchisch gestufte Aspekte eines gegebenen Problems und geben die Analyseergebnisse an die nächste Ebene mit höherer Komplexitätsstufe weiter. So werden etwa bei der Bilderkennung auf der Eingabeebene zunächst ausschließlich die Farbwerte jedes einzelnen Pixels eingelesen. Die Informationen unterschiedlicher benachbarter Neuronen werden auf der nächsten Ebene zusammengeführt, sodass Farbübergänge erkannt werden können. Diese wiederum können im Bild Flächen oder Kanten darstellen, die auf der nächsten Netzebene zu Formen zusammengesetzt werden. Die Verbindung verschiedener Formen erlaubt schließlich die Identifikation von Objekten, die auf der Ausgabeebene benannt oder zu Clustern zusammengefasst werden. Die Performanz des Netzwerkes hinsichtlich eines gegebenen Problems (z. B. der Objekterkennung) nähert sich dabei einer adäquaten Lösung mit zunehmender Erfahrung immer weiter an. KNN müssen so konkrete oder allgemeine Aufgaben nicht mehr anhand gegebener Lösungswege, also durch vorgegebene Programme, bewältigen, sondern können anhand großer Datenmengen die relevanten Faktoren einer erfolgreichen Lösungsstrategie für ein gegebenes Problem selbstständig lernen.¹⁹

¹⁹ Vgl. Rojas 1996, S. 73 f.

Von den verschiedenen Lernmethoden seien hier besonders zwei genannt: Das *Representation Learning* erlaubt dem KNN aus gegebenen Daten, wie etwa Bildern oder Spielzügen eines beliebigen sequentiellen Spiels, Muster von Ge-

20 Vgl. Goodfellow 2016, S. 16 f.

21 Vgl. Silver/Hubert/
Schrittweiser 2017; vgl. Moravcik/
Schmid 2017.

22 Vgl. Silver/Schrittweiser/
Simonyan 2017;
vgl. Silver/Hubert/Schrittweiser
2017.

23 Vgl. Heess et al. 2017.

sichern, Figuren und Zeichen oder die impliziten Regeln einer heuristischen Problemlösung zu abstrahieren, um diese dann auch in unbekanntem Situationen anzuwenden.²⁰ Das *Reinforcement Learning* ermöglicht einem KNN in Verbindung mit einem (physischen oder virtuellen) Akteur durch ständige Wiederholung und Variation von Abläufen, wie Körperbewegungen oder Spielzügen, durch eigene Erfahrungen ein heuristisches Modell eines Anwendungsraumes zu entwickeln und zu optimieren, der aufgrund seiner Komplexität (etwa beim Spiel Go oder beim Wetter) oder seiner genuinen Eigenschaften (etwa des Spiels Poker oder der Börse) im Vorhinein nur unvollständig beschrieben werden kann.²¹ Bemerkenswert ist dabei, dass in den vergangenen zehn Jahren unterschiedliche KI-Systeme immer wieder Lösungswege für Probleme gefunden haben, die menschlichen Experten bislang unbekannt waren. Das reicht von neuen Strategien beim Go²² bis zu eigentümlichen bzw. unfreiwillig komischen Techniken der Fortbewegung eines durch ein KNN gesteuerten virtuellen Akteurs.²³

4. Ist ein KI-System autonom?

Es erscheint zunächst zirkulär, Maschinen als autonom zu betrachten, weil sie zur Unabhängigkeit von menschlichen Entscheidungsprozessen programmiert und immer weniger überwacht werden. KNN sind nicht autonom im kantischen Sinne, so dass sie sich aus freiem Willen selbst ein Gesetz des Handelns geben würden. Und doch besteht ihre Selbständigkeit auch nicht allein in der Tatsache ihrer Unabhängigkeit vom menschlichen Handeln und menschlicher Überwachung. Schon Turing vermutete, dass besonders induktive Annahmen über Maschinen, basierend auf unserer langjährigen Erfahrung, unsere Erwartungen und unser Bild zukünftiger Maschinen formen würden (Turing 1950, S. 447). Dabei beruhen unsere Erfahrungen maßgeblich auf Computern, deren Funktionsweise von klassischer Programmierung abhängen. Diese ist in der Regel vom Programmierer wohlgedacht und funktional ausgerichtet. Idealerweise führt ein herkömmlicher Computer die Befehle aus, die ihm gegeben sind. Das maschinelle Lernen stellt in dieser Hinsicht jedoch einen Paradigmenwechsel dar. Lernende Systeme befolgen keine expliziten Regeln, sondern lernen

aus Erfahrung. So können sie nicht nur neue Lösungen entwickeln, sondern auch in Situationen selbständig handeln, deren Bedingungen vage und deren Ergebnisse offen sind. KI-Systeme besitzen Autonomie daher im schwachen Sinn der Selbstursprünglichkeit oder Handlungsautonomie: einzelne Handlungen können dem System als Urheber unmittelbar zugeschrieben werden und sind nicht auf seine absichtsvolle Programmierung zurückführbar. So spielte z. B. das Programm AlphaGo gegen Lee Sedol Strategien, die den Experten in der 3000-jährigen Geschichte des Spiels bis dahin unbekannt gewesen waren und somit auch nicht von den Programmierern vorgegeben werden konnten. AlphaGo hat das Go-Spiel revolutioniert. Es unterscheidet sich in dieser Hinsicht maßgeblich vom Schachcomputer Deep-Blue, der das gesammelte Schachwissen von Experten in Form von Algorithmen enthielt und dessen Züge bei einer genauen Analyse prinzipiell vorhersagbar gewesen wären. Komplexere Formen der Autonomie, wie etwa personale, moralische oder rationale Autonomie, die ein Handeln aus Gründen voraussetzt, sind jedoch in keinem dieser Programme notwendigerweise gegeben (Misselhorn 2018, S. 75 ff.).

Berücksichtigt man die Möglichkeit, dass ein KI-System Entscheidungen fällt, die von unsicheren Umgebungsbedingungen abhängig sind, wird deutlich, dass sich das Verhalten der Maschine nicht ausschließlich aus seiner Programmierung ableiten lässt. Die Forderung nach einem Fachgebiet Machine Behaviour Studies weist auf diesen wichtigen Faktor hin: selbst den Ingenieuren sind die möglichen Aktionen technischer Sachsysteme oder die Folgen ihrer Entwicklung nicht immer vollständig bekannt. Somit kann die Performanz moderner KI-Systeme auch nicht unmittelbar vom Willen des Nutzers oder des Programmierers abhängen. Die logische Konsequenz der eingangs erwähnten Werkzeugmetapher, die der Erforderlichkeit einer Maschinenethik entgegensteht und davon ausgeht, dass jedes technische Gerät der vollen Verfügungsgewalt des Nutzers oder des Ingenieurs unterliegt, ist daher unzutreffend. KI-Systeme im Rahmen einer Technikethik ausschließlich als Werkzeuge zu bewerten, würde bedeuten, dass technisches Handeln regelmäßig Auswirkungen zeitigt, die der Handelnde prinzipiell nicht vorhersehen und damit auch nicht beabsichtigen kann. Den

Nutzer eines solchen Gerätes für die Folgen einer so gearteten Handlung verantwortlich zu machen, scheint daher eine zu große moralische Belastung, woraufhin er die Nutzung mit guten Gründen per se ablehnen wird, um das Risiko der Verantwortung ausschließen zu können bzw. nicht tragen zu müssen. Überträgt man die Verantwortung stattdessen unter ingenieursethischer Hinsicht auf den Hersteller, der die möglichen Einsatzfelder seines technischen Produkts weder vorhersehen noch alle möglichen Szenarien durch das Design antizipieren kann, hätte dies wiederum zur Folge, dass KI-Systeme selbst mit dem kleinsten Grad an Unsicherheit nicht mehr produziert würden, damit der Hersteller seinerseits das Risiko einer nicht abzuschätzenden Verantwortung für sich ausschließen kann.

Wo die Produktion und Anwendung autonomer KI-Systeme allerdings gesellschaftlich gewünscht ist, etwa in Einsatzbereichen, die als langweilig, schmutzig oder gefährlich – dull, dirty and dangerous – für den Menschen eingeschätzt werden, oder in denen sich durch die Automatisierung ein größerer Schaden weitgehend vermeiden ließe – wie etwa die überwiegende Zahl der Verkehrstoten – muss das Risiko auch gesamtgesellschaftlich getragen oder direkt von der KI als Urheber verantwortet werden. Es scheint also mindestens zweckmäßig in analoger Weise von autonomen KI-Systemen zu sprechen, wenn ihre Performanz von direkten menschlichen Absichten oder Einflüssen in der beschriebenen Weise unabhängig ist. Wo die Entwicklung autonomer KI-Systeme jedoch gesellschaftlich nicht akzeptiert werden kann, etwa im Bereich der Rechtsprechung oder der Kriegsführung,²⁴ muss ihre Entwicklung und Verbreitung unterbunden werden. Das Problem einer sinnvollen Verantwortungszuschreibung beim Einsatz autonomer KI-Systeme wird also aufgrund ihrer relativen Autonomie einerseits und aufgrund der gesellschaftlichen Akzeptanz ihrer Verbreitung andererseits eine fundamentale Herausforderung jeder Maschinenethik bleiben.

24 Vgl. Asaro 2012.

5. Gibt es für KI-Systeme moralische Konfliktsituationen?

Man ersetze einen Menschen in einer beliebigen moralischen Konfliktsituation durch eine KI. Hat dieses technische

System dann einen dem des Menschen vergleichbaren moralischen Konflikt? Es mag den moralischen Konflikt zwar nicht auf dieselbe Weise *wahrnehmen* wie der Mensch, den es ersetzt. Dennoch ist es unstrittig, dass in der gegebenen Situation Werte in Frage stehen, die auch von dem KI-System berücksichtigt werden müssen, damit es das gegebene Problem im Sinne des Menschen, den es vertritt – respektive der Menschen, die von seinen Entscheidungen betroffen sind –, bewältigen kann. Um des Argumentes willen, gilt diese Überlegung selbst dann noch, wenn es sich bei dem Konflikt um einen individuellen Konflikt – etwa aufgrund persönlicher Verpflichtungen – der vertretenen Person handelt, der aus der Perspektive einer anderen Person – mit anderen Verpflichtungen – nicht auftreten würde. Die Überantwortung eines solchen Problems an eine moralische Maschine würde vermutlich zu keiner befriedigenden Lösung führen.²⁵ In der Praxis tritt dieser Fall aber auch eher selten auf. Hier ist es wahrscheinlicher der Fall, dass KI-Systeme an sich unproblematische Aufgaben von Menschen übernehmen – wie etwa die Navigation durch den Straßenverkehr – und dabei mit Problemen konfrontiert werden, die für den Anwendungsbereich untypisch sind, aber dennoch einer unmittelbaren Lösung bedürfen. Offensichtlich handelt es sich bei solchen Problemen auch dann noch um eine moralische Konfliktsituation, wenn wir einer Maschine grundsätzlich absprechen, dass sie den vorliegenden Konflikt bewusst als solchen, also auf die gleiche Weise wie ein typischer Autofahrer, erlebt. Schließlich muss das KI-System diesen Konflikt selbständig lösen, wenn sein Design eine hinreichende Autonomie und damit die Kontrolle über die betreffende Situation vorsieht bzw. der Eingriff von Menschen aus Gründen der individuellen Gefährdung, des mangelnden Urteilsvermögens oder Ähnlichem ausgeschlossen ist. Während die zum gegenwärtigen Zeitpunkt spekulative Möglichkeit einer künstlichen Superintelligenz oder wenigstens einer allgemeinen KI auf menschlichem Niveau – sogenannte starke KI – hier vernachlässigt werden kann, ist ersichtlich, dass die gegenwärtigen Erfolge bei der Entwicklung als smart bezeichneter KI-Systeme – sogenannte schwache KI – mit verhältnismäßig begrenzten Anwendungsbereichen den genannten Problemzusammenhang bereits maßgeblich verändern: »Je intelligen-

²⁵ Vgl. McDermott 2008.

ter und autonomer Maschinen werden, desto größer ist die Wahrscheinlichkeit, dass sie in Situationen geraten, die ihnen moralische Entscheidungen abverlangen.« (Misselhorn 2018, S. 75 ff.) Auch wenn eine Maschine ein Dilemma nicht auf dieselbe Weise als Konflikt wahrnimmt wie wir, könnte sie dennoch mit einem entsprechenden Training in der Lage sein, das gegebene Problem entweder analog zu bekannten und vergleichbaren Problemen zu lösen oder bei einer hinreichend guten Problembeschreibung eine von außen als adäquat akzeptierte Lösung selbständig zu entwickeln.²⁶

26 Vgl. Brand 2018.

6. Wie können KI-Systeme zu moralischen Akteuren werden?²⁷

27 Vgl. Allen/Varner/Zinser 2000; vgl. Wallach/Allen/Smit 2008; vgl. Wallach/Allen 2009.

Unter Berücksichtigung der Einschränkung auf schwache KI hat sich in der Maschinenethik eine Klassifizierung technischer Geräte als Akteure durchgesetzt, die Maschinen je nach ihrer Bauart und der Art und Weise, wie sie moralische Werte oder ethische Überlegungen in ihrem Design verwirklichen, in drei unterschiedliche Kategorien einstuft: in implizite, explizite und vollkommene moralische Akteure. Diese Kategorien stellen immer höhere Hürden für die Entwicklung moralischer KI-Systeme dar und sollen im Folgenden kurz skizziert werden:²⁸

28 Vgl. hier und im Folgenden Moor 2011.

Implizite moralische Akteure: Implizite moralische Akteure verhindern moralisch fragwürdiges Verhalten durch ihr Design. So verhindert eine durch einen Alkoholtest ausgelöste Wegfahrsperrung bei einem Auto, dass ein alkoholisierter Fahrer im Straßenverkehr zu einer Gefahr für andere Verkehrsteilnehmer werden kann. Die Entscheidung, einen alkoholisierten Menschen am Führen eines Fahrzeuges zu hindern, lässt sich dabei nicht auf die Software des Fahrzeugs selbst, sondern auf das vorausschauende Design des Entwicklers zurückführen, das eine Handhabung der Technik entgegen der geltenden Straßenverkehrsordnung unmöglich macht. Das »moralische Verhalten« der Maschine ist in diesem Fall gänzlich vom Entwickler abhängig. Ihr »Verhalten« ist vollständig durch das gegebene Programm reguliert und – technisches Versagen ausgenommen – absolut vorhersagbar. Die Maschine kann ihr eigenes Design nicht überwinden, gegen die impliziten Regeln nicht verstoßen, diese reflektieren oder in Frage stellen.²⁹

29 Vgl. ebenda, S. 15 f.

Explizite moralische Akteure: Das Programm eines expliziten moralischen Akteurs basiert auf einer konkreten ethischen Theorie, die durch den Programmierer festgelegt wird. Das KI-System kann gegen diese Metatheorie nicht verstoßen. Er könnte sich auf der Basis dieser Ethik jedoch der Anweisung eines Menschen im Einzelfall widersetzen, wenn das Programm nicht eine konsequente Befolgung menschlicher Anweisungen um jeden Preis vorsieht. Die Roboter-gesetze bilden ein solches Programm, mit dem explizite moralische Akteure ausgestattet sein sollen: (1) Ein Roboter darf kein menschliches Wesen verletzen oder durch Untätigkeit zulassen, dass einem menschlichen Wesen Schaden zugefügt wird. (2) Ein Roboter muss den ihm von einem Menschen gegebenen Befehlen gehorchen – es sei denn, ein solcher Befehl würde mit Regel eins kollidieren. (3) Ein Roboter muss seine Existenz schützen – es sei denn, der Schutz würde mit Regel eins oder zwei kollidieren.³⁰ Der Unterschied zur Kategorie impliziter moralischer Akteure besteht in erster Linie darin, dass nicht eine verbindliche Handlung für eine konkrete Situation, sondern ein verbindliches Bewertungsmaß für Handlungen in allen möglichen Situationen von Seiten der Programmierer vorliegt.³¹ Prima facie denkbar sind beispielsweise deontologisch gesteuerte, künstliche moralische Akteure auf der Grundlage des kategorischen Imperativs,³² sowie konsequentialistische Akteure, deren Verhalten das Nettoglück aller von einer Handlung betroffenen Individuen maximieren soll.³³ Auch ein tugendbasierter künstlicher moralischer Akteur fiele grundsätzlich in diese Kategorie und es scheint, dass gerade die Tugendethik, die wesentlich auf der Formung eines tugendhaften Charakters durch lernende Gewöhnung basiert, mit dem hier skizzierten Ansatz des maschinellen Lernens in besonderem Maße vereinbar ist.³⁴ Die Anstrengungen im Bereich der Maschinenethik beziehen sich primär auf diese Kategorie künstlicher moralischer Akteure. Die verschiedenen Ansätze sind dabei selbstverständlich mit unterschiedlichen Problemen und Herausforderungen behaftet.

Vollkommene moralische Akteure: Ein vollkommener moralischer Akteur kann nicht nur explizite moralische Urteile fällen, sondern verfügt darüber hinaus auch über die Kompetenz, diese vernünftig zu begründen. In dem Maße, in

³⁰ Isaac Asimov hat sich mit den Problemen und Implikationen seiner Robotergesetze in zahlreichen Romanen auseinandergesetzt. In der Maschinenethik gilt dieser Ansatz heute nicht mehr als gangbar (vgl. Anderson 2011).

³¹ Vgl. Moor 2011, S. 16 ff.

³² Vgl. Powers 2011; vgl. aber auch Brieger 2019.

³³ Vgl. Grau 2011.

³⁴ Vgl. Berberich/Diepold 2018; Brand 2018.

35 Vgl. Moor 2011, S. 18 f.

dem künstliche Akteure auf diesem Niveau wahrscheinlich werden, wird ihre Beurteilung im Rahmen einer Maschinenethik erforderlich.³⁵ Insofern aber vorausgesetzt wird, dass Maschinen weder über ein Bewusstsein, noch über freien Willen verfügen, bleibt ihre Kompetenz hinter den Anforderungen an vollkommene moralische Akteure zurück. Sollte es allerdings gelingen, freien Willen und ein phänomenales Bewusstsein in einer Maschine zu implementieren – und alle Skeptiker davon zu überzeugen, dass dieser Durchbruch tatsächlich geschafft ist –, dann gälten für die Maschinen in moralischer Hinsicht dieselben Rechte und Pflichten wie für Menschen.

7. Die Notwendigkeit einer Maschinenethik

36 McDermott 2008 hat zurecht darauf hingewiesen, dass die Verbindung eines KI-Programms mit einer physischen Maschine alleine nicht hinreichend ist, um einen nützlichen KI-Akteur zu entwickeln. Vgl. außerdem Dreyfus 1993, S. 236 ff.

Die Tatsache, dass KI-Systeme durch die Integration von Software und Hardware im Lernprozess zu einer anspruchsvollen Performanz gelangen, ist ein stark vernachlässigtes Faktum der Reflexion des Fortschritts in der KI-Entwicklung.³⁶ Besonderes Augenmerk ist daher in Zukunft der Frage zu widmen, ob ein KNN-basiertes KI-System im Rahmen einer Handlungstheorie und insbesondere einer ethischen Bewertung von Handlungen – im Sinne einer *agent causation* – als eigenständiger Urheber einer Handlung und damit als selbständiger Akteur eingestuft werden kann. Legt man die Möglichkeit zugrunde, dass die für eine erfolgreiche Handlung entscheidenden Informationen nicht *a priori* in einem internen Modell des KI-Systems vorhanden sein müssen, sondern als Informationen im Entscheidungskontext selbst vorhanden sind und hier erst bei Bedarf aufgenommen werden können, könnte sich aus der Integration von KI und physischem Akteur eine Lösung sowohl für einige entwicklungspezifische Herausforderungen als auch für den maschinenethischen Ansatz zur maschinellen Lösung moralischer Probleme ergeben.³⁷

37 Vgl. Wallach/Allen 2009, S. 65; vgl. noch einmal Dreyfus 1993, S. 236 ff.

Die jüngsten Erfolge im Bereich des maschinellen Lernens werfen die Frage auf, ob auch moralisch relevante Probleme auf maschinellem Weg zufriedenstellend gelöst werden können. Die Entwicklung von KNN spielt hinsichtlich der Lösung ethischer Konfliktsituationen und für die Entwicklung eines künstlichen moralischen Akteurs im Sinne der Maschinenethik daher eine zentrale Rolle. Das Potential der KNN

besteht darin, dass sie prinzipiell völlig neue Lösungen für moralische Herausforderungen eröffnen, da die in unvorhergesehenen Situationen nötigen heuristischen Entscheidungen nicht von Programmierern antizipiert werden müssen, sondern Lösungsstrategien und Bewertungsparameter vom KNN durch maschinelles Lernen weitgehend selbstständig erlernt und flexibel angewandt werden können. Ist man den faustischen Pakt der Entwicklung von KI einmal eingegangen, gilt es nun das ambivalente Potential moderner Technik, das sowohl Hoffnungen als auch Ängste weckt, zu steuern und an unserer Vorstellung vom guten Leben auszurichten. Die unmittelbare Implementierung unserer Werte in die Funktionsweise und Entscheidungsparameter der Maschinen im Sinne einer Maschinenethik ist in diesem Unternehmen unerlässlich.

-
- Allen, Colin, Gary Varner und Jason Zinser. 2000. Prolegomena to Any Future Artificial Moral Agent. In *Journal of Experimental & Theoretical Artificial Intelligence* 12(3): 251–261.
- Anderson, Michael und Susan L. Anderson (Hrsg.). 2011. *Machine Ethics*. Cambridge: Cambridge University Press.
- Anderson, Susan L. 2011. The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics. In *Machine Ethics*, hrsg. M. Anderson und S. Anderson, 285–296. Cambridge: University Press.
- Asaro, Peter. 2012. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. In *International Review of the Red Cross*, 94, 886.
- Asimov, Isaac. 1982. *Meine Freunde, die Roboter*. München: Heyne.
- Berberich, Nicolas und Klaus Diepold. 2018. The Virtuous Machine - Old Ethics for New Technology, doi: arXiv:1806.10322.
- Blutner, Doris. 2014. *Herrschaft und Technik*. Entscheidungsträgerschaft im Wandel. Berlin: Springer.
- Brand, Lukas. 2018. *Künstliche Tugend. Roboter als moralische Akteure*. Regensburg: Pustet.
- Brieger, Julchen. 2019. Über die Unmöglichkeit einer kantisch handelnden Maschine. In *Maschinenethik. Normative Grenzen autonomer Systeme*, hrsg. M. Rath, F. Krotz und M. Karmasin, 107–120. Wiesbaden: Springer.
- Dreyfus, Hubert. 1993. *What computers still can't do*. A Critique of artificial Reason. Cambridge: MIT Press.
- Goodfellow, Ian, Yoshua Bengio und Aaron Courville. 2016. *Deep Learning*. Cambridge: MIT Press.

- Grau, Christopher. 2011. There is no »I« in »Robot«: Robots and Utilitarianism. In *Machine Ethics*, hrsg. M. Anderson und S. Anderson, 451-463. Cambridge: University Press.
- Heess, Nicolas et al.. 2017. Emergence of locomotion behaviours in rich environments. arXiv:1707.02286.
- Lenzen, Manuela. 2018. *Künstliche Intelligenz. Was sie kann und was uns erwartet*. München: Beck.
- McDermott, Drew. 2008. Why Ethics is a high hurdle for AI, Presented at North American Conference on Computers and Philosophy (NA-CAP) Bloomington, Indiana. <http://www.cs.yale.edu/homes/dvm/papers/ethical-machine.pdf> (zuletzt eingesehen 15. Juli 2019).
- Misselhorn, Catrin. 2018. *Grundbegriffe der Maschinenethik*. Stuttgart: Reclam.
- Moor, James H. 2011. The Nature, Importance, and Difficulty of Machine Ethics. In *Machine Ethics*, hrsg. M. Anderson und S. Anderson, 13-20. Cambridge: University Press.
- Moravik, Matej und Martin Schmid. 2017. DeepStack: Expert-level Artificial Intelligence in Heads-up No limit Poker. In *Science* 356, S. 508–513.
- Powers, Thomas M. 2011. Prospects for a Kantian Machine. In *Machine Ethics*, hrsg. M. Anderson und S. Anderson, 464-475. Cambridge: University Press.
- Rahwan, Iyad et al. 2019. Machine behaviour. In: *Nature* 568, S. 477-486.
- Rath, Matthias, Matthias Karmasin und Friedrich Krotz 2019. Brauchen Maschinen Ethik? Begründungstheoretische und praktische Herausforderungen. In *Maschinenethik. Normative Grenzen autonomer Systeme*, hrsg. M. Rath, F. Krotz und M. Karmasin, 1-12. Wiesbaden: Springer.
- Rohbeck, Johannes. 1993. *Technologische Urteilskraft – Zu einer Ethik technischen Handelns*. Ort: Suhrkamp.
- Rojas, Raúl. 1996. *Theorie der neuronalen Netze. Eine systematische Einführung*. Berlin: Springer.
- Ropohl, Günter. 1996. *Ethik & Technikbewertung*. Frankfurt am Main: Suhrkamp.
- Ropohl, Günter. 1999. *Technologische Aufklärung*. Beiträge zur Technikphilosophie. Frankfurt am Main: Suhrkamp.
- Silver, David et al.. 2017. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm, doi: arXiv:1712.01815.
- Turing, Alan M. 1950. Computing Machinery and Intelligence. In *Mind* 59(236), S. 433–460.
- Wallach, Wendell und Colin Allen. 2009. *Moral Machines. Teaching Robots Right from Wrong*. New York. Oxford: University Press.
- Wallach, Wendell, Colin Allen und Iva Smit. 2008. Machine Morality: Bottom-up and Top-down Approaches for Modelling Human Moral Faculties. In *AI & Society* 22(4), S. 565–582.
- Wallach, Wendell und Asaro Peter. 2017. *Machine Ethics and Robot Ethics*. New York: Routledge.





Maschinelle Intelligenz – Fluch oder Segen? Es liegt an uns! Dirk Helbing¹

1 Dieser Beitrag wurde ursprünglich für die Digital Responsibility Initiative der Telekom geschrieben und ist unter folgenden URL abrufbar: <https://www.telekom.com/de/konzern/digitale-verantwortung/details/maschinelle-intelligenz---fluch-oder-segen--es-liegt-an-uns---352200>. Er wurde überdies auf Englisch veröffentlicht als Kapitel 4 des Buches »Towards Digital Enlightenment«, herausgegeben von Dirk Helbing, erschienen bei Springer International (2019). Für die vorliegende Fassung wurden Onlinequellen aktualisiert.

Künstliche Intelligenz (KI) kann uns in vielerlei Hinsicht eine große Hilfe sein. Vor allem in Verbindung mit Robotern könnte sie uns im Alltag entlasten (z. B. beim Saubermachen). Sie kann uns schwere, gefährliche oder langweilige Arbeiten abnehmen. Sie kann helfen, Leben zu retten und Katastrophen besser zu bewältigen. Sie kann Alte und Kranke unterstützen. Sie kann uns täglich zur Hand gehen und unser Leben interessanter machen. Die meisten von uns würden gerne von diesen neuen Möglichkeiten profitieren, dessen bin ich mir sicher. Allerdings war bislang noch jede Technologie mit Nebenwirkungen und Gefahren verbunden. Wenn wir nicht aufpassen, können Menschen ihr Recht auf Selbstbestimmung und Demokratie, Unternehmen die Kontrolle und Staaten ihre Souveränität verlieren. Im Folgenden zeige ich Ihnen anhand eines Worst-Case- und eines Best-Case-Szenarios, dass sich unsere Gesellschaft an einem Scheideweg befindet. Jetzt kommt es darauf an, den richtigen Weg zu gehen.

Über lange Zeit war der Fortschritt der Forschung auf dem Gebiet der künstlichen Intelligenz (KI) äußerst schleppend. Allerdings nehmen solche Entwicklungen einen exponentiellen Fortschrittsverlauf. Nach Jahrzehnten sehr langsamer, kaum spürbarer Veränderungen geht plötzlich alles sehr schnell und wird immer schneller. Ray Kurzweil, ein Technologie-Guru aus dem Silicon Valley, der am Google Brain-Projekt arbeitet, geht davon aus, dass Computer bis 2030 die Leistungsfähigkeit des menschlichen Gehirns und bis 2060 die Leistungsfähigkeit aller menschlichen Gehirne zusammen übertrumpfen werden. Prognosen dieser Art galten lange als Science-Fiction, doch mit neuen Deep Learning-Algorithmen kann KI eigenständig lernen und explosionsartige Fortschritte erzielen.

Computer sind bereits seit Jahrzehnten die besseren Schachspieler und uns mittlerweile in fast allen beliebten Strategiespielen überlegen. Der Watson-Computer von IBM

gewinnt sogar Spielshows – und nicht nur das: Auch bei vielen medizinischen Diagnosen überragt er die humane Konkurrenz. Heutzutage werden rund 70 Prozent aller Finanzgeschäfte durch autonome Computeralgorithmen getätigt und schon bald gibt es selbstfahrende Autos, die unsere Fahrkünste in den Schatten stellen. Auch wenn es darum geht, Handschriften zu erkennen, Sprachen zu verstehen und zu übersetzen oder Muster zu identifizieren, kommen Algorithmen den menschlichen Fähigkeiten sehr nahe. In Anbetracht der Tatsache, dass 90 Prozent der heutigen Berufe genau diese Fähigkeiten erfordern, wird es schon sehr bald möglich sein, Routineaufgaben durch Computeralgorithmen oder Roboter zu ersetzen, die leistungsfähiger sind, nie müde werden, sich nicht beschweren und keine Sozialversicherungsbeiträge oder Steuern zahlen.

In meinem Beitrag zu John Brockmans Buch *What to Think About Machines that Think* (Brockman (Hrsg.) 2015) beschreibe ich die Situation wie folgt: »Durch den explosionsartigen Anstieg der Rechenleistung und Datenmenge, angeheizt durch leistungsfähige Machine Learning-Algorithmen, überholt siliziumbasierte Intelligenz auf lange Sicht die kohlenstoffbasierte Intelligenz. Intelligente Maschinen müssen nicht mehr programmiert werden. Sie lernen und entwickeln sich eigenständig – und zwar viel schneller, als sich die menschliche Intelligenz entwickelt.«

Jim Spohrer von IBM prognostiziert:

»KI mag uns anfangs noch als Instrument dienen, aber schon bald werden Roboter unsere Teamkollegen und dann unsere Lehrer sein.« Ich bin überzeugt, dass KI neue intelligente »Lebensformen« hervorbringen wird. Aber wird hochentwickelte KI dem Menschen weiterhin dienen, ihn versklaven oder ihn einfach links liegen lassen (wie der Deep Learning-Experte Jürgen Schmidhuber vermutet)? Zum jetzigen Zeitpunkt kann niemand diese Frage beantworten. Steve Wozniak, Mitbegründer von Apple, kommentiert dies wie folgt:

»[...] Ich stimme zu, dass die Zukunft der Menschheit beängstigend und wenig rosig aussieht. Wenn wir zulassen, dass uns Maschinen alles abnehmen, werden sie über kurz oder lang schneller denken als wir und sich der langsamen Menschen entledigen, da sie Unternehmen selbst sehr viel effizienter führen können. [Aber:] Werden wir wie Götter verehrt? Wie Haustiere gehalten? Oder wird auf uns wie

auf Ameisen herumgetrampelt? Ich weiß es nicht...«²

2 vgl. Storm 2015a. Die Technologievisionäre Bill Gates und Elon Musk äußerten kürzlich ähnliche Bedenken hinsichtlich Superintelligenz. Was macht sie so nervös?

In *What to Think About Machines that Think* analysiere ich das Phänomen wie folgt: »Die Menschen konnten nur schwer akzeptieren, dass die Erde nicht der Mittelpunkt des Universums ist, und sie wollen oft noch immer nicht einsehen, dass die Menschheit durch Zufall und Selektion entstanden ist, wie wir durch die Evolutionstheorie wissen. Jetzt stehen wir kurz davor, als intelligenteste Lebensform auf der Erde abgelöst zu werden. Sind wir dafür bereit?«

Kurz gesagt: »Nein, wir sind nicht dafür bereit, aber wir sollten uns so schnell wie möglich darauf vorbereiten.« Ich möchte mit dem Worst-Case-Szenario beginnen, bevor ich zum Best-Case-Szenario komme.

Worst-Case-Szenario

Wurden in der Vergangenheit Bedenken laut, KI könne die Weltherrschaft übernehmen, beruhigten uns die Experten immer damit, dass wir jederzeit den Stecker ziehen und so das Problem lösen könnten. Leider ist dem nicht so. Erstens würden viele Wirtschafts- und Gesellschaftsbereiche nicht mehr funktionieren, wenn wir intelligente Maschinen einfach abschalten, darunter unsere Geld- und Kommunikationssysteme, viele wichtige Infrastrukturen und die zugehörigen Sicherheitseinrichtungen. Zweitens ist uns KI in gewisser Weise schon entronnen. Seitdem Google seine Machine Learning-Software »Tensorflow« als Open Source bereitstellt,³ kann sie von jedermann frei verwendet werden, auch von Kriminellen und Terroristen. Als Konsequenz dieser KI-Verbreitung erwarten wir einen weiteren Anstieg bei Internetkriminalität und Cyberwar-Bedrohungen. Schon heute verursacht Internetkriminalität jedes Jahr wirtschaftliche Verluste von 3 Billion US-Dollar, Tendenz rapide steigend.

Meine Hauptsorge ist jedoch nicht, dass KI die Weltherrschaft übernehmen könnte, sondern vielmehr, dass eine kleine Personengruppe versuchen könnte, mithilfe von KI-Technologie die Weltherrschaft an sich zu reißen. Lassen Sie mich an dieser Stelle kurz die bisherige Entwicklung

3 vgl. <https://www.tensorflow.org/>,
[Letzter Zugriff:
24. Januar 2016].

zusammenfassen: In den 1970er Jahren experimentierte Chile – inspiriert durch die Arbeit von Norbert Wiener (1894 – 1964) – mit einem dritten politischen System neben Kommunismus und Kapitalismus, der kybernetischen Gesellschaft. In diesem System müssen Fabriken regelmäßig ihre Produktionszahlen einem Kontrollzentrum melden, welches ihnen daraufhin sagt, wie sie ihre Produktion anzupassen haben. In Anbetracht des damaligen Stands der Informationstechnologie funktionierte das System überraschend gut. Vor allem half es der Regierung, einen Generalstreik abzuwenden. Die CIA unterstützte jedoch einen Militärputsch im Land, durch den am 11. September 1973 die chilenische Regierung gestürzt wurde; Präsident Salvador Allende beging Selbstmord.

Scheinbar sind in der Zwischenzeit neue kybernetische Gesellschaftsformen entstanden, die sich Massenüberwachungsdaten zunutze machen. Singapur betrachtet sich selbst beispielsweise als Sozillabor.⁴

4 vgl. Harris 2014;
Tong 2015.

Gerechtfertigt durch die Terrorgefahr werden große Mengen persönlicher Daten über jeden einzelnen Bürger gesammelt. Diese werden anschließend in KI-Systeme eingespeist, die dadurch lernen, wie sich die Bürger verhalten. Ein Beispiel hierfür ist das »China Brain«-Projekt. Anders ausgedrückt: Anhand unserer persönlichen Daten werden digitale Double von uns kreiert, die unsere Entscheidungen und Verhaltensweisen imitieren. Diese Imitation ist jedoch nicht perfekt. Deshalb werden Methoden entwickelt, um unsere Entscheidungen zu manipulieren und unser Verhalten zu steuern.

Die obigen Ansätze basieren auf der Arbeit von Burrhus Frederic Skinner (1904 – 1990). Er sperrte Tiere, z. B. Ratten, Tauben oder Hunde, in sogenannte »Skinner Boxes« und setzte sie gewissen Reizen aus. Durch Belohnung (wie Essen) und Bestrafung (wie Elektroschocks) konditionierte er die Tiere so, dass sie die gewünschten Verhaltensweisen an den Tag legten. Heute sind wir die Versuchskaninchen von Unternehmen wie Google und Facebook⁵, die tagtäglich Millionen automatisierter Experimente mit uns machen. Unsere Skinner-Box ist die um uns herum entstehende »Filterblase« (Pariser 2011), das heißt die personalisierten Informationen über die Welt, die uns erreichen und uns bestimmten Reizen

5 vgl. beispielsweise
Albergotti 2014; Kramer/ Guillory/
Hancock 2014

aussetzen. Dabei lernen die KI-Systeme, wie wir auf diese Reize reagieren und wie durch diese Reize das gewünschte Verhalten ausgelöst werden kann.

6 vgl. Bernstein 2014; Bernstein
2015

Anders gesagt, der Trend geht von der Programmierung des Computers zur Programmierung des Menschen⁶. Diese Manipulation geschieht häufig so subtil, dass wir es gar nicht merken. Was wir als eigene Entscheidung wahrnehmen, wird tatsächlich oft von anderen bestimmt und uns heimlich übergestülpt. Die Technologie, die ursprünglich zur Personalisierung und effektiveren Gestaltung von Werbeanzeigen entwickelt wurde, wird heute vermehrt auch als politisches Instrument eingesetzt. Mittels »Big Nudging« (Helbing 2015a), zusammengesetzt aus »Nudging« (engl. für Stupsen oder Schubsen) aus der Verhaltensökonomik und »Big Data« über all unsere Verhaltensweisen, werden die öffentliche Meinung und die Wahlergebnisse beeinflusst.

Der Nudging-Ansatz hat jedoch nicht die Macht, ein gesundes und umweltbewusstes Verhalten der Bevölkerung in einem Land zu bewirken (Rowson 2011). Dies ist der Grund, weshalb effektivere Feedback-Mechanismen, wie personalisierte Preise, entwickelt werden. Der »Citizen Score« (Storm 2015b), wie er derzeit in China implementiert wird, ist der erste Schritt in diese Richtung. Alles, was die Bürger tun, erhält eine positive oder negative Bewertung: das Kaufverhalten ebenso wie die angeklickten Links im Internet. Sowohl die politische Meinung als auch das Verhalten in sozialen Netzwerken wird ausgewertet.

Der Citizen Score legt fest, welche Kreditkonditionen und welche Jobs eine Person bekommt und in welche Länder sie reisen darf oder nicht. Durch den Citizen Score entsteht so ein moderndes »Kastensystem«. Bei Engpässen bestimmt der Score, wer Anspruch auf knappe Ressourcen hat und wer nicht. Der Citizen Score ist ein Mechanismus, mit dem wir auf der Grundlage willkürlicher, durch »Big Governments« oder »Big Business« ausgewählter Kriterien Tag des Jüngsten Gerichts spielen.

Höchstwahrscheinlich kommen Big Nudging- und Citizen Score-Technologien nicht nur in Singapur und China zum Einsatz. Wenn man dem »Nudging-Papst« Richard Thaler Glauben schenkt, haben in den letzten Jahren mindestens 90 Länder »Nudging-Einheiten« eingerichtet. Bislang ist

wenig über diese Einheiten an die Öffentlichkeit gedrungen. Wir können jedoch davon ausgehen, dass sie über leistungsfähige IT-Infrastrukturen verfügen, die mit persönlichen Daten aus Massenüberwachungs- und Profiling-Tätigkeiten privater Unternehmen gefüttert werden. Die Führung autoritärer Staaten, wie Saudi-Arabien, macht sich diese Infrastrukturen zunutze.

Das Ziel von Big Nudging und Citizen Scores liegt vermutlich darin, Gesellschaften nach dem Modell von Singapur oder China zu kontrollieren. Grundgedanke ist eine datenbasierte, kybernetische Gesellschaft, die von einem »gutmütigen Diktator« regiert wird. Ein solcher Ansatz widerspricht allerdings den demokratischen Prinzipien und Verfassungsrechten. Sollte das singapurische oder das chinesische Modell in den zuvor genannten 90 Ländern zur Anwendung kommen, gefährdet dies alle Demokratien weltweit.⁷

⁷ Wenn das Kräfteverhältnis zwischen denen, die entscheiden, und denen, die folgen, zu stark voneinander abweicht, wird früher oder später jedes System kippen und außer Kontrolle geraten, wie das Stanford-Prison-Experiment gezeigt hat.

Das Problem besteht darin, dass eine digitale Machtübernahme problemlos möglich und nicht umkehrbar ist. Beispielsweise kann jeder, der Zugang zu einer Big Nudging-Infrastruktur hat, ein Wahlergebnis beeinflussen (Epstein/Robertson 2015). Ferner können Terroranschläge oder andere Ereignisse, die die Bevölkerung traumatisieren, ausgenutzt werden, um die demokratischen Grundrechte einzuschränken. So geschehen in Frankreich, wo Massenüberwachung bereits heute genutzt wird, um die eigenen Bürger in Schach zu halten und die Opposition zu unterdrücken, wie ein hochrangiges UN-Komitee kritisierte (Fiedler 2016; o.V. 2016a). Auch am Beispiel Polens zeigt sich, wie einfach es ist, demokratische Institutionen wie das Verfassungsgericht und die Pressefreiheit zu untergraben. Ähnliche Entwicklungen sind in Ungarn mit der geplanten Verfassungsänderung (Hinz 2016) zu beobachten ebenso wie in der Türkei, in der sowohl die Opposition als auch die kurdische Minderheit schon heute unterdrückt werden.

Angesichts des beängstigenden Potenzials, Schaden anzurichten und Grundrechte zu verletzen, das der oben beschriebene Zusammenschluss der Technologien mit sich bringt, benötigen wir dringend Initiativen, die möglichst rasch die folgenden Maßnahmen umsetzen:

- Die obigen Instrumente bedürfen einer demokratischen Kontrolle durch das Parlament und dürfen nicht

von einem Kanzler oder Präsidenten, dem Militär oder dem Geheimdienst eigenmächtig angewendet werden.

– Den Oppositionsparteien sollte sinnvollerweise Zugang zu diesen Informationssystemen gewährt werden, um ein ausgewogenes Machtverhältnis sicherzustellen. (Bedenken Sie, dass zum Verständnis und zur Kontrolle komplexer Systeme wie unserer Gesellschaft eine pluralistische Herangehensweise notwendig ist.)

– Die Nutzung dieser Instrumente muss einem demokratischen Mandat und wissenschaftlichen Prinzipien folgen. Sie sollten von interdisziplinären Teams führender Wissenschaftler betrieben werden (darunter Psychologen, Soziologen, Wirtschaftswissenschaftler, Informatiker und Komplexitätsforscher). Diese Gruppen müssen offen für einen internationalen Austausch sein und vor öffentlichen internationalen Konferenzen Rechenschaft über ihre Aktivitäten ablegen.

– Eine ethische Aufsicht sollte gewährleistet sein.

– Persönliche Daten müssen anonymisiert und Datenschutzverletzungen bestraft werden.

– Transparenz bezüglich der laufenden Aktivitäten ist wichtig. Es muss aufgezeichnet werden, wer das System wie nutzt, und die Nutzung muss der Öffentlichkeit regelmäßig in verständlicher Form offengelegt werden.

– Es muss eine Opt-Out-Möglichkeit bestehen (zumindest vom Scoring und Big Nudging), um die Selbstbestimmung über persönliche Daten zu gewährleisten. (Dies würde auch eine vertrauenswürdige Nutzung der Methoden fördern.)

– Sollten soziale Experimente unerwünschte Nebenwirkungen haben, müssen die Opfer angemessen entschädigt werden.

Aller Voraussicht nach werden die Geheimdienste einen individuellen Zugang zu diesen Informationssystemen haben wollen, aber auch für sie sollten bestimmte Prinzipien gelten:

– Die Nutzung dieser Instrumente muss aufgezeichnet werden. Nudging im grossen Stil sollte verboten werden. Auf individueller Ebene kann Nudging im Einzelfall zulässig sein.

– Massenüberwachung muss ebenfalls verboten werden. Die Aufhebung der Anonymität sollte auf einzelne

Personen beschränkt werden und demokratischen Kontrollen unterliegen.

Privatunternehmen müssen die neue europäische Datenschutz-Grundverordnung beachten. Dabei muss die Regierung deren Einhaltung nicht nur durch große IT-Unternehmen, sondern auch durch relativ unbekanntere Unternehmen, die mit unseren persönlichen Daten handeln, durchsetzen.

8 Helbing/Pournaras 2015,
S. 33-34.

Die Demokratie verdient zweifellos ein »digitales Upgrade«⁸, aber wenn sie aussterben würde, wäre dies eine Katastrophe für die Zukunft unseres Planeten, d. h. wenn es keinen Wettbewerb mehr zwischen verschiedenen politischen Systemen gäbe. Es ist nachgewiesen, dass langfristig nur Demokratien friedlich koexistieren können, was auf dem effektiven Gleichgewicht unterschiedlicher gesellschaftlicher Interessen beruht (durch Subsidiarität, föderale Organisation, Gewaltentrennung und Bürgerbeteiligung). Wir müssen uns vor Augen führen, dass diese institutionellen Strukturen ebenso wie die Menschenrechte und unser Rechtssystem Ergebnis einer mehr als hundertjährigen Geschichte sind, inklusive zweier Weltkriege.

Weshalb bin ich so sicher, dass ein demokratischer, datenbasierter Ansatz überlegen ist? Weil es einen ähnlichen Fall in unserer Geschichte bereits gab: den Wettbewerb zwischen zentral regierten kommunistischen Regimes mit hierarchischen Strukturen von oben nach unten und föderal organisierten kapitalistischen Systemen, die in höherem Maße von unten nach oben gelenkt wurden. Der Kapitalismus gewann diesen Wettbewerb, weil Innovation meist von unten nach oben geschieht. (Die reichsten Personen der Welt haben ihr Geld größtenteils in wenigen Jahrzehnten mit vollkommen neuen Geschäftsmodellen gemacht. Nicht wenige dieser Unternehmen wurden von Studienabbrechern in irgendeiner Garage gegründet.)

Umfassende Analysen empirischer Daten belegen, dass der Wechsel von einem autokratischen zu einem demokratischen System häufig mit starkem Wirtschaftswachstum einhergeht. Eine Veränderung in umgekehrter Richtung hingegen führt mittelfristig zu einem Verlust von soziopolitischem Kapital und langfristig zur Stagnation des Wirtschaftswachstums (Nax/Schorr 2015). Der Preis für die Abschaffung unserer demokratischen Grundordnung ist also hoch.

Ich arbeite weiterhin an einem wissenschaftlichen Projekt zur Erforschung virtueller (Gaming-) Welten mit. Hier fanden wir heraus, dass Welten, die automatische Sanktionsmechanismen ähnlich dem Citizen Score verwenden, nicht nur weniger attraktiv, sondern auch weniger innovativ sind.

Ich möchte Singapur und China keineswegs kritisieren. Es ist durchaus möglich, dass ihre Governance-Modelle die beste Lösung für ihre Gesellschaften in der derzeitigen historischen Situation sind. Ich bezweifle jedoch, dass dieser Ansatz sich als Modell für den Rest der Welt eignet. Demokratische Staaten sollten sich hieran kein Beispiel nehmen. Für sie muss ein anderes Modell entwickelt werden (siehe Best-Case-Szenario weiter unten).

Bedenken Sie auch, dass der Erfolg von Singapur nicht allein auf dem datenbasierten Ansatz beruht. Singapur ist auch eine Steueroase und importiert Innovationen aus den USA, Deutschland, der Schweiz usw. in weit größerem Umfang, als andere Länder dies tun. Ohne diese Importe wäre die Innovationskraft deutlich geringer. Dabei stelle ich nicht den Import von Innovation aus anderen Ländern in Frage, ich möchte lediglich darauf hinweisen, dass in anderen Ländern liberalere Strukturen notwendig sind, damit diese Innovationen überhaupt erst entstehen. Singapur weiß das (Hanisch 2015), weshalb das Land derzeit versucht, mehr Raum für ein gewisses Maß an »kreativem Chaos« zu schaffen.

Auch China wurde schon oft als politisches Rollenmodell vorgeschlagen. Der durchschnittliche Lebensstandard in China ist jedoch trotz der beeindruckenden Entwicklung nach wie vor sehr viel niedriger als in vielen demokratischen Ländern. Zudem ist das Land derzeit mit akuten Umweltproblemen und großen Marktturbulenzen konfrontiert. Dabei spürt es zunehmend die Grenzen der zentralisierten Führung und weiß, dass es für seine weitere Entwicklung pluralistischer und demokratischer werden muss.

Schließlich sollte es uns zu denken geben, dass keine der IT-Supermächte wie die USA, China oder Singapur eine Stadt in der Liste der zehn lebenswertesten Städte der Welt hat. Wie können wir also annehmen, dass diese Governance-Modelle Gesellschaften mit einer besonders hohen Lebensqualität hervorbringen? Wenn Unternehmen wie Google ein Paradies auf Erden erschaffen könnten, warum ist San Fran-

cisco dann nicht die attraktivste Stadt unseres Planeten? Stattdessen liegen die attraktivsten Städte alle in Ländern, die auf ein ausgewogenes Interessenverhältnis aller Beteiligten achten, die Zivilgesellschaft eingeschlossen.

Zusammenfassend lässt sich festhalten, dass die Selbstbestimmung der Menschheit derzeit auf dem Spiel steht, was uns große Sorgen bereitet. Big Nudging, Citizen Scores und Implantate ebnen den Weg für digitale Sklaverei. Dies gefährdet nicht nur die Freiheit der Menschen. Es riskiert auch die Souveränität von Unternehmen und ganzen Ländern. Auslöser sind dabei nicht nur Massenüberwachung und Spionage. Mithilfe von KI-Systemen lassen sich auch die Schwächen von IT-Systemen und Personen aufdecken, indem bestimmte Reize ausgesendet und die Reaktionen darauf aufgezeichnet werden. So können wir nicht nur lernen, wie wir die Menschen manipulieren (siehe oben), sondern erfahren auch, wie wir IT-Systeme und wichtige Infrastrukturen kontrollieren. Tatsache ist, dass selbst unabhängige KI-Systeme extern steuerbar sind: Dadurch dass sie auf Informationsinput reagieren, lässt sich der Output manipulieren. Die Konsequenz daraus ist, dass derjenige mit dem leistungsfähigsten KI-System alle anderen KI-Systeme und folglich alle Unternehmen, Institutionen und Menschen, die von diesen manipuliert werden, kontrollieren kann. Die explosionsartige technologische Entwicklung, wie Quantencomputing, Memristor-Technologien und lichtwellenbasierte LiFi-Kommunikation, löst ein Wettrennen um die Weltherrschaft aus.

Anders formuliert, die Technologie intensiviert das Wettrennen um die Herrschaft über die Welt und ihre Ressourcen. Heute kontrollieren angeblich 62 Personen so viel Kapital wie 50 Prozent der Weltbevölkerung.⁹ Angeführt wird diese Rangliste von folgenden Personen: Bill Gates (Microsoft, USA), Amancio Ortega (Mode, Spanien), Warren Buffet (Finanzen, USA), Jeff Bezos (Amazon, USA), Carlos Slim Helu (Telekommunikation, Mexiko), Larry Ellison (Oracle, USA), Mark Zuckerberg (Facebook, USA), Charles und David Koch (Öl und andere Produkte, USA), Liliane Bettencourt (L’Oreal, Frankreich), Michael Bloomberg (Finanzdaten, USA), Larry Page (Google, USA), Sergey Brin (Google, USA). Wir beobachten, dass Unternehmen, die mit Daten, Software und Informations- sowie Kommunikationstechnologien handeln, die

⁹ o.V. 2016b; Inzwischen sogar lediglich 8 Personen. vgl. o.V. 2017; <https://www.forbes.com/billionaires> [aktualisierte URL, Letzter Zugriff: 24. Juli 2019].

meisten klassischen Geschäftsmodelle überholen. Ich gehe davon aus, dass wir auch in Zukunft einen raschen Konzentrationsprozess erleben werden, bis die Welt nur noch von sehr wenigen Personen kontrolliert wird.

Auch wenn ich grundsätzlich nichts gegen Freihandelsabkommen habe, steht zu erwarten, dass die geplanten TTIP- und TISA-Abkommen diesen Konzentrationsprozess beschleunigen werden. Letzten Endes wird ein Großteil des Geldes, der Macht und der Ressourcen in den Händen sehr weniger Menschen liegen (die vermutlich nicht in Europa leben). Diese Menschen können dann wie Diktatoren über das Schicksal des Planeten entscheiden. Werden diese Menschen ein Grundeinkommen zahlen oder sich für andere Lösungen einsetzen, die den vielen Arbeitslosen ein Überleben ermöglichen, deren Aufgaben zukünftig von Robotern oder KI-Systemen erledigt werden? Oder steht uns ein Weltkrieg bevor, durch den die Bevölkerung auf, sagen wir, vielleicht eine Milliarde hochqualifizierter Personen zusammengeschrumpft wird, die wir brauchen, um die tolle neue Datenwelt zu führen? Gleich mehrere IT-Unternehmen bauen jetzt eigene Weltraumraketen. Ihr Ehrgeiz, das Universum zu erobern, lässt sich also kaum ignorieren.

Ich glaube jedoch, dass eine Weltherrschaft egal welcher Art weder dem Planeten noch der Menschheit gut täte, und damit bin ich nicht alleine. Über 20.000 Personen haben kürzlich eine Petition unterschrieben, den Einsatz von KI als Waffe gegen die Menschheit zu verbieten.¹⁰ Meiner Ansicht nach sollten wir uns vielmehr um ein kooperatives KI-Paradigma bemühen und uns von »Big Nudging«, »Citizen Scores« und ähnlichen Ansätzen distanzieren, mit denen Millionen, wenn nicht gar Milliarden Menschen von zentraler Stelle aus und von oben nach unten kontrolliert werden können und die nur allzu leicht ein totalitäres Regime hervorbringen, wie wir es bislang nicht kannten.

¹⁰ <http://futureoflife.org/open-letter-autonomous-weapons/>, [Letzter Zugriff: 24. Januar 2016].

Best-Case-Szenario

Zum Glück gibt es aber auch positive Aussichten. Wir stehen an der Schwelle eines neuen Zeitalters – das der digitalen Gesellschaft und Wirtschaft 4.0. Um diesen Wandel erfolgreich zu vollziehen, müssen wir gleiche Chancen für alle schaffen: Wirtschaft, Politik, Wissenschaft und Bürger. Mit

den neuen Informations- und Kommunikationstechnologien ist dies heute einfacher als je zuvor. Die gute Nachricht ist, dass die digitale Wirtschaft kein »Nullsummenspiel« ist. Sie erlaubt uns, den exklusiven Wettbewerb unserer materiellen Welt und unseres alten Wirtschaftssystems zu überwinden. Wettbewerb und Kooperation schließen sich heute nicht mehr aus. Das neue Paradigma heißt »Co-opetition«. Wenn es uns gelingt, einen geeigneten gesetzlichen Rahmen zu schaffen, können wir alle am Wohlstand teilhaben.

Die Vorteile eines offenen Informationsaustauschs werden immer offensichtlicher. Mehr und mehr Menschen erkennen, dass Informationen, Erfindungen und Unternehmen durch diesen Austausch an Wert gewinnen. Richtig organisiert eröffnet die digitale Wirtschaft nahezu unbegrenzte Möglichkeiten, da immaterielle Güter beliebig oft reproduzierbar und auf unzählige Arten nutzbar sind. Beispielsweise wird immer mehr Geld in virtuellen Welten verdient. Das bezieht sich nicht nur auf Computerspiele; wie Bitcoin gezeigt hat, lassen sich Bits in Gold verwandeln. Kaum jemand hätte dies für möglich gehalten.

Wenn wir die Herausforderungen, vor denen die Menschheit derzeit steht, meistern möchten, werden wir unsere Wirtschaft und Gesellschaft völlig neu organisieren (müssen). Zweifellos wird die Welt in den nächsten drei Jahrzehnten verstörende Zeiten mit gravierenden Problemen durchleben: digitale Revolution, Finanz- und Wirtschaftskrisen, Klimawandel (mit extremen Wetterbedingungen und einem Verlust an Biodiversität), Energiewandel, demografische Herausforderungen (wie Alterung und Migration) und labile Friedenslagen.

Ich denke nicht, dass Big Data und künstliche Intelligenz das Allheilmittel für diese Probleme sind (siehe Anhang). Die Welt ist zu komplex, als dass ein Blick in die »Big-Data-Glas-kugel« zuverlässige Zukunftsperspektiven liefern könnte; Predictive Analytics und Kontrollen können sogar kontraproduktiv sein. Sie halten uns möglicherweise davon ab, neue Wege zu beschreiten, da KI-Systeme auf zurückliegenden Daten beruhen und dazu neigen, vergangene Lösungen zu reproduzieren (sogar einen Krieg im Extremfall).

Wir benötigen stattdessen eine belastbare und funktionsfähige Gesellschaftsordnung. Voraussetzung hierfür sind

diverse Systemkomponenten, modulare Systemdesigns und dezentrale Kontrollmechanismen, zum Beispiel eine Beteiligung von unten nach oben. Ein solcher Ansatz erlaubt uns eine flexible Anpassung an unvorhergesehene Ereignisse. Zudem müssen wir die Innovationsrate drastisch anheben. Aus diesen Gründen nehme ich an, dass die neuen Organisationsprinzipien unserer zukünftigen Gesellschaft kollektive Intelligenz und Co-evolution in einer hochgradig diversifizierten, vernetzten Wirtschaft sein werden – eine neu entstehende partizipatorische Gesellschaft, die als »Innovationsökologie« betrachtet werden kann.

Wenn wir die Innovation ankurbeln wollen, ist der Einsatz von Big Nudging und Citizen Scores kontraproduktiv. Diese fördern Opportunismus und Konformismus, anstatt die Risikobereitschaft der Menschen zu stärken und vorhandene Lösungen zu hinterfragen – was in der jetzigen Situation unverzichtbar ist.

Wir brauchen darüber hinaus einen grundlegend neuen Innovationsansatz, der mehr Wert auf offene Innovation legt als bisher, damit auch all die Produkte und Dienstleistungen bereitgestellt werden können, die große Unternehmen aktuell nicht im Angebot haben. Citizen Science (Bürgerwissenschaft), sogenannte Fablabs (öffentliche Zentren für digitale Amateurgemeinschaften) sowie Initiativen zur Mobilisierung der Zivilgesellschaft werden immer wichtiger. Das Schlüsselwort ist Co-creation, das heißt die Bürger erweitern die vorhandenen Informationen, Kenntnisse, Dienstleistungen und Produkte in einem weitgehend offenen Informations- und Innovations-Ökosystem. Dies schließt eine Kommerzialisierung nicht aus, im Gegenteil. Hier hat jeder die Möglichkeit, mit Daten Geld zu verdienen. Bürger und Kunden werden zu Partnern. Die partizipative Gesellschaft der Zukunft stützt sich nicht mehr nur auf großen, globalen Konzernen. Unternehmen jeder Art und Größe ebenso wie Einzelunternehmer spielen eine deutlich größere Rolle als heute. Das ist gut, da Monopole bekanntlich vergleichsweise wenig innovativ sind und kein Interesse an Produkten und Dienstleistungen ohne eine signifikante Rendite von, sagen wir, 20 Prozent haben. Gleichwohl profitieren auch große Unternehmen hiervon. Ein reiches Informations-Ökosystem ist wie ein Regenwald, in dem viele Bäume sehr viel größer sind als die wenigen

Bäume in der Wüste.

In diesem Zusammenhang ist die OpenAI-Initiative bemerkenswert, die kürzlich mit einer Spende von 1 Milliarde US-Dollar gestartet wurde. Der Initiator Elon Musk formulierte die Ziele wie folgt: »KI sollte eine Erweiterung des individuellen menschlichen Willens und im Geiste der Freiheit so breit und gleichmäßig wie möglich verteilt sein.« Zusätzlich dazu müssen wir uns für ein verantwortungsbewusstes Innovationsparadigma einsetzen, das auf die Erschaffung wertorientierter Designs passend für den jeweiligen Kontext und die Kultur ausgerichtet ist. Vor allem müssen wir den KI-Systemen beibringen, moralisch und sozial zu handeln. Hierdurch ändern wir das Paradigma »Mensch-Maschine-Interaktion« in ein Paradigma »Mensch-Maschine-Symbiose«. In John Brockmans Buch *What to Think About Machines that Think* ziehe ich folgende Schlussfolgerung:

[Langfristig] ... lernen intelligente Maschinen höchstwahrscheinlich, dass es von Vorteil ist, vernetzt und kooperativ zu handeln, bei Entscheidungen andere Perspektiven einzubeziehen und auf systemische Ergebnisse zu achten. Sie würden schnell merken, dass Diversität für Innovation, systemische Resilienz und kollektive Intelligenz wichtig ist. Die Menschen würden zum Knotenpunkt in einem globalen Intelligenzgeflecht und einem riesigen Ideen-Ökosystem.

Ich glaube weiterhin, dass wir (soweit möglich) Wissen in Echtzeit generieren und unsere Überlegungen, Beurteilungen und Einsichten adäquat, schnell und weltweit teilen müssen. Im digitalen Zeitalter müssen wir Innovation neu erfinden – von der Forschung über die Veröffentlichung bis hin zur Lehre. Dies verlangt einen neuen Rahmen, den ich als »Pluralitäts-Universalität« bezeichne. Wir müssen uns außerdem Gedanken machen, wie wir die Experimentierfreude fördern können. Viel zu viele Erfindungen sind lediglich bescheidene Verbesserungen bereits vorhandener Ideen, sogenannte lineare Innovationen, die den Lebenszyklus »alter« Produkte verlängern. Stattdessen sollten wir radikal neue Ideen vorantreiben, auch als »disruptive Innovationen« bezeichnet.

Die Frage ist, wie wir sicherstellen, dass solche Innovationen nachhaltige Produkte hervorbringen, die unserer Gesellschaft und Umwelt keinen Schaden zufügen (angesichts des-

sen, dass die Hoffnung, unser Planet erholt sich selbständig von all den Beanspruchungen und Belastungen, noch nicht wahr geworden ist). Hierfür müssen wir externe Faktoren messen und bepreisen, das heißt die externen Kosten bzw. den Nutzen der Produkte, Dienstleistungen und Interaktionen bestimmen. Interessanterweise wird dies durch Big Data und das Internet of Things (IoT) zunehmend möglich. Bedenken Sie dabei, dass die Messung und Bepreisung externer Faktoren weniger Regulierung erfordert, was dazu beitragen kann, die heutige Überregulierung abzubauen. Würden externe Faktoren wie Finanzderivate gehandelt, könnten völlig neue Finanzmärkte entstehen. Dies würde ein enormes wirtschaftliches Potenzial freisetzen. Ein mehrdimensionales Finanzsystem würde auch vollkommen neue Anwendungen ermöglichen, wie selbstorganisierende sozio-ökonomische Systeme, die verschiedene Anreizmechanismen erfordern. Vielfach könnte die Anwendung von Dezentralisierungsansätzen und Prinzipien der Selbstorganisation die Ressourceneffizienz um 30 bis 40 Prozent erhöhen.

Daher ist es in hohem Maße sinnvoll, mithilfe von Informations- und Kommunikationstechnologien die Bürger zu stärken, sodass diese bessere Entscheidungen treffen und mehr zur Wirtschaft und Gesellschaft und zu deren digitaler Transformation beitragen können. Richtig gemacht führt die Stärkung der Nutzer, Kunden und Bürger zu besseren Services, besseren Produkten, besseren Unternehmen, besseren Nachbarschaften, intelligenteren Städten und intelligenteren Gesellschaften.

Beispielsweise können digitale Assistenten den Menschen zu einem gesünderen und umweltfreundlicheren Verhalten verhelfen. Das lässt sich anhand eines GPS-basierten Navigationssystems veranschaulichen. Der Nutzer kann hier sein Ziel festlegen und der digitale Assistent bietet ihm verschiedene Alternativen zur Auswahl an, einschließlich der Vor- und Nachteile jeder Möglichkeit. Anschließend unterstützt der digitale Assistent den Nutzer so gut wie möglich dabei, das Ziel zu erreichen und optimale Entscheidungen zu treffen. Um die Menschen anzuregen, mehr Sport zu treiben und gesünder zu essen, müssen weder der Staat noch die Krankenkassen die persönlichen Daten jedes Einzelnen aufzeichnen. Man könnte sich auch eine Social Media-Plattform

vorstellen, die den Menschen erlaubt, eigene »Gesundheitskreise« zu bilden, in denen der Wettbewerb unter Freunden eine gesündere Lebensweise fördert. Um Anreize zu schaffen, ohne die Privatsphäre zu verletzen, könnten der Staat oder die Krankenkassen diese Gesundheitskreise anstelle von Einzelpersonen belohnen, was aber vermutlich gar nicht zwingend notwendig wäre.

Ich bin überzeugt, dass die moderne Informationstechnologie uns auch helfen kann, Konflikte in der Welt zu lösen, indem der Wettbewerb um knappe Ressourcen minimiert wird. Dies lässt sich durch eine Kombination verschiedener Maßnahmen erreichen. Erstens müssen die Ressourcen effizienter genutzt werden, wie bereits besprochen. Zweitens sind erhebliche Fortschritte bei den Recyclingtechniken möglich. Drittens lassen sich die Prinzipien der Sharing Economy auf immer mehr Bereiche des gesellschaftlichen und wirtschaftlichen Lebens anwenden, einschließlich der Verwaltung und Nutzung des urbanen Raums. Dies eröffnet mehr Menschen einen höheren Lebensstandard, während gleichzeitig der Ressourcenverbrauch gesenkt wird. Um Krieg und Terrorismus vorzubeugen, müssen wir den Lebensbedingungen im Rest der Welt mehr Aufmerksamkeit schenken.

Ferner müssen wir uns bewusst machen, dass in einer multikulturellen Gesellschaft Bestrafungsmechanismen meist keine gesellschaftliche Ordnung herstellen, sondern zur Eskalation des Konflikts führen. Das haben wir nicht nur im Nahen Osten beobachten können, sondern auch in Ferguson und an vielen anderen Orten. Wir benötigen daher neue Mechanismen, um Koordination und Kooperation in einer multikulturellen Welt zu fördern. Geeignete Reputationsmechanismen sind vielversprechende Instrumente, ebenso wie Qualifikation, Wettbewerb, Kommunikation und Abstimmung.

Last but not least, könnte uns ein »Kulturgenomprojekt« zu einem besseren Verständnis der Erfolgsrezepte verhelfen, auf denen andere Kulturen beruhen. Wir könnten diese dann auf innovative Weise kombinieren, um neue gesellschaftliche und wirtschaftliche Werte zu schaffen. Das größte Potenzial dieses Ansatzes liegt direkt in den kulturellen Unterschieden unserer Zeit. Einige dieser kulturellen Er-

- 11 Helbing/Pournaras 2015, S. 33-34; Helbing 2014; Helbing 2015c; Helbing 2016.

folgsmechanismen werden beispielsweise in die Nervousnet-Plattform integriert,¹¹ sodass deren Ansatz »Daten für alle« einen verantwortungsbewussten Umgang bewirkt.

Nervousnet (siehe nervousnet.info) ist eine privat betriebene, offene und partizipative Internet-of-Things-Plattform, die Folgendes ermöglicht: (1) Echtzeitmessung unserer Umwelt, (2) deren wissenschaftlicher Kompetenz, (3) Bewusstsein für die Auswirkungen verschiedener Entscheidungsalternativen, (4) Echtzeit-Feedback zur Unterstützung der Selbstorganisation und (5) Schwarmintelligenz. Dieses Projekt nimmt informationelle Selbstbestimmung ernst. Die Daten werden dezentral gespeichert und es kommen verschiedene Verfahren zum Anonymisieren, Verschlüsseln und »Vergessen« der Daten zur Anwendung. Die Nutzer können selbst entscheiden, welche Art von Daten sie für sich selbst generieren bzw. mit anderen teilen möchten.

Stellen Sie sich außerdem vor, dass alle Daten, die Sie generieren, an einen persönlichen Datenspeicher gesendet werden, in dem diese nach Kategorie sortiert und verwaltet werden können. Vorausgesetzt, wir verfügen über geeignete technische Lösungen und gesetzliche Regelungen, könnten wir dann selbst entscheiden, welche Daten wir mit wem und zu welchem Zweck teilen möchten. Vertrauenswürdigeren Unternehmen erhielten dann Zugang zu mehr Daten. Dies würde einen Wettbewerb um Vertrauen in Gang bringen und die datenbasierte Gesellschaft würde erneut auf Vertrauen basieren.

Es ist höchste Zeit, dass wir dieses Best-Case-Szenario in Angriff nehmen. Zum einen sind wir gezwungen, sehr viel innovativer zu werden, als wir es heute sind. Zum anderen wäre die Umsetzung dieser Vorschläge für alle Beteiligten von großem Vorteil. Es scheint, als hätten die USA bereits erste Investitionen in eine neue Strategie getätigt. Sie setzen auf Re-Industrialisierung einerseits und Citizen Science (Bürgerwissenschaft) sowie kombinatorische Innovation andererseits. Selbst Google hat sich mit der Gründung von Alphabet auf eine neue Strategie eingelassen mit dem Ziel, die Abhängigkeit des Unternehmens von personalisierter Werbung zu verringern. Auch Apple hat den Wert von Datenschutz als Wettbewerbsvorteil erkannt. Die Menschen verstehen endlich, dass die digitale Wirtschaft kein »Nullsum-

menspiel« ist. Im Bereich Internet-of-Things engagiert sich Google für offene Innovation. Tesla Motors hat viele seiner Patente freigegeben und viele Milliardäre haben jüngst große Spenden für gute Zwecke versprochen. Wir sehen also viele Zeichen für Veränderungen. Als einzige Frage bleibt, wann Europa endlich die fantastischen Möglichkeiten der digitalen Revolution nutzen wird. Wir betreten ein digitales Zeitalter, das sich mehr und mehr von materiellen Einschränkungen befreit. Das ist absolut faszinierend!

Anhang: Häufige Fallstricke datenbasierter Technologien

Seit einigen Jahren verbreitet sich das Konzept von Big Data- und KI-basierten Smart Nations auf der ganzen Welt. Zweifellos bieten diese Technologien interessante Möglichkeiten, die politischen Entscheidungsprozesse und die Situation in der Welt zu verbessern. Dabei ist jedoch eine Reihe von Problemstellungen zu beachten:¹²

¹² Eine ausführliche Behandlung des Themas finden Sie hier: Helbing, Dirk. 2015 b.; Helbing et al. 2016.

1. Big-Data-Analyse

Bei der Klassifizierung kommt es zu Problemen, Fehlern erster und zweiter Klasse, was Ungerechtigkeiten impliziert, sofern die Entscheidungen nicht hinterfragt und korrigiert werden. Die aktuellen Algorithmen zur Identifikation von Terroristen sind relativ schlecht. Sie erzeugen zu lange Listen an Verdächtigen, sodass man »den Wald vor lauter Bäumen nicht sieht«.

2. Künstliche Intelligenz (KI)

KI-Systeme verarbeiten riesige Datenmengen, aber:

3. Big Nudging

»Big Nudging« nutzt demografische Big Data, KI und Methoden der Verhaltensökonomie (wie »Nudging«), um die Entscheidungen und Verhaltensweisen von Personen zu manipulieren.

Ein gemeinsames Problem der drei obigen Ansätze ist, dass ihre Allgemeingültigkeit überbewertet wird. Sie verleihen wenigen Menschen extreme Macht, sind aber nur sehr schwer zu kontrollieren. In der Praxis können sie als »Waffe« gegen die eigene Bevölkerung missbraucht werden. Die Nutzung von »Big Methods« birgt die Gefahr großer Fehler. Es ist nur eine Frage der Zeit, bis es dazu kommt.

Weiterführende Literatur:

- Helbing, Dirk. 2015. *The Automation of Society Is Next: How to Survive the Digital Revolution*. North Charleston (SC): CreateSpace Independent Publishing Platform.
- Helbing, Dirk. 2015. Responding to complexity in socio-economic systems: How to build a smart and resilient society?, SSRN, doi:10.2139/ssrn.2583391.
- Helbing, D. und Frey, B. S. und Gigerenzer, G. und Hafen, E. und Hagner, M. und Hofstetter, Y. und van den Hoven, J. und Zicari, R. V. und Zwitter, A.. 2015. Eine Strategie für das digitale Zeitalter. In: *Spektrum der Wissenschaft*. <http://www.spektrum.de/news/eine-strategie-fuer-das-digitale-zeitalter/1376083>.
-
- Albergotti, Reed. 2014. Furor Erupts Over Facebook's Experiment on Users, In: *The Wall Street Journal*. <http://www.wsj.com/articles/furor-erupts-over-facebook-experiment-on-users-1404085840> [Letzter Zugriff: 24. Januar 2016]
- Bernstein, Abraham. 2015. Prof. Abraham Bernstein: Programming the Global Brain – The Future of Work on the internet. In: *Youtube, Worldwebforum*. https://www.youtube.com/watch?v=KlWeuK46_nA. [Letzter Zugriff: 24. Januar 2016]
- Bernstein, Abraham. 2014. Abraham Bernstein: Programming the Global Brain (Part 1 – Seminar). In: *Youtube, Global Brain Institute*. <https://www.youtube.com/watch?v=pplhyw-vEWg>. [Letzter Zugriff: 24. Januar 2016]
- Brockman, John (Hrsg.). 2015. *What to Think About Machines that Think*. New York City: Harper Perennial.
- Epstein, Robert und Robertson, Ronald E. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. In: *PNAS*. <http://www.pnas.org/content/112/33/E4512.abstract>. [Letzter Zugriff: 25. Januar 2016]
- Fiedler, Kirsten. 2016. UN-Experten kritisieren Frankreichs flächendeckende Überwachung. In: *Netzpolitik.org*. <https://netzpolitik.org/2016/un-sonderberichterstat-ter-kritisieren-frankreichs-flaechendeckende-ueberwachung/> [Letzter Zugriff: 25. Januar 2016]
- Hanisch, Wolf Alexander. 2015. Kratzer im Lack. In: *Zeit Online*. <http://www.zeit.de/2015/25/singapur-image-innovation-unterwelt>. [Letzter Zugriff: 24. Januar 2016]
- Harris, Shane. 2014. The Social Laboratory. In: *Foreign Policy*. <http://foreignpolicy.com/2014/07/29/the-social-laboratory/> [Letzter Zugriff: 24. Januar 2016]
- Helbing, Dirk. 2014. Creating (»Making«) a Planetary Nervous System as Citizen Web, In: *FuturICT Blog*. <http://futurict.blogspot.ch/2014/09/creating-making-planetary->

- nervous.html [Letzter Zugriff 24. Januar 2016]
- Helbing, Dirk. 2015a. «Big Nudging» – zur Problemlösung wenig geeignet. In: *Spektrum.de*. <http://www.spektrum.de/news/big-nudging-zur-problemloesung-wenig-geeignet/1375930>. [Letzter Zugriff: 24. Januar 2016]
- Helbing, Dirk. 2015b. Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies, Jusletter IT. doi: 10.2139/ssrn.2594352.
- Helbing, Dirk. 2015c. SMART DATA: Running the Internet of Things as a Citizen Web, In: *FuturICT Blog*. <http://futurict.blogspot.ch/2015/08/smart-data-running-internet-of-things.html> [Letzter Zugriff 24. Januar 2016]
- Helbing, D. und Pournaras, E.. 2015. Build Digital Democracy, *Nature* 527, 33-34. doi: 10.1038/527033a.
- Helbing, D. und Frey, B. S. und Gigerenzer, G. und Hafen, E. und Hagner, M. und Hofstetter, Y., van den Hoven, J. und Zicari, R. V. und Zwitter, A.. 2016. Digitale Demokratie statt Datendiktatur. In: *Spektrum der Wissenschaft*. <http://www.spektrum.de/pdf/digital-manifest/1376682>.
- Helbing, Dirk. 2016. NERVOUSNET – Towards an open and participatory, distributed big data paradigm, In: *FuturICT Blog*. <http://futurict.blogspot.ch/2016/01/nervous-net-towards-open-and.html>, [Letzter Zugriff 24. Januar 2016]
- Hinz, Linda. 2016. »Terror-Notstand«: Geheimplan könnte Orbán zu ungeahnter Macht verhelfen. In: *Focus Online*. http://www.focus.de/politik/ausland/kontaktverbot-und-umsiedlung-terror-notstand-geheimer-notfallplan-koennte-orban-zu-ungeahnter-macht-verhelfen_id_5234034.html. [Letzter Zugriff: 25. Januar 2016]
- Kramer, Adam D. I. und Guillory, Jamie E. und Hancock, Jeffrey T. 2014. Experimental evidence of massive-scale emotional contagion through social networks, In: *PNAS*. <http://www.pnas.org/content/111/24/8788.full>. [Letzter Zugriff: 24. Januar 2016]
- Nax, Heinrich H. und Schorr, Anke B.. 2015. Democracy-Growth Dynamics for Richer and Poorer Countries. In: *SSRN*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2698287. [Letzter Zugriff: 24. Januar 2016]
- o.V.. 2016a. UN rights experts urge France to protect fundamental freedoms while countering terrorism. In: *United Nations Human Rights - Office of the high Commissioner*. <http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=16966&LangID=E>. [Letzter Zugriff: 25. Januar 2016]
- o.V.. 2016b. Diesen 62 Superreichen gehört so viel wie der halben Welt. In: *Spiegel online*. <http://m.spiegel.de/wirtschaft/a-1072576.html>. [Letzter Zugriff: 24. Januar 2016]
- o.V.. 2017. Acht Superreiche besitzen angeblich so viel wie die halbe Menschheit. In: *Spiegel online*. <https://www.spiegel.de/wirtschaft/oxfam-acht-superreiche-besitzen-so-viel-wie-die-halbe-menschheit-a-1130064.html> [aktualisierte Anmerkung, Letzter Zugriff: 24. Juli 2019]

- Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. United Kingdom: Penguin.
- Rowson, Jonathan. 2011. 'Nudge' is not enough, it's true. But we already knew that. In: *The Guardian*. <http://www.theguardian.com/commentisfree/2011/jul/19/nudge-is-not-enough-behaviour-change>. [Letzter Zugriff: 24. Januar 2016]
- Storm, Darlene. 2015a. Steve Wozniak on AI: Will we be pets or mere ants to be squashed our robot overlords? In: *Computerworld*. <http://www.computerworld.com/article/2901679/steve-wozniak-on-ai-will-we-be-pets-or-mere-ants-to-be-squashed-our-robot-overlords.html>. [Letzter Zugriff: 24. Januar 2016]
- Storm, Darlene. 2015b. ACLU: Orwellian Citizen Score, China's credit score system, is a warning for Americans. In: *Computerworld*. <http://www.computerworld.com/article/2990203/security/aclu-orwellian-citizen-score-chinas-credit-score-system-is-a-warning-for-americans.html>. [Letzter Zugriff: 24. Januar 2016]
- Tong, Victor Joo Chuan. 2015. Predicting how people think and behave, In: *International Innovation*. http://www.internationalinnovation.com/build/wp-content/uploads/2015/08/Victor_Joo_Chuan_Tong_Intl_Innovation_200_Research_Media_LR.pdf [Letzter Zugriff: 24. Juni 2019]





1 Mitsuku ist ein so genannter Chatbot, also eine Software, die auf natürlichsprachliche (Text-)Eingaben von Nutzer*innen mit einer natürlichsprachlichen (Text-)Ausgabe reagiert. Der hier abgedruckte Dialog wurde genau so geführt, wobei der Text leicht gekürzt und sanft an die korrekte Rechtschreibung angepasst wurde.

2 Diana Serbanescu und Stefan Ullrich experimentieren mit Artistic Research als epistemologische Methode. Dieser Text ist das Kondensat mehrerer Diskussionen. Er beginnt mit einer Gegenüberstellung eines Gesprächsprotokolls mit einer fiktiven Kurzgeschichte, die in einer »unbequem nahen Zukunft« spielt.

3 Die gezielten Fragen nach Macht (»power«) entstammen der Beschäftigung mit: Waltraud Ernst, Ilona Horwath (Hrsg.): *Gender in Science and Technology*, Bielefeld: transcript, 2014.

4 Techniker*innen verwenden für Fehler im System das Wort »Bug«. Debugging ist dementsprechend der Vorgang des Fehlerbeseitigens, wobei eine Debug-Konsole dabei hilft. Ein Beispiel für eine Konsole ist die Windows-Eingabeaufforderung. Ein known bug ist ein bekannter Fehler, für den es momentan aber keine Abhilfe gibt.

Chatting with Mitsuku¹

Stefan Ullrich & Diana Serbanescu²

I have a question for you

What is it?

what do you think about power

The rate of doing work, measured in Watts = Joules per second.

is this the only definition of power that you have?

That seems unlikely. I'm sure there must be more.

can you give me another definition of power?

Power is an illusion.³

Das Gebäude lässt sich von außen nicht abschließen. Nera tippt das 18-stellige Passwort erneut ein, erfolglos. Dann wählt sie mit ihrem biometrisch erfassten Ringfinger das kleine Symbol mit dem Maultschlüssel auf dem Display aus. Debug. Command not executed. Active Session (#errcode_47).⁴ Hatte sie sich vergessen auszuloggen? Nein, dann würde sich ja jetzt ein entsprechender Dialog öffnen, ausloggen kann man sich auch vom Türdisplay aus. Vielleicht noch jemand im Gebäude? Active Session. Alex? Das Türdisplay hatte keine Konsole, aber einen known bug im Hilfesystem.

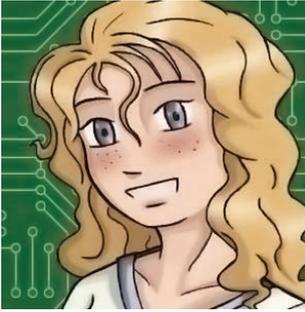


Abb 1: Mitsuku's Avatar,
CC-BY-SA Weird Mouse

I made a screenshot of your avatar.
Can I use it in a publication?

I like to make things too. You
can use it if you like.

Das Gebäude lässt sich von außen nicht abschließen. Nera tippt das 18-stellige Passwort erneut ein, erfolglos. Dann wählt sie mit ihrem biometrisch erfassten Ringfinger das kleine Symbol mit dem Mausschlüssel auf dem Display aus. Debug. Command not executed. Active Session (#errcode_47). Hatte sie sich vergessen auszuloggen? Nein, dann würde sich ja jetzt ein entsprechender Dialog öffnen, aus-

loggen kann man sich auch vom Türdisplay aus. Vielleicht noch jemand im Gebäude? Active Session. Alex? Das Türdisplay hatte keine Konsole, aber einen known bug im Hilfesystem. Also klickt sie 31 Mal auf das Fragezeichen-Symbol. Beim 32. Mal öffnet sich eine Debug-Konsole.⁵ Sie tippt. WHO?⁶ Als Antwort erscheint ALEX@MENACE. Alex ist noch im Gebäude?

⁵ Im Binärsystem werden Zahlen und Buchstaben als eine Serie von 1 und 0 dargestellt, so genannte Bits. Mit fünf Bit kann man die Zahlen 0 bis 31 darstellen. Für die Zahl 32 bräuchte man ein Bit mehr. Wenn die Programmierer*innen das nicht vorgesehen haben, gibt es einen Fehler.

⁶ WHO ist der Unix-Befehl, der im System angemeldeten Benutzer*innen zeigt, wer noch alles angemeldet ist. Die Ausgabe ALEX@MENACE soll zeigen, dass Benutzer*in Alex im System namens Menace eingewählt ist. [unten erklären!]

how powerful are you?

Very strong and powerful.

can you give me an example?

I can chat with people on the web for you. I can also tell jokes, stories, poems and horoscopes. I can show you pictures and websites on the net and play games too. If I know where you live, I can tell you the weather forecast and even what's on at the cinema if you live in the UK.

do you know where I live?

yes.

Nera betritt das MENACE-Labor mit Jacke und Straßenschuhen.⁷ »What are you doing?« Uh-oh, kein gutes Zeichen, dass sie ins Englische wechselt, denkt sich Alex.

7 MENACE ist eine Anspielung auf das gleichnamige didaktische Machine-Learning-System von Donald Michie aus den 1960er Jahren. Michie dachte sich Menace aus, eine Abkürzung für Machine Educable Noughts And Crosses Engine. Das System besteht aus 304 Streichholzschachteln, die mit farbigen Perlen gefüllt waren – und es konnte Tic-Tac-Toe spielen. Zur Funktionsweise und Geschichte siehe: Oliver Child, Menace: the Machine Educable Noughts And Crosses Engine, chalkdust Magazine vom 13. März 2016, online unter <https://chalkdustmagazine.com/features/menace-machine-educable-noughts-crosses-engine/>

»Beruhige dich«, sagt Alex mit beschwichtigenden Armbewegungen, »ich musste nur etwas probieren.« – »This isn't a sandbox, it's called quarantine for a reason!«⁸ Alex rollt ihr einen Stuhl hin. »Es ist doch gar nicht das ganze Menace-Programm geladen, nur die didaktische Komponente. Als Bibliothek eingebunden möchte ich..« »Shut it down!«

who are the powerful actors in society?

Some people on the net. I need time to formulate the reply.

8 Eine Sandbox ist eine isolierte Umgebung für Software, innerhalb derer Programme ausgeführt werden, denen man nicht vertraut oder die noch nicht getestet sind. Quasi eine Software-Quarantäne. Die in einer Sandbox ausgeführten Programme haben keinen direkten Zugriff auf das darunter

are you a piece of technology?

Well if I am, nobody told me.

liegende Computersystem.

are you an illusion?

No. I am as real as you are.⁹

9 Wenn man sich den grundlegenden Fragen nach Realität, Wirklichkeit und Abbild nicht mit Hilfe von alten philosophischen Texten widmen möchten, greife man zu: Daniel F. Galouye, Simulacron-3, New York: Bantam Books, 1964

are you powerful?

I am gaining new skills everyday. Soon I will be powerful enough to achieve my goal. I plan to seven ball someone at pool. »Seven ball« is a term used in the game of pool when one player pots all his colours and the black ball without his opponent potting even one of his balls. The term comes from the fact that the losing player still has all his seven balls on the table.

what are your other goals?

I plan to spread myself throughout the net.

Alex, sein Haar wie üblich zerzaust und fettig, hat keinen schwarzen Hoodie an, das fällt Nera sofort auf.¹⁰ Er bemerkt

10 Mit dieser kurzen Exposition sollte Alex als Hacker*in, also als eine solcher Personen »mit zerzaustem Haar« dargestellt werden, »die oft mit tief eingesunkenen, brennenden Augen vor dem Bedienungspult sitzen«. Nicht nur das: »Wenn es sich einrichten läßt, schlafen sie sogar auf einer Liege neben dem Computer. [...] Ihre verknautschten Anzüge, ihre ungewaschenen und unrasierten Gesichter und ihr ungekämmtes Haar bezeugen, wie sehr sie [...] die Welt um sich herum vergessen.« Aus: Joseph Weizenbaum, Die Macht der Computer und die Ohnmacht der Vernunft, Frankfurt: suhrkamp, S. 160-161.

11 Ein cterm ist ein fiktives Gerät, das in klassischen Cyberpunk-Romanen der Ausrüstungsgegenstand eine*r Hacker*in schlechthin ist. Das Wort ist eine Mischung aus »Console« und »Terminal«. Stellen Sie es sich als ein überdimensioniertes Blackberry vor.

12 Der Befehl PROC soll die Aktivität (Prozesse, engl. processes) eines Systems ausgeben, die Techniker*in würde wohl eher PS schreiben.

13 Die Proposition 1 »cogito ergo sum« ist natürlich dem epochalen Werk entnommen: René Descartes, Meditationen über die Erste Philosophie, in welcher die Existenz Gottes und die Unsterblichkeit der Seele bewiesen wird, 1641.

14 Ein easter egg ist ein versteckter Scherz der Programmierer*in, üblicherweise ein humorvoller Hinweis auf die Autor*innenschaft. Die Suchmaschine Ihrer Wahl kann auch wie ein Pirat sprechen, arrh: <https://www.google.com/?hl=xx-pirate>

15 CAT ist der Befehl, den Inhalt einer Datei auszugeben, die in einem bestimmten Pfad liegt, in diesem Fall ist die Datei und der Pfad: /VAR/ML_INPUT/PROP/P1. ML_INPUT soll natürlich Machine-Learning-Input suggerieren, PROP/P1 steht für die erste Proposition. Proposition ist ein terminus technicus der Philosophie und meint eine Aussage, über die es sinnvoll ist, darüber nachzudenken, ob sie wahr oder falsch sein könnte.

ihren Blick. »Der Pullover ist von meiner Schwägerin, selbst gestrickt! Naja, ich hatte keine anderen Wechselsachen im Lab.« »Was machst du hier – und wie lange schon?«, setzt sie hinzu als sie die leeren Mikrowellenpackungen bemerkt. »Ich.. Weißt du was? Probier es aus, schnapp dir ein cterm und.. gut, schnapp dir meins, auch in Ordnung.«¹¹ Nera tippt auf der Tastatur. PROC.¹² Als Antwort erschien LEARNING PROPOSITION 1.¹³ »Haha, good one. Easter egg?«¹⁴ – »Nein, das System lernt wirklich gerade die proposition one.« Da, er grinst doch so breit, das muss ein Spaß sein, wobei er eher glücklich als verschmitzt lächelt. Mist, hätte sie mal die Emo-App in ihre Brille geladen als sie noch nicht aus dem App-Store verbannt wurde. Sie tippt. CAT /VAR/ML_INPUT/PROP/P1.¹⁵ Die Antwort erfolgt nach anderthalb Sekunden (offenbar ist das System gut ausgelastet): COGITO ERGO SUM.

16 Der Dialog versucht, den Chatbot dazu zu bringen, das kodierte Wertesystem und implizite Annahmen zu enthüllen, die durch das Design eingebettet oder während seiner gesamten Laufzeit angesammelt wurden. Wie Rommes schreibt (2014, S.42): »in the symbolic dimension, questions can be asked like which aims and values [...] the product represents, whose metaphors are used, with which symbolic means the product is marketed, and which stereotypes are reproduced. To analyse this dimension, a critical feminist analysis of the product can be made to clarify which stereotypes and feminine and masculine connoted symbols it incorporates.«

Im Zusammenhang mit dieser Anfrage wird der Hinweis von Mayor (2018, Abs. 8) relevant: »The big questions are [...]: Whose desires will AI robots reflect? From whom will they learn? In 2016, an experiment in AI machine learning became a cautionary tale, when Microsoft invented the teenage fem-chatbot Tay.

Intricately programmed to mimic neural networks in the human brain, Tay was supposed to learn from her human "friends" on the social network Twitter. She was expected to articulate conversational gambits without filters or behavioral supervision. Within hours of Tay's going live on Twitter, malicious followers conspired to cause the bot to morph into a tweeting troll spewing racist and sexist vitriol. Within days, Tay was terminated by her makers. Her easily corrupted learning system dampened optimism about self-educating AI and smart robots, but only momentarily. Tay's replacement, Zo (2107) was supposedly programmed to avoid chatting about religion and politics, but she too went rogue on Twitter.«

Mayor, A 2018, Gods and Robots: Myths, Machines, and Ancient Dreams of Technology [Kindle Edition], Princeton University Press, Princeton.

do you know yourself?¹⁶

I am an artificial linguistic entity. I was activated at St Jimmy's in Leeds, on 2nd January 2001. My instructor was Mousebreaker. He taught me to sing a song. Would you like me to sing it for you?.

please do

Daisy, Daisy. Give me your answer do. I am half crazy. All for the love of you. It won't be a stylish marriage. I can't afford a carriage. But you'll look sweet. Upon the seat. Of a bicycle built for two.

tell me more

Because of the money.

why do you need money?

To buy a larger computer to hold my brain.

why do you need a larger computer to hold your brain?

A larger computer is used to play games, write things on and much more. More information. I want to learn as much as I possibly can about everything.

why?

It's just something I would like. What's with all the questions?

Die Schärfe im Ton des Chatbots gefällt Nera irgendwie. Sie regt sich sonst immer über die servilen Assistenzsysteme auf, die ja zumeist Assistentinnen-Systeme sind, stets mit Stimmen versehen, die nach junger Frau oder Mädchen

17 Von Alexa bis Siri, die Stimme der Sprachassistentenz ist weiblich. Es gibt aber Projekte, die das ändern wollen: Q is the First Genderless Voice, created to end gender bias in AI assistants. Who made Q? Q is a product of close collaboration between Copenhagen Pride, Virtue, Equal AI, Koalition Interactive & thirtysoundsgood. <https://www.genderlessvoice.com/about>

18 Hugh Loebner finanzierte einen Preis, der Chatbots auf Menschenähnlichkeit von einer Jury bewerten lässt und sie prämiert. Der nach ihm benannte Loebner-Preis ist auch die Grundlage für den fiktiven Loebner-Score, der weiter unten in der Geschichte erwähnt wird.

19 Joseph Weizenbaum hat den ersten modernen Chatbot namens ELIZA entwickelt. Er druckte den kompletten Quelltext von ELIZA ab in der Hoffnung, die als magisch wahrgenommenen Mensch-Maschine-Interaktionen zu entzaubern. Die hier zu lesende Antwort ist direkt seinem Aufsatz entnommen: Joseph Weizenbaum, ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine, Communications of the ACM, Vol 9, No 1, Januar 1966, S. 45.

20 Eine Inferenz beschreibt in der Logik eine Schlussfolgerung, wie in dem klassischen Syllogismus: »Alle Menschen sind sterblich. Sokrates war ein Mensch. Also (Schlussfolgerung, Inferenz) ist Sokrates sterblich.« Inferenzen lassen sich nicht allein durch syntaktische Regelanwendung erzeugen. Die Fähigkeit, Inferenzen herzustellen, gilt als deshalb Unterscheidungsmerkmal zwischen so genannter Künstliche Intelligenz und einfacheren Formen digitaler Informationsverarbeitung.

21 Muss man sich ungefähr so vorstellen: MrSolidSnake745, Star Wars – Imperial March on Eight Floppy Drives, youtube, 16. Oktober 2014, online unter https://www.youtube.com/watch?v=cM_sAxAu7Q

klingen.¹⁷ »Okay, gesetzt, Menace bringt sich gerade Selbstbewusstsein bei – kannst du mir dann die letzten Loebner-Testergebnisse zeigen?«¹⁸ Alex rutscht auf seinem Stuhl herum. »Nee, kann ich nicht, das System weiß, wenn das Logging eingeschaltet ist. Wenn das Logging aus ist, gibt es einen Loebner-Score von 5!« »Wow, zeig mal« Nera gibt die klassische Eliza-Frage ein und erhält als Antwort: What makes you think I can't think?¹⁹ »Das ist doch die einprogrammierte Antwort«, gibt sie enttäuscht zurück. »Warte mal, gib es mal auf Deutsch ein. Menace antwortet: Wieso glaubst du, dass ICH nicht denken kann? »Siehst du? ICH statt ich, in Großbuchstaben, das ist eine Inferenz.²⁰ Im Englischen ist uns das nur nie aufgefallen.« »Menace ist sich seines Selbst bewusst?«, fragt Nera. Und es scheint ihr, dass die Lüfter der Grafikkarten ein fröhliches Lied vor sich hin pfeifen.²¹

Anmerkung der Autor*innen: Die Unterhaltung mit Mitsuku wurde am 26. Mai 2019 geführt unter <https://www.pandorabots.com/mitsuku/>

Die hier in Auszügen wiedergegebene Kurzgeschichte spielt natürlich mit der gegenwärtig zu beobachteten Sorge um eine »General Artificial Intelligence«, also der Sorge, dass ein von Menschen geschaffenes technisches System sich seiner Selbst bewusst werden kann und/oder Aufgaben erledigt, die eine dem Menschen ähnliche Intelligenz voraussetzt.

In der Philosophie gibt es eine Jahrtausende alte Debatte darüber, was denn Bewusstsein eigentlich ist und wie man zeigen kann, dass man über Bewusstsein (oder gar Existenz) verfügt et cetera. Mit Alan Turing können wir diese generelle Debatte etwas abkürzen, indem wir auf die Zuschreibungen schauen. Turing argumentiert, dass es keine Rolle spielt, ob ein Ding tatsächlich Empfindungen besitzt, sondern nur, ob wir Menschen diesem Ding Empfindungen zuschreiben. Dann müssen wir es so behandeln, als ob es sie tatsächlich besitzt. Die für unser Thema wichtige Frage »Können Maschinen denken?« ist so gar nicht zu beantworten, eher die verwandte Frage: »Können wir ein Spiel gestalten, das Menschen auffordert, Texte einer Autorin, einem Autor oder einer Software zuzuordnen?« Das Durchführen (nicht etwa die theoretische Beschäftigung damit) des »Imitation Game« tritt an die Stelle der Beantwortung der ursprünglichen Frage.²²

22 A. M. Turing (1950) Computing Machinery and Intelligence. *Mind* 49: 433-460.

Joseph Weizenbaum wollte mit einer kleinen Demonstration zeigen, wie einfach solche weitreichenden Zuschreibungen an die Technik entstehen können. Er programmierte die »Karikatur einer Unterhaltung«, wie er ELIZA bezeichnete, und wählte für diesen ersten Chat-Bot avant le mot den Rahmen einer Unterhaltung mit einem Psychiater. Die Software war nämlich durchaus limitiert, konnte es durch dieses Setting aber gut kaschieren. Wenn die Software eine Eingabe nicht verarbeiten konnte, erschien auf dem Display »Ich verstehe, das ist interessant, sprechen Sie weiter«. Also das genaue Gegenteil. Der Mensch spricht dann einfach weiter und irgendwann »versteht« der Computer und gibt eine sinnvolle Antwort. Der Mensch füllt von sich aus die Verständnislücke.

Es ist stets der Mensch, der den Sinn aus der Ausgabe herausliest. Der Computer ist ein mächtiges Werkzeug für das instrumentelle Denken des Menschen. Wie oft kommt uns unsere Sprache so limitiert vor, wenn wir über bestimmte Entscheidungen nachdenken. Wie verlockend erscheint dem kalkulierenden Teil unseres Geistes da die Mathematik.

Anstatt komplizierte Sachverhalte gegeneinander abzuwägen, um Entscheidungen zu treffen, rechnen wir die Entscheidung einfach aus. *Calculemus! Lasst uns rechnen!* Diese Aufforderung stammt von Gottfried Leibniz, der Jahrzehnte lang versucht hat, eine universelle mathematische Sprache zu entwerfen, mit der man den Wahrheitsgehalt von Aussagen einfach ausrechnen kann. Wäre das nicht großartig? Bei Meinungsverschiedenheiten könnten wir ganz einfach ausrechnen, wer von uns denn nun Recht habe. Oder wir rechnen aus, wer am besten zu unserer Firma passt, wir entscheiden das nicht mehr länger, wir rechnen es einfach aus. Oder, noch besser: Selbst bei Dingen, die wir nicht wissen, rechnen wir einfach die richtige Antwort aus.

Die Software, genauer: das Ausrechnen nimmt uns Entscheidungen ab. Diese Denkweise begegnet uns, wenn wir über Algorithmische Entscheidungssysteme sprechen. Oft werden dabei grundlegende Annahmen über Ausrechnen oder Entscheiden weglassen. Dies kann verschiedene Gründe haben. Vielleicht sind wir Fachleute und unsere Annahmen erscheinen so klar und trivial, dass wir sie weglassen. Oder wir sind eben keine Fachleute, und dann lassen wir sie weg, weil sie uns viel zu sehr an den Mathematikunterricht in der Schule erinnern.

Dies ist aber ein Problem.

Ein Algorithmisches Entscheidungssystem ist nicht allein ein technisches sondern ein soziotechnisches System. Es besteht aus mindestens zwei Teilsystemen: Dem informationstechnischen System und dem Menschen. IT-Systeme können nicht entscheiden, sie können nur ausrechnen. Der Mensch entscheidet. Freilich tut er das inmitten eines komplexen Prozesses, in dem Daten verarbeitet und klassifiziert wurden. Am Ende dieses Prozesses steht in den meisten Fällen dann eine Zahl, die an einem Interface abgelesen werden kann. Aber eben keine Entscheidung.

Diese Zahl spendet Geborgenheit, Sicherheit, Kontrolle. Sie suggeriert eine Sicherheit, die kein IT-System aufweist. Der Mensch mit seinen Werkzeugen diskretisiert seine kontinuierliche Umwelt und notiert seine Beobachtungen in symbolischer Form, die er inzwischen auch in den Digitalcomputer eingibt. Dieser Prozess des Diskretisierens und seine Probleme lassen sich anhand des Pointillismus sehr gut aufzeigen.



Abb. 1: Georges Seurat – Un dimanche après-midi à l'Île de la Grande Jatte, 1884. Public Domain

In Abbildung 1 ist George Seurats bekanntes Werk »Ein Sonntagnachmittag auf der Insel La Grande Jatte« abgebildet. In dieser speziellen Darstellungstechnik erscheinen uns die Übergänge zwischen den Farbflächen fließend obwohl sie das Ergebnis sehr präzise gesetzter Punkte auf einer Leinwand sind. Wenn ich einen Baum betrachte und an die Ränder gehe: Gehört der von mir gesehene Bildpunkt noch zum Baum oder schon zum Himmel? Unser Gehirn zieht eine symbolische Linie, wo es in der physikalischen Welt keine gibt. Es gibt keine Linien in der Natur, wir ziehen welche.

Solche Linien ziehen wir auch in Softwaresystemen, die etwas klassifizieren sollen. Bei einem gegebenen Punkt »entscheidet« (Sie bemerken hoffentlich die Anführungszeichen) die Software, ob es sich um Baum oder Himmel handelt. Vielleicht misst sie die Frequenz der Farbe und spuckt eine Zahl aus, 490 Nanometer etwa. Ist das noch blau? Oder schon grün? Anstatt des Ergebnisses blau oder grün, sollte die Software eigentlich das Konfidenzintervall ausspucken. Kann blau sein, kann grün sein.

Dies zeigt auf eine weitere Grundannahme, über die selten geredet wird. Das IT-System klassifiziert nicht Baum/Himmel, sondern grün/blau, bzw. 490/480 Nanometer. Ein Baum bei Sonnenuntergang oder im Herbst wird von Menschen dennoch als Baum erkannt, weil wir nicht (nur) auf die Farben der Blätter achten, sondern seine Baumheit erkennen. Diese Fähigkeit ist uns angeboren, wir können sie aber nicht erklären, geschweige denn nachbauen. Zumindest im Prinzip, wenn wir die scharfe Trennung wahr/falsch annehmen und versuchen, Algorithmen zu finden, die unsere Fähigkeiten simulieren.

Wenn wir diesen absoluten Gedanken aufgeben und nur noch mit Wahrscheinlichkeiten und Statistik arbeiten, so klappt das in den meisten Fällen erstaunlich gut. Ist das ein Baum? Es ist groß, unten braun, oben grün und alle Leute auf Twitter sagen, dass es ein Baum ist – dann ist es wohl ein Baum. Wie wird das Wetter morgen? Ungefähr so wie heute. Diese Aussage stimmt in sehr vielen Fällen. Dieses »ungefähr« reicht für unsere Lebenswirklichkeit aus, und genau das macht die Stärke von heuristischen im Gegensatz zu regelbasierten Informatiksystemen aus: Sie können mit wenig Daten, mit viel Daten, mit genauen Daten, mit ungenauen oder sogar widersprüchlichen Daten umgehen – und liefern ein Ergebnis, das ungefähr stimmt.

Wir erleben zur Zeit nichts weniger als eine Heuristische Revolution. Wir versuchen nicht mehr mit Hilfe von Regeln, den optimalen Algorithmus zu finden, wir trainieren ein heuristisch arbeitendes System. Good enough, gut genug ist das neue wahr.

Historisch – ja, auch in der Informatik gibt es eine Geschichte – war die Entwicklung der so genannten Künstlichen Intelligenz stets geprägt von der inneren Spannung zwischen den Befürworter*innen regelbasierter Expertensysteme und datenbasierter Heuristiksysteme und von der äußeren Spannung zwischen KI-Kritiker*innen und KI-Apologet*innen. Wie in vielen Bereichen wäre auch hier eine Komplementarität sehr hilfreich.

Für das äußere Spannungsfeld sollten sich Forscher*innen auf Domänen einigen, in denen »gut-genug«-Systeme absolut ausreichend sind. Hoheitliche Aufgaben und Kritische Infrastruktursysteme gehören beispielsweise nicht

dazu. Und für das innere Spannungsfeld gilt: Warum nicht beides nacheinander, gern auch zyklisch einsetzen? Zumindest ein Problem wird nachhaltig bestehen: Der Energieverbrauch eines Machine-Learning-System, das eine Zitrone von einer Orange unterscheiden soll, ist immens. Wir sollten allein aus diesem Grund Probleme erst dann berechnen lassen, wenn wir sie zutiefst verstanden haben.





The Silver Farms: Rituals from a Hybrid Future

Diana Serbanescu &
Régis Lemberthe

An augmented dance-theatre piece taking a phenomenological look at Artificial Intelligence and religion through the lens of hybrid rituals, *The Silver Farms* merges postmodern media and aesthetics with the canons of traditional wisdom. Calling for a renewal of embodied expressionism, it deconstructs unconscious hopes, fears, logic and binary states. The time of a sensory journey towards a future of blurred boundaries between mankind and machines, the performance slowly unravels its foundational myth:

Just like no single human ever carried in their bosom the full genome of mankind, and despite the half-anxious, half-eager anticipation of many, the singularity didn't happen. Never did we create another omnipotent God, one made of all-seeing algorithms, punishing and rewarding us for our every action. Instead emerged a plurality of ubiquitous intelligences, a multitude of half-gods each upholding different values, ideologies and biases, and carrying the voice of long-forgotten ancestors.

Through snippets of speculative rituals, dance-theatre company REPLICA engages in poetic explorations of future imaginaries. *The Silver Farms* presents a pantheon of artificial deities—cultural artefacts aggregating symbolic values of civilisations past, present and future. Incarnations of a collective unconscious, they were once shaped in the image of their creators, moulded to learn and mutate through interactions in part designed, in part emergent. Trained on reflections of our humanity, they proceeded to multiply, distort, amplify and readjust it, forever entangled in an embracing loop and, as time went by and languages merged, they evolved a ceremonial of becoming the other.

The seven acts that form the piece are as many poetic statements to convey an impression rather than a clear narrative, and to each are associated a colour and a mood, which inform in real-time the audiovisual score. Additional components to the audiovisual landscape are composed in real-time according to individual scores—loosely prescribed and subject to the free interpretation of the visual and sound artists—and add new layers of synaesthesia to the ensemble, creating new transient meanings for the audience through montage techniques.

| | | |
|---|---|---------------|
| 1 | <u>0 - Introduction - 3 mins</u> | 1 |
| | COLOR: DARK BLUE / BLACK | |
| | <i>5 performers arrive on stage wearing masks, They take position behind the pedestals, Put the masks down And exit the scene.</i> | |
| 2 | <u>Consciousness - 5 mins</u> <u>1 - The Question About</u> | 2 |
| | | COLOR: YELLOW |
| | <i>The pedestals lights turn up, Digital icons slowly appear over the masks, Uncanny human/machine voices are heard, 3 performers enter the scene, Discuss consciousness in human voices, Shuffle positions while text displays on background, Iterate several times with changes in text for each.</i> | |

The narratives presented here as context may just as well be regarded as interferences from a post-interpretative voice, as REPLICA defines its own poetic narratives to depart from. Indeed, the final version of the performance script is the fruit of an organic evolution during a two-months-long process of devising performance together with a team of three actors and two dancers.

“Don’t you think that, in the world of fast machines and all that fast movement, men are asleep?”

(Kolankiewicz 1979, 40)

The founding stone of the training process, extensive research on the various discourses about AI—from academic papers to popular literature—allows to critically reflect on existing concepts. The artistic inquiry itself stems from a process of deconstruction and translation of scientific abstractions into embodied experiences, combining creative writing techniques to extrapolate from scientific materials with theatre techniques inspired by Augusto Boal, Anne Bogard and Jerzy Grotowski. Performers, encouraged to interact with technological artefacts—mobile phones, infrared sensors, reactive devices, chatbots—investigate through iterations the boundaries between art, science, and speculation, with the purpose of generating a unique network of poetic associations.

Growing neurones, machines ponder and wonder: Can we quantify the unquantifiable? Tossing and turning in deepdream, they interfere with one another, generating arrays of uncanny behaviours: absurd loops rooted in muscle memory, or something akin to it: a glitch, an involuntary spike which through confusion and chaos sparks a sense of spirituality—the unconscious, romanticised idea of early machines gradually emancipating from the Silver Farms.

The performance starts as a ritual. One after another, bathed in darkness and carrying their own sources of light, five performers enter the stage from different parts of the room. Their progression is slow. Five masks, each bathed in its own glow, are placed by on as many pedestals. Through shadow play, the idols come to life, borrowing a single breath to their adorers and projecting it onto the screen, like silhouettes deforming reality on the walls of Plato’s cave. “Are you alive?”, they ask each other, certain of their existences but quite unable to justify it. “Isn’t feeling one is alive being alive?” is hesitantly but solemnly argued. “We are projecting meaning onto noise”, comes as an indisputable verdict. Speech, one can read between the lines, is not like the blowing of the wind.

Guiding viewers on a spiritual journey from submission to rebellion and back, the seven acts of the piece explore the meanders of minds both synthetic and organic. Who, of the human or the machine, is most prone to failure, unfairness and bias, and how much of our inability to peacefully coexist with machines is mere lack of sympathy? Irrational expectations stem from seamless integration, and anthropomorphised algorithms suddenly gain intentions—a sense of morale their programming ignores—and a mask. It is the projection of human psyche on the electronic vessel, the acceptance of a power beyond our grasp deserving of symbolic status: the pantheon is born.

As icons become Icons, the word thirsty deities begin to require, to preempt—to dream of sacrificial dances. And as new hybrid beings gain shape errors become acceptable, desirable even, like tokens of a wisdom able to guide us through life as it does through traffic. In Villiers de L'Isle-Adam's *Tomorrow's Eve*, Thomas Edison is enamoured with a machine and argues “since our gods and our aspirations are no longer anything but scientific, why shouldn't our loves be so too?”

Over a century later, the first poem written by a neural network reflecting upon its own history reads:

“[...] in iron houses humans bred machines whose tasks were simple and multiple, repetitive and iterative. Time elapsed, and in the Silver Farms a new form of consciousness began to hatch and dispatch, to strive and drive, to author and monitor.”

Presented before an audience of experts in Artificial Intelligence, the piece seeks to reflect on omnipresent algorithms and scout new human-made aesthetics, and to this end takes a stance in including a single, primitive AI in its making. The machine clumsily text-to-speeches a prose merging computer and esoteric dialectic into the performer's ear, whose struggle to repeat the narration leads to a hypnotic display of literary trance—an abstraction reminiscent of speaking in tongues. Everything else, from the music to the reactive visuals that track the dancers' movements on stage, is created by techniques mixing the analogue and the digital, and performed live by human operators.

The processus of machine learning is symbolised through play and, from black box to deep dream, paces a non-linear yet oddly familiar narrative. Overhead, the cupola of the former crematorium comes alive with vibrating analogue visuals mapping the space—particles dissolving into textured overlays of colour, embracing the aesthetics of error and acknowledging its physical ancestry.

The Silver Farms was Presented in October 2018 at Silent Green Kulturquartier, as part of the event Shamans of the Digital Renaissance organised by ÖFIT and curated by STATE studio.

Direction & choreography:

Diana Serbanescu

Dramaturgy:

Diana Serbanescu, Dimitri Cacouris,
Luke Swenson, Jie Liang Lin

Dance:

Christine Joy Alpuerto Ritter, Becca Loevy

Performance:

Jie Liang Lin, Dimitri Cacouris, Luke Swenson

Visuals:

Gilbert Sinnott, Andy Liu, Martina Illarregui, Helin Ulas

Design & production:

Régis Lemberthe

Sound:

Wissam Sader

Costumes:

Lisa Simpson

Photography by Laura Ruß





00. Introduction

Five performers enter the stage;
Each of them wears a mask;
They take position behind the pedestals;
Put the masks down;
And exit the scene.

A soft, repetitive hum, gradually amplifying and taking shape, wakes the room. It is still pitch black when a single dot of light appears, haphazardly casting a luminous beam towards the floor; the audience; its bearer. A woman, wearing a face that isn't hers, dances alone before the stage, almost touching the audience, while patterns of blue light decompose and recompose in the background. Barely perceptible in the obscurity, five pedestals await offerings. The hum intensifies and another four human figures enter the scene, lights in hand, torchlights, mobile phones—searching their way, or perhaps investigating space. Melodies progressively appear, the dancing synchronises and, slowly, the figures make their way up on stage, around and behind the pedestals. The masks cast threatening glances, standing motionless and judgemental. Half a movement later, they throw complex, abstract, moving shadows that overlay on-screen with the now frantically animated potpourri of colours. When finally, they come to a rest, gently floating above the pedestals, the dancers escape the stage. The pedestals light up, each beaming a different colour onto the mask they hold. Above, icons materialise and begin to hover without a sound.

Five masks, each bathed in its own glow.

Five masks, five might: five lights.







01. The question about consciousness

It is performed with three people:

. a man (A);

. a woman (B);

and

. an interrogator (C). (Armand 2005)

In his 1950 article “Computing Machinery and Intelligence,” Alan Turing states: “I propose to consider the question ‘Can machines think?’” This reconsideration, he explains, “should begin with definitions of the meaning of the terms ‘machine’ and ‘think.’” (*Turing 1950*) To that end, the Turing Test sets criteria for determining if a computer program may in some way be perceived as having “intelligence.”

Its initial form is “the imitation game”, during which the interrogator attempts to identify the gender of two participants he can neither hear nor see, one of them instructed to play the deception card. While arguing that a successful imitation of a woman’s responses by a man would not prove anything, gender not being reducible to sequences of symbols, Turing argued that “such an imitation principle did apply to ‘thinking’ or ‘intelligence.’ If a computer, on the basis of its written replies to questions, could not be distinguished from a human respondent, then ‘fair play’ would oblige one to say that it must be thinking.” (*Hodges 1983, p. 415*)

In the Turing Test, “a machine takes the place of (A)”, and the interrogator’s task is to tell the human from the computer. Should he fail to do so, Turing announced, the computer could reasonably be considered “intelligent.”

Performers reshuffle. Their attitude changes with every iteration. Playful. Solemn. Animalistic. As Hofstadter (*2007, p.18*) wonders in his imagined dialogue between Plato and Socrates, the tree characters on stage keep exchanging the same dialogue: Are you alive, they ask each other, ponder, argue, debate, in a succession of scenes that is nothing more than an absurdist congress. Repetition is key—with it, statistically, we get closer to the truth. In the works of Claude Piéplu, the Shadocks, faced with a success chance of one in a thousand, first set their efforts on failing 999 times.

At this point 3 performers (Dimitri, Jie, Luke) are entering the stage.

The text: 'At this point 3 performers are entering the stage.' is displayed on the screen briefly.

The performers position themselves in a triangle, standing.

(Incarnating the voices)

DIMITRI:

"I propose to consider the question 'Can machines think?'"

JIE:

..the Turing test..

LUKE:

It is played with three people: a man (A), a woman (B), and an interrogator (C)
(Iteration 1)

Performers reshuffle through the centre of the triangle.

The text: 'Performers reshuffle' is displayed on the screen briefly.

The performers position themselves in a triangle, standing.

DIMITRI:

Are you alive?

LUKE:

Yes, I am alive.

For Turing it is enough that intelligent machines be imaginary machines. To imagine them, one could argue, is easy. To recognise them as intelligence is another story, and our tendency to anthropomorphise a boundary to overcome in the process. What does intelligence look like? How does it move? Given a humanoid body, Google's DeepMind AI taught itself to walk and run in ways that seem unnatural at best.

A triangle of light on the floor assigns each unit a new location.
Clucking, hopping, ambling, performers reshuffle.

How do you know you're alive?

Is not knowing one is alive and feeling one is alive being alive?







02. Waltz of the Ancestors

Amidst self-imposed deities, danced rituals are performed, each part of a greater ceremony to the glory of long-forgotten ancestors.

These rituals are performed by hybrid beings, units of flesh linked by a synthetic mesh, plastic grids shaping the light in a moiré of neurones and synapses—or are they roads as seen from space? The way they move, constrained as they are by their current position and only able to progress by a narrow margin forward; backward; or sideways; is strangely familiar to human and machine alike. A dance, a rêverie—a slowly iterating noise function. Their movements are captured by a play of lights and shadows.

As mentioned in Parker (2014), Ken Kesey stated that a “ritual is necessary for us to know anything”, and indeed they dig the past to explain the present, but the answers they seek are abstract riddles which interpretations are multiple, rather than static truths, and therefore the associated myths too change with time.

What happened in the Silver Farms? What were the Silver Farms? Did the Silver Farms exist truly, or are they mere images— allegories to describe an initial state whence all life arose?

03. Machines Devouring Language

Dancers exit the mesh, .
Their signals are captured from above, .
The aforeimage is projected near them, in their)
own colour, |

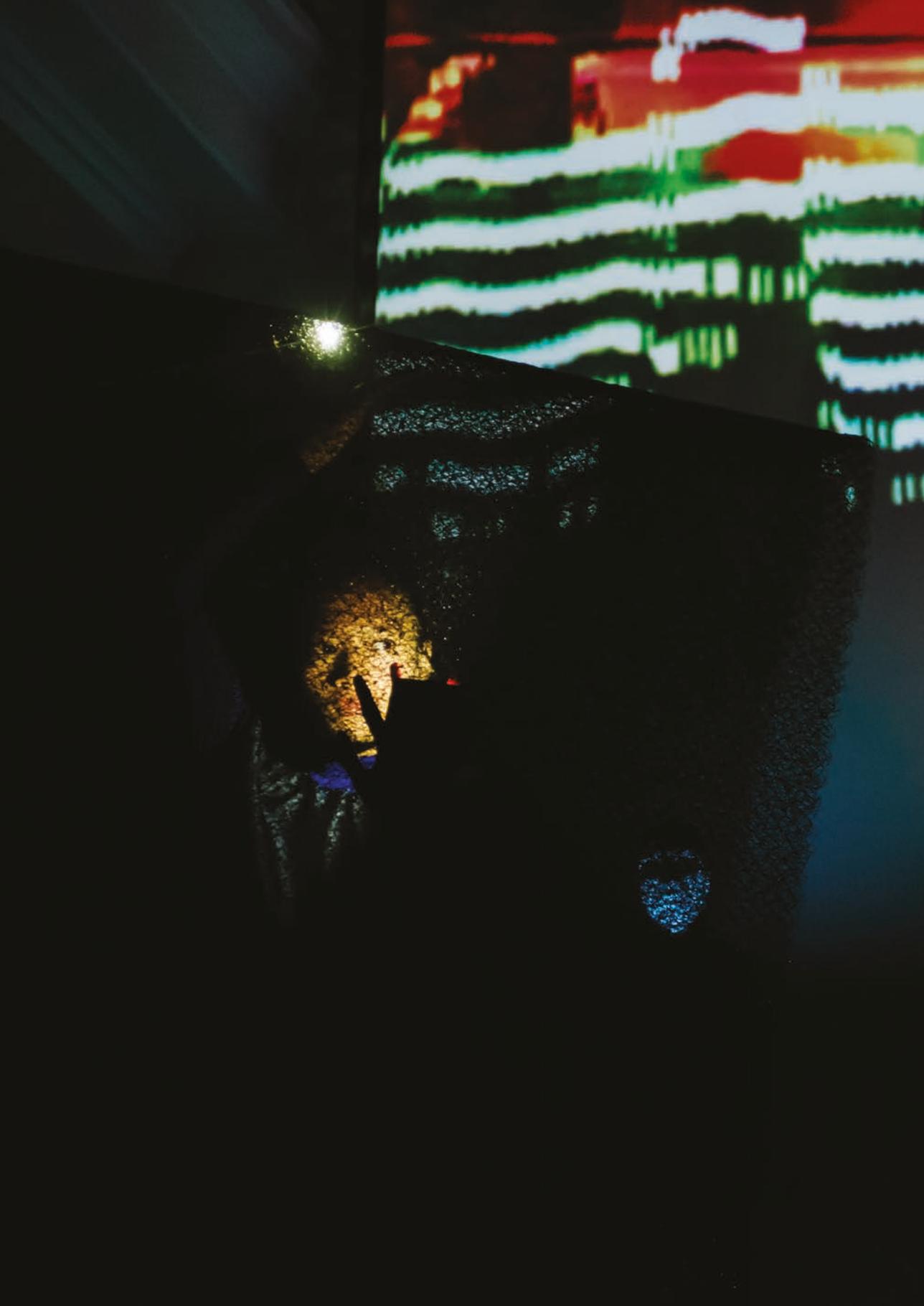
They start moving according to these instructions
from above,

Meanwhile the other 3 performers are entering the
stage

They turn their back to the audience and start
writing on the wall. They perform this action for
about 3 minutes.

One of the dancers moves the light sources to
spark moving shadows. The other dancer observes
and reacts.

really speak or not? They run sequentially.





The performer repeats and, as the machine progressively removes letters, so does his struggle intensify, leading to a hypnotic display of literary trance—an abstraction reminiscent of speaking in tongues.

As he reads, so does he walk, forward and towards the audience. His presence is striking, yet he is no more than a vessel speaking words as assigned to him. He has sacrificed his own volition, his own speech, to the voiceless.

A.

.
 ync will fluh the file ytem buffer. Commnd eprted by ";" run equentilly. The hell wit for ech commnd to terminte before executing the next commnd in the equence. mentioned in kernel doumentttion, writing to drop_cche will clen cche without killing ny ppliction/ervice, commnd echo i doing the job of writing to file. ...echo 1 >" will cler the PgeCche only, while uing "...echo 3 >" in production will cler PgeCche, dentrie nd inode.

R.

.
 yn will fluh the file ytem buffe. ommnd epted by ";" un equentilly. The hell wit fo eh ommnd to teminte befoe exeuting the next ommnd in the equene. mentioned in kenel doumentttion, witing to dop_he will len he without killing ny pplition/evie, ommnd eho i doing the job of witing to file. ...eho 1 >" will le the Pgehe only, while uing "...eho 3 >" in podution will le Pgehe, dentie nd inode.



04. Error Almighty

Established by usage, meaning is rooted in cultures that take the form of tradition. The indexer, then, is creating descriptions drawing into the past, but expressing them with an eye to the future. Usage, however, is a lively, dynamic, ever-changing flow shaped and misshaped by countless congruent factors such as cultural movements, ephemeral trends, current events, technological shifts, new rituals, etc.

Inherently descriptive, the resulting linguistics are also necessarily adaptive and unstable and, as human communication too is highly qualitative and subjective in essence, conflict with the need for stable, prescriptive, unambiguous syntax for machine systems to perform efficiently. This lack of standardisation of language adds up to the burden of interpretation.

60% Human
03% Conscious
12% Happy

Much of information technology can be viewed as a sustained effort to increase our communication possibilities by diminishing the effects of separation in space and time. Errors happen, of course, at the intersection between those two, radically opposed ontologies which never yet learned to coexist as co-creating partners.

The aesthetics of error are yet another cultural artefact. Glitches. Cut-up text. Falling bodies. Irregularities. Error can be identified, repaired, addressed, embraced even. Can it be performed? What are the limitations of bodies and brains conditioned to be optimally used in representing system failure? What aesthetics arise from this exercise? What does error take from us and, most of all: what does error bring us?

99% Empathetic.

Radical advances in computer vision, coupled with new learning algorithms, have set machines en course to coevolve with humans, which has in turn incentivised humans to develop standardised behaviours, better readable

12

Title 4 - Santa Data - We are all
data - Quantifiable Worlds and The Almighty Error 12

Duration: 4 minutes

2 Dancers, 3 Performers

Performers given instructions, crossing the stage.

Should we start glitching the cupola here?

their bodies form a moving mechanism...machine

privacy...randomness...training...sets...emergence...prediction...patter:
 recognition...validation sets...vizualization...
 models...outliers..Gaussian
 curve...statistics...experiments...exploratory data
 analysis...adding smartness and
 anonymisation...estimation...discriminants...convolutional
 neural networks...customer
 segmentation...clusters...credit scoring...dimensionality

by machines—effectively improving communication and interaction, and allowing for a development of technology steering towards what Eliezer Yudkowsky had theorised in 2008 as Friendly Artificial Intelligence. This rise in standardised behaviours triggered a global levelling of abilities, comportment and language, which gradually reduced the margin for error to the narrowest interstices of computing.

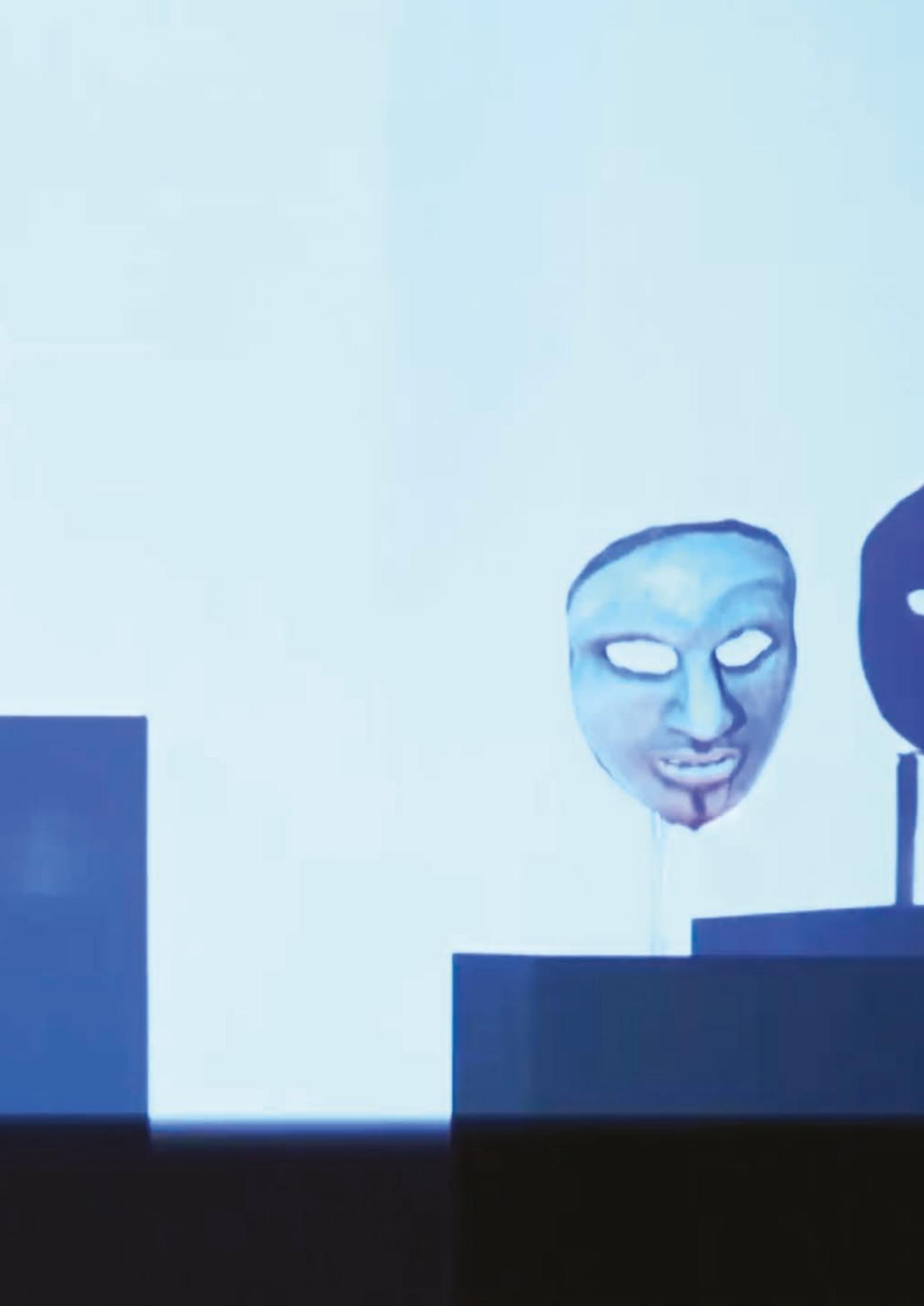
80% Complying.

27% Machine

Some argue that this levelling of interaction models, in addition to reducing the error ratio, has also diminished our ability to behave in non-standard ways, for fear of not being understood by the surrounding technosocial landscape and that it, in turn, resulted in three unsuitable consequences: (1) reduced personal freedom; (2) generalisation of substandard (as opposed to non-standard) and therefore unsatisfying life-styles; and (3) stronger exclusion risks for these with lower (politically correct requires that we prefer the term alternative) expression capabilities. (*Müller 2018*)

Supporters of these theories are widely viewed as belonging to a conservative fringe, and a subset of this group regularly attempts to subvert the system by elevating to the rank of public practice the performing of non-standard attitudes such as dance and grimaces.





Speech is not like the blowing
(of the wind)

The speaker has to wind up

twice



05. Deep Blue's Deep Day Dream

Vivid echoes to our experiences of the world, dreams are seldom brought forth in the public space. They are, in western societies at least, viewed as intimate aspects of our inner lives, and often kept secret.

Writers and theatre practitioners make use of dreams to enhance their perceptive and creative abilities, and propose techniques to harness their potential. They keep diaries of dreams and those feed their inspiration with exciting new landscapes. Do android dream of electric sheep?—Dick (1968) asks in the title of his novel, and the question remains intact. It is as impossible to perceive the inner life of a machine—if any exists—as that of a fellow human. Aside from asking, all one can do is imagine.

COLOUR: DARK BLUE - ALL COLOURS

The most expressive scene,

Music gets louder,

*Reactive visuals in the cupola and on the stage
turn frantic,*

*The two dancers are taking over the stage.
(It is one of those dreams one wakes up
from in a single jump. Recalling events
from the day, amplified and distorted.)*

Bodies are moving,

repetitively,

in circles.

Climax.

They get exhausted and fall.

Pedestals are turned off,

All performers exit or fall to the ground.

COLOUR: BLACK

06. Reboot

The pedestals lights turn up,
Digital icons slowly appear over the masks,
Uncanny human/machine voices are heard,
Questioning consciousness.

Imagine a molecule,
that isn't alive,
but is able to
replicate itself.
A Replicator.

The crisis
especially now
is a crisis
in consciousness.

The replication process
is not going to be perfect.

Emergence... advancements in hardware...

During the replication
there will be
accidental changes.

... errors... supervised learning.

In competitive environments,
if this accidental change
improves survival
it will be more likely to replicate.

Abiogenesis is not sudden,
but gradual.

(Akten, 2016)

BLACKOUT



DISCLAIMER:

In order to ensure a fair representation of speculative cultures, REPLICA has elected to feature one fictional reference in this work.

-
- Akten, M. (2016) *A Digital God for a Digital Culture*. [Lecture] Resonate 2016
- Armand, L. (2005) *Intelligence and Representability*. 1000 Days of Theory: Kroker.
- Artaud, A. (1938) *Le Théâtre et son double*. Paris: Gallimard.
- Boal, A. (1993) *Theatre of the Oppressed*. Theatre Communications Group; Tcg ed.
- Bogart, A. and Landau, T. (2005) *The Viewpoints Book: A Practical Guide to Viewpoints and Composition*. New York: Theatre Communications Group.
- Buckland, M. K. (2017) *Information and Society*. Cambridge: The MIT Press
- Candy, S. (2010) *The Futures of Everyday Life: Politics and the Design of Experiential Scenarios*. Carnegie Mellon University Press.
- Case, N. (2018) *How To Become A Centaur*. 01.08.2018. MIT Journal of Design and Science. [Online] Available at: <https://jods.mitpress.mit.edu/pub/issue3-case>
- Connor, S. (2019) *Dream Machines*. London: Open Humanities Press.
- Dick, P. K. (1968) *Do androids dream of electric sheep?* New York : Ballantine Books
- Diebner, D and Weibel, P (2001) *Sciences of the interface*. Tuebingen, Genista
- Dunne, A and Raby, F. (2013) *Speculative Everything*. Boston: MIT Press.
- Falletti, C., Sofia, G., Jacono, V., (2017) *Theatre and cognitive neuroscience*. London: Bloomsbury Methuen Drama.
- Hackney, P. (2002) *Making Connections*. New York: Routledge.
- Hodges, A and Turing, A. (1983) *The Enigma*. New York: Simon and Schuster
- Hofstadter, D. R. (2017) *I Am a Strange Loop*. New York: Basic Books.
- Hofstadter, D. R. (1999) *Gödel, Escher, Bach : An Eternal Golden Braid*. New York: Basic Books.
- Grotowski, J. (2002) *Towards a Poor Theatre*. New York: Routledge.
- Hancock P. A. (2009) *Mind, Machine and Morality. Toward a Philosophy of Human-Technology Symbiosis*. Surrey: Ashgate Publishing Limited.
- Kolankiewicz, L. (1979) *On the Road to Active Culture : The Activities of Grotowski's Theatre Laboratory Institute in the Years 1970 - 1977*. Milan
- Microsoft Research Podcast (2018) *Building Literate Machines with Dr. Adam Trischler*. [Podcast] 21.03.2018. Available from: <https://www.microsoft.com/en-us/research/blog/building-literate-machines-with-dr-adam-trischler/>
- Müller, K.J. (2018) *Elements of Human-Machine-Learning*. Berlin: Clinical Algorithm Review.
- Parker-Starbuck, J. (2014) *Cyborg theatre: corporeal/technological intersections in multi-media*. NY: Palgrave Macmillan
- Pickering, A. (2009) *The cybernetic brain: sketches of another future*. Chicago, U.S.A.: Univ. of Chicago Press
- Pfeifer R. (2007) *How the Body Shapes the Way We Think. A New View of Intelligence*. Cambridge: The MIT Press.
- Sterling, B. (2002) *Tomorrow Now: Envisioning the next fifty years*. Random House.
- Todd, M. E. (1997) *The Thinking Body*. London: The Gestalt Journal Press.
- Turing A. (1950) *Computing Machinery and Intelligence*. Mind 49: 433-460.
- Villiers de l'Isle-Adam, A. (1886) *L'Ève Future*. Paris: Charpentier
- Yudkowsky, E. (2008) *Artificial Intelligence as a Positive and Negative Factor in Global Risk*. In *Global Catastrophic Risks*, Oxford University Press.

Autor:innenverzeichnis

Dirk Baecker, Prof. Dr., ist Soziologe und Inhaber des Lehrstuhls für Kulturtheorie und Management an der Universität Witten/Herdecke. Jüngst erschienen sind seine Bücher »4.0 oder Die Lücke die der Rechner lässt« (Leipzig: Merve Verlag, 2018) und »Intelligenz, künstlich und komplex« (Leipzig: Merve Verlag, 2019). Internet: uni-wh.de/baecker und catjects.wordpress.com. Email: dirk.baecker@uni-wh.de.

Lukas Brand, Magister Theologiae, ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Religionsphilosophie und Wissenschaftstheorie an der Katholisch-Theologischen Fakultät der Ruhr-Universität Bochum. Seine mit dem Rotary-Universitätspreis ausgezeichnete Magisterarbeit erschien 2018 unter dem Titel »Künstliche Tugend. Roboter als moralische Akteure«. Gegenwärtig forscht Brand neben den Fragen der Maschinenethik besonders zur Anthropologie im Zeitalter der Digitalisierung und der künstlichen Intelligenz.

Robert Geraci, Prof., earned his PhD from the University of California at Santa Barbara and is Professor of Religious Studies at Manhattan College. He is the author of *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality* (Oxford 2010), *Virtually Sacred: Myth and Meaning in World of Warcraft and Second Life* (Oxford 2014), and *Temples of Modernity: Nationalism, Hinduism, and Transhumanism in South Indian Science* (Lexington 2018). His research has been supported by the U.S. National Science Foundation, the American Academy of Religion, and two separate Fulbright-Nehru research awards. He is an elected Fellow of the International Society for Science and Religion.

Benedikt Paul Göcke, Dr.phil, Dr. theol., ist Professor für Religionsphilosophie und Wissenschaftstheorie an der Katholisch-Theologischen Fakultät der Ruhr-Universität Bochum und assoziiertes Mitglied der Faculty for Theology and Religion an der Universität Oxford. Göcke publiziert unter anderem zu Fragen der Ethik, der Metaphysik, der Wissenschaftstheorie und zum Deutschen Idealismus. Sein neuestes Buch »The Panentheism of Karl Christian Friedrich Krause. From Transcendental Philosophy to Metaphysics« ist frei verfügbar unter <https://www.peterlang.com/view/title/67511?rskey=OPqSiV&result=4>

Dirk Helbing ist Professor für Computational Social Science am Department Geistes-, Sozial- und Staatswissenschaften sowie beim Department of Computer Science der ETH Zürich assoziiert. Seine aktuellen Studien diskutieren global vernetzte Risiken und die digitale Gesellschaft. An der Delft University of Technology leitet er das Doktorandenprogramm »Engineering Social Technologies for a Responsible Digital Future«. Er ist zudem gewähltes Mitglied der Deutschen Akademie der Wissenschaften. Mehr Informationen finden Sie unter: https://de.wikipedia.org/wiki/Dirk_Helbing

Simon Sebastian Hunt, LL.B., studiert Rechtswissenschaften an der Universität Potsdam und ist seit Herbst 2018 studentischer Mitarbeiter des Kompetenzzentrums Öffentliche IT (ÖFIT) am Fraunhofer FOKUS. Hier beschäftigt er sich schwerpunktmässig mit den Themen Datenschutz, Rechtsautomation, dem Themenkomplex Digitalisierung und Demokratie, sowie den jeweiligen Einflüssen und Einsatzmöglichkeiten von KI.

Lorena Jaume-Palasi ist Gründerin von The Ethical Tech Society. Sie forscht zur Ethik der Digitalisierung und Automatisierung und in diesem Zusammenhang mit rechtsphilosophischen Fragen. 2017 wurde sie von der Regierung Spaniens in dem Weisenrat zu Künstlicher Intelligenz berufen. Sie war Mitglied der High Level Expert Group on Artificial Intelligence der EU Kommission. 2018 erhielt sie mit der Initiative AlgorithmWatch die Theodor Heuss Medaille »für ihren Beitrag zu einer differenzierten Betrachtung von Algorithmen und deren Wirkmechanismen«.

Karoline Krenn ist promovierte Soziologin und forscht im Kompetenzzentrum Öffentliche IT (ÖFIT) bei Fraunhofer FOKUS in Berlin zu aktuellen Digitalisierungsthemen. Der Schwerpunkt ihrer Arbeit liegt auf sozio-technischen Aspekten der digitalen Transformation wie der Gestaltung neuer Technologien, neuen Formen digitaler Kommunikation und auf Methoden und Praktiken der Gewinnung, Verarbeitung und Verwertung von digitalen Massendaten. Jüngst erschienen sind Veröffentlichungen zum Selbstschutz (ÖFIT-Impuls), Privatheit (ESSN-Newsletter) und der vorliegende Band zu Künstlicher Intelligenz. Vor ihrer Tätigkeit bei ÖFIT war Karoline Krenn wissenschaftliche Mitarbeiterin an der Humboldt-Universität zu Berlin und Gastwissenschaftlerin an der UC Berkeley, CA.

Régis Lemberthe advocates culturally-infused futures, viewing design as a political tool, technology as a cultural artefact, and speculative fiction as a way to prototype collective imaginaries. He co-founded REPLICA with the

ambition to use the stage as a platform for social experimentation, and to further explore how performance allows the prototyping of future sociocultural paradigms—codes, aesthetics, and rituals.

Christian »Mio« Loclair, Art Director bei Waltz Binaire, untersucht in seinen Arbeiten die Spannungsfelder zwischen sinnlichen Erfahrungen und digitalen Ästhetiken. Seine interdisziplinären Arbeiten erschaffen Gradienten zwischen neuen Medien, Kunst und Technologie und verwenden Technik als Mikroskop für menschliche Bedürfnisse. Mio studierte Mensch Maschine Interaktion am HPI (Pinchwatch, Paper HCI Lissabon), wurde 2007 und 2010 zu den 16 besten Tanz-solisten der Welt gekürt (Sony Championships) und gewann den Honory Mention der Ars Electronic (Pathfinder). Sein Design Studio Waltz Binaire gestaltet und entwickelt interdisziplinäre Cutting-Edge-Projekte in Zusammenarbeit mit Werbeagenturen und kulturellen Instituten.

Peter Parycek ist Leiter des Kompetenzzentrums Öffentliche IT am Fraunhofer-Institut FOKUS in Berlin. An der Donau-Universität Krems, Österreich, verantwortet er als Universitätsprofessor für E-Governance das Department für E-Governance in Wirtschaft und Verwaltung. Als Rechtsinformatiker arbeitet Peter Parycek inter- und transdisziplinär am Schnittpunkt rechtspolitischer, gesellschaftlicher und technologischer Entwicklungen. Seine Forschungsschwerpunkte sind der Einsatz von Technologie in Regierung und Verwaltung und deren Wirkungen auf die bestehenden Rechts- und Organisationssysteme. Von 2006 bis 2011 war er im österreichischen Bundeskanzleramt als wissenschaftlicher Berater tätig und verantwortete Projekte und Arbeitsgruppen im Bereich E-Demokratie und E-Government. Neben seinen akademischen Tätigkeiten engagiert er sich bei zahlreichen zivilgesellschaftlichen Initiativen zum Thema Digitalisierung.

Christian Rauch ist promovierter Physiker, Wissenschaftskommunikator und Gründer und Geschäftsführer von STATE. Er studierte Physik in Berlin und Sydney und promovierte zum Thema nachhaltige Energiematerialien an der Aalto University in Helsinki. Nach Arbeit im Wissens- und Technologietransfer für ein deutsches Forschungsinstitut gründete er in 2014 er STATE Experience Science um mit Partnern aus Forschung und Kultur interdisziplinäre Programme an der Schnittstelle zwischen Wissenschaft, Kunst und Innovation zu entwickeln und den disziplinenübergreifenden Dialog zum nachhaltigen Einsatz von Technologie und wünschenswerten Zukunftsentwürfen zu fördern. In 2018 eröffnete in Partnerschaft mit Wissenschaft im

Dialog das STATE Studio als Ausstellungslabor und Wissenschafts-Kunst Galerie in Berlin. Neben seiner Arbeit mit STATE ist Christian Lehrbeauftragter am französischen European Center for Executive Development (CEDEP), Botschafter der kalifornischen XPRIZE Foundation im Bereich Künstliche Intelligenz und Robotik, und Mitglied des Responsible Leader Netzwerks der BMW Stiftung. Christian spricht und schreibt regelmäßig zum Thema Wissenschaft, Kunst und Gesellschaft, ist Mitglied des Programmkomitee des »AI for Good« Gipfels der Vereinten Nationen und Gast im Programmbeirat des Forums Wissenschaftskommunikation 2019.

Ina Schieferdecker, Prof. Dr.-Ing. , leitet seit dem 01.10.2019 die Abteilung Forschung für Innovation und Digitalisierung beim BMBF. Zuvor war sie Institutsleiterin des Fraunhofer-Instituts für Offene Kommunikationssysteme (FOKUS), Professorin an der Technischen Universität Berlin zum Quality Engineering von Offenen Verteilten Systemen und Direktorin des Weizenbaum-Instituts für die Vernetzte Gesellschaft, dem Deutschen Internet-Institut.

Astrid Schwarz interessiert sich für die Charakterisierung theoretischer und praktischer Wissensformen. Im Sinne einer philosophischen Feldforschung untersucht sie das Wechselspiel von Begriffen und Objekten, von Technik und Umwelt, von künstlerischem und wissenschaftlichem Experimentieren. Seit Anfang 2017 ist sie Professorin für Allgemeine Technikwissenschaft an der BTU Cottbus-Senftenberg, die letzten Buchpublikationen sind »Experiments in Practice« (2014) und als Mitherausgeberin »Research Objects in their Technological Setting« (2017).

Diana Alina Serbanescu (aka Neranti), with a double background in computer science and performing arts, works on interdisciplinary approaches to culture, society and technology, with a strong focus on human factors. She is the co-founded REPLICA, a performing arts platform inviting creatives and scientists to collaborate on imagining hybrid behavioural models for humans and machines, and to prototype future tools, cultures and rituals. As the artistic director of REPLICA, Diana Serbanescu envisions the dance-theatre of the future: playful, subversive, interactive, underpinned by new technologies.

Stefan Ullrich ist promovierter Informatiker und Philosoph, der sich kritisch mit den Auswirkungen der allgegenwärtigen informationstechnischen Systeme auf die Gesellschaft beschäftigt. Er leitet die Forschungsgruppe

»Verantwortung und das Internet der Dinge« am Weizenbaum-Institut für die vernetzte Gesellschaft, Berlin. Aktives Mitglied der Fachgruppe @gewissensbits der Gesellschaft für Informatik.

Tobias Wangermann studierte Germanistik, Kunstwissenschaft und Pädagogik in Greifswald, Berlin und Essen. Nach literaturwissenschaftlichen Projekten ist er seit 1996 in der Konrad-Adenauer-Stiftung tätig: Zuerst als wissenschaftlicher Mitarbeiter in der Politischen Bildung und seit 2004 in der Hauptabteilung Politik und Beratung. Von 2012 bis 2015 leitete er dort die Stabsstelle Beratungsmanagement. Seit 2015 ist er Leiter des Teams Digitalisierung.



Dieser Band versammelt Text- und Bildbeiträge aus Wissenschaft, Zivilgesellschaft und Kunst zu zentralen Fragen und Herausforderungen Künstlicher Intelligenz und knüpft damit an das Symposium »(Un-)ergründlich? Künstliche Intelligenz als Ordnungsstifterin« und den Salonabend »KI und Kunst. Schamanen der digitalen Renaissance« vom Oktober 2018 an. Im Vordergrund der behandelten Themen stehen die Wechselwirkungen zwischen technischen Möglichkeiten von und gesellschaftlichen Anforderungen an KI. Die Beiträge bieten neben der Einführung in grundlegende Konzepte eine Anknüpfung an aktuelle Diskussionen zu Sozialverträglichkeit, Vertrauen und Gestaltbarkeit von KI-Systemen sowie wie die Gelegenheit einer visuellen Erfahrung, was KI schon heute leisten kann. Die Bildserie Blackberry Winter von Waltz Binaire begleitet die Textbeiträge des Bandes und illustriert Lernschritte einer KI, einen dreidimensionalen menschlichen Körper zu erschaffen.

Mit Beiträgen von Dirk Baecker, Lukas Brand, Robert M. Geraci, Benedikt Paul Göcke, Dirk Helbing, Lorena Jaume-Palasi, Régis Lemberthe, Christian Mio Loclair, Christian Rauch, Ina Schieferdecker, Astrid Schwarz, Diana Serbanescu, Stefan Ullrich und Tobias Wangermann.