

BEITRÄGE AUS DER FORSCHUNG

Band 218

Hartmut Hirsch-Kreinsen & Thorben Krokowski

„False Promises“ der Künstlichen Intelligenz



Impressum

Beiträge aus der Forschung, Band 218

ISSN: 0937-7379

Dortmund 2023

Sozialforschungsstelle Dortmund (sfs)

Fakultät Sozialwissenschaften | Technische Universität Dortmund

Evinger Platz 17

D-44339 Dortmund

Tel.: +49 (0)2 31 – 755-1

Fax: +49 (0)2 31 – 755-90205

Email: information.sfs@tu-dortmund.de

www.sfs.sowi.tu-dortmund.de

„False Promises“ der Künstlichen Intelligenz

Zusammenfassung

Die Entwicklungsdynamik jeder neuen Technologie und im Speziellen der Künstlichen Intelligenz ist zu-meist mit Versprechungen über ihre besondere Leistungsfähigkeit und völlig neue Anwendungsmöglichkeiten verbunden. Vor diesem Hintergrund zielt die Fragestellung des Beitrags darauf ab, zu ergründen, welche Besonderheiten das Technologieversprechen der KI aufweist. Es wird gefragt, woher die offensichtliche und scheinbar unwiderstehliche Überzeugungskraft, aber auch die Faszination und die Ängste der Öffentlichkeit gegenüber dieser Technologie kommen, gerade auch aktuell im Hinblick auf neue Sprachsysteme wie Chat-GPT. Die These ist, dass das Technologieversprechen eine sehr unpräzise Rhetorik aufweist. Sie erscheint einfach, klar und überzeugend, ist grundsätzlich nur schwer bestreitbar und eröffnet vielfältige Anschluss-möglichkeiten. Um dies zu erreichen, basiert sie auf einer unscharfen begrifflichen Metaphorik, sehr unspezifischen Einschätzungen, impliziten Fehlannahmen und übergeht grundlegende Anwendungs- und Funktionsprobleme der KI. Insofern kann auch von den „False Promises“ (Noam Chomsky) der KI im Allgemeinen und der neuen, KI-basierten Sprachsysteme im Speziellen gesprochen werden. Gleichwohl wird ferner angenommen, dass das Technologieversprechen der KI seine dauerhafte Überzeugungskraft letztlich nicht aus der zuvor genannten Opazität und fehlenden Substantialität speist, sondern aus dem alten *Mythos der intelligenten Maschine* zieht, die der menschlichen Intelligenz ebenbürtig, wenn nicht gar überlegen sei.

Abstract

The development dynamics of any new technology, and in particular of artificial intelligence, is usually associated with promises of its special performance and completely new application possibilities. Against this background, the question of this article aims to find out which special features the technological promise of AI has. It is asked where the obvious and seemingly irresistible persuasiveness, but also the fascination and fears of the public towards this technology come from, especially currently with regard to new speech systems like ChatGPT. The thesis is that the technology promise has a

¹*Danksagung

Die Mitarbeit Thorben Krokowskis am vorliegenden Beitrag entstand im Rahmen des Graduiertenkollegs GRK 2193 „Anpassungsintelligenz von Fabriken im dynamischen und komplexen Umfeld“ gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 276879186.

very imprecise rhetoric. It appears simple, clear, and convincing, is fundamentally difficult to dispute, and opens up multiple possibilities for connection. To achieve this, it relies on fuzzy conceptual metaphors, very unspecific assessments, implicit misconceptions, and overlooks fundamental application and functionality problems of AI. In this respect, one can also speak of the „False Promises“ (Noam Chomsky) of AI in general and of the new, AI-based language systems in particular. Furthermore, it is also assumed that the technological promise of AI ultimately does not feed its lasting persuasive power from the aforementioned opacity and lack of substantiality, but rather from the old *myth of the intelligent machine*, which is equal to, if not superior to, human intelligence.

Schlüsselbegriffe

Künstliche Intelligenz · Technologieversprechen · Innovations- und Technikforschung · Maschinenintelligenz · Innovationspolitik · Schlüsseltechnologie

Keywords

Artificial Intelligence · Promising Technology · Technology and Innovation Research · Machine Intelligence · Innovation Policy · Key Technology · Bottleneck Technology

1. Technologieversprechen

Jede neue Technologie ist mit Versprechungen über ihre besondere Leistungsfähigkeit sowie völlig neue Anwendungsmöglichkeiten verknüpft. So werden für die Zukunft beispielsweise hohe Gewinne in Aussicht gestellt und oftmals sollen durch eine neue Technologie drängende gesellschaftliche Herausforderungen bewältigt werden. Ein solcher Diskurs reicht mit seinen Versprechungen oftmals weit in die Zukunft, wobei die mit den Versprechungen einhergehenden Dynamiken nicht selten als eine wesentliche Triebkraft von technologischen Innovationen sowie generell der ökonomischen Dynamik angesehen werden (vgl. Beckert 2016). Dazu werden seit längerem im Feld der Science and Technology Studies (STS) Studien vorgelegt, die zeigen, dass die Dynamiken wissenschaftlichen Erkenntnisgewinns und technologischer Innovationen eng an Narrative, Visionen und Erwartungen über die Perspektiven von Innovationen und den damit zu er-

reichenden gesellschaftlichen Fortschritt gekoppelt sind (vgl. Brown et al. 2017). Konkret kann dieser Zusammenhang mit dem Konzept der *Promising Technology* gefasst werden (vgl. Van Lente 1993; Van Lente & Rip 1998; Bender 2005; Borup et al. 2006). Das Kernargument lautet, dass Akteure bei ihren Entscheidungen, an einer Technologieentwicklung teilzunehmen, sich zunächst noch an einem sehr allgemeinen Versprechen über diese erst noch zu entwickelnde Technologie orientieren. Im Prinzip offeriert das Versprechen interessierten Akteuren eine Entwicklungsperspektive, stellt zukünftige Nutzenpotenziale in Aussicht, eröffnet einen einzuschlagenden Weg für die Innovation und begründet auch die Möglichkeit für die Durchsetzung eigener Interessen. Es bietet die Voraussetzung dafür, weitere Akteure anzusprechen, sie in den Innovationsprozess einzubinden, ihr Handeln zielgerichtet zu koordinieren sowie Wissens- und Innovationsressourcen zu mobilisieren und Investitionen in Forschung und Entwicklung zu initiieren oder anders formuliert: Das Technologieversprechen hat die Funktion eines Leitmotivs für heterogene Akteure mit ihren im Grunde sehr unterschiedlichen Interessen, die in einen Innovationsprozess eingebunden werden. In diesem Sinn lässt sich festhalten, dass Versprechen „do not strive for truth or accuracy, but are meant to influence specific social processes in technological developments“ (Geels & Smit 2000: 880).

Dabei muss allerdings das Technologieversprechen für interessierte Akteure in einsichtiger Weise mit dem Stand der Forschung begründet sein und als aussichtsreich einzuschätzende Innovationspotenziale aufzeigen. Zugleich muss es das stets gegebene Ungewissheitsproblem einer Innovation glaubhaft überbrücken. Gerade angesichts fehlender Erfahrungen mit einer neuen Technologie muss argumentiert werden, dass es sich dieses Mal tatsächlich um neue, noch nie dagewesene, technologische Möglichkeiten samt großer Entwicklungs- und Gewinnpotenziale handelt. Insofern muss ein Technologieversprechen stets Aufmerksamkeit über *Neuigkeitsdramatisierung* erzeugen. Einerseits geht es um die Verheißung einer besseren Zukunft, andererseits sollen Befürchtungen bei Skeptikern:innen zerstreut werden, die Sorge haben könnten, beispielsweise im ökonomischen Wettbewerb abgehängt zu werden oder negativen gesellschaftlichen Konsequenzen anheimfallen zu können (vgl. Faust 2021: 71).

2. KI-Hypes

Im Fall der KI sind diese Zusammenhänge besonders ausgeprägt. Ihre Genese und Diffusion waren immer wieder von hochschlagenden Versprechungen begleitet. Diese haben sich jedoch in der Vergangenheit in den meisten Fällen als unrealistisch erwiesen, sodass ihre weitere Entwicklung zwischenzeitlich gar ins Stocken zu geraten drohte (vgl. Ahrweiler 1995; Hirsch-Kreinsen 2023). Historisch lassen sich solche Phasen, die auch als „AI Winter“ (Bostrom 2014; Teich 2020) bezeichnet werden, gegen Ende der 1960er und Anfang der 1990er Jahre ausfindig machen.

Dass Versprechungen, Erwartungen und Narrative die KI-Entwicklung prägen, zeigen eine ganze Reihe kritischer Studien über die Entwicklungsgeschichte der KI (vgl. u. a. Cyranek & Coy 1994; Brödner 1997; Buchanan 2006; Hirsch-Kreinsen 2023). Einerseits handelt es sich um Versprechungen, die sich angesichts nur beschränkter Leistungsfähigkeit und hoher Anwendungsprobleme der jeweils verfügbaren KI-Systeme und Methoden als überzogen erweisen und als Mythen abgetan werden (vgl. Larson 2021). Resultat sind oftmals „Erwartungsenttäuschungen“ (Ahrweiler 1995) wichtiger beteiligter Akteure und ihr Rückzug aus dem Technologiefeld. Andererseits aber bietet die gleichwohl schnelle und dynamische Entwicklung der KI, ihrer Grundlagen und ihrer Anwendungspotenziale, stets wieder Anlass für neuerliche Versprechungen, die mit großer Überzeugungskraft die Erwartungen und das Interesse vieler Akteure und der Öffentlichkeit an dieser Technologie wecken. Auf den Punkt bringt diese von Versprechen getriebene KI-Dynamik ein früher Protagonist der amerikanischen KI und Gründungsmitglied der American Association for Artificial Intelligence, Bruce G. Buchanan: „The history of AI is a history of fantasies, possibilities, demonstrations, and promises“ (Buchanan 2006: 53). Ähnlich lautet die Feststellung eines deutschen KI-Wissenschaftlers aus dem Jahr 1994, der nicht erreichte Ziele und ein ständiges Hinausschieben und Erneuern der Perspektiven der KI-Entwicklung betont: „Das ist ein altes Spiel in der AI. Man sagt: In 10 Jahren werden wir das und das haben, und in 10 Jahren stellt man dann die gleiche Frage und sagt: In 20 Jahren werden wir das erreicht haben. In 20 Jahren stellt man wieder die gleiche Frage, und auch in 50 Jahren werden wir die gleiche Frage stellen“ (Cyranek & Coy 1994: 259).

Spätestens seit Beginn der 2010er Jahre nun scheint die KI-Entwicklung einen ungebrochenen Aufschwung zu erfahren und sie trifft auf ein schnell wachsendes öffentliches und politisches Interesse. Dieser Aufschwung ist von weitreichenden Versprechungen begleitet, die sich nicht nur auf technologische Entwicklungspotenziale richten, sondern geradezu gesellschaftspolitischen Charakter aufweisen (vgl. Katzenbach 2021; Hirsch-Kreinsen 2023). Avisiert werden beispielsweise Perspektiven durchgreifender Produktivitätssteigerungen bei der wirtschaftlichen Anwendung von KI und damit eines weiteren Wachstums, des Abbaus des Fachkräftemangels, eine weitreichende Verbesserung der medizinischen Diagnose und Therapiefähigkeiten oder gar ein maßgeblicher Beitrag zur Bewältigung der Klimakrise. Aus deutscher und europäischer Sicht wird damit vor allem auch das Ziel verfolgt, in der globalen Technologiekonkurrenz gegenüber den übermächtigen Ländern China und USA sowie deren international agierenden Internet- und KI-Konzernen nicht endgültig ins Hintertreffen zu geraten (vgl. u. a. Bundesregierung 2018).

Nach der Vorstellung des Großen Sprachsystems „ChatGPT“ (Generative Pre-Trained Transformer) (Heckel 2023) durch die US-Firma OpenAI und deren Investor Microsoft oder auch der artverwandten Konkurrenzprodukte von Google namens „LaMDA“ (Language Model for Dialog Applications) und „Bard“ gewinnt der Diskurs über die Potenziale und Chancen der KI aktuell den Charakter eines überbordenden Hypes: „Wieder einmal werden wir gegenwärtig durch viele vollmundige Versprechungen rund um die Künstliche Intelligenz [...] verrückt gemacht. Wer etwas älter ist, der kennt diese Wellen, die von Zeit zu Zeit aufflackern. Vor 20 Jahren war es das ‚Deep Learning‘, vor zehn Jahren das autonome Fahren – und heute sind es eben diese Textroboter“ (Otto 2023). Ohne Frage haben sich Sprachmodelle in den zurückliegenden Jahren kontinuierlich verbessert. Die amerikanische Firma OpenAI, die ChatGPT ins Leben gerufen hat, hat ein sehr gutes Sprachmodell für die breite Öffentlichkeit zugänglich und leicht nutzbar gemacht. Ähnlich wie das iPhone vielfältig genutzt werden kann, weist ChatGPT mannigfaltige Verwendungsmöglichkeiten auf, beispielsweise um Gute-Nacht-Geschichten für Kinder zu erfinden, Computercodes zu schreiben oder neue Themen zu lernen (vgl. Heckel 2023). Der vorherrschende Tenor ist daher die Faszination über die erwarteten Perspektiven,

diese indes gepaart mit Ängsten, Warnungen und Skepsis über die mitunter schwer absehbaren sozialen Folgen (vgl. Roose 2013). Allerdings werden auch immer wieder Funktionsdefizite und Anwendungsrisiken der neuen Systeme hervorgehoben, die keineswegs so perfekt seien, wie vielfach unterstellt. Wiederholt als gravierend angesehene Aspekte wie Unzuverlässigkeit, Falschinformationen oder auch konkrete Anwendungsprobleme sowie Risiken und Gefahren einer unreflektierten Nutzung werden besonders im Fall von ChatGPT und ähnlicher Systeme angesprochen. Ein Beispiel hierfür sind sogenannte „KI-Halluzinationen“ (Gurevych 2023), also das reine Phantasieren von Inhalten durch die Software.

Ungeachtet dessen wird unisono erwartet, dass die gegenwärtig erkennbaren Anwendungsgrenzen eben jener KI-Systeme weiter hinausgeschoben werden können und die diskutierten Probleme durch eine weitere und schnelle technologische Entwicklung reduziert werden. Für viele Beobachter:innen scheint es nur eine Frage der Zeit zu sein, bis diese Defizite überwunden sind. Systeme wie ChatGPT seien in der Lage, zu „lernen“ und sie würden kontinuierlich „nachjustiert“ werden. Damit soll etwa die Reproduktion von Biases, falschen Stereotypen und diskriminierenden Texten vermieden werden (vgl. Krüger 2023; Otte 2023). Vor allem wird auch immer wieder die Erwartung formuliert, dass derzeit noch existierende Probleme über kurz oder lang von „Durchbrüchen“ bei der Technologieentwicklung überwunden werden könnten (vgl. Ford 2018; Barthelmeß & Furbach 2021; Armbruster 2022). So ist es unter vielen Experten:innen unstrittig, dass „wir mit den großen KI-Modellen am Anfang einer potentiell exponentiellen Entwicklung“ (Krüger 2023) stünden.

3. Versprechungen der KI

Daran anknüpfend soll im Folgenden gefragt werden, welches die Besonderheiten des Versprechens der KI sind und worin sich seine offensichtliche Überzeugungskraft und die generelle Faszination dieser Technologie konkret begründen. Festhalten lässt sich generell, dass das Technologieversprechen eine rhetorische Architektur aufweist, die durch Aspekte wie Hypostasierung der neuen Technologie, kommunikative Generalisierungen und Reduktion gesellschaftlicher Komplexität, De-Kontextualisierung der Argu-

mente, unabweisbare Aktualität, Quantifizierbarkeit und alltagsweltliche Relevanz charakterisiert ist (vgl. Hirsch-Kreinsen 2023: 243 ff.). Die Vereinfachungen und Generalisierungen erleichtern die pointierte Vermittlung einer wünschenswerten technologischen Zukunft, indem sie sowohl den Bezug zu den Erfahrungen und traditionellen Orientierungen vieler Akteure als auch an schon existierende Diskurse über zukünftig notwendige technologische und gesellschaftliche Entwicklungsperspektiven herstellen. Mehr noch: Mit dem Verweis auf technologisch lösbare gesellschaftliche Herausforderungen wird zugleich an allgemeinpolitische Debatten über unverzichtbare und drängende Modernisierungsnotwendigkeiten angeknüpft und es werden damit mögliche diskursive Anschlussmöglichkeiten für Akteure jenseits der zunächst angesprochenen Fachöffentlichkeit eröffnet (vgl. Dickel & Schrape 2015). Zudem resultiert die Überzeugungskraft des Versprechens der KI daraus, dass ganz offensichtlich das „Timing“ stimmt. Denn eine technologische Vision ist nur dann überzeugend, wenn sie den „Nerv der Zeit“ (Kieser 1996: 26) trifft. So schließt der KI-Diskurs an die seit Jahren wachsende gesellschaftliche Fokussierung auf digitale Technologien und Internet sowie an die vorherrschende Überzeugung an, dass es sich hierbei um einen gesellschaftlichen Megatrend handelt. Insofern trifft das Thema KI den vorherrschenden Zeitgeist über die unvermeidbare Modernisierung der Gesellschaft durch neue Technologien (vgl. Madsen 2019). Darüber hinaus bezieht es sich auf die laufenden gesellschaftspolitischen Diskurse über die zunehmend drängenderen gesellschaftlichen Herausforderungen wie den Fachkräftemangel, die Verkehrsproblematik oder Ökologie- und Klimaprobleme. Mehr oder weniger explizit offeriert das Technologieversprechen auch hierfür schnelle und situationsadäquate Lösungen. Dies gilt sowohl für KI im Generellen (vgl. Hirsch-Kreinsen 2023) als auch besonders für die aktuelle Auseinandersetzung mit ChatGPT im Speziellen (vgl. u. a. Fruhlinger 2023; Krüger 2023).

Darüber hinaus wartet das Technologieversprechen mit einer unpräzisen Rhetorik auf. Sie basiert auf sehr unspezifischen Einschätzungen, impliziten Fehlannahmen und übergeht grundlegende Anwendungs- und Funktionsprobleme der KI. Es handelt sich um eine Rhetorik, die die KI-Entwicklung von ihrem Anbeginn an grundlegend prägt (vgl. Hirsch-Kreinsen 2023) und in Fällen wie ChatGPT und seinen artverwandten Systemen

kulminiert. Insofern kann in Anschluss an den prominenten Linguisten Noam Chomsky auch von den „False Promises“ (Chomsky et al. 2023) der KI und insbesondere von ChatGPT und seinen verwandten Systemen gesprochen werden.² Zuspitzen lässt sich diese These auf drei Argumentationslinien, die für den gegenwärtigen zugleich von Faszination und Befürchtungen geprägten Hype um KI und die neuen Entwicklungen großer Sprachsysteme zentral sind.

3.1 Mehrdeutigkeit des Begriffs

Der rhetorische Kern des Versprechens der KI besteht vordergründig in der Mehrdeutigkeit des Begriffs „Künstliche Intelligenz“, der wenig mehr als ein „very loose umbrella term“ (Kalthener 2021: 23) ist. Dies spiegelt sich auch in dem Umstand wider, dass es sich bei der definitorischen Begriffsbestimmung, obgleich seit Jahrzehnten intensiv diskutiert, bis heute um ein auch für Fachleute nur schwer überschaubares und häufig umstrittenes Themenfeld handelt, welches durch zahlreiche, teils divergierende Definitionen gekennzeichnet ist. Angesichts der Vielfältigkeit des Technologiefelds wird KI so oftmals als unspezifischer Sammelbegriff für sowohl Anwendungen im Sinne technischer Artefakte und Algorithmen als auch Technologien im Sinne technischer Prozeduren gebraucht. Das führt dazu, dass der Begriff häufig bloß als Platzhalter für allgemeine Themen der Digitalisierung herangezogen wird (vgl. Bauberger et al. 2021: 908). Als begriffliche Gemeinsamkeit wird häufig die Fähigkeit eines Systems angesehen, selbst gestellte Ziele zu erreichen. Allerdings wird auch hier immer wieder kontrovers diskutiert, inwieweit eine Maschine sich selbst Ziele setzen darf oder ob diese Ziele vom Menschen vorgegeben werden müssen (vgl. Koehler 2021: 3).

Das Technologieversprechen und der daran orientierte Diskurs abstrahieren dabei von konkreten Techniken und Systemen. Anpassungserfordernisse an konkrete Anwendungsfelder werden weitgehend ausgeblendet. Notwendige Differenzierungen etwa zwischen verschiedenen Methoden und Konzepten wie *symbolischer* oder konnektionistischer und neuronaler KI sowie ihren verschiedenen Varianten (vgl. Görz et al. 2021)

² Man kann auch davon sprechen, dass im gegenwärtigen Hype über KI und ChatGPT der „Elefant im Raum“ (Padmanabhan 2023) übersehen wird.

spielen im öffentlichen Diskurs kaum eine Rolle. Allenfalls wird der Begriff KI seit den 2010er-Jahren und im gegenwärtigen Hype mit maschinellem Lernen und künstlichen neuronalen Netzen gleichgesetzt, ohne deren je individuelle Grundlagen und Funktionsweisen zu präzisieren. Ferner werden die hoch spezialisierten Anwendungen realisierter KI-Systeme dabei vielfach nicht als solche eingeschätzt, sondern mit ihnen oftmals nur die weitreichenden Versprechungen und Visionen einer Artificial General Intelligence (AGI) assoziiert (vgl. Ford 2018: 186).

Unklare Metaphorik

Damit lässt sich das Technologieversprechen KI, wie oftmals neue Technologie- und Organisationskonzepte, als eine „raffinierte Mischung von Einfachheit und Mehrdeutigkeit“ (Kieser 1996: 24) charakterisieren. Die technologische Vision erscheint einfach, klar und überzeugend und ist daher grundsätzlich nur schwer bestreitbar. Voraussetzung hierfür ist eine leicht zugängliche mit KI verknüpfte Metaphorik, die auf gängigen Kategorien wie „Information“, „intelligent“, „lernend“ oder auch „autonom“ basiert. Es handelt sich um eine Gleichsetzung von menschlicher und maschineller Ebene, die, so kritische Experten:innen, als „irriges Ontologie von Mensch und Maschine“ (zit. n. Hirsch-Kreinsen 2023: 248) verworfen wird. Denn bloße Zuschreibungen werden dabei mit der tatsächlichen Funktionsweise von Computern verwechselt und „fundamentale Unterschiede zwischen Mensch und Maschine werden so ignoriert. Kompetentes Handeln von Menschen wird auf algorithmisch gesteuertes Verhalten von Maschinen reduziert; zugleich entstehen Illusionen über die tatsächliche Leistungsfähigkeit von Computern“ (ebd.) oder wie ein KI-Experte den dominanten Diskurs treffend beschreibt: Nur schwer durchschaubare Berechnungsverfahren würden mithilfe durchweg anthropomorphisierender Metaphern erklärt und es würden völlig fehlgeleitete Vorstellungen von Intelligenz vorgetragen. Verfahren maschinellen Lernens würden mit „lernenden Maschinen“ gleichgesetzt, programmgesteuerte, automatische Verfahrensabläufe avancierten zu „[a]utonomen Systemen“ und ein Artefakt wie ChatGPT werde zu einem System, das vermeintlich „Sprache versteht“. Ferner würden aus sorgfältigen Problemanalysen entstammenden Berechnungsverfahren der Status „Künstlicher Intelligenz“ angedichtet

(vgl. Brödner 2023). Ein weiteres Beispiel hierfür ist der weit verbreitete, geradezu mythisch aufgeladene Gebrauch des Terminus „Algorithmus“. Ein Algorithmus bezeichnet kein im humanen Sinn intelligentes Vorgehen, sondern ist ein rein ideelles mathematisches Objekt bzw. ein präzise determiniertes Berechnungsverfahren, welches aus Eingabedaten nach endlich vielen Schritten ein Ergebnis erzeugt und anhält (vgl. ebd.).

Moment der generellen Computereuphorie

Indes sind diese Mehrdeutigkeit und unklare Metaphorik ein wesentlicher Grund für die Faszination, die von dieser Technologie ausgeht, denn: Angeknüpft wird damit an gängige Alltags-, aber auch Wunschvorstellungen über die Zukunft technologischer Entwicklungen. Insofern kann schon die Erfindung des Begriffs „Artificial Intelligence“ im Jahr 1956 bei der in Dartmouth USA abgehaltenen „Gründungskonferenz der KI“ durch John McCarthy, der als der „Gründungsvater“ (Konrad 1998; Marquis et al. 2020) der KI gilt, als „genialer PR-Coup“ (zit. n. Hirsch-Kreinsen 2023: 248) bezeichnet werden. McCarthy schließt vor allem an die schon in den 1950er-Jahren beginnende Computereuphorie an und trägt zu einer weiteren Begründung des späteren technologischen und gesellschaftlichen Fortschrittsoptimismus bei (vgl. Katzenbach 2021). Insofern ist der KI-Diskurs mit seinen hochgesteckten Erwartungen eng mit der technikfaszinierten Sicht auf Computer verknüpft. Phänomene wie Elektronengehirn, papierloses Büro, Telearbeit, intelligente Bildschirmarbeitsplätze, menschenleere Fabrik, Cyberspace, Datenautobahn und Multimedia sind nur einige Interesse weckende bildhafte Begriffe aus der Welt dieser Technologie, um die sich weitreichende Visionen ranken. Die Technologieschübe der IT und die damit einhergehenden Spekulationen, Zukunftsvorstellungen und initiierten Metaphoriken lösten schon in der Vergangenheit stets immer neue Phasen der Euphorie aus, denen auch immer wiederkehrende Phasen der Ernüchterung sowie nicht realisierte Erwartungen bis heute nichts anhaben können (vgl. Schwarte-Amedick 2005, Natale & Ballatore 2020).

In dieses Diskursmuster passt auch die Rede darüber, dass wir angesichts von ChatGPT und Co. aktuell Zeuge „einer revolutionären Entwicklung“ (Heil 2023) seien – so etwa erst jüngst prominent durch den aktuellen Bundesarbeitsminister Hubertus Heil ver-

kündigt. Übersehen wird dabei jedoch, dass die großen Sprachsysteme als Ergebnis einer inkrementellen Entwicklung anzusehen sind, die auf einer ganzen Reihe von Vorarbeiten wie sogenannten Sprache-zu-Text-Modellen wie Whisper oder den KI-Sprachmodellen GPT-2 und GPT-3 aufbauen (vgl. Gurevych 2023; Heckel 2023). So formuliert ein KI-Experte dezidiert: „Was ChatGPT liefert, ist nicht revolutionär, sondern evolutiv. Wir haben die Entwicklung vorher nur nicht mitbekommen. Erst ab Version GPT-3 hat das Unternehmen das sehr geschickt vermarktet. Plötzlich wurde es für Massen erlebbar. Das hat einen Buzz erzeugt und unter anderem auch die Milliardeninvestition von Microsoft reingeholt“ (Breher & Lehmann 2023). Obendrein wird betont, dass generative KI ein sehr altes Konzept sei und schon der 1966 am MIT entwickelte und damals prominente Chatbot ELIZA, welcher Gespräche mit Psychotherapeuten:innen simulieren sollte, dieser Kategorie zuzurechnen sei (vgl. Fruhlinger 2023).

Die Unklarheiten und generell die Computereuphorie sind die zentralen Voraussetzungen dafür, dass das Technologieversprechen in hohem Maße Verknüpfungen mit unterschiedlichsten Orientierungen und Interessen sowie einen weiten Definitions- und Interpretationsspielraum für diese eröffnet. Insofern kann das Technologieversprechen im Sinne von Susan Leigh Star and James Griesemer als *Boundary Object* verstanden werden, das mit verschiedensten Kontexten verknüpft und von unterschiedlichsten Akteuren aufgegriffen wird. Eine Vielzahl von lokalen Interpretationen und Praktiken lässt sich damit vergleichsweise einfach auf einen Gemeinsamkeiten stiftenden Kern, eben die nicht weiter spezifizierten und vieldeutigen Besonderheiten von KI, beziehen (vgl. Faust 2021: 72).

3.2 Technologiefixierter Diskurs

Die damit metaphorisch stilisierte, besondere Leistungsfähigkeit der Technologie erlaubt es der KI, sich im öffentlichen und politischen Diskurs eine überaus große Bedeutung für die weitere gesellschaftliche Entwicklung, sowohl mit wünschenswerten als auch mit negativen, ja, dystopischen Konsequenzen, zuschreiben zu lassen. Dabei sind der Diskurs über KI, ihre Entwicklungsperspektiven und vor allem auch ihre sozialen Folgen hochgradig technologiefixiert und vernachlässigen komplexe sozio-technologische Zusammenhänge der Diffusion und Implementation neuer Technologien.

Unausweichlichkeit

Ein wesentliches Moment dieser Technologiefixierung ist die Auffassung, dass die Diffusion und breite Nutzung der KI einen geradezu unausweichlichen Charakter habe. Als Ursache gilt ihre dynamische Entwicklung (vgl. Cath et al. 2018; Enquetekommission 2020; Bareis & Katzenbach 2022; Krüger 2023). Entwicklungstreiber sind in etwa die kontinuierliche Entwicklung hochkomplexer KI-Methoden wie statistische Verfahren und die Wahrscheinlichkeitsrechnung, Methoden künstlicher neuronaler Netze, die massive Zunahme global verfügbarer Daten und die Möglichkeiten, diese zu nutzen (Big Data). Hinzu kommt eine exponentiell steigende und sich verbilligende Rechnerleistung. Damit sei die Voraussetzung für die Anwendung von KI in weiten gesellschaftlichen Bereichen wie der Landwirtschaft, Produktion oder im Gesundheitswesen gegeben (vgl. Horowitz 2018; Krüger 2023). Dies bringt auch die Bundesregierung in der Begründung ihrer KI-Strategie schon 2018 auf den Punkt: „Künstliche Intelligenz hat in den letzten Jahren eine neue Reifephase erreicht und entwickelt sich als Basisinnovation zum Treiber der Digitalisierung und autonomer Systeme in allen Lebensbereichen, Staat, Gesellschaft, Wirtschaft“ (Bundesregierung 2018: 10). Auf ähnliche Art zu interpretieren, lässt sich das Zitat eines renommierten Informatikers, wonach ein „Big Bang of Deep Learning“ (Görz et al. 2021: 9) die KI-Dynamik in den letzten Jahren massiv vorantreibe. Betont wird dabei vor allem auch die enge Wechselwirkung der Diffusion von KI mit dem generellen Prozess der Digitalisierung, d. h. der zunehmenden Durchdringung der Gesellschaft mit digitalen Systemen. Die Vernetzung existierender digitaler Technologien wie das sogenannte Internet of Things (IoT) mit KI sei ein treibender Faktor der zukünftigen Entwicklung. Etabliert werde damit ein „Virtuous Cycle“ zwischen beiden: Einerseits stelle IoT Milliarden von gesammelten Daten zur Verfügung und verbessere damit kontinuierlich die Voraussetzungen für KI. Andererseits entfaltet IoT-Anwendungen erst durch KI ihre tatsächlichen Anwendungspotenziale (vgl. McClelland 2017), weshalb KI „actually the ultimate enabler“ (Horowitz 2018: 41) für einen weitreichenden Digitalisierungsschub werde.

Diese technologiefixierte Auffassung wird besonders auch daran greifbar, dass die Innovationspolitik die KI als eine unverzichtbare und notwendige *Schlüsseltechnologie* für

die Verbesserung der globalen Konkurrenzfähigkeit und des erreichten Wohlstandsniveaus sowie für kommende gesellschaftliche Transformationsprozesse ins Zentrum vieler Maßnahmen rückt (vgl. Enquetekommission 2020; EFI 2022; Kroll et al. 2022).³ Dem ungeachtet ist auch der Zusatzbegriff „Schlüsseltechnologie“ ebenso vage wie irreführend. Eben jener suggeriert, dass KI als generelle Querschnittstechnologie problemlos unter den verschiedensten Anwendungsbedingungen nutzbar sei und sich tendenziell autonom und lernend an die unterschiedlichsten Bedingungen anpassen könne. Übersehen wird dabei, dass allenfalls Hardware-Komponenten von Computersystemen wie integrierte Schaltsysteme oder Systeme zum Speichern und Übertragen von Daten universell nutzbar sind. Dies trifft aber für diejenigen steuernden Algorithmen und Programme nicht zu, deren Funktionalität für jeden spezifischen Einsatzfall unter Umständen neu entwickelt werden muss. Es ist immer erforderlich, auf der Grundlage von Daten bestimmter Qualität ein Modell der Objekte, Strukturen und Prozesse eines spezifischen Anwendungsbereichs zu bilden und dieses anschließend zu formalisieren und zu algorithmisieren (s. u.). Dieser Umstand wird mit dem Begriff der Schlüsseltechnologie weit hin ausgeblendet. Zudem negiert der Begriff, dass die Rede von *der* KI, im Allgemeinen Sinn, unzutreffend ist. Wie nicht zuletzt die erwähnten Definitionsprobleme von KI belegen, handelt es sich dabei, genau genommen, um ein Bündel verschiedenster Methoden, die nur schwer unter einem Dachbegriff subsumiert werden können (vgl. Bauberger et al. 2021).

Mit Blick auf seine Unbestimmtheit eröffnet gerade auch der kolportierte Status als Schlüsseltechnologie zahlreiche Bezugspunkte für unterschiedlichste Akteure, Sichtweisen, Technologieorientierungen und Anwendungsperspektiven. Im Zuge der Distanzierung von konkreten Standpunkten lassen sich vielfältige Betrachtungsweisen und oftmals durch Allgemeinplätze geprägte Positionen ausfindig machen. Dadurch, dass konkrete Perspektiven (tendenziell) bewusst vermieden werden, entsteht ein breites Spektrum an Interpretations- und Auslegungsspielräumen, wodurch wiederum möglichst allgemeine Anknüpfungspunkte geschaffen werden, aus denen eine große Vielfalt

³ Bereits in der zweiten Hälfte der 1970er Jahre wurde Künstlicher Intelligenz vom damaligen Forschungsminister, Hans Matthöfer, die Rolle einer Schlüsseltechnologie zugeschrieben (vgl. Ahrweiler 1995: 85).

an Akteuren ihren Nutzen für ihre je spezifische Situation und ihre Interessen ziehen können. In diesem Zusammenhang ist es wichtig darauf hinzuweisen, dass eben jene mutmaßliche universelle Anknüpfungsbasis des Gebrauchs des Begriffs Schlüsseltechnologie sich allerdings allein auf die Entwicklungsseite einer Technologie bezieht. Ihre Diffusion in konkrete Anwendungsfelder, die damit verbundenen Anpassungserfordernisse oder gar Nutzungsprobleme oder Anwendungsgrenzen werden vollends ausgeblendet. So wird in der einschlägigen Debatte hervorgehoben, dass „es für die Einordnung einer Technologie als Schlüsseltechnologie und für die technologische Souveränität eines Landes nur eine untergeordnete Rolle spielt, in welchen Branchen und Produkten eine Schlüsseltechnologie zur Anwendung kommt“ (EFI 2022: 43). Damit rekurriert der Begriff auf die seit langer Zeit mit neuen Technologien verknüpfte kollektive Überzeugung, dass eben jene bislang ungeahnte Entwicklungsmöglichkeiten offerierten. Umgekehrt wird die Gefahr beschworen, dass bei einem Rückstand im internationalen Technologiewettbewerb um Schlüsseltechnologien massive Wohlstandsverluste zu erwarten seien. In Hinblick auf die Entwicklung digitaler Technologien wie insbesondere der KI wird diese Gefahr für die Bundesrepublik seit Jahren beschworen (vgl. EFI 2018; 2022). Künstliche Intelligenz wird in der Bundesrepublik mit den Basisinnovationen der Industrialisierung des 19. Jahrhunderts wie Dampfmaschine und Elektrizität gleichgesetzt (vgl. Bundesregierung 2018). Das implizite Argument ist dabei, dass die früheren Basistechnologien sich inzwischen natürlich überlebt hätten und nun endlich mit der KI die für einen neuen gesellschaftlichen Entwicklungsschub erforderliche neue Basistechnologie nicht nur zur Verfügung steht, sondern unbedingt und schnell genutzt werden müsse.

Technikdeterminismus

Indes verbindet sich mit den zuvor erörterten Perspektiven ein grundlegend verkürztes Verständnis der Diffusions- und Implementationsbedingungen technologischer Innovationen wie eben auch der KI. Das Technikversprechen und die daran orientierten Erwartungen sind von einem kruden Technikdeterminismus geprägt, wonach KI mehr oder weniger eindeutig bestimmbare und für die Zukunft voraussagbare soziale Konsequenzen

nach sich ziehen wird. Diese Sicht vernachlässigt jedoch Basiserkenntnisse der sozialwissenschaftlichen Innovations- und Technikforschung, denn: Danach besteht zwischen der Entwicklung neuer Technologien und ihren möglichen Anwendungspotenzialen sowie ihrer Diffusion und den konkreten Formen ihrer Nutzung keinesfalls ein eindeutiger Zusammenhang. Vielmehr handele es sich um einen komplexen und wechselseitigen Zusammenhang, der von einer Vielzahl nicht-technischer, ökonomischer, sozialer und politischer Faktoren geprägt werde. Deren Einfluss entscheide darüber, welche Entwicklungsphasen einer Technologie im Einzelnen durchlaufen würden, in welcher Weise die Nutzungspotenziale der Technologien tatsächlich ausgeschöpft würden und welche Konsequenzen für die gesellschaftliche Entwicklung sich manifest einspielten (vgl. Rogers 2003; Fagerberg 2005; Erixon & Weigel 2016; Pfeiffer 2018). Die verkürzte technikdeterministische Sicht mündet in den erwähnten weitreichenden Versprechungen, wonach eine schnelle Nutzung von KI nicht nur absehbar weitreichende Gewinne, ökonomische Vorteile und auch vielfältige politische und auch gesellschaftlichen Fortschritt mit sich bringen wird, sondern eben auch viele Gefahren, Risiken und unerwünschte soziale Konsequenzen nach sich ziehen wird. Diese technikzentrierte Sicht kulminiert in den aktuellen Rufen nach einem Moratorium und den Warnungen vieler prominenter Entwickler:innen vor den gesellschaftlichen Folgen von ChatGPT und ähnlichen Systemen.⁴

Zweifelsfrei sind die zahlreichen Warnungen und Forderungen nach Regulierung immer auch zugleich Ausdruck gesellschaftspolitisch indizierter Gestaltungsmöglichkeiten und -notwendigkeiten. Diesen Warnungen implizit ist jedoch auch gleichermaßen die technikdeterministisch geprägte Überzeugung, dass die weitere Entwicklung und Verbreitung von KI unausweichlich sei und der Technologie derart gewaltige Potenziale obliegen würden, dass man vor diesen nicht nur warnen, sondern diese auch regulieren müsse. Freilich schwingt bei diesen Äußerungen zugleich das unüberhörbare Interesse mit, weitere absehbare Entwicklungsmöglichkeiten und damit auch Investitionen und

⁴ So besonders prominent der CEO von OpenAI, der vor den von seiner Firma entwickelten Sprachsystemen, zuletzt GPT-4, nachdrücklich warnt (Ordóñez et al. 2023, Online: <https://abcnews.go.com/Technology/openai-ceo-sam-altman-ai-reshape-society-acknowledges/story?id=97897122> [21. Juli 2023]).

Gewinne nicht durch zu weitreichende Regulationsvorgaben zu behindern. Vor allem jedoch soll wachsender Kritik und Skepsis an den Systemen der Wind aus den Segeln genommen werden. Ein Kritiker bezeichnet diese und ähnliche Statements als „Nebelkerzen, sie lenken die Sorge geschickt auf die Technik selbst [...]. Wir müssen uns nicht vor einer KI schützen, sondern vor Menschen und Organisationen [...]. Ein Siliziumkristall in einem Computer [...] ist [...] tot, ein Siliziumkristall ist einem Stein näher als einer primitiven Amöbe, ein Siliziumkristall wird – genauso wie ein Stein im Garten – niemals etwas wollen“ (Otte 2023).

3.3 Übersehene Leistungsgrenzen

Die Mehrdeutigkeit des Begriffs und der technikfixierte Diskurs implizieren, dass mit dem Technologieversprechen die derzeitige und zukünftige Leistungsfähigkeit von KI überzogen wird, Anwendungsgrenzen vernachlässigt und die grundsätzlich begrenzte und widersprüchliche Funktionsweise der Systeme ausgeblendet werden.

Unabdingbare Spezialisierung

Aktuell wird dieser Umstand vor allem durch die unkritische Verwendung des Attributs *generativ* für die Charakterisierung des Sprachsystems ChatGPT belegt. Unter Rückgriff auf jenes Attribut im Kontext der aktuell vieldiskutierten Sprachsysteme wird eben jenen die Fähigkeit attestiert, in unterschiedlichsten Kontexten gleichermaßen nutzbar und autonom kontextunspezifisch funktionsfähig zu sein (vgl. Abeba et al. 2022; Padmanabhan 2023). Generativ ist ein Sammelbegriff für KI-Systeme, mit denen auf scheinbar professionelle und kreative Weise alle möglichen Ergebnisse produziert werden können, wie etwa Bilder, Video, Audio, Text, Programmcodes, 3D-Modelle und Simulationen. Ziel ist es, die menschliche Leistungsfähigkeit in mannigfaltigen gesellschaftlichen Bereichen zu erreichen oder gar zu übertreffen.⁵ Unterstellt wird hier ein Entwicklungssprung dieser Systeme im Vergleich zu den früheren bekannten, lediglich als spezialisiert angesehenen, Systemen wie IBM Watson oder Alphazero (vgl. ebd.). Danach erreicht die KI jenes Entwicklungsstadium, das schon vor längerem als *General Purpose*

⁵ Bendel, Oliver (2023): *Generative KI*. Online: <https://wirtschaftslexikon.gabler.de/definition/generative-ki-124952> [23. Mai 2023].

Technology charakterisiert wurde (vgl. McAfee & Brynjolfsson 2017). Eine so verstandene Generalisierung der Technologie stellt ein zentrales Moment des laufenden Diskurses und des Technologieversprechens dar. Daher werden immer wieder dezidierte Thesen formuliert, gemäß welcher die neuen Systeme deshalb so beeindruckend seien, weil sie Merkmale einer allgemeinen Intelligenz aufwiesen (vgl. Chalmers 2023).

In der Tat kann man davon ausgehen, dass KI-Systeme eine besondere funktionale Qualität aufweisen. So werden Computer durch deren Methoden befähigt, ihr Verhalten mittels Sensorsignalen per algorithmisch gesteuerter Funktionsapproximation an die durch Daten beschriebenen wechselnden Gegebenheiten der Umgebung anzupassen, wodurch ihre Informationsbestände und ihre Nutzung somit in gewisser Weise generalisierbar wird. Gleichwohl gilt selbst für fortgeschrittene künstliche neuronale Netze, dass sie für ihre jeweils konkreten Aufgaben im Rahmen ihres Entwurfs passend strukturiert und ihre Prozessoren mit spezifischen Lern-Algorithmen auf Basis unterschiedlicher Methodiken gesteuert werden müssen. Hinzu kommt, dass diese Systeme vor allem von der Qualität ihrer Trainingsdaten abhängig sind, mit denen die Algorithmen für bestimmte Zwecke trainiert werden. Ein neuronales Netz als Ganzes wird über diese Festlegungen hinaus allerdings nicht programmiert, sondern passt sich nach Maßgabe der vorab definierten Funktionsweise des Algorithmus an spezielle Aufgabenstellungen an. Denn die für die Bewältigung einer spezifischen Aufgabe trainierten Funktionen lassen sich auf neue Datenobjekte ähnlicher Art anwenden, vorausgesetzt, der infolge prinzipieller Kontingenz sozialer Praktiken veränderliche Kontext bleibt erhalten (Brödner 2022: 34). Daher gelingt es, für spezielle, auch schwierige Anwendungen, sogenannte adaptive KI-Systeme zu realisieren, die – freilich immer nur in einem begrenzten Rahmen – auf andere Aufgaben übertragbar sind (vgl. Brödner 2019; Dyer-Witthof et al. 2019; Bischoff 2023; Otte 2023a).

Auch für die genannten Sprachsysteme lassen sich die erörterten Dynamiken und Abläufe identifizieren. Dialogbasierte Chatsysteme wie ChatGPT können für eindeutige Anwendungsdomänen wie etwa die Erstellung von Computerprogrammen oder vermeintlich Kreativität benötigende Aufgaben wie die Erstellung von Versreimen (vgl. Chomsky et al. 2023) sehr nützlich sein. So bestimmen auch in diesem Fall die Daten, auf deren

Basis ein Sprachmodell trainiert wird, was es kann. Wird ein solches System beispielsweise mit einem speziellen Code und Fragen und Antworten zum Programmieren trainiert, erhält es die Fähigkeit, Codes zu generieren und Fragen zum Code zu beantworten (vgl. Heckel 2023). Ungeachtet dessen folgt jedoch auch die Produktion von Texten sehr unterschiedlichen Regeln und ist in hohem Maße domänenabhängig. So unterscheiden sich wissenschaftliche Papiere in Stil und Inhalt substantiell von solchen in Nachrichtenmedien, die wiederum anders sind als Schulaufsätze oder Social Media Posts. Insofern lässt sich festhalten, dass

die KI-Modelle eng in die jeweilige Arbeitsumgebung integriert werden und eben nicht autonom eine beliebig komplexe Aufgabe übernehmen. Menschliche Kontrolle und Überwachung werden immer noch auf lange Zeit und in vielen Fällen notwendig sein, um qualitativ akzeptable Ergebnisse zu erzielen und die Auswirkungen von Fehleinschätzungen und Fehlaussagen von KI-Systemen einzudämmen. Im Moment kann man diese Fehleinschätzungen nicht systematisch verhindern, da das Verhalten großer KI-Modelle schwer vorherzusehen und kaum erklärbar ist (Krüger 2023).

Dennoch ist es möglich und nicht abzustreiten, initial den Eindruck zu haben, dass in Fällen wie ChatGPT von einer Verallgemeinerbarkeit der Funktionen gesprochen werden könnte. So greift das System ChatGPT auf eine riesige Menge von Daten zurück, die, zumindest auf den ersten Blick, domänenspezifische Normen und Kriterien gut imitieren. Demungeachtet erweist sich ein solcher Eindruck als überzogen, da er die Fehleranfälligkeit der teilweise genutzten stochastischen Methoden unterschätzt: „Das Programm erinnert [...] an einen Wahrsager, der seine eigentliche Prognose in einem Strauß offensichtlicher Wahrheiten präsentiert und es der fragenden Person überlässt, Fakt von Fiktion zu trennen“ (Gurevych 2023). Anders formuliert: „This arguably relies on view-from-nowhere objectivity and propagates an idea that being structurally tone-deaf to the nuances of separate domains of operation”⁶ (Padmanabhan 2023).

Grundlegende Grenzen

Daneben existieren diverse Funktionsgrenzen der genannten Systeme, die im laufenden Diskurs, jenseits der Fachöffentlichkeit, nur sehr verhalten thematisiert werden. So

⁶ Siehe hierzu Padmanabhan (2023) mit Bezug auf Nagel (1989).

handelt es sich selbst bei den gegenwärtig mit großem PR-Aufwand begleiteten Sprachsystemen um wenig mehr als eine „collection of paradigms of statistical pattern discovery and exploitation“ (Padmanabhan 2023). Das bedeutet, dass ChatGPT oder anderweitige KI-basierten Programme grundsätzlich keinerlei „Bewusstsein“ oder „Vernunft“ besitzen – und dies auch nicht in rudimentärer Form, wie Ilya Sutskever, Mitbegründer von Open AI, suggeriert (vgl. Collier 2023) – und obendrein auch erst recht kein ethisches Verständnis von „richtig“ und „falsch“ besitzen, wie beispielsweise durch das KI-basierte Tool „Ask Delphi“ (Noor 2021) kolportiert wird. Die Programme nutzen Statistiken, um aus Wortbausteinen eine Antwort zusammenzustellen, deren Korrektheit seine Algorithmen die größte Wahrscheinlichkeit zumessen (vgl. Gurevych 2023; Otte 2023a). Dies wiederum impliziert grundlegende Funktionsgrenzen von KI-Systemen:

- Zum einen kann den Systemen nur eine mangelnde Fähigkeit der Systeme zur Bewältigung von „offenen Welten“ attestiert werden, d. h. neuer und nur schwer ex ante kalkulierbarer Kontexte und Situationen. Wie schon angesprochen, können sie zwar in begrenzten, standardisierten Anwendungsbereichen als scheinbar „lernende“ Systeme erfolgreich sein und sich an verändernde Situation anpassen. Treten dann allerdings unerwartete Ereignisse und nicht kalkulierbare Sondersituationen auf, wird die „Robustheit“ eines KI-Systems zum kaum bewältigbaren Problem (vgl. Görz et al. 2021: 10). Beispielsweise ist im Fall der Robotik bis heute eines der größten Probleme „die Wahrnehmung der Welt“ (Lenzen 2018: 86), dass umso größer wird, je größer das Repertoire der Handlungen eines Roboters ist.
- Zum Zweiten gibt es bislang keine Möglichkeit, dass das, was man als Erfahrung und Alltagswissen bezeichnet, in die Systemprozesse einzubeziehen. Dazu gehört nicht nur die Fähigkeit der KI, „zu einem gewissen Grad Abstraktionen zu leisten und Kausalitäten maschinell nachzuvollziehen, sondern auch das zu approximieren, was Menschen in besonderer Weise auszeichnet, nämlich Handlungen zu verstehen und zu erklären“ (Görz et al. 2021: 10). Dies gilt offensichtlich auch für die großen Sprachsysteme, die zwar in der Lage sind, Muster relativ präzise zu beschreiben und mehr oder weniger oberflächliche Vorhersagen zu treffen, jedoch keine Erklärungen für bestimmte Phänomene liefern können (vgl. u. a. Larson 2021; Chomsky et al. 2023).

Nicht erfasst werden daher auch kultur- und kontextspezifische Besonderheiten, insbesondere ethische Normen, die stets nur auf der Basis erfahrungsgesättigter Interpretationsleistungen und entsprechender Schlussfolgerungen erbracht werden können (vgl. Gill 2023).

- Zum Dritten geht es um das schon seit den 1990ern diskutierte grundlegende Problem der sogenannten *Explainable Artificial Intelligence*. Für die Gestaltung neuronaler Netze existieren keine theoretisch fundierten Erkenntnisse. „Folglich müssen für jede Aufgabe passende Netzwerkstrukturen und Nutzenfunktionen mühsam mit großem Trainingsaufwand und ohne Erfolgsgarantie ausprobiert werden“ (Brödner 2019a: 76). Oder anders: Die Prozesse subsymbolischer KI-Systeme weisen alle Merkmale eines „theorielosen Probierens“ (Brödner 2022: 34) auf. Je unverzichtbarer deren Entscheidungen für menschliches Handeln allerdings werden, desto wichtiger wird es sein, zu verstehen, was die Systeme eigentlich tun. Das viel zitierte Beispiel hierfür ist die Nutzung eines KI-unterstützten Diagnosetools im medizinischen Bereich, in welchem Diagnoseentscheidungen mitunter mit weitreichenden, gar existenziellen Konsequenzen verbunden sein können. Die hier eingesetzten Systeme sollten daher in der Lage sein, ihre Therapievorschlüsse zu erklären, was in der bisherigen praktischen Realität jedoch nicht der Fall ist, denn: Die Systeme lernen anhand von Beispielen und nicht mithilfe von (theoretischen) Regeln, die die Voraussetzung für eine Erklärung von Entscheidungen wären (vgl. Science Media Center 2021; Krüger 2023). Indes ist zu bezweifeln, dass das „Problem der Explainability“ künstlicher neuronaler Netze wirklich lösbar ist und dass im Zuge ihrer Realisierung auch tatsächlich die gewünschte Transparenz von KI-basierten Entscheidungsprozessen erreicht wird (vgl. Gill 2023; Padmanabhan 2023).
- Zum Vierten verkörpert die Frage nach der Datenqualität und der für bestimmte Anwendungsfälle verfügbaren begrenzten Datenmengen ein kaum lösbares Problem. Schon durch geringfügige Veränderungen der eingegebenen Daten kann die Funktionstüchtigkeit von Systemen des maschinellen Lernens stark beeinträchtigt werden. So existieren viele Fehlleistungen und zahlreiche dokumentierte Beispiele hierfür,

wie z. B. bei der Klassifizierung von Bildern.⁷ Zudem wird immer wieder berichtet, dass auf den ersten Blick zwar die für lernende Systeme erforderlichen großen Mengen von Daten zur Verfügung stehen. Bei genauerem Hinsehen wird jedoch deutlich, dass diese Daten aufgrund ihrer schlechten Qualität und eines mangelhaften Formats nicht nutzbar sind (Stichwort: *Dirty Data*) und sehr aufwendig angepasst werden müssen (vgl. Uppenkamp 2020). Insbesondere im medizinischen Bereich stellt die mangelnde Validität der Daten oftmals ein nicht lösbares Problem dar (vgl. Kaltheuner 2021: 35). Eine weiterhin bestehende Fehleinschätzung besteht auch in der Fehlannahme, infolge einer möglichst umfänglichen „Reparatur der Datensätze“ unverzerrte, fehlerfreie Datensätze, die eine objektive Wahrheitserschließung ermöglichen, zu erhalten. Kritiker:innen erwidern: „Diese Reparaturen können Datensätze definitiv verbessern, sogar massiv, und bestimmte Diskriminierungsvektoren ausblenden. Sie werden durch die Anpassungen aber niemals komplett eliminiert. Und manchmal erzeugen diese kleinen Reparaturen sogar neue Fehler, die der Datensatz vorher nicht hatte“ (Geuter 2018). Schließlich geht es um das Problem, dass, etwa in industriellen Prozessen, nur geringe Mengen oft multimodaler, d. h. qualitativ sehr unterschiedlicher Daten, für einen effizienten Einsatz von lernenden Systemen verfügbar sind. Stichwort ist hier *Small Data* statt Big Data. Wie auch unter den Bedingungen geringer Datenmengen lernende Systeme genutzt werden können, ist ein bislang allenfalls in ersten Ansätzen gelöstes Problem (vgl. Hirsch-Kreinsen 2023: 158 f.). So werden im Zuge der Erzeugung künstlich generierter Daten, sogenannte *Synthetic Data*, in den vergangenen Jahren Fortschritte erzielt, wenn es um die Kompensation von Datensätzen mit einer geringen Menge an realen Daten geht (vgl. u. a. Alkhalifah et al. 2022; Savage 2023). Freilich bleibt dabei die Frage unbeantwortet, inwieweit mithilfe per Simulation erzeugter Daten damit alle prinzipiell möglichen und u. U. risikoträchtigen Situationen, etwa im Fall des autonomen Fahrens, auch erfasst werden können.

⁷ Vgl. Brödner (2019a: 76) und die dort angegebene Literatur.

Das Technologieversprechen der KI impliziert jedoch die schon angesprochene Erwartung, dass die aufgeführten Grenzen in absehbarer Zeit von „Durchbrüchen“ bei der Technologieentwicklung überwunden werden können (vgl. Ford 2018; Barthelmeß & Furbach 2021; Armbruster 2022). Konträr zu dieser Erwartungshaltung sind auch im aktuellen Fall der großen Sprachsysteme keine signifikanten Quantensprünge erkennbar und darüber hinaus in Zukunft absehbar. Es handelt sich dabei eben nur um einen weiteren evolutionären Entwicklungsschritt der KI, mit dem verfügbare Methoden und neue Erkenntnisse der Computerlinguistik, auf der Basis massiv angestiegener Rechnerleistungen und Speicherkapazitäten sowie der Verfügbarkeit über große Datenmengen, weiterentwickelt und jetzt in breiteren Kontexten nutzbar gemacht worden sind (vgl. Breher & Lehmann 2023).

4. Weiterungen: Der Mythos der intelligenten Maschine

Jenseits aller Visionen und Befürchtungen, die sich mit KI verbinden, bezieht das Technologieversprechen seine dauerhafte Überzeugungskraft letztlich aus dem „Mythos der intelligenten Maschine“, die der menschlichen Intelligenz ebenbürtig, wenn nicht gar überlegen sei. Kern dieses Mythos ist, dass Maschinen geschaffen werden können, mit denen die körperlichen und intellektuellen Unzulänglichkeiten des *Mängelwesens Mensch* (Arnold Gehlen) überwunden werden können (vgl. Marquis et al. 2020). Seine Faszination begründet sich „in der unfassbaren Kühnheit des Anspruchs“, Maschinen Denkleistungen und intelligentes Verhalten vollbringen zu lassen, die den *antiquierten Menschen* (Günther Anders) seine ganze Inferiorität gegenüber seinen Machwerken empfinden lässt (zit. n. Brödner 1997: 187). In dieser Logik wird KI als eine Voraussetzung dafür angesehen, dass Zivilisation und Kultur gewissermaßen durch Maschinenintelligenz in alle Zukunft gesichert werden könnten, da Menschen dazu nicht in der Lage seien. Die Verheißung ist, dass KI „Kultur auf noch nie da gewesene Weise zu bewahren“ ermögliche und „ein Gefäß der menschlichen Geschichte und Kultur“ werde. Während Zivilisationen zerbrechlich seien und „menschliches Fleisch ein brüchiges Medium“ sei, werde KI „eine tektonische Verschiebung in unserer Fähigkeit bewirken, Kultur zu schaffen, zu vermitteln und zu bewahren“ (Rahwan & Powers 2022).

Der Mythos der intelligenten oder auch denkenden Maschine geht bekanntermaßen bis weit in die Antike zurück und hat seit jeher die Visionen von Philosophen:innen, Schriftstellern:innen, Filmemachern:innen und vielen anderen Künstlern:innen stimuliert. Er steht im Kontext einer in der Menschheitsgeschichte lange zurückreichenden Tradition von fantasievollen Geschichten über die intelligente Maschine, die in der KI-Literatur zugespitzt auch als *AI Narratives* gefasst werden (vgl. Nilsson 2010; Cave et al. 2020). Sie reichen von der griechischen Mythologie, etwa der bronzene Riese Talos zum Schutze von Kreta, über intelligent erscheinende künstliche Wesen in der Neuzeit wie dem „Schachtürken“ von Wolfgang von Kempelen bis hin zum Computer HAL aus Stanley Kubricks Film *A Space Odyssey*.⁸

Diese Visionen prägen den Gründungsmythos der KI, welcher sich bis heute als übergeordnetes und generelles Entwicklungsziel wie ein roter Faden, auch gerade durch die verschiedenen Entwicklungsphasen der KI, zieht (vgl. u. a. Natale & Ballatore 2020). Human Level Artificial Intelligence, so Nils Nilsson in seinem Werk „The Quest for Artificial Intelligence: A History of Ideas and Achievements“ aus dem Jahr 2010, sei schon das Ziel der Gründer:innengeneration und vieler anderer früher KI-Forscher:innen gewesen und es sei bis heute die Vision vieler Forscher:innen, die mit der erwähnten Formel der Artificial General Intelligence gefasst wird (vgl. Nilsson 2010: 525 ff.). Die KI-Forschung zielt von ihrem Anbeginn an darauf ab, Modelle menschlicher kognitiver und perzeptiver Fähigkeiten zu entwickeln und diese Fähigkeiten auf Computern zu reproduzieren. Dies wird explizit im Gründungsproposal von McCarthy formuliert: Kognitive Leistungen seien als Informationsverarbeitung zu verstehen, die daher computertechnologisch automatisiert werden könnten (vgl. McCarthy et al. [1955] 2006). Ein fundamentaler und holistischer Ausdruck dieses Ziels ist auch der sogenannte General Problem Solver von Herbert Simon und Allen Newell aus den 1950ern, der auf die Automatisierung von Common-Sense-Reasoning und die Entwicklung lernender Systeme zielt (vgl. Simon & Newell 1958). Als weiteres frühzeitliches Beispiel kann das 1958 von Frank Rosenblatt vorgestellte Konzept eines neuronalen Netzes (vgl. Olazaran 1996) angesehen werden. Das

⁸ Vgl. z. B. Dittmann (2005: 133 ff.) und Marquis et al. (2020: 4 ff.).

Erscheinen des Konzepts veranlasste die New York Times damals zu der Einschätzung, dass jenes System derart zu verstehen sei, dass „the embryo of an electronic computer that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence“ (zit. n. Hardt & Recht 2022: 6).

Sehr viel später dann erneuern in Deutschland die renommierten Informatiker um Barth et al. im Jahr 1991, nach dem Auslaufen des Booms über Expertensysteme gegen Ende der 1980er Jahre, diese Vision der KI. Sie knüpften explizit an den Mythos der intelligenten Maschine und seine Bedeutung für die zukünftige gesellschaftliche Entwicklung an und hobten das „Intelligenzphänomen“ als entscheidendes Thema für die Perspektive der Disziplin hervor (Barth et al. 1991: 202). Wie die weitreichenden Versprechungen einer AGI zeigen, ist dieses Thema insbesondere auch im aktuellen Diskurs nicht wegzudenken. Angesichts der erreichten technologischen Fortschritte der KI, insbesondere der wachsenden Rechnerleistungen, erwarten viele KI-ler:innen eine beschleunigte Entwicklung hin zu Systemen der *starken KI*. Kritisch formuliert dies der amerikanische Tech-Unternehmer und Informatiker Erik J. Larson: „The myth of artificial intelligence is that its arrival is inevitable, and only a matter of time – that we have already embarked on the path that will lead to human-level AI, and then superintelligence“ (Larson 2021: 1).⁹ Auch wenn im geschichtlichen Entwicklungsverlauf der KI-Dynamik bis heute anwendungsorientierte Konzepte einer sogenannten *schwachen KI* im Vordergrund der Entwicklung stehen, hat der Mythos einer intelligenten Maschine ganz offensichtlich den Charakter einer faszinierenden und übergreifenden Orientierung für zukünftige Forschungsanstrengungen der KI nicht verloren. Den Kern dieser Perspektiven repräsentieren vielfach erwartete Durchbrüche bei der KI-Entwicklung in den nächsten Jahren und die daran gekoppelten Hoffnungen und Erwartungen, neue Möglichkeiten für die Erreichung des Zieles des Konzepts einer AGI zu generieren und diesem Ziel näher zu kommen (vgl. u. a. Ford 2018; Barthelmeß & Furbach 2021). Insgesamt nimmt der Mythos der intelligenten Maschine auch durch Kritik und Skepsis keinen Schaden, ganz im Gegen-

⁹ Vgl. hierzu insbesondere die überaus instruktiven Interviews mit global führenden KI-Wissenschaftlern:innen in Ford (2018).

teil: Wie historische Studien belegen, werden technologische Mythen infolge der Konfrontation mit Kritik und Skepsis gar noch verstärkt (vgl. Natale & Ballatore 2020). Dies gilt ohne Frage für die in der Vergangenheit überzogenen und nicht eingelösten Versprechen, die zu einem hohen Maß an Enttäuschungen und Skepsis führten und die AI-Winter nach sich zogen.

Jedoch gilt es hier auch zu konstatieren, dass in der bisherigen Historie bislang auf jeden „Winter“ stets ein erneuter Aufschwung, ja ein neuer KI-Hype folgte (vgl. Bostrom 2014; Teich 2020; Hirsch-Kreinsen 2023). Dies gilt auch für die kritische Diskussion über die skizzierten grundlegenden Entwicklungs- und Anwendungsgrenzen der KI, die die Geschichte der KI seit ihrem Anbeginn an begleitet. Schon frühe fundamentale KI-Kritiker wie Dreyfus oder Weizenbaum sprechen der KI über eine formalisierte Wissensrepräsentation hinausgehende Fähigkeiten zu einem Kontext- und Weltverständnis ab (vgl. Dreyfus 1972;1992; Weizenbaum 1978). Ähnlich klingen aktuelle Argumentationen, gemäß welcher gerade auch künstliche neuronale Netze wenig mehr als lediglich adaptive Systeme für spezifisch programmierte Anwendungsfelder seien. Eines der grundlegenden kritischen Argumente lautet, dass menschliche Intelligenz und Denken sich nicht mit der Fähigkeit des Berechnens gleichsetzen oder nachahmen ließen. KI-Systeme „denken“ nicht, sondern vollziehen nur von Menschen entwickelte und programmierte Algorithmen. Intuition, „Abduktion“ (Charles Sanders Peirce), Begreifen und Urteilen seien menschliche kognitive Leistungen, die weit über maschinelles Berechnen und Schließen hinausgingen. Auch wird mit solchen prinzipiellen Argumenten die Erwartung von absehbaren Durchbrüchen bei der weiteren KI-Entwicklung hin zu einer AGI vehement in Abrede gestellt (vgl. Dickson 2021; Larson 2021; Otte 2023a).

Ohne Frage haben die Entwicklung und die Anwendung der KI beachtliche Erfolge gezeitigt. Es ist unübersehbar, dass die Leistungsfähigkeit der IT-Systeme kontinuierlich steigt und fraglos werden in vielen Anwendungsbereichen menschliche Fähigkeiten von der Leistungsfähigkeit der KI deutlich überschritten. Auch die Highlights der KI, die Siege im Schachspiel oder beim Go-Spiel, sind eindrucksvoll, weshalb die intelligente Maschine gegenwärtig erstmalig in der Menschheitsgeschichte von einigen Vertretern:innen als realisierbar erscheint. Mit diesen kontinuierlichen Fortschritten wird der

Mythos der ultimativen Möglichkeit, den Geist als Maschine zu implementieren, stets aufs Neue bestärkt (vgl. Brödner 1997: 238). Aktuelle Belege hierfür sind eben auch Systeme wie ChatGPT, die den Eindruck erwecken, man kommuniziere nicht mit einer Maschine, sondern mit einem verständnisvollen Gesprächspartner. Die Faszination dieses Mythos begründet sich zu einem Gutteil in dem Umstand, dass KI für Viele, gerade auch für jene aus dem Bereich der Innovationspolitik und die begeisterten Laien, in ihren Grundlagen, ihrer Funktionsweise, aber auch in Hinblick auf ihre Grenzen nur wenig transparent ist und letztlich rätselhaft bleibt. Zugleich jedoch erlauben die Mehrdeutigkeit und die Genialität des Begriffs Künstliche Intelligenz sowie die eingängigen Metaphern des Technologieversprechens, dass weitreichende Fähigkeiten mit dieser Technologie verbunden werden können.

Moment des Mythos sind zudem technikutopische Verheißungen. Das meint, dass mit KI für viele Protagonisten die Erwartung einhergeht, die zunehmende und unübersichtliche gesellschaftliche Komplexität steuernd und regulierend bewältigen und damit Unsicherheiten und drohende Kontrollverluste vermeiden zu können. Eine technikutopische Vision von einer dadurch möglichen rationalen Planung und Steuerung gesellschaftlicher Entwicklung durch Technologieanwendung, jenseits von intransparenter und bestenfalls schwer regulierbarer sozialer Komplexität und aufwendigen demokratischen Prozeduren, scheint nun endgültig in Reichweite zu liegen. Es handelt sich um jene Erwartung, die der Internet-Kritiker Evgeny Morozov als „Solutionismus“ (Morozov 2013: 5) gefasst hat. Anders formuliert: Politik, demokratische Prozesse und gesellschaftliche Herausforderungen werden im Sinne der Informationsverarbeitungs-These als Problem der Modellbildung, Algorithmierung und einer optimalen Datenauswertung begriffen. Diese Idee formuliert präzise schon Edward Fredkin, ein KI-Forscher in den USA in den 1950er-Jahren: „The idea was to view the world as a great computer and to write a program [the ›global algorithm‹] that, if methodically executed, would lead to peace and harmony“ (zit. n. Noble 1997: 165).

Der Mythos KI gewinnt damit einmal mehr seine Bedeutung, weil die Technologie erwarten lässt, dass sie Situationen bewältigt, die Unsicherheit und Angst vor Kontrollverlust erzeugen. Insofern hat der Mythos der intelligenten Maschine eine doppelte Funktion:

Zum einen ist er eine wirksame Form der Komplexitätsreduktion und zum anderen hat er angesichts der großen gesellschaftlichen Herausforderungen und dem notwendigen weitreichenden Wandel eine entlastende Funktion, und: Die Highlights der KI scheinen zu belegen, dass die zu meisternden Aufgaben bewältigt werden können. Technologisch werden damit kollektive Orientierungen und Zuversicht geschaffen, die ganz wesentliche kognitive und emotionale Ressourcen für gesellschaftliches Handeln bilden (vgl. Münkler 2017: 11).

Wichtige Voraussetzung für die anhaltende Faszination der KI sind freilich stets auch rituelle und erneuerte Inszenierungen des Mythos, mit denen ihm neue Überzeugungskraft und vor allem auch erweiterte Erwartungshorizonte verliehen werden und sich die treibenden Protagonisten im Zuge dessen ihrer Erwartungen auch selbst vergewissern. Im Fall der KI-Dynamik spielen die verschiedenen Festveranstaltungen, Publikationen, Seminare und Konferenzen der einflussreichen KI-Community und vor allem auch innovationspolitische Inszenierungen der verschiedensten Art, wie etwa die jährlichen Digitalgipfel der Bundesregierung, eine einflussreiche Rolle. Auf politisch höchster Ebene werden die technikutopischen Erwartungen hier nicht nur erneuert, sondern auch fortgeschrieben und neue Horizonte für eine erwartbare Zukunft rhetorisch aufgeschlossen. Mit Rückgriff auf die Mythostheorie kann daher auch vom stetigen „Basteln am Mythos KI“ (ebd.: 22) gesprochen werden –, d. h., er ist immer wieder Gegenstand von Diskursen und seine Überzeugungskraft wird durch stets wieder aufkeimende Hypes erneuert. In diesem Sinne ist der Mythos der KI weder richtig noch falsch. Vielmehr ist er mal mehr und mal weniger vital, wird jedoch kontinuierlich am Leben erhalten (vgl. Natale & Ballatore 2020). Einerseits erfordert dies, dass der Mythos kontinuierlich weiterentwickelt wird. Andererseits aber droht stets die Veralltäglichsung und Verwässerung des Narrativs. Darüber hinaus droht der Umstand, was richtig ist und geglaubt werden sollte, an Schärfe zu verlieren. Diese Gefahr ist im Fall von Technologiemythen nicht auszuschließen, da, wie im Fall der KI-Dynamik, die zunehmend breitere Nutzung der Technologie zu ihrer Entmythisierung führen kann. Im Hinblick auf KI ist jedoch davon auszugehen, dass die Wirkmächtigkeit des in der Menschheitsgeschichte tief verankerten Mythos der

intelligenten Maschine angesichts der weiteren wissenschaftlichen und technologischen Entwicklungen durchaus erneut unter Beweis gestellt werden kann. Auf dieser Basis erhält das Technologieversprechen KI seine Überzeugungskraft und wird wohl auch in Zukunft kontinuierlich fortgeschrieben.

Literatur

- Abrahamson, Eric (1996): *Management Fashion*. In: *Academy of Management Review*, Vol. 21, No. 1, 254–285.
- Ahrweiler, Petra (1995): *Künstliche Intelligenz-Forschung in Deutschland. Die Etablierung eines Hochtechnologie-Fachs*. Münster/New York: Waxman.
- Alkhalifah, Tariq/ Wang, Hanchen & Ovcharenko, Oleg (2022): *MLReal: Bridging the Gap between Training on Synthetic Data and Real Data Applications in Machine Learning*. In: *Artificial Intelligence in Geosciences*, Vol. 3, 103–114.
- Armbruster, Alexander (2022): *Große Künstliche Intelligenz*. Interview mit Jonas Andruslis. *Frankfurter Allgemeine Zeitung*, Stand: 17. Mai 2022. Online: <https://zeitung.faz.net/faz/unternehmen/2022-05-16/75949fd8e1ead69fbb32e0132dd78f8c/?GEPC=s5> [17. Mai 2022].
- Bareis, Jascha & Katzenbach, Christian (2021): *Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and their Performative Politics*. In: *Science, Technology, & Human Values*, Vol. 47, No. 5, 855–881.
- Barth, Gerhard/ Christaller, Thomas/ Cremers, Armin B./ Neumann, Bernd/ Radermacher, Franz J./ Radig, Bernd /Richter, Michael M./ Siekmann, Jörg H. & Von Seelen, Werner (1991): *Künstliche Intelligenz – Perspektiven einer wissenschaftlichen Disziplin und Realisierungsmöglichkeiten*. In: *Informatik Spektrum*, Vol. 14, No. 4, 201–206.
- Barthelmeß, Ulrike & Furbach, Ulrich (2021): *Computer auf dem Weg zum Bewusstsein*. *Frankfurter Allgemeine Zeitung*, Stand: 06. Juli 2021. Online: <https://www.faz.net/aktuell/wirtschaft/digitec/computer-auf-dem-weg-zum-bewusstsein-17421543.html?premium> [27. Juli 2021].
- Bauberger, Stefan/ Beck, Birgit/ Burchardt, Aljoscha & Remmers, Peter (2021): *Ethische Fragen der Künstlichen Intelligenz*. In: Görz, Günther/ Schmid, Ute & Braun, Tanya (Hrsg.): *Handbuch der Künstlichen Intelligenz*. 6. Aufl. Berlin/Boston: De Gruyter Oldenbourg, 905–934.

- Beckert, Jens (2016): *Imagined Futures: Fictional Expectations and Capitalist Dynamics*. Cambridge: Harvard University Press.
- Bendel, Oliver (2023): *Generative KI*. Stand: 2023. Online: <https://wirtschaftslexikon.gabler.de/definition/generative-ki-124952> [23. Mai 2023].
- Bender, Gerd (2005): *Technologieentwicklung als Institutionalisierungsprozess*. In: *Zeitschrift für Soziologie*, Vol. 34, No. 3, 170–187.
- Birhane, Abeba/ Kalluri, Pratyusha/ Card, Dallas/ Agnew, William/ Dotan, Ravit & Bao, Michelle (2022): *The Values encoded in Machine Learning Research*. 2022 ACM Conference on Fairness, Accountability, and Transparency.
- Bischoff, Manon (2023): *Wie man einem Computer das Sprechen beibringt*. Spektrum, Stand: 09. März 2023. Online: https://www.spektrum.de/news/wie-funktionieren-sprachmodelle-wie-chatgpt/2115924?utm_source=flipboard&utm_content=Spektrumverlag%2Fmagazine%2FIT%2FTechnik [05. Juni 2023].
- Borup, Mads/ Brown, Nik/ Konrad, Kornelia & Van Lente, Harro (2006): *The Sociology of Expectations in Science and Technology*. In: *Technology Analysis & Strategic Management*, Vol. 18, No. 3/4, 285–298.
- Bostrom, Nick (2014): *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Breher, Nina & Lehmann, Hendrik (2023): *Vordenker zu ChatGPT: Hype oder Revolution?* Tagesspiegel, Stand: 26. März 2023. Online: <https://www.tagesspiegel.de/chancen-und-gefahren-der-kunstlichen-intelligenz-das-ist-schon-ziemlich-revolutionar-9558341.html> [20. Juli 2023].
- Brödner, Peter (1997): *Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen*. Berlin: edition sigma.
- Brödner, Peter (2019): *Coping with Descartes' Error in Information Systems*. In: *AI & SOCIETY*, Vol. 34, No. 2, 203–213.
- Brödner, Peter (2019a): *Grenzen und Widersprüche der Entwicklung und Anwendung ›Autonomer Systemen‹*. In: Hirsch-Kreinsen, Hartmut & Karacic, Anemari (Hrsg.): *Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt*. Bielefeld: Transcript Verlag, 69–100.
- Brödner, Peter (2022): *Die Illusionsfabrik der KI-Narrative*. In: *FLjF-Kommunikation*, Vol. 22, No. 2, 32–36.
- Brödner, Peter (2023): *Künstliche Intelligenz: Dichtung und Wahrheit Einblicke in die Technik des Berechnens und in Mythen um Intelligenz*. [Unveröffentl. Manuskript].
- Brown, Nik/ Rappert, Brian & Webster, Andrew (2016): *Introducing Contested Futures: From Looking into the Future to Looking at the Future*. In: Brown, Nik & Rappert, Brian

- (Hrsg.): *Contested Futures. A Sociology of Prospective Techno-Science*. New York: Routledge, 3–20.
- Buchanan, Bruce G. (2006): *A (Very) Brief History of Artificial Intelligence*. In: *AI Magazine*, Vol. 26, No. 4, 53–60.
- Bundesregierung (2018): *Strategie Künstliche Intelligenz der Bundesregierung*. Stand: 16. November 2018. Online: <https://www.bmwi.de/Redaktion/DE/Publikationen/Technologie/strategie-kuenstliche-intelligenz-der-bundesregierung.html> [01. Juli 2021].
- Cath, Corinne/ Wachter, Sandra/ Mittelstadt, Brent/ Taddeo, Mariarosaria & Floridi, Luciano (2017): *Artificial Intelligence and the ›Good Society‹: the US, EU and UK Approach*. In: *Science and Engineering Ethics*, Vol. 24, No. 2, 505–528.
- Cave, Stephen/ Dihal, Kanta & Dillon, Sarah (2020): *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press.
- Chalmers, David (2023): *Wie ist es, ChatGPT zu sein?* Frankfurter Allgemeine Zeitung, Stand: 09. März 2023. Online: <https://www.faz.net/aktuell/feuilleton/david-chalmers-kann-chatgpt-ein-bewusstsein-entwickeln-18714811.html> [15. Juni 2023].
- Chomsky, Noam/ Roberts, Ian & Watumull, Jeffrey (2023): *The False Promise of ChatGPT*. The New York Times, Stand: 08. März 2023. Online: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html> [10. März 2023].
- Christmann, Karin & Fiedler, Maria (2023): *Ab 2035 wird es keinen Job mehr geben, der nichts mit KI zu tun hat*. Tagesspiegel, Stand: 29. April 2023. Online: <https://www.tagesspiegel.de/politik/arbeitsminister-uber-kunstliche-intelligenz-ab-2035-wird-es-keinen-job-mehr-geben-der-nichts-mit-ki-zu-tun-hat-9733466.html> [29. April 2023].
- Collier, Kevin (2023): *What is Consciousness? ChatGPT and advanced AI might redefine our Answer*. NBC News, Stand: 28. Februar 2023. Online: <https://www.nbcnews.com/tech/tech-news/chatgpt-ai-consciousness-rcna71777> [27. Juni 2023].
- Cyranek, Günther & Coy, Wolfgang (Hrsg.) (1994): *Die maschinelle Kunst des Denkens – Perspektiven und Grenzen der Künstlichen Intelligenz*. Braunschweig/Wiesbaden: Vieweg.
- Dickel, Sascha & Schrape, Jan-Felix (2015): *Dezentralisierung, Demokratisierung, Emanzipation – Zur Architektur des digitalen Technikutopismus*. In: *Leviathan*, Vol. 43, No. 3, 442–462.

- Dickson, Ben (2021): *Abductive Inference: The blind Spot of Artificial Intelligence*. TechTalks, Stand: 20. September 2021. Online: <https://bdtech-talks.com/2021/09/20/myth-of-artificial-intelligence-erik-dduniekon/> [14. Februar 2022].
- Dittmann, Frank (2005): *Maschinenintelligenz zwischen Wunsch und Wirklichkeit*. In: Pias, Claus (Hrsg.): *Zukünfte des Computers*. Zürich/Berlin: Diaphanes, 133–154.
- Dreyfus, Hubert (1972): *What Computers can't do: A Critique of Artificial Reason*. New York: Harper & Row.
- Dreyfus, Hubert (1992): *What Computers still can't do*. New York: Harper & Row.
- Dyer-Witheford, Nick/ Kjösen, Atle Mikkola & Steinhoff, James (2019): *Inhuman Power: Artificial Intelligence and the Future of Capitalism*. London: Pluto Press.
- EFI (Expertenkommission Forschung und Innovation) (2018): *Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands*, Berlin.
- EFI (Expertenkommission Forschung und Innovation) (2022): *Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands*, Berlin.
- Enquetekommission (2020): *Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale*. BT-Drucks, 19/23700, Vorabfassung, Berlin.
- Erixon, Fredrik & Weigel, Björn (2016): *The Innovation Illusion: How so Little is created by so Many working so hard*. New Haven/London: Yale University Press.
- Fagerberg, Jan (2005): *Innovation: A Guide to the Literature*. In: Fagerberg, Jan/ Mowery, David & Nelson, Richard R. (Hrsg.): *The Oxford Handbook of Innovation*. Oxford: Oxford University Press, 1–27.
- Faust, Michael (2021): *Finanzialisierung und Digitalisierung von Unternehmen: Konzeptionelle Überlegungen und empirische Annäherungen*. In: Buss, Klaus-Peter/ Kuhlmann, Martin/ Weißmann, Marliese/ Wolf, Harald & Apitzsch, Birgit (2021): *Digitalisierung und Arbeit*. Frankfurt/New York: Campus Verlag, 67–106.
- Ford, Martin (2018): *Architects of Intelligence: The Truth about AI from the People building it*. Birmingham: Packt Publishing.
- Fühlinger, Josh (2023): *Was ist Generative AI?* Computerwoche, Stand: 17. März 2023. Online: <https://www.computerwoche.de/a/was-ist-generative-ai,3614061> [10. Juni 2023].
- Geels, Frank W. & Smit, Wim A. (2000): *Failed Technology Futures: Pitfalls and Lessons from a Historical Survey*. In: *Futures*, Vol. 32, No. 9, 867–885.

- Geuter, Jürgen (2018): *Nein, Ethik kann man nicht programmieren*. Zeit Online, Stand: 27. November 2018. Online: <https://www.zeit.de/digital/internet/2018-11/digitalisierung-mythen-kuenstliche-intelligenz-ethik-juergen-geuter> [26. Juni 2023].
- Gill, Karamjit S. (2023): *Algorithmic Agency: Technological Exuberance or a Cultural Crisis*. Vortrag, Conference Industry 4.0 Ten Years Later, Modena, May 19, 2023.
- Görz, Günther/ Braun, Tanya & Schmid, Ute (2021): *Einleitung*. In: Görz, Günther/ Schmid, Ute & Braun, Tanja (Hrsg.): *Handbuch der Künstlichen Intelligenz*. 6. Aufl. Berlin/Boston: De Gruyter Oldenbourg, 1–26.
- Görz, Günther/ Schmid, Ute & Braun, Tanja (Hrsg.) (2021a): *Handbuch der Künstlichen Intelligenz*. 6. Aufl. Berlin/Boston: De Gruyter Oldenbourg .
- Gurevych, Iryna (2023): *So begrenzt ist ChatGPT – noch*. Frankfurter Allgemeine Zeitung, Stand: 02. März 2023. Online: <https://www.faz.net/aktuell/wirtschaft/digitec/chatgpt-die-kuenstliche-intelligenz-ist-noch-sehr-begrenzt-18690464.html> [15. April 2023].
- Hardt, Moritz & Recht, Benjamin (2022): *Patterns, Predictions, and Actions: Foundations of Machine Learning*. Princeton: Princeton University Press.
- Heckel, Reinhard (2023): *Was jeder jetzt über ChatGPT & Co. wissen muss*. Frankfurter Allgemeine Zeitung, Stand: 19. Juli 2023. Online: <https://www.faz.net/aktuell/wirtschaft/digitec/ki-was-jeder-ueber-chatgpt-wissen-sollte-19037664.html> [20. Juli 2023].
- Hirsch-Kreinsen, Hartmut (2023): *Das Versprechen der Künstlichen Intelligenz. Gesellschaftliche Dynamik einer Schlüsseltechnologie*. Frankfurt/New York: Campus Verlag.
- Horowitz, Michael C. (2018): *Artificial Intelligence, International Competition, and the Balance of Power*. In: *Texas National Security Review*, Vol. 1, No. 3, 36–57.
- Kaltheuner, Frederike (2021): *Fake AI*. Manchester: Meatspace Press.
- Katzenbach, Christian (2021): *“AI will fix this” – The Technical, Discursive, and Political Turn to AI in Governing Communication*. In: *Big Data & Society*, Vol. 8, No. 2, 1–8.
- Kieser, Alfred (1996): *Moden & Mythen des Organisierens*. In: *Die Betriebswirtschaft*, Vol. 56, No. 1, 21–39.
- Koehler, Jana (2021): *Zum Begriff der Künstlichen Intelligenz*. In: Catani, Stephanie & Pfeiffer, Jasmin (Hrsg.): *Handbuch Künstliche Intelligenz und die Künste*. [Preprint Version], 1–15.
- Konrad, Erhard (1998): *Zur Geschichte der Künstlichen Intelligenz in der Bundesrepublik Deutschland*. In: Seifkes, Dirk/ Stach, Heike/ Eulenhöfer, Peter & Städtler, Klaus

- (Hrsg.): Sozialgeschichte der Informatik. Studien zur Wissenschafts- und Technikforschung. Wiesbaden: Deutscher Universitätsverlag, 287–296.
- Kroll, Henning/ Berghäuser, Hendrik/ Blind, Knut/ Neuhäusler, Peter/ Scheifele, Fabian/ Thielmann, Axel & Wydra, Sven (2022): *Schlüsseltechnologien*. In: Studien zum deutschen Innovationssystem, Nr. 7, Berlin.
- Krüger, Antonio (2023): *Worauf es in der Künstlichen Intelligenz jetzt ankommt*. Frankfurter Allgemeine Zeitung, Stand: 17. Mai 2023. Online: <https://www.faz.net/aktuell/wirtschaft/digitec/worauf-es-in-der-kuenstlichen-intelligenz-jetzt-ankommt-18875988.html> [27. Mai 2023].
- Larson, Erik J. (2021): *The Myth of Artificial Intelligence: Why Computers can't think the Way we do*. Cambridge: Harvard University Press.
- Lenzen, Manuela (2018): *Künstliche Intelligenz: Was sie kann & was uns erwartet*. München: Verlag C. H. Beck.
- Madsen, Dag Ø. (2019): *The Emergence and Rise of Industry 4.0 viewed through the Lens of Management Fashion Theory*. In: Administrative Sciences, Vol. 9, No. 3, 71.
- Mager, Astrid & Katzenbach, Christian (2021): *Future Imaginaries in the Making and Governing of Digital Technology: Multiple, contested, commodified*. In: New Media & Society, Vol. 23, No. 2, 223–236.
- Marquis, Pierre/ Papini, Odile & Prade, Henri (2020): *Elements for a History of Artificial Intelligence*. In: Marquis, Pierre/ Papini, Odile & Prade, Henri (Hrsg.): A Guided Tour of Artificial Intelligence Research. Knowledge Representation, Reasoning and Learning. Cham: Springer, 1–44.
- McAfee, Andrew & Brynjolfsson, Erik (2017): *Machine – Platform – Crowd: Harnessing our Digital Future*. New York/London: Norton.
- McCarthy, John/ Minsky, Marvin L./ Rochester, Nathaniel & Shannon, Claude E. ([1955] 2006): *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955*. In: AI Magazine, Vol. 27, No. 4.
- McClelland, Calum (2017): *The Difference Between Artificial Intelligence, Machine Learning, and Deep Learning*. Medium, Stand: 4. Dezember 2017. Online: <https://medium.com/iotforall/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning-3aa67bff5991> [19. Mai 2021].
- Morozov, Evgeny (2013): *To Save Everything, Click here: Technology, Solutionism and the Urge to Fix Problems that don't exist*. New York: Penguin Books Limited.
- Münkler, Herfried (2017): *Die Deutschen und ihre Mythen*. 4. Aufl. Berlin: Rowohlt.
- Nagel, Thomas (1989): *The View from Nowhere*. Oxford: Oxford University Press.

- Natale, Simone & Ballatore, Andrea (2020): *Imagining the thinking Machine: Technological Myths and the Rise of Artificial Intelligence*. In: *Convergence: The International Journal of Research into New Media Technologies*, Vol. 26, No. 1, 3–18.
- Nilsson, Nils J. (2010): *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge: Cambridge University Press.
- Noble, David F. (1997): *The Religion of Technology: The Divinity of Man and the Spirit of Invention*. New York: Penguin Books.
- Noor, Poppy (2021): *'Is it OK to ...': The Bot that gives you an instant Moral Judgment*. *The Guardian*, Stand: 02. November 2021. Online: <https://www.theguardian.com/technology/2021/nov/02/delphi-online-ai-bot-philosophy> [27. Juni 2023].
- Olazaran, Mikel (1996): *A Sociological Study of the official History of the Perceptrons Controversy*. In: *Social Studies of Science*, Vol. 26, No. 3, 611–659.
- Ordonez, Victor/ Dunn, Taylor & Noll, Eric (2023): *OpenAI CEO Sam Altman says AI will reshape Society, acknowledges Risks: 'A little bit scared of this'*. *abcNews*, Stand: 16. März 2023. Online: <https://abcnews.go.com/Technology/openai-ceo-sam-altman-ai-reshape-society-acknowledges/story?id=97897122;%20Zugriff%2030.05.2023%20sowie%20https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html> [21. Juli 2023].
- Otte, Ralf (2023): *Die große KI-Illusion*. *Frankfurter Allgemeine Zeitung*, Stand: 27. Mai 2023. Online: <https://www.faz.net/aktuell/wirtschaft/digitec/die-grosse-ki-illusion-eine-gegenrede-gegen-den-hype-um-chatgpt-co-18908863.html> [27. Mai 2023].
- Otte, Ralf (2023a): *Es gibt kein Auslöschungsrisiko durch diese Künstliche Intelligenz*. *Frankfurter Allgemeine Zeitung*, Stand: 19. Juni 2023. Online: <https://zeitung.faz.net/faz/unternehmen/2023-06-19/1c282217517431440fd441390886a000?GEPC=s5> [20. Juni 2023].
- Padmanabhan, Deepak (2023): *ChatGPT is not OK! That's not (just) because it lies*. In: *AI & SOCIETY*, 2023.
- Pfeiffer, Sabine (2018): *Technisierung von Arbeit*. In: Böhle, Fritz G./ Voß, Günter & Wachtler, Günther (Hrsg.): *Handbuch Arbeitssoziologie*. 2. Aufl., Bd. 1: Arbeit, Strukturen und Prozesse. Wiesbaden: VS Verlag für Sozialwissenschaften, 321–357.
- Rahwan, Iyad & Powers, William (2022): *KI, die Weise*. In: *Frankfurter Allgemeine Zeitung*, Stand: 20. Mai 2022. Online: <https://www.faz.net/aktuell/wirtschaft/digitec/kuenstliche-intelligenz-und-die-aufbewahrung-unserer-kultur-18042029.html> [21. Mai 2022].
- Rogers, Everett M. (2003): *Diffusion of Innovations*. 5. Aufl. New York: Free Press.

- Roose, Kevin (2023): *GPT-4 Is Exciting and Scary*. The New York Times, Stand: 15. März 2023. <https://www.nytimes.com/2023/03/15/technology/gpt-4-artificial-intelligence-openai.html> [15. März 2023].
- Savage, Neil (2023): *Synthetic Data could be better than Real Data*. Nature, Stand: 27. April 2023. Online: <https://www.nature.com/articles/d41586-023-01445-8> [27. Juni 2023].
- Schwarte-Amedick, Margret (2005): *Von papierlosen Büros und menschenleeren Fabriken*. In: Pias, Claus (Hrsg.): *Zukünfte des Computers*. Zürich/Berlin: Diaphanes, 67–86.
- Science Media Center (2021): *Transkript: Künstliche Intelligenz – Stand der Forschung und Förderung in Deutschland*. [Press Briefing] Science Media Center, Online: https://www.sciencemediacenter.de/fileadmin/user_upload/Press_Briefing_Zubehoer/Transkript_KI-Situation-Deutschland_SMC-Press-Briefing_2021-03-17.pdf [04. Februar 2022].
- Simon, Herbert A. & Newell, Allen (1958): *Heuristic Problem Solver: The Next Advance in Operation Research*. In: *Operation Research*, Vol. 6, No. 1, 1–10.
- Sloane, Mona & Moss, Emanuel (2019): *AI's Social Sciences Deficit*. In: *Nature Machine Intelligence*, Vol. 1, No. 8, 330–331.
- Teich, Irene B. (2020): *Meilensteine der Entwicklung Künstlicher Intelligenz*. In: *Informatik Spektrum*, Vol. 43, No. 4, 276–284.
- Uppenkamp, Max (2020): *Data Cleaning: Fallstricke und Lösungen*. Inform Datalab, Stand: 26. Mai 2020. Online: <https://www.inform-datalab.de/data-cleaning-fallstricke-und-loesungen/> [23. Juli 2023].
- Van Lente, Harro & Rip, Arie (1998): *Expectations in Technological Developments: An Example of Prospective Structures to be filled in by Agency*. In: Van Lente, Harro/ Rip, Arie/ Disco, Cornelis & Van der Meulen, Barend J. R. (Hrsg.): *Getting New Technologies Together: Studies in Making Sociotechnical Order*. Berlin: Walter de Gruyter, 203–229.
- Weizenbaum, Joseph (1978): *Die Macht der Computer und die Ohnmacht der Vernunft*. Frankfurt am Main: Suhrkamp.

Band 218
Beiträge aus der Forschung

sfs