

Künstliche Intelligenz in der Zukunftsforschung

Axel Zweck, Thomas Werner

1. Einleitung¹

Seit über 100 Jahren verfolgen Denker:innen, Forscher:innen und Entwickler:innen das Ziel, Menschen und Maschinen miteinander zu verbinden. Mit zunehmender Anwendungsbreite steht heute nicht mehr die gesellschaftliche Utopie, sondern die Rolle, die Künstliche Intelligenz auf unserem Weg in eine digitale Wirtschaft und Gesellschaft einnimmt, im Mittelpunkt. Künstliche Intelligenz ändert gegenwärtig unsere Lebens- und Arbeitswelt wie jede grundlegend neue Technologie, die sich in kurzer Zeit so ubiquitär durchsetzt. Bisher ist die Frage, welche Auswirkungen dieser Wandel haben wird, noch nicht entschieden. KI gilt als zentraler Treiber der jetzigen Phase der Digitalisierung und als entscheidende Zukunftstechnologie für die wirtschaftliche Entwicklung nicht nur unseres Landes, sondern als ein künftiger Wirtschaftsmotor aller Industrienationen. Für das Jahr 2025 wird geschätzt, dass mit KI-Anwendungen weltweit ein Umsatzvolumen von ca. 30 Milliarden USD erreicht werden wird.² Länder wie die USA, China und auch Deutschland fördern KI-Forschung und deren Umsetzung in die Anwendung mit Milliardenbeträgen. China sieht in der KI eine »strategische Gelegenheit«, um die Wirtschaft des Landes in eine wissensbasierte Ökonomie zu transformieren (Groth et al. 2018). Die KI-Strategie Pekings sieht vor, dass China bis 2030 eine Vormachtstellung einnimmt und die mit KI verbundenen Industriezweige einen Wert 1,2 Billionen Euro erlangen (ebd.). Auch Deutschland will im Rahmen der »Strategie Künstliche Intelligenz« bis 2025 insgesamt etwa fünf

1 Der Redaktionsschluss für diesen Beitrag war Ende 2022, weswegen die rasante Entwicklung von ChatGPT nicht mehr berücksichtigt werden konnte.

2 Statistiken zum Thema Künstliche Intelligenz, Statista 2021, <https://de.statista.com/themen/3103/kuenstliche-intelligenz/#dossierKeyfigures>, abgerufen am 02.12.2021.

Milliarden Euro für die Umsetzung zur Verfügung stellen.³ Dies macht offensichtlich, dass dieses Thema auch für die Zukunftsforschung von besonderem Interesse ist. Wobei es in den weiteren Ausführungen weniger um die Frage geht, wie sich KI aus Perspektive der Zukunftsforschung entwickelt, als um die Frage, wie der Einsatz von KI Zukunftsforschung und Zukunftsforscher:innen beeinflussen wird.

Im vorliegenden Text wird gezeigt, dass sowohl der Begriff der Künstlichen Intelligenz als auch der Begriff der Intelligenz inhärent unscharf sind. Nach Gardners Theorien der multiplen Intelligenzen, hat menschliche Intelligenz zahlreiche Dimensionen, sodass der Intelligenzbegriff differenziert und im Kontext des jeweiligen Anwendungsfeldes betrachtet werden muss. So kann erahnt werden, wie wir Leistungen von KI-Systemen hinsichtlich ihrer »Intelligenz« zu bewerten haben. Es reicht nicht aus, nur Chancen und Risiken beim Einsatz von KI kritisch zu hinterfragen, sondern vor allem unsere eigene Erwartungshaltung einer solchen Technologie gegenüber.

Einem Computersystem bzw. einer Künstliche Intelligenz wird schnell intelligentes Verhalten zugeschrieben, sobald es menschliches Verhalten erfolgreich nachahmt. Dass diese Projektion auf KI (und das Verhalten) historische Wurzeln hat, zeigen wir anhand des »Siegeszuges der schwachen KI«.

Während »starke KIs« noch Gegenstand der Forschung sind, dringen »schwache KIs« (also spezifisch problemlösende Intelligenzen) zunehmend in menschliche Problemdomänen vor, für deren Lösungen bisher Kreativleistungen oder gar »echte Intelligenzleistungen« vermutet wurden.

Künstliche Intelligenz ist ein Instrument, um bisher als unlösbar oder sehr aufwändig gesehene Probleme über neuartige Wege zu lösen und zu denken. Sie führt in neue, auch disruptive Nutzungskontexte und verändert zugleich den technischen und methodischen Kern wissenschaftlicher Arbeitsweisen. Damit geht nicht nur ein Wandel des Wissenschaftsverständnisses einher, sondern auch des wissenschaftlichen Erkenntnisprozesses selbst.

KI macht Komplexität handhabbarer und schafft zugleich neue Komplexität. Sie erhält durch ihren bisher vorherrschenden Black Box-Charakter (Entscheidungs-)Autonomie und wandelt sich zunehmend vom Gegenstand der Betrachtung zum Instrument für Wissenschaft und Forschung. Sie ändert hierdurch auch ihre Rolle vom Instrument zur Akteur:in.

3 KI-Strategie der Bundesregierung, 2020, <https://www.bmwi.de/Redaktion/DE/Artikel/Technologie/kuenstliche-intelligenz.html>, abgerufen am 02.12.2021

KI-Systeme in Wissenschaft und Forschung charakterisieren sich immer weniger ausschließlich in ihrer Werkzeugfunktion. Sie ermöglichen fundamental neue Denkansätze und eröffnen neue Zugänge des Denkens und des Forschens. Auch die heute verfügbare schwache KI führt bereits zu neuen, sogar disruptiven Nutzungskontexten und dringt in die Zweckdimension forschenden Handelns ein. Der Mensch und auch die Wissenschaftler:in sind damit nicht mehr alleinige Akteur:in. Sie definieren ihre Rolle als handelnde Akteur:in im Zentrum von Wissenschaft und Forschung neu. Dennoch werden KIs nicht als eigenständige Akteur:innen in Begründungsdiskursen auftreten, weil ihnen die Merkmale einer Handlungsurheberschaft fehlen. Dies kann mit dem Angebot Whites für ein gestuftes Identitätsmodell nachvollziehbar gemacht werden (White 2008; Schmitt/Fuhse 2015), mit dem eine genaue Nachvollziehbarkeit der graduellen Übertragung von Handlungsträgerschaften (als Träger von identifizierbaren Kontrollprojekten, die nicht allein aus dem KI-Verfahren stammen) an KI-Verfahren möglich wird. Sie zeigt sich in einer anderen Positionierung der technischen Bestandteile in Netzwerken und Erzählungen, ohne auf die Komplexitätsstufe der Person (als integriertes Bündel von multiplen Positionierungen und Erzählungen) zu gelangen. Kontrolle wird in dieser Analyse dann an KI-Verfahren übertragen, die in deren Umsetzung dann auch mit steigenden Freiheitsgraden ausgestattet sind.

2. Der Weg zur Künstlichen Intelligenz

Doch was steckt hinter dieser wichtigen Schlüsseltechnologie? Wie wird KI einzelne Branchen, künftige Arbeitsweisen, Forschung und auch die Zukunftsforschung verändern? Wie ist eine Technologie bzw. die dahinterliegenden Algorithmen und trainierten Modelle⁴ in der Lage, mit nahezu

4 Der klassische Weg (Rechen-)Probleme mit dem Computer zu lösen, ist die Analyse des Problems derart, dass zunächst Regeln und Zusammenhänge identifiziert werden. Diese werden dann genutzt, um ein Schema zu entwickeln – den Algorithmus –, mit dem (wiederholbare, also deterministische) Rechenvorgänge durchgeführt werden, die das Ausgangsproblem lösen. Machine Learning-Modelle werden nicht programmiert, sondern zunächst mit Trainingsdaten trainiert und berechnen zu einer Eingabe auf Basis des trainierten Modells eine passende Ausgabe – auch wenn das zugrundeliegende Machine Learning-System durch Algorithmen abgebildet wird, arbeiten KI-Systeme in ihrer Wirkweise daher nicht wie klassische Algorithmen.

übermenschlicher Genauigkeit, die Leistung des Menschen in vielen Bereichen zu übertreffen? Definitionen für Künstliche Intelligenz haben eine große inhaltliche Streubreite. Im vorliegenden Text wird Künstliche Intelligenz als die Fähigkeit eines Computers verstanden, Aufgaben zu lösen, für die üblicherweise Intelligenz erforderlich ist. Solche Systeme sind in der Lage, sich wie ein Mensch »intelligent« zu verhalten und selbständig zu lernen. Die KI-Forschung hat mittlerweile eine Leistungsstärke erreicht, die selbst bei komplexeren Aufgabenstellungen beeindruckende Ergebnisse erzielt: Smartphones übersetzen Fremdsprachen in Echtzeit in die eigene Sprache und autonomes Fahren soll die Mobilität verändern sowie Sicherheit wie auch Komfort auf ein neues Niveau heben. Im Gesundheitssektor ist durch KI eine Verbesserung in Diagnostik, Früherkennung von Krankheiten und bei der Entwicklung neuer Medikamente zu erwarten (Häußling et al. 2021). Auch die rasche Entwicklung von mRNA-Impfstoffen gegen das Covid-19-Virus wurde durch den Einsatz von KI erst ermöglicht (Baidu 2020). Facebook entwickelt eine KI, die eine eigene, für Menschen unverständliche »KI-Sprache« erfundet und Wissenschaftler:innen entwickeln einen Algorithmus, der automatisch Fake News generiert, sich aber aus ethischen Gründen weigern, diesen zu veröffentlichen – gleichzeitig kann dieser Algorithmus dazu eingesetzt werden, um Fake News aufzudecken (Giansiracusa 2021).

Doch nicht überall, wo KI draufsteht, ist auch KI drin und nicht alles was intelligent erscheint (sei es durch die geschickte Nutzung regelbasierter Systeme oder das Erzeugen überraschender Ergebnisse), ist wirklich intelligent im menschlichen Sinne. Nach einer Untersuchung von MMC Ventures (Kelnar/Kostadinov 2019) haben 40 Prozent aller europäischen Neugründungen, die als »KI-Unternehmen« klassifiziert werden, in Wirklichkeit nichts mit Künstlicher Intelligenz zu tun. Sie nutzen das Label KI zur Erlangung von Fördermöglichkeiten und Kund:innen. Zur besseren Einordnung solcher Beobachtungen und wie künstliche mit menschlicher Intelligenz verglichen werden kann, lohnt ein vertiefter Blick auf die Geschichte und die Technologie hinter der Künstlichen Intelligenz.

Was ist Intelligenz?

Die inhärente Unschärfe in der Definition des Begriffes »Künstliche Intelligenz« liegt zum einen in der Historie, zum anderen im Verständnis des Begriffes von Intelligenz im Allgemeinen. Der Psychologe und Erziehungswissenschaftler Howard Gardner (Gardner 1983) vertritt eine Theorie multipler

Intelligenzen. Über die Beobachtung von Menschen mit Inselbegabungen und der Untersuchung historisch herausragender Persönlichkeiten beschreibt Gardner Intelligenz als Zusammenspiel von acht Dimensionen: Die *sprachlich-linguistische Intelligenz* umfasst die Sensibilität für gesprochene und geschriebene Sprache sowie die Fähigkeit, Sprachen zu lernen und zu bestimmten Zwecken zu gebrauchen. Berufsgruppen wie Rechtsanwält:innen, Redner:innen, Schriftsteller:innen und Dichter:innen wird eine hohe sprachlich-linguistische Intelligenz nachgesagt. Zur *logisch-mathematischen Intelligenz* zählt die Fähigkeit, Probleme nach logischen Gesichtspunkten zu analysieren und mathematische Operationen einzusetzen, um wissenschaftlichen Fragen nachzugehen. Mathematiker:innen, Logiker:innen, Programmierer:innen und Naturwissenschaftler:innen machen von dieser Form der Intelligenz besonders Gebrauch. *Musikalische Intelligenz* bedeutet die Begabung zu Musizieren, zu Komponieren und das Verständnis für musikalische Prinzipien. Der *räumlichen Intelligenz* wird der theoretische und praktische Sinn für die Strukturen großer Räume zugeordnet. Seeleuten und Pilot:innen, Bildhauer:innen, Chirurg:innen, Schachspieler:innen, Ingenieur:innen, Graphiker:innen oder Architekt:innen nutzen besonders diese Form der Intelligenz. Die *körperlich-kinästhetische Intelligenz* beschreibt die Fähigkeit den Körper und Teile davon zur Problemlösung oder zur Gestaltung von Produkten einzusetzen. Neben Vertreter:innen technischer Berufe wie Handwerker:innen, Chirurg:innen und Mechaniker:innen zählen hierzu auch Tänzer:innen, Schauspieler:innen und Sportler:innen. Die *naturalistische Intelligenz* beschreibt die Fähigkeit, Naturphänomene zu beobachten, zu unterscheiden und zu erkennen. Typische Vertreter:innen für den Einsatz dieser Fähigkeiten sind Naturforscher:innen, Umweltspezialist:innen, Tierärzt:innen und Köch:innen. Die *interpersonale Intelligenz*, die nach den Psychologen Edward Lee Thorndike und David Wechsler auch als soziale Intelligenz (Wechsler 1964) bezeichnet wurde, bezeichnet die Fähigkeit, unausgesprochene Gefühle und Motive anderer Menschen nachzuempfinden und deren Emotionen zu beeinflussen. Eine Fähigkeit, die wesentliche Voraussetzung für den erfolgreichen Umgang mit anderen Menschen ist und laut Gardner bei politischen oder religiösen Führer:innen, Eltern und Lehrer:innen sowie bei beratenden oder heilenden Berufen besonders stark ausgeprägt ist.

Als letzte der acht Dimensionen der Intelligenz wird die *intrapersonelle Intelligenz* angeführt. Die Fähigkeit, die eigenen Gefühle und Motive zu verstehen und letztlich auch zu beeinflussen. Eine Art »interne Selbsterkenntnis« die Menschen zu richtigen Entscheidungen verhilft und besonders bei Schriftstel-

ler:innen, Schauspieler:innen und Künstler:innen ausgeprägt ist. Gerade die beiden letzten Intelligenztypen sind Basis der Theorien der Psychologen John D. Mayer und Peter Salovey zur »Emotionalen Intelligenz« (Mayer et al. 1990: 772f.; Mayer/Salovey 1990: 185f.), die später auch durch Daniel Golemans Buch »Emotionale Intelligenz« (Goleman 1995) popularisiert wurde.

Obgleich Gardners Theorie der multiplen Intelligenz von der wissenschaftlichen Intelligenzforschung aufgrund empirischer Mängel kritisch gesehen wird (Süß 2011: 97f.), zeigt dieser Exkurs die Vielfältigkeit des Intelligenzbegriffs. Es wird deutlich, dass es keine universelle Antwort auf die Frage nach der Messbarkeit von Intelligenz geben kann. Die zahlreichen existierenden Testverfahren zur Messung von Intelligenz und die damit einhergehende Kritik an den einzelnen Verfahren stützen diese Einschätzung.

Eine weit größere Herausforderung scheint die Beurteilung, ob ein Computersystem intelligent ist. Der britische Mathematiker Alan Turing hat 1950 ein Verfahren – den Turing-Test – vorgeschlagen (Turing 1950: 433f.), um zu entscheiden, ob ein System sich intelligent verhält bzw. ob eine Maschine ein dem Menschen gleichwertiges Denkvermögen hat. Der Testablauf kann im Wesentlichen wie folgt beschrieben werden: Eine menschliche Fragesteller:in führt über eine Tastatur und einen Bildschirm, also ohne Sicht- und Hörkontakt, eine Unterhaltung mit zwei ihm unbekanntem Gesprächspartner:innen. Ein(e) Gesprächspartner:in ist ein Mensch, die andere eine Maschine. Kann die Fragesteller:innach intensiver Befragung nicht entscheiden, welcher der Gesprächspartner:innen die Maschine ist, hat die Maschine den Turing-Test bestanden. Es wird der Maschine dann ein dem Menschen gleichrangiges Denkvermögen unterstellt. Turing prophezeite in seiner Schrift von 1950, dass es in 50 Jahren nicht mehr als 70 Prozent der Fragesteller:innen möglich sei, zwischen Programm und Mensch zu unterscheiden. Aus der Interpretation dieses Postulats von Turing resultiert die mutmaßliche 30-Prozent-Grenze, nach der der Turing-Test als bestanden gilt, wenn mindestens 30 % der menschlichen Teilnehmer:innen getäuscht werden können. Der klassische Turing-Test und die bisherigen Versuche, ihn zu bestehen, lässt allerdings vermuten, dass dieses Verfahren eher das Potenzial hat, erfolgreiches Täuschen zu bewerten, er teste »eher menschliche Leichtgläubigkeit als echte künstliche Intelligenz« (Tegmark 2017: 138f.).

Legendär wurde auch das von Joseph Weizenbaum 1966 entwickelte Computerprogramm ELIZA. Der Informatikprofessor am Massachusetts Institute of Technology (MIT) wollte mit einem einfachen, regelbasierten System die Möglichkeiten der Kommunikation zwischen Mensch und Maschine über

natürliche Sprache demonstrieren. Er simulierte (aus verschiedenen Gründen der Vereinfachung und Generalisierung) einen Psychotherapeuten als Gesprächspartner. Zu Weizenbaums Überraschung war ein Großteil der Versuchspersonen überzeugt, dass ihr Gesprächspartner ein tatsächliches Verständnis für ihre Probleme entwickelt habe. Da das Programm aufgrund der zugrundeliegenden Regeln konstruiert wurde, um menschlich, intelligent und einfühlungsvermögend zu wirken – also zu täuschen –, weigerten sich viele Proband:innen nach dem Experiment zu akzeptieren, dass ihr Gesprächspartner lediglich ein Computersystem war. Weizenbaum, der ursprünglich zeigen wollte, dass ein echter, vertiefter Dialog zwischen Mensch und Maschine nicht möglich sei, sondern stets oberflächlich bliebe, war von diesem Ergebnis erschüttert (Schanze 2010). Das Programm löste statt einer kritischen Haltung einen gegenteiligen Effekt aus: Menschen begannen der Maschine Gefühle und Verständnis zuzuschreiben. Der so genannte ELIZA-Effekt war geboren. Weizenbaum schrieb dazu später: »I had not realized [...] that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people.« (Weizenbaum 1976: 7f.)

2008 traten bei einem Experiment an der University of Reading sechs Computerprogramme an, um den Turing-Test zu bestehen. Dem besten Programm gelang es, 25 Prozent der menschlichen Versuchsteilnehmer:innen zu täuschen. 2011 konnte die Software Cleverbot beim »Technique 2011« am indischen Institut IIT Guwahati sogar 59 Prozent der Teilnehmer:innen täuschen.

2014 titulierte die internationale Presse, dass der Chatbot Eugene Goostman auf einer Veranstaltung der Royal Society in London den Turing-Test bestanden hat. 33 Prozent der Jury-Mitglieder:innen konnten die Handlungen des Chatbots nicht von den Handlungen einer menschlichen Person unterscheiden. Ein denkwürdiger Meilenstein in der KI-Forschung galt damit als überwunden. Doch das Experiment warf Fragen auf. So wurde Kritik am Versuchsaufbau wie auch »kognitiven Kniffen und Programmiertricks«, die verwendet wurden, um die Prüfer:innen zu täuschen, laut (Kühl 2021).

Mit Blick auf das primäre Ziel zu täuschen, den grundsätzlichen Mängeln des Turing-Test zu begegnen und die bereits beschriebenen multiplen Dimensionen von Intelligenz zu berücksichtigen, wurden im Folgenden Konzepte entwickelt, den Turing-Test zu erweitern. Zu diesen gehören der Lovelace-Test (Bringsjorden et al. 2003) mit der Forderung, eine KI müsse »Kreativität beweisen und originäre Leistungen erbringen«. Beim Metzinger-Test (Met-

zinger 2001), erfüllt die KI den Intelligenzanspruch, wenn sie mit eigener Argumentation in die Diskussion um künstliches Bewusstsein eingreift und überzeugend für ihre eigene Theorie des (eigenen) Bewusstseins argumentieren könne. Bemerkenswert ist zudem, dass heute auch »umgekehrte« Turing-Tests eingesetzt werden. Mit den sogenannten CAPTCHAs (»Completely Automated Public Turing test to tell Computers and Humans Apart«) (von Ahn et al. 2003) werden Menschen im Internet bei der Nutzung von Eingabefeldern aufgefordert den Beleg zu erbringen, dass sie keine Maschine bzw. keine Künstliche Intelligenz sind.

Es scheint ein fundamentaler Denkfehler des Menschen zu sein, einen Computer wie einen Menschen zu betrachten und zu behandeln, ihn nach menschlichen Maßstäben messen zu wollen und ihm ein menschenähnliches Verständnis der Welt zuzuschreiben. Laut dem Philosophen John Searle kann die formale Struktur der Mathematik niemals den informellen Charakter menschlicher Gedanken erfassen, oder, einfacher ausgedrückt, menschliche Intelligenz kann grundsätzlich nicht durch ein Computerprogramm simuliert werden. Searle illustrierte diese Auffassung anhand eines Gedankenexperiments mit dem Namen »Das chinesische Zimmer« (Searle 1980). In diesem hypothetischen Experiment sitzt eine Person mit einigen Texten auf Chinesisch in einem geschlossenen Raum. Diese Person kann Chinesisch weder sprechen oder schreiben noch ist sie in der Lage, die chinesischen Schriftzeichen als solche zu verstehen. Durch einen Schlitz in der Wand werden der Person auf Papier Geschichten auf Chinesisch zugeschickt, ergänzend erhält sie Fragen zu der Geschichte, ebenfalls auf Chinesisch. Außerdem erhält diese Person ein »Handbuch« in ihrer Muttersprache. Diese Anleitung erlaubt es der Person anhand der erhaltenen Symbole, also der Geschichte und der Fragen, eine Antwort ebenfalls auf Chinesisch zu schreiben. Sie folgt hierbei aber ausschließlich den Anweisungen, also den Regeln aus der Anleitung und versteht die Antworten, die sie anschließend durch den Schlitz wieder nach draußen schiebt, nicht. Außerhalb des Raumes nimmt eine chinesische Muttersprachler:in die Antworten auf ihre Fragen zur Geschichte entgegen und kommt zwangsläufig zu dem Schluss, dass sich im Raum jemand befindet, der ebenfalls chinesisch spricht. Anhand dieses Aufbaus argumentiert Searle, dass ein Programm, das den Turing-Test besteht, nicht zwangsläufig auch intelligent sein muss, es ist nur scheinbar intelligent. Ein korrektes Ausführen von Anweisungen setzt daher weder Verständnis noch Bewusstsein voraus.

Siegeszug der schwachen KI

Was verleitet nun den Menschen dazu, Computer hinsichtlich unserer Wahrnehmung wie Menschen zu behandeln und Intelligenz in sie hineinzuprojizieren? Diese Frage ist auch interessant in Zusammenhang mit dem Wunsch des Menschen, ebenbürtige künstliche Systeme zu schaffen.

Historisch lässt sich ein Muster des Strebens nach künstlichen Kopien des Menschen erkennen. Sei es der Homunculus, der Golem, Frankenstein, später mechanische Automaten zur Mechanisierung des Menschen und seines Verhaltens bis hin zum Human Brain Project⁵, einem Forschungsprojekt der Europäischen Kommission, welches das gesamte verfügbare Wissen über das menschliche Gehirn zusammenfassen und mittels computerbasierter Modelle in Simulationen nachbilden will. Also eine ultimative Reproduktion und Imitation menschlichen Denkens.

Die Frage ist aber: Ist eine Abbildung menschlicher Denkweisen in einer Maschine überhaupt möglich? Ein Blick auf die Entstehungsgeschichte der KI zeigt, dass hierfür mindestens zwei Aspekte notwendig sind: eine formale Sprache (Denkweise/Logik/Algorithmus), um kognitive Prozesse zu beschreiben, und ein Medium (Rechenmaschine), um diesen Formalismus zu vollziehen. Gemäß dem Motto »Wer über die Zukunft reden will, sollte die Vergangenheit kennen« (Zweck 2021: 69), lohnt daher ein detaillierterer Blick auf die Historie der Künstlichen Intelligenz:

Meilensteine in der Entwicklung der Künstlichen Intelligenz

17. Jhrd.: Wilhelm Leibniz vertrat als erster die Auffassung, dass alles in der Welt, einschließlich Gehirn und Denken, den physikalischen Gesetzen folgt. Er baute als erster ein mechanisches Gerät, das geistige Operationen simulieren sollte.

1763: Das Bayes-Theorem von Thomas Bayes wird veröffentlicht; es handelt sich um eine Formel, die für das Schließen unter unsicheren Wissensbedingungen für die KI bedeutsam ist.

5 Human Brain Project, <https://www.humanbrainproject.eu/en/>, abgerufen am 20.03.2022

1769: Wolfgang von Kempelen formuliert die Idee einer intelligenten Maschine und entwickelt den »Schachtürken«, einen scheinbar, mechanischen Schachspielerroboter, in dessen Inneren jedoch ein menschlicher Schachspieler versteckt war.

19. Jhrd.: Philosophen und Mathematiker wie Boole oder Frege entwerfen mit der Logik einen formalen Apparat, mit dem sich auch kognitive Prozesse abbilden lassen.

1921: Das Deutsche Patentamt segnet eine »Vorrichtung zur Beantwortung von Fragen« ab; Carl Spierer entwarf darin ein durchdachtes Logiksystem für das was heute Künstliche Intelligenz heißt.

1931: Kurt Gödel zeigt die Grenzen der Logik (Berechenbarkeit) auf und weist gleichzeitig nach, dass jedes berechenbare Problem lösbar und damit das menschliche Gehirn mit Computern abbildbar ist. Es gibt jedoch fundamentale Grenzen dessen, was für Rechner entscheidbar ist.

1936: Alan Turing stellt sein universelles Rechenmodell vor, mit dem jedes berechenbare Problem lösbar ist. Die Turing-Maschine bietet eine Basis, auf der heutige Computer noch arbeiten.

1948: Norbert Wiener prägt mit seinem Werk »Cybernetics or Control and Communication in the Animal and the Machine« den Begriff der Kybernetik und schuf damit die mathematischen Grundlagen für die Kontrolltheorie und Regelungstechnik zur Berechnung der Dynamik und Stabilität von rückgekoppelten Systemen.

1950: Alan Turing veröffentlicht in der englischen Philosophiezeitschrift »Mind« den Artikel »Computing Machinery and Intelligence« (Turing 1950) und stellt die Frage, ob Computer denken können. In seinem Aufsatz beschrieb Turing eine Prüfung ihrer geistigen Fähigkeiten und entwickelt den beschriebenen Turing-Test.

1952: Arthur Samuel (IBM) entwickelt Computerprogramme, die Dame spielen. Im Zuge der Übertragung einer Partie im Fernsehen, stieg der Kurswert der IBM-Aktie um 15 Punkte.

1956: Auf der Dartmouth-Konferenz prägt J. McCarthy den bis heute verwendeten Begriff »Künstliche Intelligenz«.

1957: F. Rosenblatt entwickelt das Perzeptron, ein vereinfachtes künstliches neuronales Netzwerk⁶, das nach Training Muster erkennen kann.

1957: Noam Chomsky entwickelt eine Universalgrammatik, eine mathematische Beschreibungsform für die natürliche Sprache und postuliert, dass sich menschliche Sprache streng regelbasiert, wie ein Computerprogramm/Algorithmus verhält.

1960: Karl Steinbuch erfindet die Lernmatrix, ein frühes neuronales Netz mit analoger Elektronik.

1964: Joseph Weizenbaum programmiert den schon beschriebenen Chatbot Eliza.

1965: Edward Feigenbaum entwickelt das weltweit erste Expertensystem.

1969: Marvin Minsky weist nach, dass Perzeptrons keine nichtlinearen Zusammenhänge darstellen können und leitet damit den so genannten KI-Winter ein, eine Phase in der die Auseinandersetzung mit KI in einem quasi-Winterschlaf erstarre und kaum ernst zu nehmende Fortschritte erreicht werden.

1980: Fukushima schlägt mit dem Neocogitron das erste Convolutional Neural Network (CNN) vor. Die Struktur von CNNs ist von biologischen Prozessen inspiriert und gilt als Grundlage heutiger Deep Learning-Verfahren für maschinelles Lernen.

1984: David Rumelhart und James McClelland erzielen mit Backpropagation, einem Verfahren zur Fehlerrückführung bei künstlichen neuronalen Netzen, einen Durchbruch bei der Entwicklung von neuronalen Netzen.

1986: Geoffrey Hinton setzt Backpropagation zur Mustererkennung ein und begründet den Konnektionismus, also die Entwicklung und Analyse informationsverarbeitender Systeme, die wesentliche Eigenschaften kognitiver Prozesse nachbilden.

1987: Sepp Hochreiter entwickelt das LSTM (Long short-term memory). Ein Kurzzeitgedächtnis für neuronale Netze, das den Durchbruch der künstlichen Intelligenz, wie wir sie heute verstehen, erst ermöglichte.

1987: Yann LeCun setzt Backpropagation erfolgreich beim Training von Convolutional Networks ein.

2000: Jürgen Schmidhuber und Sepp Hochreiter verbessern RNN (»Rekurrente neuronale Netzwerke«) mit Hilfe des LSTM und bezeichnen diese RNNs als Deep Learning-Netzwerke⁷.

2014: Yoshua Bengio et al. stellen das Konzept der Generativen Adversarialen Netzwerke (Goodfellow et al. 2014) vor – die Methode wird auch als Turing-Lernen bezeichnet.

Der Blick auf diese Historie wesentlicher Meilensteine zeigt, dass die Entwicklung der Künstlichen Intelligenz mehr als nur eine technische Realisierung informationsverarbeitender Maschinen ist. Es geht um ein Zusammenspiel von Mathematik/Logik, Sprache und Philosophie. Die Philosophie leistete und leistet zur KI-Entwicklung insofern einen wichtigen Beitrag, als sie die Auffassung vertritt, dass Denken etwas Maschinelles ist, das auf Wissen basiert. Wissen, das in einer internen Sprache repräsentiert ist und das Denkprozesse ermöglicht, die gewünschte Handlung zu vollziehen. Mathematik und Logik liefern Werkzeuge zur Handhabung von sicheren wie von unsicheren, wahrscheinlichkeitstheoretischen Aussagen. Sie schufen erst Grundlagen

6 Künstliche neuronale Netzwerke bilden, entsprechend ihrem biologischen Vorbild, Netze aus künstlichen Neuronen nach. Sie stellen, als Forschungsgegenstand der Neuroinformatik, einen Zweig der künstlichen Intelligenz dar.

7 Deep Learning ist eine Methode zur Informationsverarbeitung und verwendet künstliche neuronale Netze (KNN) die mehrere Zwischenschichten zwischen Eingabe- und Ausgabeschicht einsetzen.

für das Schlussfolgern mit Hilfe von Algorithmen, d.h. regelbasierten Abläufen. Aus der Psychologie kommt die Vorstellung, dass Menschen und Tiere als informationsverarbeitende Maschinen betrachtet werden können – eine Betrachtung, die auch von der Linguistik geteilt wird. Elektronik und vor allem die Computertechnik lieferten die Hardwaregrundlagen für KI-Anwendungen und seit einigen Dekaden die notwendigen Infrastrukturen, um die hinsichtlich Speicher und Rechenzeit immer größer werdenden KI-Programme ausführen zu können. Die historische Entwicklung zeigt die Fortschritte in der KI-Forschung und scheint auf den ersten Blick kontinuierlich zu verlaufen. Tatsächlich verlief sie eher wellenartig in Form von Fortschritten und Rückschlägen. So waren die Jahre 1950–1970 geprägt von Aufbruchstimmung und Euphorie. Die Jahre 1965–1975 waren zunehmend von Rückschlägen und ungelösten Herausforderungen mit der konkreten Realisierung einer KI geprägt. In der Zeit von 1980–1990 entstanden völlig neuartige Zugänge, Anwendungen und KI-Industrien. Mit der Verfügbarkeit günstigerer Speicher und Rechenzeiten erfuhren die neuronalen Netze seit 1985 eine Renaissance bis zur Gegenwart. Fortschritte in jüngster Zeit ermöglichen ein besseres Verständnis der theoretischen Grundlagen von Intelligenz und brachten verbesserte Fähigkeiten realer Systeme mit sich. Heute ergeben sich daraus zwei grundsätzlich unterschiedliche Perspektiven bezüglich des Verständnisses und der Zielsetzung für KI. Die eine Strategie versucht zu verstehen, wie das menschliche bzw. das biologische Denken funktionieren und diese Einsichten nach bionischen Prinzipien abzubilden. Die andere stellt die Frage, ob von einer grundlegenden Neuformulierung des KI-Konzeptes ausgegangen werden muss, die bezogen auf das Verständnis von Intelligenz einen gänzlich neuen, eigenständigen Weg geht.

Der vorsichtige Versuch, die bisherige Historie in die Zukunft fortzuschreiben, lässt absehbar längerfristig weitere Fortschritte erwarten. Doch bereits gegenwärtig sind KI-Systeme – insbesondere künstliche neuronale Netze – zu außergewöhnlichen (Kreativ-)Leistungen fähig. Das gilt, obwohl heutige Künstliche Intelligenz lediglich den Kriterien einer sogenannten »schwachen KI« entsprechen. »Schwache KIs« besitzen per definitionem keine Kreativität nach menschlichem Verständnis und keine explizite Fähigkeit, autonom im universellen Sinne zu lernen. Die Lernfähigkeiten solcher Systeme sind in der Regel auf das Trainieren von Erkennungsmustern oder das Abgleichen und Durchsuchen großer Datenmengen reduziert. Sie bewältigen komplexe, aber klar definierte Aufgabentypen, sie folgen einer festgelegten Methodik und sind in der Lage, wiederkehrende und genau

spezifizierte Probleme zu lösen. Zu den Einsatzbereichen schwacher KI zählen die Automatisierung von Prozessen aber auch Spracherkennung und -verarbeitung. Weitere Beispiele sind Text- und Bilderkennung, Spracherkennung, Übersetzungen, Navigationssysteme etc., aber auch verbreitete digitale Assistenzsysteme wie Alexa, Siri oder Google Assistant.

Im Gegensatz zur spezifisch problemlösenden Intelligenz schwacher KIs kann eine starke KI Aufgabenstellungen selbstständig erkennen und definieren und hierfür »eigenständig« Wissen aus der entsprechenden Anwendungsdomäne zusammentragen. Sie untersucht und analysiert Probleme, um adäquate neue oder kreative Lösungen für Herausforderungen zu finden. Obwohl heutige KI-Systeme den Anschein erwecken, bereits in die Domäne der starken KI vorgedrungen zu sein, ist ihre Realisierung keineswegs in greifbarer Nähe und noch Gegenstand der Forschung.⁸

Frei nach Arthur C. Clarke formuliert: »Jede hinreichend fortschrittliche Technologie ist von Zauberei nicht zu unterscheiden.« (Clarke 1962: 14) zeigt die »Magie« heutiger, spezifisch problemlösender Intelligenzen im Sinne schwacher KIs aber, wie schnell ihnen nicht nur Kreativleistungen, sondern unberechtigterweise und verfrüht »echte Intelligenzleistungen« zugesprochen werden. Dies geschieht insbesondere, wenn sie menschliche Fähigkeiten in bestimmten Aufgabenstellungen übertrumpfen. Sie fallen dann auch mehr und mehr aus dem Bereich bio-physikalischer Regularitäten heraus, den White ja auch explizit aus sozialwissenschaftlicher Analyse ausschließen möchte.

In den letzten 50er Jahren des vorherigen Jahrhunderts besiegten Computerprogramme menschliche Spieler:innen im Brettspiel Dame. Nicht einmal 50 Jahre danach, 1977, wurde der amtierende Schachweltmeister Garri Kasparov vom IBM-Supercomputer DeepBlue im Schach geschlagen. Die verwendete Technik – ein speziell für diesen Zweck gebautes »Rechenmonster«, dass in jeder Sekunde des Spiels durchschnittlich 126 Millionen Stellungen durchrechnet – besiegte den Menschen in einem Spiel, dem von je her nachgesagt⁹

8 Generalisiert anwendbare KI: Hype und Stand der Forschung, science media center germany, URL: <https://www.sciencemediacenter.de/alle-angebote/science-responses/details/news/generalisiert-anwendbare-ki-hype-und-stand-der-forschung/>, abgerufen am 04.12.2021.

9 The mathematical complexity of Go, <https://senseis.xmp.net/?ComplexityOfGo>, abgerufen am 04.12.2021.

wurde, nur mit logischem Denken, Intelligenz, Wagemut und strategischem Scharfsinn beherrschbar zu sein.

Wiederum 20 Jahre später besiegte das auf maschinellem Lernen basierende Computerprogramm AlphaGo den amtierenden Go-Weltmeister in einem Spiel, das als das älteste und komplexeste Brettspiel der Welt gilt. Go ermöglicht mehr Züge als Atome im bekannten Universum existieren und mit traditionellen Strategien/Algorithmen (wie sie noch DeepBlue verwendet hatte) wäre kein Computer in der Lage gewesen, dem Menschen auch nur annähernd ebenbürtig zu sein. AlphaGo siegte mit völlig neuen Spielweisen und Lösungsstrategien, die im Nachhinein als kreativ empfunden wurden. In einem berühmten Spiel gegen den möglicherweise besten Go-Spieler aller Zeiten (den Südkoreaner Lee Sedol) verwendete AlphaGo einen Spielzug, den menschliche Beobachter:innen zunächst als Fehler abtaten (Menick 2016), der sich jedoch im weiteren Spielverlauf als für den Sieg entscheidend entpuppte. Der später als »Zug 37« bekannt gewordene Zug, hat viele Expert:innen überrascht, neben der internationalen Go-Gemeinschaft auch die Erschaffer:innen von AlphaGo selbst. Dieser Spielzug lag so weit außerhalb dessen, was seit über tausend Jahren gewohnte Wege waren, Go zu spielen, dass er zum Inbegriff von Kreativität in der KI wurde. Zahlreiche wissenschaftliche Abhandlungen untersuchen seitdem diesen Spielzug und das spezifische Verhalten von AlphaGo in der damaligen Spielsituation.¹⁰

2017, zwei Jahre nach AlphaGo, stellte das Unternehmen DeepMind den Nachfolger von AlphaGo, AlphaZero vor. War AlphaGo noch auf das Spielen des Brettspiels Go trainiert (zur Erinnerung, schwache KIs sind in der Regel auf eine dedizierte Problemklasse trainiert), ist AlphaZero ein autodidaktisches Computerprogramm, dessen Algorithmus mehrere komplexe Brettspiele einzig anhand der Spielregeln und Siegbedingungen sowie durch intensives Spielen gegen sich selbst erlernt. AlphaZero ist in der Lage, auch andere Strategiespiele wie Schach oder Shogi zu beherrschen und wurde mit Methoden des sogenannten *Reinforcement Learning* (kurz RL und zu Deutsch: bestärkendes Lernen oder verstärkendes Lernen) entwickelt. Es handelt sich um einen Ansatz des maschinellen Lernens, bei dem die KI eigenständig eine Strategie erlernt, um »Belohnungen« aus einer programmierten Belohnungsfunktion zu maximieren. Anders als beim überwachten Lernen (Supervised Learning) sind beim RL im Vorfeld keine Daten erforderlich. Stattdessen erfolgt die

10 AlphaGo's Games, <https://www.alphago-games.com/view/eventname/leesedol/game/1/move/37>, abgerufen am 09.01.2022.

Bildung der Datenbasis durch ausführliche Trial-and-Error-Abläufe innerhalb eines eigens angelegten Simulations-Szenarios. Die Trainingsdauer, um AlphaZero das Schachspielen beizubringen, betrug nur wenige Stunden – auf das komplexe Brettspiel Go wurde das System mit weniger als 35 Stunden Rechenzeit trainiert.

Die Reaktionen verschiedener Schachweltmeister:innen war entsprechend überschwänglich. Schachweltmeister Garry Kasparow kommentierte: »It's a remarkable achievement, even if we should have expected it after AlphaGo. We have always assumed that chess required too much empirical knowledge for a machine to play so well from scratch, with no human knowledge added at all.« (Kasparov 2017)

Andere Großmeister:innen beschrieben die ungewohnte Spielweise als »insane«, »bahnbrechend« oder »phänomenal«. Professor Wattenhofer von der ETH Zürich kommentierte die Leistungen mit:

»AlphaZero spielt sehr ungewöhnlich, nicht wie ein Mensch, aber auch nicht wie ein typischer Computer, sondern mit »echter künstlicher« Intelligenz. [...] AlphaZero's Leistung zeigt, dass wir die Künstliche Intelligenz ernst nehmen müssen. Sie beweist, dass wir Zeitzeugen eines gewaltigen Wandels sind. [...] AlphaZero & Co. werden die Gesellschaft nachhaltig verändern.« (Wattenhofer 2018)

Ähnlich erstaunliche Leistungen können beim Einsatz von KI im Bereich der pharmazeutischen Wirkstoffforschung beobachtet werden. Zu den fundamentalsten Herausforderungen der Molekularbiologie zählt das Proteinfaltungsproblem. Proteine bestehen aus Hunderten von Aminosäuren, die sich zu komplexen Strukturen zusammenfalten. Welche Aminosäuren zu einem Protein gehören, wird zwar in der DNS codiert, aber diese Codierung gibt keinen Aufschluss über die räumliche Struktur, in die sich die Aminosäurekette zu einem Protein mit spezifischer Form faltet. Diese 3D-Struktur des Moleküls ist jedoch entscheidend für seine Funktion im Organismus. Sofern bekannt, wird die Funktionsweise des Proteins nachvollziehbar. Die Frage mit welchen Wirkstoffen (Medikamenten) das Protein ggf. beeinflussbar ist, ist dann deutlich leichter zu beantworten. Ein bloßes Ausprobieren dauert bei der Strukturvorhersage von Proteinen sehr lange, obwohl sich Proteine in der Natur innerhalb weniger Minuten formieren. Im Dezember 2020 überraschte DeepMind – eine KI-Tochter von Google – die Fachwelt der Biologie, als sie mit AlphaFold, einem KI-Tool zur Vorhersage von Proteinstrukturen,

den schon seit 50 Jahren laufenden wissenschaftlichen Wettbewerb CASP (»Critical Assessment of Techniques for Protein Structure Prediction«) gewann. AlphaFold schnitt bei dem alle zwei Jahre stattfindenden Event so gut ab, dass der Mitbegründer des Wettbewerbs erklärte, dass »[...] in gewisser Weise das Problem als gelöst betrachten werden kann.« (Callaway 2020). Das Fachmagazin Science schätzte den neuen Ansatz für so bedeutsam ein, dass es ihn zum wissenschaftlichen Durchbruch des Jahres kürte (Service 2021). Eine Künstliche Intelligenz erzielt mit enorm hoher Geschwindigkeit und Genauigkeit einen wissenschaftlichen Durchbruch und revolutioniert die Wirkstoffforschung. Ein Erfolg jedoch ohne die Mechanismen geklärt zu haben, nach denen die KI dies bewerkstelligt, geschweige denn die Natur dieses komplexe Problem in nur wenigen Minuten lösen kann. So lässt sich vermuten, dass wir eine Wissenschaftsvorstellung ohne Theorie erleben und künftig möglicherweise auf ein massives Theoriedefizit zusteuern. Die Nutzer:inerkennt, dass die Dinge funktionieren, verliert aber die Fähigkeit zu erklären, warum dies so ist. Wissenschaft produziert hier also weniger eine Entzauberung (Weber 1919) als vielmehr eine erneute Verzauberung der Welt – um erneut auf Clarke zu rekurrieren und den Kreis zu schließen: »Jede hinreichend fortschrittliche Technologie ist von Zauberei nicht zu unterscheiden.« (Clarke 1962) Unklarheiten wie diese Ergebnisse beim Einsatz von KI zu Stande kommen, stellen eine neue Dimension des Umgangs mit wissenschaftlichen Verfahren dar und erfordern eine intensivere Auseinandersetzung mit den daraus sich ergebenden Fragen, insbesondere in Bezug auf unser Wissenschaftsverständnis.

3. KI in der Wissenschaft

KI-Anwendungen sind im privaten wie industriellen Umfeld bereits ubiquitär im Einsatz. Auch in Wissenschaft und Forschung finden sich zahlreiche Einsatzmöglichkeiten, bei denen KI als Werkzeug etabliert ist, wie

- Mustererkennung und Simulationen in der Klimaforschung
- Data Mining, wie z.B. beschrieben in »Big Data für die Sozialforschung« (Scheiermann/Zweck 2014)
- Wirkstoffforschung und Medikamentenentwicklung in der Pharmazie
- Prognosen und Simulationen in der Epidemiologie

- Theorieentwicklung in Mathematik und Teilchenphysik unter anderem mit Hilfe automatisierter Theorembeweiser.

Der gesteigerte Einsatz von KI lässt neue Möglichkeiten für die Forschung erwarten. So können bestimmte Forschungsaktivitäten, die bislang als nicht praktikabel galten, erst mit Hilfe von KI effizient durchgeführt werden. Hierzu zählt die Analyse großer Textmengen oder (Forschungs-)Datenbanken, um per Extraktion und Korrelation aus Daten neue Erkenntnisse zu gewinnen. Außerdem erlaubt der Einsatz von KI Analysen, die bislang grundsätzlich nicht möglich waren. Die bereits beschriebene Proteinfaltung kann hier ebenso als Beispiel dienen, wie die individualisierte Genomanalyse. KI unterstützt bei Modellbildung und Hypothesenfindung, sie deckt Muster auf, die über bisher verborgene kausale Erklärungen zur Erkenntnisbildung beisteuern.

Wie verändert sich das Wissenschaftsverständnis durch KI- und Machine Learning-Verfahren? Steht eine methodologische Revolution bevor, oder wird die bisherige Kontinuität im Wissenschaftsbetrieb – nur mit modifizierten Instrumenten – fortgesetzt?

Neue datengetriebene Verfahren dienen hier quasi als Katalysatoren. KI verändert die Wissensgenese und zugleich das, was Wissenschaft ist. Aktuell zeichnet sich für das künftige Wissenschaftsverständnis eine Verschiebung des wissenschaftlichen Arbeitsstils (im Sinne Whites als einer Veränderung von stochastischen Verteilungen hin zu neuen Dominanzen mit erkennbarer Signalwirkung, vgl. Schmitt/Fuhse 2015) weg von einer Erklärungsorientierung hin zu einer pragmatischen Fokussierung auf die reine Erkenntnisgewinnung ab. Und zwar hin zu Nutzen und Anwendbarkeit sowohl für die Forschung als auch für die Gesellschaft. Das Beispiel AlphaFold zeigt, dass das Leistungsmerkmal der Prognose (hier die Vorhersage der räumlichen Struktur eines Eiweißes auf der Basis bekannter Aminosäuresequenzen) in den Mittelpunkt gestellt wird, indem bei KI- und Machine Learning-Verfahren (verborgene) schwache Kausalitäten oder Regelmäßigkeiten und starke Korrelationen zugelassen werden, um diese für praktische Zwecke nutzbar zu machen. KI- und Machine Learning-basierte Prognoseverfahren basieren auf statistischen bzw. probabilistischen Prognosen, die auf Basis einer Datenanalyse vorgenommen werden. Sie arbeiten auf einer Datenbasis, ohne deren kausale Regelmäßigkeiten zu kennen, sie »wetten« auf Kausalität, die in den Mustern der Daten verborgen sind.

Traditionell zielt Wissenschaft darauf ab, Phänomene durch Formulierung fundamentaler Theorien, Modelle oder Gesetze zu erklären und deduktiv ab-

zuleiten. Vorrangiges Ziel angewandter Forschung ist es, Anwendungsoptionen zu realisieren, wobei Erklärungen ein Mittel hierfür sein können. KI zeigt in ihren Möglichkeiten eine Nähe zu den Zielen angewandter Wissenschaften.

2008 schrieb Chris Anderson unter dem Titel »The End of Theory: The Data Deluge Makes the Scientific Method Obsolete«, dass vor allem Korrelationen ein hinreichendes Kriterium seien, während Kausalität nicht als notwendig gesehen werden müsse. Das ist insofern interessant, als dass in vielen anwendungsrelevanten KI-Anwendungen wenig über Ursache-Wirkungs-Zusammenhänge bekannt ist – man hat es mit einer »Black Box« zu tun (Heckwolf/Schmitt 2022). Für die Medizin liegt hier eine große Herausforderung, denn für Diagnose und Therapie vieler Krankheiten ist solides, konsistentes Kausalwissen über Ursache-Wirkungs-Zusammenhänge elementar, um die Wahl der Therapie kausal belegen zu können (Häußling et al. 2021). Gleichzeitig wohnt solchen Forschungsfeldern eine natürliche Komplexität inne, die traditionelle, lineare Ursache-Wirkungsprämissen kaum mehr als analytisches Kriterium zulassen. Vor diesem Hintergrund können KI- und Machine Learning-Verfahren als methodische Ansätze zur Bewältigung von Komplexitätsproblemen betrachtet werden. Sie eröffnen so einen (wie auch immer gearteten) Zugang zu überkomplexen Objektsystemen. Es liegt in der Natur komplexer nicht-linearer dynamischer Systeme, dass für sie aufgrund von Nichtlinearitäten, Instabilitäten und Bifurkationspunkten keine weitreichenden (theorie- und gesetzesbasierten) Prognosen möglich sind.

Nun ist es wesentliches Ziel von KI-Systemen, Formen der Regelmäßigkeit, also verborgene Muster oder Abhängigkeiten in komplexen (vermeintlich regelschwachen) Objektsystemen zu finden, und auch schwache Kausalitäten zu erkennen. Ziel ist dabei prognostischen Nutzen zu ziehen, ohne die zugrundeliegenden deterministischen oder probabilistischen Gesetze zu kennen oder verstanden zu haben.

Wird KI als methodischer Ansatz verstanden, um den Umgang mit komplexen Systemen zu ermöglichen, stellt sich die Frage, wie transparent nachvollziehbar und wie reproduzierbar das Verhalten von KI-Systemen selbst ist. Der Philosoph Paul Humphreys beschrieb 2007 in seinem Vortrag »Methodological Novelties in Computer Simulations« (Humphreys 2007) eine tiefgreifende Veränderung in der Bedeutung von »Techniken« im Wissenschaftseinsatz. Computermethoden bedeuten eine grundlegende (arbeits-)philosophische Neuerung in folgendem Sinne: Sie schieben den Menschen aus dem Zentrum des Forschungsprozesses der Erkenntnisgewinnung heraus, allerdings ohne, dass er vollständig verdrängt wird.

Zugleich führe der zunehmende Einsatz computerbasierter Verfahren, von Algorithmen und Simulationen naturgemäß zu einer Intransparenz. Die Forscher:in sei, nach Humphreys, kognitiv nicht mehr in der Lage, zu durchdringen, was in Simulationen geschehe und wie sich die Ergebnisse aus den Daten ableiten. Einzeloperationen des Algorithmus' ließen sich noch nachvollziehen, aber das Verhalten eines komplexen Modells in seiner Gesamtheit entzieht sich dem Verständnis des Forschenden. Eine Erklärung dessen, was innerhalb der »Black Box« von Computermodellen vor sich gehe, sei kaum mehr möglich, womit zwangsläufig Probleme der Validierung entstehen (Knight 2017). Richtiges oder falsches Verhalten dieser Modelle können allenfalls beobachtet, aber kaum bewiesen werden. Es entsteht das beunruhigende Gefühl von unkontrollierbarer Handlungsautonomie bei solchen Systemen. Dieses Problem verschärft sich bei der Verwendung tiefer neuronaler Netze. Bei ihnen kann von außen nur noch mit großem Aufwand reproduziert werden, auf welche Muster das System reagiert und welche Merkmals-Korrelationen gezogen werden. Kommt bei den Eingangsdaten ein geringes Rauschen hinzu, führt dies zu strukturellen Instabilitäten des Modells. Kleinste Störungen oder Artefakte in den Eingangsdaten können zu vollständig anderen Resultaten führen. Das hat in der Praxis weitreichende Konsequenzen: Bei Bilderkennungssystemen autonomer Fahrzeuge wird aus einem Luftballon ein Stopp-Schild oder Geschwindigkeitsschilder werden bei überlagerten Artefakten nicht als solche erkannt. Der Ruf nach sich erklärenden Systemen »Explainable AIs« (Holzinger 2018: 138), die offenlegen können wie Entscheidungen z.B. in neuronalen Netzen entstanden sind, werden lauter (Beuth 2017). Dies gilt insbesondere, wenn sich durch diese KI-Entscheidungen direkte Konsequenzen für den Menschen ergeben. Daraus ergibt sich mit Bezug auf White die Frage nach der Neuverteilung von Kontrollbeziehungen. Wer kontrolliert hier wie durch welche Verfahren und ergeben sich aus den Kontrollschleifen neue Positionierungen in den Netzwerken, die mit dem Autofahren oder medizinischen Entscheidungen vergleichbar sind.

Doch die Forderung nach Transparenz stößt an Grenzen, wenn die Erklärung des Entscheidungsprozesses ihrerseits nur noch durch einen Computer überprüft werden kann. Ein rekursives Paradoxon: Um mit den neuen komplexen Herausforderungen in der Wissenschaft umgehen zu können, werden Verfahren und Instrumente mit ebenfalls hinreichender Komplexität benötigt. Komplexität soll durch Komplexität beherrschbar gemacht werden. Die modernen Instrumente, die neue Technik der KI-Verfahren, bekommen so einen »Eigensinn«, ein eigentümliches, scheinbar magisches Verhalten (vgl. im vor-

hergehenden Kapitel das Verhalten von AlphaGo). In der Mathematik und der theoretischen Physik haben sich Machine Learning-Verfahren als Beweisverfahren bereits etabliert. So konnte das »Vier-Farben-Theorem«, für das bisher kein analytischer Beweis gefunden wurde, mit KI-Hilfe zumindest ein numerisch-algorithmischer Beweis gefunden werden. In den 1970er Jahren proklamierte der Mathematiker und Träger der Fields-Medaille, Paul Cohen, dass »[...] zu einem unbestimmten Zeitpunkt in der Zukunft Mathematiker durch Computer ersetzt werden«. Cohen war davon überzeugt, dass sich die gesamte Mathematik aufgrund ihrer formalisierten Natur automatisieren ließe, einschließlich des Führens von Beweisen.

Mit diesem Wandel wissenschaftlicher Instrumente verändert sich auch die Form der Zusammenarbeit der Forscher:in mit ihrem Analyseinstrumentarium und damit auch der wissenschaftliche Erkenntnisprozess selbst. Es zeichnen sich neuartige Herangehensweisen, unkonventionelle Lösungen außerhalb etablierter Denkmuster ab. Es entstehen neue Forschungskonzepte, Denkweisen und Denkstile (Fleck 1980). Vor dem Hintergrund der beispielhaft beschriebenen Anwendungsfälle von KI-Systemen, wird diesem neuen Techniktyp sowohl in der Gesellschaft als auch in der Wissenschaft ein Quasi-Akteursstatus zuerkannt. Dadurch zeichnen sich grundlegende Verschiebungen in Forschungspraktiken, wie auch beim Selbstverständnis von Erkenntnisprozessen, ab. In ihnen entstehen bei der wissenschaftlichen Wissensproduktion neue und möglicherweise weitreichende Interaktionsformen zwischen menschlichen und nicht-menschlichen Akteur:innen.

4. Die Rolle der KI in der Zukunftsforschung

Es ist offensichtlich, dass KI die Experimentieroptionen und Handlungsmöglichkeiten der Forscher:in erweitern. Ihr Einsatz führt schon heute zu einer veränderten Rolle der wissenschaftlichen Wissensarbeiter:innen und der Forscher:innen. Sie verändern auch das prognoseorientierte Wissenschaftsverständnis – und zwar zunehmend in Richtung einer Prognostik ohne Gegenstandswissen. Es scheint gar nicht mehr notwendig zu sein, Wissen zu erwerben. Prognosen scheinen auch ohne Gegenstandswissen (theoriefrei) möglich, Daten sind hinreichend. Big Data (Scheiermann/Zweck 2014; Bosse

et al. 2018) und die Datengetriebene Prognostik¹¹ sind auch Bestandteil der Arbeit der Zukunftsforscher:in geworden. Mit ausreichend großen und qualitativ hochwertigen Daten scheint Realität abbildbar zu werden. Mit einem dazu passenden geeigneten Daten- und Prognosemodell werden Vorhersagen möglich, wie die automatisierten Prognosen von Maschinenausfällen, Prozessabweichungen oder sogar das Verhalten von Menschen, Märkten und Gesellschaften. Dies erinnert an Isaac Asimovs Psychohistorik (Nevala-Lee 2018), einem fiktionalen, hypothetischen Konzept, nach dem auf Basis mathematischer, statistischer und soziologischer Verfahren allgemeine und dennoch präzise Voraussagen über das zukünftige Verhalten großer Gruppen von Menschen gemacht werden können. Wie ein ganzes Land mit Hilfe einer computergestützten Planwirtschaft gesteuert werden sollte, wurde bereits in den 1970er Jahren untersucht. Unter der Regierung von Salvador Allendes unternahm Chile den ehrgeizigen Versuch, mit dem Projekt Cybersyn («cybernetic synergy») die Verwaltungswirtschaft des Landes in Echtzeit durch einen zentralen Computer zu kontrollieren. Die verstaatlichten Fabriken und Unternehmen sollten sich selbst steuern, wie Zellen und Organe eines Lebewesens, indem Lagerstände, Absatzzahlen und Produktionsmengen im ganzen Land in Echtzeit erfasst werden. So sollte zentral mit Algorithmen berechnet werden, wo Knappheiten drohen und wie Ressourcen am effektivsten verteilt werden können. Nach den Prinzipien des Konnektionismus und der Kybernetik, also der Wissenschaft von der Steuerung und Regelung von Systemen, sollte sich das Wirtschaftssystem Chiles selbst im Gleichgewicht halten. Ziel dieses technologisch-politischen Projekts war eine gerechtere Gesellschaft auf Basis einer rechnergestützten Wirtschaft (Beer 1973). Das Projekt konnte erste Erfolge nachweisen, wurde jedoch 1973 im Zuge eines Militärputsches zerstört.

Heute, ein halbes Jahrhundert später, erfahren Konzepte wie Kybernetik, System Dynamics und einige der ursprünglichen Ideen hinter Projekt Cybersyn durch Künstliche Intelligenz, dem Machine Learning und den Möglichkeiten von Big Data eine Renaissance. Methoden wie »Predictive Analytics« haben aktuell Hochkonjunktur (Zweck/Braun 2021). Die prädiktive Analyse wird dabei nicht nur eingesetzt, um künftige Bedarfe zu prognostizieren und frühzeitig die zielgerichtete industrielle Produktion zu planen, sondern auch

11 Prognose zum Volumen der jährlich generierten digitalen Datenmenge weltweit in den Jahren 2018 und 2025, <https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/>, abgerufen am 20.03.2022.

als Methode zur »(Voraus-)Berechnung« einer mehr oder weniger einfachen, linearen Zukunft mit geringem Zeithorizont. In »Predictive Analytics: Sind Zukunftsforscher:innen ein Auslaufmodell?« (ebd.) zeigen die Autoren, inwieweit KI bereits heute Zukunftsforschende unterstützen kann und plädieren dabei auch für eine stärkere Vernetzung der Zukunftsforschung mit anderen Disziplinen, die sich schon länger mit Big Data Analysen beschäftigen. Zugleich bezweifeln sie ausdrücklich, dass wir über Big Data und Predictive Analytics die Zukunft auch nur annähernd im Griff haben. Die zugrundeliegenden Analyseverfahren und Algorithmen aus Machine Learning und Statistik prognostizieren ein Bild der Zukunft, das auf Daten basiert, die die Vergangenheit abbilden oder im Idealfall sogar auf Echtzeitdaten der Gegenwart beruht. Vorhersagen auf Basis vergangener Ereignisse gelingen jedoch nur in stabilen und abgrenzbaren Umwelten einigermaßen gut: in Systemen, in denen keine komplexen Einflüsse oder gravierende interne wie externe Ereignisse eintreten, die das Gesamtsystem wesentlich beeinflussen können. Sie funktionieren also nur in den Fällen, in denen komplexe Wechselwirkungen gesellschaftlicher Dynamik wie auch unwahrscheinliche Ereignisse mit hoher Wirkungskraft (Wildcards) keinen Faktor darstellen. Extrapolationen und Prognosen gelingen umso schlechter, je dynamischer die betrachteten Systeme sind und je stärker exogene Ereignisse Einfluss wirken – wie beispielsweise der Ausbruch einer Pandemie. Im Frühjahr 2020, also im ersten Jahr der Corona-Pandemie, hatten viele Prognosealgorithmen (insbesondere aus dem Bereich der Predictive Analytics) erhebliche Probleme mit Strukturbrüchen in Folge des Corona-Lockdowns (Moorstedt 2020). Das bisher prognostizierte Konsumverhalten der Kund:innen in Supermärkten hatte sich grundlegend verändert und die massiv ansteigenden Verkaufszahlen bei Toilettenpapier und Teigwaren führten dazu, dass die Absatzprognosen nach einem Einbruch den Trend vollkommen überzeichneten. Extrapolationen und Prognosen sind also nur geeignet, um Aussagen über Systeme zu gewinnen, die sich wie in einem Laborexperiment unter klar definierten und oft künstlich geschaffenen experimentellen Konditionen abspielen. Letztlich geht dieses Konzept von einer Fortführung von Tendenzen aus, die sich aus den bisherigen Daten im Sinne »einer prognostizierten Zukunft« ableiten lässt.

Um die Zukunft in Situationen hoher Unsicherheit oder komplexen Umfeldern zu antizipieren, sind andere Methoden der Zukunftsforschung erforderlich, die keinen Anspruch von Prognosen im Sinne einer vorausberechneten Zukunft haben, sondern in einer Vielzahl möglicher Zukünfte denken. Es geht hier um Prospektionen, d.h. einer Auslotung möglicher

Zukunftsoptionen, um strategische Entscheidungen mit größerem Zeithorizont zu unterstützen. Dementsprechend wird hier ein Strauß von möglichen Entwicklungspfaden einer offenen Zukunft angeboten (Geschka/Hammer 1997). Dabei geht es weniger darum, einzelne Ereignisse mit maximaler Präzision vorherzusagen, wie bei den o.g. operativen Prozessen, als darum, strategische Handlungsalternativen anhand möglicher Zukunftsszenarien aufzuzeigen. Zu diesen Methoden gehört z.B. die Szenarioanalyse (Kosow/Gaßner 2008), eine der bekanntesten Methoden der Zukunftsforschung. Ziel der Szenariomethode ist es, mögliche Entwicklungen der Zukunft zu analysieren und zusammenhängend und konsistent darzustellen. Nach der Definition und Abgrenzung von Betrachtungsgegenstand und Zeithorizont werden zunächst Einflussfaktoren ermittelt, die das Thema charakterisieren. Durch Befragungen werden paarweise Wechselwirkungen bewertet, anschließend per Wechselwirkungsanalyse Schlüsselfaktoren ermittelt, zu denen dann, bezogen auf den betrachteten Zeithorizont, Projektionen durchgeführt werden. Die Projektionen werden über die Schlüsselfaktoren paarweise hinsichtlich ihrer Konsistenz bewertet. Rechnergestützt werden dann jene Projektionsbündel (Rohszenarien) berechnet, die eine ausreichend hohe Konsistenz aufweisen. Auf Basis der Rohszenarien werden dann alternative zukünftige Situationen (Szenarien) sowie Pfade, die zu diesen zukünftigen Situationen führen können, ausformuliert. Szenarien stellen evidenzbasierte, hypothetische Folgen von Ereignissen auf, um auf kausale Prozesse und Entscheidungsmomente aufmerksam zu machen. Szenarien stellen mögliche Zukunftsbilder einer offenen Zukunft dar und geben Hinweise darauf, welche Faktoren wie ausgeprägt sein müssen, damit ein bestimmtes Szenario möglich werden könnte. Neben der Darstellung, wie eine hypothetische Situation in der Zukunft zustande kommen kann, werden also ebenso mögliche Varianten und Alternativen dargestellt und aufgezeigt, welche Entscheidungs- und Handlungsoptionen es für verschiedene Akteur:innen gibt.

Durch das Einbeziehen von Wildcards, kann darüber hinaus das Eintreten auch unvorhergesehener oder unwahrscheinlicher Ereignisse mit hoher Wirkungsintensität auf das Gesamtsystem in die Szenario-Methode integriert und modelliert werden. Meistens werden in Verbindung mit Risikomanagementinitiativen Worst-Case-Szenarien wie Kriegausbrüche, Naturkatastrophen oder Pandemien herangezogen, um vorbereitende Strategien für solche Herausforderungen zu entwickeln. Unternehmen haben, wenn eines dieser Ereignisse von geringer Wahrscheinlichkeit aber großer Reichweite eintritt, bereits eine passende Strategie in der Schublade. Sie können sich so in Kri-

sensituation rasch von der taumelnden Masse im Markt abheben. Eine umfassende Vorbereitung wichtiger politischer Entscheidungen erfordert nicht nur Analysen auf Basis bestehenden Wissens, sondern auch eine Auseinandersetzung mit möglichen Zukunftsperspektiven.

Ein Gedankenspiel ist, dass bei der Prädiktion der Zukunft »nur« spezielle, isolierte Teilprobleme berechnet werden, die durch neuronale Netze gezielt trainiert werden können. So liefert die KI Watson von IBM Antworten auf Basis von Wahrscheinlichkeiten – Prädiktion per Wahrscheinlichkeit. Auch Klassifikationsalgorithmen (wie sie beispielsweise in Spam-Filtern verwendet werden) liefern Ergebnisse nach dem Muster: Antwort A ist mit 87 % Wahrscheinlichkeit wahr, Antwort B mit 9 %, Antwort C mit 2 % usw. Für viele Anwendungen ist das ausreichend. Hinsichtlich der kausalen Argumentation mutet dies unvollständig an und erscheint im Kontext evidenzbasierter Forschung unbefriedigend, denn die Forderung nach Nachvollziehbarkeit, Erklärbarkeit und Transparenz sind ihrerseits Gütekriterien seriöser Zukunftsforschung (Gerhold et al. 2015). Es existieren aber Problemklassen, deren Lösungen zwar »schwierig« (im mathematischen Sinne z.B. der NP-Vollständigkeit, Exponentialität usw.) zu finden, aber leicht (in konstanter Zeit) zu überprüfen sind. Zu diesen Problemen zählen beispielsweise bestimmte Optimierungsprobleme der Statik, Logistik, Proteinfaltung oder Existenzbeweise. Ähnlich verhält es sich mit bestimmten Methoden der Zukunftsforschung. Beispielsweise erklärt die Szenario-Methode selten (das ist oft auch gar nicht notwendig) die Ursachen und Gründe bzw. die historischen Entwicklungspfade, die zu einem zu betrachteten Szenario geführt haben. Es wird vielmehr ein Szenario als konsistentes Gesamtereignis innerhalb eines zukünftigen Möglichkeitsraumes formuliert. Der Wahrheitswert eines solchen Szenarios steht nicht zur Diskussion, solange die Konsistenz ausreichend und das Bild insgesamt schlüssig ist. KI als Akteurin der Prädiktion wäre hier keine Rechenschaft pflichtig, sondern präsentiert der Entscheider:in eigenständig ermittelte, konsistente Zukunftsszenarien, die betrachtet und berücksichtigt werden sollten. Wie dies funktionieren kann, wird in der Folge an einem Beispiel gezeigt.

Ansätze wie das Projekt Cassandra des Bundesverteidigungsministeriums und des Weltethos-Instituts zur Krisenfrüherkennung und Gewaltprävention zeigen¹², wie immer häufiger nicht nur algorithmenbasierte Modelle für

12 <https://weltethos-institut.org/forschung/cassandra-projekt/>, abgerufen am 20.03.2022.

Prognosen eingesetzt werden, sondern auch bereits existierende Weltbilder zielgerichtet für neue Erkenntnisse ausgewertet werden können. In diesem Fall wird bestehende Literatur ausgewertet, um auch das Unerwartete, wie Pandemien, Klimakrisen, Kriminalistik oder geopolitische Spannungsfelder zu durchdenken. Im Projekt Cassandra dient Literatur als besondere Form des Wissens. Die Literatur zeigt die verdeckten Seiten der Wirklichkeit und dient hier als Instrument für das »Worldbuilding« (Gavins 2016), für die Prospektion und für das, was eintreten könnte.

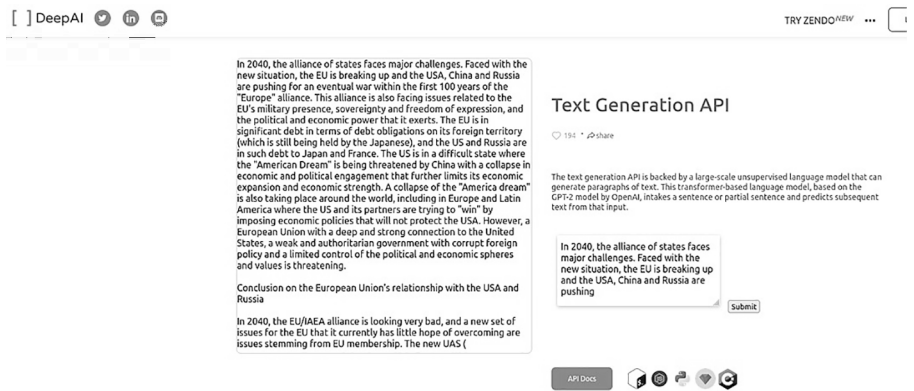
Wie Künstliche Intelligenzen bzw. Machine Learning-Verfahren im Detail als Instrumente in der Zukunftsforschung eingesetzt werden können, überstiege den Umfang dieses Textes und bedürfte einer umfassenderen Untersuchung, die an dieser Stelle nicht möglich ist. Dennoch kann an einem Beispiel gezeigt werden, wie Sprachmodelle als Instrument für eine fiktive Narration dienen können, indem sie auf bestehendem Weltwissen (wie z.B. Literatur oder wissenschaftlichen Texten) trainiert wird. Künstliche Intelligenzen verwenden in den meisten Fällen statistische Verfahren, die auf Trainingsdatensätze angewendet werden, um auf die in Daten enthaltenen Strukturen und Muster zu reagieren (auch wenn diese augenscheinlich versteckt sind). KI-Systeme werden trainiert – und zwar ausschließlich mit dem »Weltbild« der Trainingsdaten. Das ist zwar ein Bias, aber gleichzeitig kann diese Prägung auch Teil der Lernstrategie sein. Eine Art »narrative Prospektion« auf Basis von Sprachmodellen (Mayer 2021; Knight 2021) liegt nahe. So stellt das Trainieren von KI-Systemen in gewisser Weise die Schaffung eines Abbildes der Vergangenheit dar – die Gegenwart oder die Zukunft werden auf die Daten der Vergangenheit normiert. Ein KI-basiertes Instrument für Worldbuilding stellen schon heute umfangreiche Sprachmodelle (Heaven 2017) wie GPT-2, GPT-3 oder Aleph Alpha dar.

Auch derartige Sprachmodelle werden auf Grund ihres jeweiligen Trainings in ihrer Eigenschaft die Formalität der Sprache abzubilden nicht nur zu Instrumenten der Narration im Sinne eines erzählenden deskriptiven Charakters, sondern zu einem Werkzeug der Vermittlung dahinterstehender offen oder nicht offen gelegter impliziter Narrative. Die »narrative Prospektion« (Eggert/Zweck 2021: 424f.) hat in diesem Sinne also einen doppelzüngigen Charakter.

Wie Sprachmodelle und Textgeneratoren für die narrative Prospektion bzw. für das Szenario-Building eingesetzt werden können und zu welchen Leistungen diese Modelle heutzutage in der Lage sind, auch um neue, unkonventionelle Denkwege zu beschreiten, zeigt das folgende Beispiel: Es wird als

Grundannahme definiert, dass in einer alternativen, hypothetischen Zukunft im Jahr 2040 die EU gescheitert ist. Das Staatenbündnis bricht auseinander. Welches Szenario zeichnet ein KI-basiertes Sprachmodell für eine solche Zukunft? Der Textgenerator Deep AI von OpenAI, wird mit dieser hypothetischen Situation in der Zukunft »konfrontiert«. Mit welcher Geschichte und mit welchen Erklärungen »konfrontiert« uns die KI in der Folge? Abbildung 2 zeigt die Situation, wie sie auch die Leser:in ausprobieren kann.¹³

Abbildung 2: Texterstellung fiktionaler Texte mit Hilfe eines KI-gestützten Textgenerators



Der folgende Eingangstext wurde dem Textgenerator als Starttext gegeben, die KI vervollständigt die Geschichte dann selbstständig.

»In 2040, the alliance of states faces major challenges. Faced with the new situation, the EU is breaking up and the USA, China and Russia are pushing[...].«

Von der KI produzierte Geschichte oder in diesem Kontext, das produzierte Szenario:

»In 2040, the alliance of states faces major challenges. Faced with new situations, the EU is breaking up and the USA, China and Russia are pushing for

13 DeepAI, Textgenerator, <https://deepai.org/machine-learning-model/text-generator>, abgerufen am 20.03.2022.

an eventual war within the first 100 years of the ›Europe‹ alliance. This alliance is also facing issues related to the EU's military presence, sovereignty and freedom of expression, and the political and economic power that it exerts. The EU is in significant debt in terms of debt obligations on its foreign territory (which is still being held by the Japanese), and the US and Russia are in such debt to Japan and France. The US is in a difficult state where the ›American Dream‹ is being threatened by China with a collapse in economic and political engagement that further limits its economic expansion and economic strength. A collapse of the ›America dream‹ is also taking place around the world, including in Europe and Latin America where the US and its partners are trying to ›win‹ by imposing economic policies that will not protect the USA. However, a European Union with a deep and strong connection to the United States, a weak and authoritarian government with corrupt foreign policy and a limited control of the political and economic spheres and values is threatening.

Conclusion on the European Union's relationship with the USA and Russia: In 2040, the EU/IAEA alliance is looking very bad, and a new set of issues for the EU that it currently has little hope of overcoming are issues stemming from EU membership. The new UAS [...].«

Der Textgenerator beendet diesen spannenden Ausblick auf das Jahr 2040 mit einem wahren »Cliffhanger«, in Form eines nicht weiter erwähnten Staatenbündnisses namens »UAS«. Doch so »kreativ« oder »intelligent« KI-Modelle und insbesondere die modernen Sprachmodelle wie GPT-3 und Aleph Alpha auch wirken, am Ende sind sie doch nur »stochastische Papageien«, wie es amerikanische Forscherinnen genannt haben (Bender et al. 2021): »Sie verstehen nicht, sondern plappern nach, besser oder schlechter, je nachdem, wie man sie füttert«. Dennoch, neuronale Netze finden Lösungen zu Problemen, die Expert:innen überraschen und erstaunen. Sie gehen Wege, verwenden Strategien, und liefern Lösungen, formulieren approximierete Wahrheitswerte in Wahrscheinlichkeiten und entwickeln Zukunftsentwürfe, die verblüffen. Wie bereits festgestellt, sind wahrheitsgetreue (im Sinne der Aussagenlogik) Aussagen über die Zukunft inhärent unscharf. Diese naturgegebene Unschärfe bei der Prognostik muss daher auch Künstlichen Intelligenzen bei der Berechnung von Zukünften zugestanden werden. Auch aus dieser Perspektive lässt sich die Einschätzung rechtfertigen, dass KI eine hilfreiche Unterstützung für die Zukunftsforschung sein kann. Der daraus entstehende Wechselbezug zwischen KI und Zukunftsforschung und wie die gegenseitige Beeinflussung wirkt (Rückkopplungseffekte und Technikfolgenabschätzung), wird in der

Zukunftsforschung selbst intensiv diskutiert (Steinmüller et al. 2022) und kann an anderer Stelle weiter vertieft werden. Es kann beobachtet werden, dass, wenn KI bzw. Prognostik verstärkt zum Einsatz kommen, dies auch das Umfeld verändert – wie das Beispiel Predictive Policing oder flächendeckende Videoüberwachung zeigen (Gierke 2013) – und zu wechselseitigen, fast schon rekursiven, selbstreferenziellen Überlegungen führt. KI-Prognostik verändert das Umfeld und damit auch die Zukunftsforschung.

5. Neuere Künstliche Intelligenz – weder Akteur:in noch reines Instrument?

Es ist deutlich geworden, dass KI-Systeme in Wissenschaft und Forschung weit mehr als nur Werkzeugcharakter haben. Sie ermöglichen fundamental neue Denkansätze und eröffnen neue Zugänge des Denkens und des Forschens. Es liegt daher nahe zu fragen, wie die Zukunft der KI aussehen könnte. Der vorliegende Text jedoch befasst sich mit der Perspektive des Einsatzes von KI in der Zukunftsforschung. KI führt zu neuen, auch disruptiven Nutzungskontexten, und dringt in die Zweckdimension forschenden Handelns ein. Auf der anderen Seite wird die o.g. besondere Leistungsfähigkeit von KI-Systemen neue, bislang kaum zugängliche Forschungsfelder erschließbar machen. KI wird Teil des Forschens, sie wird Impulsgeber:in, Lösungsgehilfin und Sparringspartner:in. Aus dieser Perspektive erscheint es berechtigt und zweckmäßig, KI sowohl in der Soziologie als auch in der Zukunftsforschung nicht nur als Gegenstand der Betrachtung zu sehen, sondern als agierende Akteur:innen. Ob und in welcher Weise Computer – wenn überhaupt – »denken« können, rückt zunehmend in den Hintergrund gegenüber der Frage in welchem Sinne Computer (z.B. als humanoide Roboter) »handeln« können. Diese Aufmerksamkeitsverschiebung verläuft parallel zur technischen Entwicklung, die intelligente Systeme nicht nur als reine Computersysteme mit hoher Informationsverarbeitungskapazität versteht, sondern als Quasi-Akteur:in, die menschliches Handeln nicht nur nachahmen oder übertreffen, sondern zunehmend »autonom« Handeln. Es stellt sich somit die Frage, in welchem Sinne diese neue Form der Akteur:innen überhaupt handelt und in welchem Umfang »autonome« Systeme für ihr Handeln verantwortlich gemacht werden können. Eine vertiefte Reflektion hierüber ist an dieser Stelle nicht möglich, aber es wird deutlich, dass mit der digitalen, autonom handelnden Akteur:in die Akteursrolle in das Instrument selbst verlegt wird, das

begrifflich genau in diesem Moment aufhört, nur noch Instrument zu sein. Es ist nachvollziehbar, dass eine Handlungsurheberschaft, also die Empfindung alleine und autonom verantwortlich für die Veränderung des eigenen Wesens oder der umgebenen Umwelt zu sein – der bewusste Wille – solchen autonomen Systemen nicht zugestanden werden kann.

Bruno Latour wies auf eine zunehmend unklare Rollenverteilung zwischen Mensch und Maschine hin (Latour 2001), nach der die Welt voller Hybride, Quasi-Subjekte ist, die nicht nur konstruiert sind, sondern Welt durch Vermittlung und Delegation selbst konstruieren und in Folge dessen weder beherrschbar noch zwangsläufig nachvollziehbar sind. Latour stellt die Außengrenze des Menschen in Frage, sodass beim »Duell« Mensch gegen Maschine die Rollen nicht mehr klar verteilt sind.

Die Relationale Soziologie um Harrison White konzipiert Netzwerke als zwischenmenschliche Konstruktionen. Als Ausgangspunkt seiner Betrachtung sieht White sogenannte Transaktionen, soziale Prozesse in denen sich Netzwerke und andere soziale Formationen ergeben (White 1995: 1035f.). Der Begriff lenkt damit den Blick weg von den Akteur:innen und hin zu den Prozessen, die zwischen den Akteur:innen bzw. »Identitäten« ablaufen. Bei White sind Identitäten die Einheiten, die in den Netzwerken um gegenseitige Kontrolle ringen (White 1992).

Andererseits verweist der Identitätsbegriff auch auf die konstruierende Beobachtung dieser Einheiten im Netzwerk. Der Begriff Identität darf den Blick nicht auf menschliche Individuen verengen, da White unter dem Begriff der Identität auch kollektive Einheiten wie Staaten, Unternehmen, soziale Bewegungen versteht. Da (soziale) Netzwerke im Sinne von White sowohl zwischen Personen als auch zwischen Organisationen und anderen kommunikativ konstruierten Identitäten zu finden sind, ist also der Identitätsbegriff für eine theoretische Grundlegung sozialer Netzwerke besser geeignet als der der Person – dies auch, wenn das Personenkonzept genauer zwischen den internen Prozessen von psychischen Systemen und deren Beobachtung als Person in der Kommunikation unterscheidet. Nach White sind soziale Strukturen notwendig, denn individuelle Identitäten werden erst durch soziale Strukturen oder Netzwerke konstituiert und die sozialen Strukturen werden die Umgebung für Identität, die deren Entstehen und Entwicklung formieren und kontrollieren (vgl. White 1992; White et al. 1976; White et al. 2007).

6. Ausblick

Zukunftsforschung reduziert sich in der Praxis von Foresight, Technologiefrüherkennung oder Technikfolgenabschätzung (Zweck 2002: 55f.) nicht auf einen rein rechnerbasierten, technokratisch-organisierten Beratungsprozess (Zweck/Braun 2021). In der Praxis handelt es sich um einen sozialen Prozess: um einen in wohldefinierten, mehreren Schritten verlaufenden Einsichtsprozess der Beteiligten. Die Schritte dienen aber zugleich dazu, eine iterative Interaktion zwischen den Akteur:innen, seien es Experten:innen, Betroffene, oder anderweitig Beteiligte sicher zu stellen. Denn während des Zukunftsforschungsprozesses ist es neben dem Erkenntnisgewinn essenziell, eine begleitende und über das Projekt hinausgreifende Akzeptanz und Verbreitung gewonnener Einsichten sicher zu stellen. Vor dem Hintergrund, der im vorliegenden Beitrag reflektierten Fragen, bleibt offen, ob, wann und in welchem Grad sich KI über die diskutierten Vorteile bezüglich Datenanalyse etc. hinaus zu einer im Prozess eigenständigen und sozialen Akteur:in entwickeln könnte, einer Akteur:in die im Zukunftsforschungsprozess zukünftig Gehör findet. Gegenwärtig erreicht KI diese Qualität im sozialen Prozess in keiner Weise. KI tritt allenfalls als Unterstützung in zweiter Reihe an, also als Hintergrundakteur:in zur neuen impulsgebenden Verbesserung des Wissensstandes einzelner menschlicher Beteiligter im Zukunftsforschungsprozess, die die entsprechenden Ergebnisse einbringen. Wenn auch gegenwärtig noch aus zweiter Reihe, wird dieser datengetriebene Beitrag der KI zum Zukunftsforschungsprozess zunehmend wichtiger werden. Wie der vorliegende Beitrag zeigt, zeichnet sich dies sowohl was kreative Impulse anbelangt, in dem unerwartete Zusammenhänge oder übersehene Wildcards erfasst werden, wie auch bei der Analyse großer Datenbestände komplexer Fragestellungen ab. In einer KI-getriebenen Zukunftsforschung wird die Vernetzung der Akteur:innen aus verschiedenen Disziplinen zu einem noch zentraleren Erfolgsfaktor. Die Erschließung der Vielseitigkeit disziplinspezifischer Erfahrungen und Datensammlungen durch KI-gestützte Vernetzung und Auswertung dürften zukünftig wichtige Impulse für eine bessere Erfassung und Charakterisierung des Umfeldes des gewonnenen Zukunftsbildes bieten und neue und verbesserte Möglichkeiten der Qualitätssicherung eröffnen.

Technische Systeme, die mit sog. künstlicher »Intelligenz« ausgestattet sind, können grundsätzlich nicht als Akteur:in in Begründungsdiskursen auftreten, da ihnen die Merkmale von Handlungsurheberschaft und Zurechenbarkeit fehlen und allenfalls vom Menschen zugerechnet werden. Die

wissenschaftliche Forschung ist und wird auch im Zeitalter von KI, Machine Learning und Big Data weiterhin auf den Menschen als Wissenschaftler:in und Forscher:in angewiesen sein. In der bestehenden Phase schwacher KI bleibt der Mensch als handelnde Akteur:in im Zentrum von Wissenschaft und Forschung. Ein etwaiger Ersatz des Menschen durch »starke KI« ist nicht zu erwarten, da diese zumindest auf absehbare Zeit nicht erreichbar erscheint. Immerhin ist es mit den abgestuften Konzepten von Handlungsträgerschaften – wie bei White oder Latour – nachvollziehbar, an welchem Punkt der soziotechnische Entwicklungsprozess steht und welche Implikationen damit für die sozialen Formationen einhergehen, etwa wenn Eliza tatsächlich für Therapiesitzungen geeignet wäre oder Programme den Turing-Test bestehen. Forscher:innen müssen sich auch im Zeitalter von KI als Urheber:in des Forschungshandelns begreifen. Die Frage, ob es gelingt, dieser Einsicht treu zu bleiben, wenn Wissenschaftler:innen vor Bewunderung von durch schwache KI gefundenen Erkenntnissen – wie im Falle von Eliza – intelligente und menschliche Eigenschaften und Fähigkeiten hineinprojizieren, wird die Zukunft zeigen.

Seit der Endredaktion dieses Beitrages hat sich die Entwicklung in der KI, insbesondere im Bereich der großen Sprachmodelle rasant beschleunigt. Nur wenige Wochen nach Fertigstellung dieses Beitrages präsentierte OpenAI mit ChatGPT sein heute bekanntestes Sprachmodell zur Verarbeitung natürlich sprachlicher Computeranweisungen, das inzwischen von mehreren hundert Millionen Menschen weltweit benutzt wird. In den darauf folgenden Monaten entstanden zahlreiche weitere, leistungsfähige Sprachmodelle von Google, Amazon, Meta, Nvidia, Tencent, Aleph Alpha und vielen anderen großen Akteur:innen/Plattformen. Basierend auf einigen dieser KI-Modelle wurden mehrere frei verfügbare Open Source Modelle entwickelt, die in zahlreichen Anwendungsbereichen (wie z.B. Question Answering, Information Retrieval oder Logical Reasoning) auch und vor allem in verschiedenen Forschungsbereichen zur Unterstützung der Forschenden eingesetzt werden. Mit Stand 13.02.2024 sind ca. 505.000 KI-Modelle verfügbar, d.h. täglich entstehen weltweit ca. 1.000 neue Modelle – eine Entwicklungsgeschwindigkeit, mit der Publikationen in gedruckter Form kaum noch mehr Schritt halten können.

7. Literatur

- Baidu (2020): »How Baidu is bringing AI to the fight against coronavirus«, in: MIT Technology Review vom 11. März 2020, <https://www.technologyreview.com/2020/03/11/905366/how-baidu-is-bringing-ai-to-the-fight-against-coronavirus/>.
- Beer, Stafford (1973): *Fanfare for Effective Freedom: Cybernetic Praxis in Government*, Wales: Cwavel Isaf Institute.
- Bender, Emily M./Gebru, Timnit/McMillan-Major, Angelina/Shmitchell, Shmargaret (2021): »On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?«, in: *Conference on Fairness, Accountability, and Transparency (FACCT '21)*, New York: Association for Computing Machinery, S. 610–623.
- Beuth, Patrick (2017): »Die Automaten brauchen Aufsicht«, in: *Die Zeit* vom 25.10.2017, S. 3.
- Bosse, Christian K./Hoffmann, Judith/van Eist, Ludger (2018): »Potenzialeinschätzung von Big Data Mining als methodischer Zugang für Foresight«, in: *Zeitschrift für Zukunftsforschung* 2018(1), S. 2195–3155.
- Bringsjorden, Selmer/Bello, Paul/Ferrucci, David (2001): »Creativity, the Turing Test, and the (Better) Lovelace Test«, in: *Minds and Machines* 11, S. 3–27.
- Callaway, Ewen (2020): »It will change everything: DeepMind's AI makes gigantic leap in solving protein structures«, in: *nature* 588, S. 203–204.
- Clarke, Arthur C. (1962): *Hazards of Prophecy: The Failure of Imagination*, London: Gollanz.
- Eggert, Michael/Zweck, Axel (2021): »Die Analyse naturwissenschaftlich-technischer Narrative als Element zukunftsbezogener Beratungs- und Entscheidungsprozesse«, in: Jürgen Gausemeier (Hg.)/Wilhelm Bauern (Hg.)/Roman Dumitrescu (Hg.), *Vorausschau und Technologieplanung: 16. Symposium für Vorausschau und Technologieplanung*. 2. und 3. Dezember 2021, Berlin/Paderborn: Universität Paderborn, S. 424–448.
- Fleck, Ludwig (1980): *Entstehung und Entwicklung einer wissenschaftlichen Tatsache*, Frankfurt a.M.: Suhrkamp.
- Gardner, Howard (1983): *Frames of Mind. The theory of multiple intelligences*, New York: Basic Books.
- Gavins, Joanna/Lahey, Ernestine (2016): *World Building: Discourse in the Mind*, London: Bloomsbury.

- Gerhold, Lars/Holtmannspötter, Dirk/Neuhaus, Christian/Schüll, Elmar/Schulz-Montag, Beate/Steinmüller, Karlheinz/Zweck, Axel (2015): Standards und Gütekriterien der Zukunftsforschung: Ein Handbuch für Wissenschaft und Praxis, Salzburg: Springer.
- Geschka, Horst/Hammer, Richard (1997): »Die Szenario-Technik in der strategischen Unternehmensplanung«, in: Dietger Hahn/Bernard Taylor (Hg.), Strategische Unternehmensplanung/Strategische Unternehmensführung, Heidelberg: Physica, S. 464–489.
- Giansiracusa, Noah (2021): How Algorithms Create and Prevent Fake News. Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More, Acton/MA/USA: Apress.
- Gierke, Sebastian (2013): »Wie intelligent sind Videoüberwachungssysteme?«, in: Süddeutsche Zeitung vom 31.06.2013, S. 1–2.
- Goleman, Daniel (1995): Emotional Intelligence. Why It Can Matter More Than IQ, New York: Bantam.
- Goodfellow, Ian/Pouget-Abadie, Jean/Mirza, Mehdi/Xu, Bing/Warde-Farley, David/Ozair, Sherjil/Courville, Aaron/Bengio, Yoshua (2014): »Generative Adversarial Nets«, in: NIPS'14, S. 2672–2680.
- Groth, Olaf J./Nitzberg, Mark/Zehr, Dan (2018): Vergleich nationaler Strategien zur Förderung von Künstlicher Intelligenz, Berlin: Konrad-Adenauer-Stiftung.
- Häußling, Roger/Franke, Tim/Härpfer, Claudius/Roth, Philip/Schmitt, Marco/Strüver, Niklas/Zantis, Sascha (2021): Mendelian und das Erklärungspotential der Theorie von Identität und Kontrolle: Ein techniksoziologischer Blick auf Recommender-Systeme, Aachen: RWTH-Aachen IFS.
- Heaven, Will D. (2021): 2021: »Das Jahr der KI-Monstermodelle«, in: heise online vom 31.03.2021, <https://www.heise.de/hintergrund/2021-Das-Jahr-der-KI%20Monstermodelle6305269.html>.
- Heckwolf, Christoph/Schmitt, Marco (2022): KI zwischen Blackbox und Transparenz– Das Koppeln und Entkoppeln von Kontrollprojekten, Aachen: RWTH-Aachen IFS.
- Holzinger, Andreas (2018): »Explainable AI (ex-AI)«, in: Informatik-Spektrum 41, S. 138–143.
- Humphreys, Paul (2009): »The Philosophical Novelty of Computer Simulation Methods«, in: Synthese 169, S. 615–626.
- Kasparov, Garry (2017): »AlphaZero AI beats champion chess program after teaching itself in four hours«, in: Garry Kasparov vom 12.08.2017, <https://>

- www.kasparov.com/blogpost/alphazero-ai-beats-champion-chess-program-after-teaching-itself-in-four-hours/.
- Kelnar, David/Kostadinov, Asen (2019): *The State of AI 2019: Divergence*, London: MMC Ventures Research, S. 99.
- Knight, Will (2017): »The U.S. Military Wants Its Autonomous Machines to Explain Themselves«, in: MIT Technology Review vom 4. März 2017, <https://www.technologyreview.com/2017/03/14/243295/the-us-military-wants-its-autonomous-machines-to-explain-themselves/>.
- Knight, Will (2021): »This AI Can Generate Convincing Text – and Anyone Can Use It«, in: Wired vom 29.03.2021, <https://www.wired.com/story/ai-generate-convincing-text-anyone-use-it/>.
- Kosow, Hannah/Gaßner, Robert (2008): *Methoden der Zukunfts- und Szenarioanalyse Überblick, Bewertung und Auswahlkriterien. Werkstattbericht, Bericht Nr. 103*, Berlin: Institut für Zukunftsstudien und Technologiebewertung.
- Kühl, Eike (2014): »Ein Trickser namens Eugene Goostman«, in: Die Zeit vom 10.06.2014, <https://www.zeit.de/digital/internet/2014-06/turing-test-eugene-goostman-kritik>.
- Latour, Bruno (2001): *Das Parlament der Dinge: Für eine Ökologie der Dinge*, Frankfurt a.M.: Suhrkamp.
- Mayer, Boris (2021): »Textgenerator GPT-3 auf Deutsch getestet.«, in: golem.de IT-New für Profis vom 03.11.2021, <https://www.golem.de/news/textgenerator-gpt-3-auf-deutsch-getestet-genau-wahrscheinlich-sie-sind-wie-die-ameisen-2111-160468.html>.
- Mayer, John D./Salovey, Peter (1990): »Emotional intelligence«, in: *Imagination, Cognition, and Personality* 9, S. 185–211.
- Mayer, John D/DiPaolo, Maria/Salovey, Peter (1990): »Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence«, in: *Journal of Personality Assessment* 54, S. 772–781.
- Menick, John (2016): »Move 37: Artificial Intelligence, Randomness, and Creativity«, in: *Mousse Magazine* 55/53, <https://johnmenick.com/writing/move-37-alpha-go-deep-mind.html>
- Metzinger, Thomas (2001): »Postbiotisches Bewusstsein: Wie man ein künstliches Subjekt baut – und warum wir es nicht tun sollten«, in: *Paderborner Podium/20 Jahre Heinz Nixdorf Museums Forum*, S. 87–113.
- Moorstedt, Michael (2020): »Vorhersage-Algorithmen: Wie Corona künstliche Intelligenzen verwirrt«, in: *Süddeutsche Zeitung* vom 19.05.2020, <https://>

- www.sueddeutsche.de/digital/corona-kuenstliche-intelligenz-vorhersage-algorithmus-1.4910460.
- Nevala-Lee, Alec: »What Isaac Asimov Taught Us About Predicting the Future«, in: *The New York Times* vom 31.10.2018, <https://www.nytimes.com/2018/10/31/books/review/isaac-asimov-psychohistory.html>.
- Scheiermann, A./Zweck, Axel (2014): »Big Data für die Sozialforschung. Innovations- und Technikanalyse«. Kurzstudie, VDI-Technologiezentrum (Hg.), Schriftenreihe Zukünftige Technologien Consulting Nr. 98, Düsseldorf.
- Schmitt, Marco/Fuhse, Jan (2015): Zur Aktualität von Harrison White. Einführung in sein Werk, Wiesbaden: Springer Fachmedien.
- Searle, John R. (1980): »Minds, Brains, and Programs«, in: *Behavioral and Brain Sciences* 3(3), S. 417–424.
- Service, Robert (2021): »2021 Breakthrough of the year – Protein structures for all«, in: *Science* vom 16.12.2021, <https://www.science.org/content/article/breakthrough-2021>.
- Steinmüller, Karlheinz/Burchardt, Aljoscha/Gondlach, Kai/von der Gracht, Heiko/Kisgen, Stefanie/Ellermann, Kai/Martini, Melanie/John, Marcus (2022): »Kann Künstliche Intelligenz Zukunftsforschung? – Ein spekulativer Impuls«, in: *Zeitschrift für Zukunftsforschung* 2022(1), S. 2195–3155.
- Süß, H.-M.; Beauducel, A. (2011): »Intelligenztests und ihre Bezüge zu Intelligenztheorien«, in: Lutz F. Hornke/Manfred Amelang/Martin Kersting (Hg.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik*, Göttingen/Bern/Toronto/Seattle: Hogrefe, S. 97–234.
- Tegmark, Max (2017): *Leben 3.0: Mensch sein im Zeitalter Künstlicher Intelligenz*, Berlin: Ullstein.
- Turing, Alan M. (1950): »I. Computing Machinery and Intelligence«, in: *Mind* LIX(236), S. 433–460.
- von Ahn, Luis/Blum, Manuel/Hopper, Nicholas J./Langford, John (2003): »CAPTCHA: Using Hard AI Problems For Security«, in: Eli Biham (Hg.), *EUROCRYPT 2003: Advances in Cryptology*, Berlin/Heidelberg: Springer, S. 294–311.
- Wattenhofer, Roger (2018): »AlphaZero will doch nur spielen!«, in: *ETH Zürich* vom 23.02.2018, <https://ethz.ch/de/news-und-veranstaltungen/ethnews/news/2018/02/wattenhofer-alphazero.html>.
- Weber, Max (1919): *Wissenschaft als Beruf*, Berlin: Duncker & Humblot.
- Wechsler, David (1964): *Die Messung der Intelligenz Erwachsener*, Stuttgart: Hueber.

- Weizenbaum, Joseph (1976): *Computer Power and Human Reason: From Judgment to Calculation*, New York: W.H. Freeman and Company, S. 7.
- White, Harrison C. (1992): *Identity and Control: A Structural Theory of Social Action*, Princeton: Princeton University Press.
- White, Harrison C. (1995): »Network Switchings and Bayesian Forks: Reconstructing the Social and Behavioral Sciences«, in: *Social Research* 62, S. 1035–1063.
- White, Harrison C./Boorman, Scott/Breiger, Ronald (1976): »Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions«, in: *American Journal of Sociology* 81, S. 730–780.
- White, Harrison C./Fuhse, Jan/Thiemann, Matthias/Buchholz, Larissa (2007): »Networks and Meaning: Styles and Switching«, in: *Soziale Systeme* 13(1-2), S. 543–555.
- White, Harrison C. (2008): *Identity and Control. How social formations emerge*. Second edition, Princeton: Princeton University Press.
- Zweck, Axel (2002): »Three perspectives for one future in economy and society«, in: *Futures Research Quarterly* 18(1), S. 55–66.
- Zweck, Axel (2012): »Gedanken zur Zukunft der Zukunftsforschung«, in: Reinhold Popp (Hg.), *Zukunft und Wissenschaft: Wege und Irrwege der Zukunftsforschung*, Salzburg: Springer, S. 59–80.
- Zweck, Axel/Braun, Matthias (2021): *Predictive Analytics: Sind Zukunftsforscher*innen ein Auslaufmodell?*, Düsseldorf: VDI Technologiezentrum GmbH.

8. Filme

PLUG & PRAY (2010) (D, R: Jens Schanze)

