Rainer Rehak | Essay | 14.12.2023

Zwischen Macht und Mythos

Eine kritische Einordnung aktueller KI-Narrative

Einleitung

Bei der Veröffentlichung des interaktiven Chatbots *ChatGPT* durch das US-amerikanische Unternehmen OpenAI im November 2022 wurde das erste Mal eine Anwendung der künstlichen Intelligenz (KI) vorgestellt, die auch für Laien beeindruckende Ergebnisse liefert und als textbasierte Webanwendung quasi voraussetzungslos benutzt werden kann. Dadurch war es sowohl der Medienwelt als auch interessierten Akteuren aus Politik, Wirtschaft und Zivilgesellschaft – ebenso wie Privatpersonen – möglich, diese Art von Technik unmittelbar selbst auszuprobieren. Diskussionen zur aktuellen Leistungsfähigkeit und möglichen Entwicklungen dieser Technologie wurden so in der Breite der Gesellschaft anschlussfähig und das Thema KI war quasi über Nacht brennend aktuell.

Obwohl es viele andere Arten künstlicher Intelligenz gibt, zeichnen sich an den Diskursen über die Möglichkeiten und Grenzen von *ChatGPT* und ähnlichen Programmen einige wiederkehrende Narrative ab, die die gesellschaftliche Auseinandersetzung mit KI aktuell prägen. Der Artikel beginnt mit einer kurzen historischen und technischen Einordnung von KI-Systemen und widmet sich anschließend einer Beschreibung und kritischen Reflexion der gängigen Narrative, wie etwa Entscheidungs-, Wissens- und Wahrheitsfähigkeit, apolitische Neutralität, Demokratisierungspotenzial sowie Arbeitsplatzvernichtung und gesellschaftliche Utopien/Dystopien durch KI. Der Artikel schließt mit einem Gesamtfazit zur gesellschaftlichen Auseinandersetzung mit dem Thema künstlicher Intelligenz.

Eine kurze Geschichte der KI

Der Terminus Artificial Intelligence (AI) wurde im Jahre 1955 durch US-Informatiker um John McCarthy und Marvin Minsky eingeführt, um ein Forschungsprogramm im Rahmen der Dartmouth Conference vorzustellen und dafür Forschungsgelder einzuwerben.[1] In dem Zusammenhang klang „Artificial Intelligence“ natürlich vielversprechender als Automatentheorie, wie das Feld zuvor hieß. In Abgrenzung zur traditionellen Automatentheorie etablierte sich alsbald ein Forschungsfeld mit dem Ziel, Computer *intelligence* simulieren zu lassen. Während die grundsätzliche Idee intelligenter Computer

schon älter ist und sich etwa auf die Kommunikation mit Menschen bezog,[2] beschrieb *intelligence* nunmehr das Vorhaben, Computer so zu programmieren, dass sie alle möglichen Aufgaben erledigen können, die bislang nur dem Menschen vorbehalten waren.[3]

Um diesem Ziel näher zu kommen, bildeten sich in den folgenden Dekaden im Wesentlichen zwei Strömungen innerhalb der Informatik heraus, die unterschiedliche Ansätze bei der Entwicklung von KI-Systemen verfolgten. Einerseits entstanden die sogenannten *symbolischen* Systeme, bei denen Informationen explizit abgespeichert und miteinander verknüpft werden (etwa „Hauptstadt(Deutschland)=Berlin“). Die eingegebenen Daten und Zusammenhänge können dann strukturiert durchsucht und logisch kombiniert werden. Schachcomputer etwa funktionierten lange Zeit auf diese Weise – während des Spiels wurden Datenbanken mit den Verläufen vergangener Schachpartien durchsucht, um einen Zug zu ermitteln, der in Anbetracht der Spielsituation die eigene Gewinnwahrscheinlichkeit erhöht. Dieser Ansatz ermöglichte später die sogenannten Expertensysteme und Wissensdatenbanken, mit denen beispielsweise komplexe medizinische Probleme vereinfacht werden sollten, indem etwa aus Symptomen und anderen Krankendaten automatisiert individuelle (Differenzial-)Diagnosen erstellt werden.[4]

Andererseits und parallel dazu wurden sogenannte *subsymbolische* Systeme entwickelt, auch Machine Learning (ML) genannt, bei denen Informationen und Zusammenhänge nur implizit vorgehalten werden. Entsprechende Systeme sind so gestaltet, dass sich die Programmkonfiguration (etwa bestimmte Parameter im System) anhand von vielen Eingabedaten automatisiert so lange verändert, bis das Ergebnis in der gewünschten Qualität vorliegt. Wie das Programm letztendlich zu diesem Ergebnis gelangt, ist zweitrangig und lässt sich auch nicht ohne Weiteres nachvollziehen.

Während in symbolischen Systemen logische Zusammenhänge explizit modelliert und bei der Entwicklung in das System eingeschrieben werden, verläuft die Konfiguration subsymbolischer Systeme in vielen kleinen automatisierten Schritten. Dabei enthält das Programm am Ende eine Vielzahl komplexer statistischer Beziehungen, aber keine expliziten semantischen Zusammenhänge: Insofern die eingegebene Datengrundlage das hergibt, gibt es etwa bei einer textbasierten KI dann beispielsweise eine abstrakt-statistische Beziehung zwischen den Worten „Hauptstadt“, „Deutschland“ und „Berlin“, aber diese ist nicht direkt aus den unzähligen Parametern des Systems auslesbar.

Ein aktuell viel genutzter subsymbolischer Ansatz sind die Künstlichen Neuronalen Netze

(KNN), die in den 1960er-Jahren erdacht wurden und der Vernetzung von Neuronen im menschlichen Gehirn nachempfunden sind. Im Vergleich zu diesen sind KNN sehr stark vereinfacht, sie werden im Computer als mathematische Gleichungen mit vielen Variablen abgebildet. Sie sind also komplexe statistische Modelle, die mit sehr vielen Beispieldaten vorkonfiguriert – „trainiert" – werden müssen, bevor sie eingesetzt werden können. Ein so vorkonfiguriertes KNN heißt „Modell", kann auf neue Daten angewendet werden und darin automatisiert Muster detektieren oder gar neue Muster generieren. In Bezug auf Bilddaten heißt das etwa, dass ein dafür erzeugtes Modell bei Eingabe eines Fotos hinreichend gut detektieren kann, ob darauf ein Affe, ein Pferd, ein Mopshund oder keines der drei Tiere zu sehen ist. Oder das Modell könnte Bilder von Affen, Pferden beziehungsweise Mopshunden generieren. So etwas ist mit symbolischen Ansätzen der künstlichen Intelligenz nie gelungen, denn dafür müssten sämtliche Kriterien für bestimmte Tiere auf Fotos explizit definiert werden, was praktisch unmöglich ist.

Weil in den 1980er-, 1990er- und 2000er-Jahren trotz vereinzelter methodischer Neuerungen konkrete Durchbrüche ausblieben, wurde es in dieser Phase immer wieder ruhig um die künstliche Intelligenz – oft ist in diesem Zusammenhang vom „KI-Winter" die Rede. Durch gesteigerte Rechenleistung, massenhaft verfügbare Daten (Stichwort *Big Data*) und das Aufkommen des Web 2.0, das es Nutzer:innen zunehmend ermöglichte, selbst Inhalte zu „generieren" (Stichwort *User-generated content*), konnte künstliche Intelligenz ab den 2010er-Jahren jedoch wieder von sich reden machen. Inzwischen kann KI nicht nur Bilder, sondern auch Texte erzeugen (man denke an automatische Übersetzungen von Aufsätzen und Transkriptionen von Tonaufnahmen oder eben Chatbots). Die Qualität vieler Systeme ist mittlerweile so gut, dass sie breit genutzt werden.[5]

Seit jeher wurde die technische Entwicklung von künstlicher Intelligenz stets von einem Diskurs bezüglich der Möglichkeiten und Grenzen der jeweiligen Systeme begleitet. Im Jahr 1966 entwickelte der Informatiker Joseph Weizenbaum ein simples Chatprogramm namens ELIZA – heute würden wir es als symbolische KI-Anwendung bezeichnen –, das zeigen sollte, wie einfach es ist, menschliches Verständnis zu simulieren.[6] Das Programm sollte einen Gesprächspsychotherapeuten nachahmen, indem es Begriffe aus den Texteingaben der Nutzer:innen in vorkonfigurierte Fragekonstruktionen einfügte („Was meinen Sie, warum X so ist?") oder aber generische Erwiderungen ausgab („Führen Sie das gern etwas aus."). Die für Weizenbaum verblüffende Reaktion bestand darin, dass dem Chatprogramm über die Fachwelt hinaus tatsächlich Empathie und Persönlichkeit zugeschrieben wurden. Diese Reaktionen machten ihn zu einem der vehementesten Kritiker mythischer Technikgläubigkeit, übersteigerter Erwartungen an Automatisierung und der Tendenz,

Computer zu vermenschlichen.[7] Seitdem beschreibt der „Eliza-Effekt“ den Hang, KI-Systemen menschliche Eigenschaften wie Intelligenz oder Bewusstsein zuzuschreiben, wenn sie nur gut genug typisch menschliche Redewendungen, Aussehen oder Verhaltensweisen imitieren.

Auch die wirtschaftlichen Folgen von KI wurden schon vor fünfzig Jahren diskutiert. So schrieb der Philosoph Hubert Dreyfus bereits im Jahr 1972 kritisch über den damaligen KI-Hype: „Jeden Tag lesen wir, dass digitale Computer Schach spielen, Sprachen übersetzen, Muster erkennen und bald in der Lage sein werden, unsere Jobs zu übernehmen“[8] – eine Prognose, die schon damals regelmäßig zu hören war und auch in den jüngeren Debatten eine wichtige Rolle spielt.[9]

Diese überaus grobe Zusammenfassung soll zwei Dinge herausstellen: Einerseits hat das Feld der künstlichen Intelligenz mittlerweile so große Erfolge vorzuweisen, dass alle, die mit dem Computer arbeiten, auch regelmäßig KI-Systeme nutzen (ob indirekt in Suchmaschinen oder direkt zum Übersetzen). Andererseits zeigt sich, dass die knapp 70-jährige Geschichte dieser Technologie von Anfang an geprägt war von der großen Erzählung einer angeblich kurz bevorstehenden KI-Revolution, die alles für immer verändern werde,[10] oft mit apokalyptischem Unterton.[11]

Um in der teilweise unübersichtlichen Gemengelage von Gegenwartsdiagnosen und Vorhersagen Orientierung zu bieten, werden im Folgenden aktuell prominente Narrative zu künstlicher Intelligenz umrissen und diskutiert. Auf dieser Grundlage können dann sachgerechte Diskussionen über die gesellschaftspolitischen Implikationen von KI-Systemen angestoßen werden – über Potenziale, Gefahren, Regulierung und Gestaltung dieser Technologien.

Zwei Arten von KI und ein Diskurswerkzeug

Um einzelne Narrative analysieren zu können und die Bedingungen zu verstehen, unter denen künstliche Intelligenz aktuell gesellschaftlich verhandelt wird, müssen noch zwei Arten von KI-Systemen unterschieden werden. Diese werden in Diskursen zu künstlicher Intelligenz regelmäßig aufgegriffen und oft vermischt, obwohl sie jeweils verschiedene Eigenschaften und Implikationen haben.

Erstens gibt es domänenspezifische KI-Systeme (*artificial narrow intelligence*, ANI), manchmal auch schwache KI genannt, die für einen Aufgabenbereich konzipiert und

entwickelt werden. Sie sind für bestimmte Abläufe optimiert, aber auch nur in diesem Rahmen überhaupt zu gebrauchen.[12] Ein Programm zur energieeffizienten Klimatisierung von Rechenzentren etwa wird mit ausgewählten Lastdaten der Vergangenheit vorkonfiguriert, sodass es den Einsatz von Kühlaggregaten je nach Bedarf wirtschaftlich regelt; es kann jedoch keine Musik-Playlist kuratieren. Ebenso wenig kann ein Schachprogramm Bilder erzeugen oder eine Bildgenerator-KI mathematische Regressionsanalysen durchführen. Darum sind vermenschlichende Begriffe wie „selbstlernend“ unpassend.[13] Alle aktuell existierenden KI-Systeme fallen zweifelsfrei in die Kategorie domänenspezifischer künstlicher Intelligenz, vom modernen Go-Computer über Bilderkennungs- oder Übersetzungssoftware bis hin zu den großen Sprachmodellen (large language models, LLMs) wie OpenAIs GPT, Googles PaLM/Bard, BAAIs WuDao oder Metas LlaMa.

Zweitens gibt es das Konzept universeller KI-Systeme (*artificial general intelligence*, AGI), manchmal auch starke KI genannt, die eigenständig lernfähig seien und abstrakt denken könnten, dazu Kreativität, Motivation sowie Bewusstsein und Emotionen besäßen.[14] In manchen Vorstellungen haben diese AGIs geradezu übermenschliche Fähigkeiten.[15] Allerdings gibt es bislang weder eine funktionierende AGI noch belastbare Anhaltspunkte, dass ein solches System mit aktuellen Computerarchitekturen überhaupt entwickelt werden kann.[16] Zwar gibt es seit Jahrzehnten Forschungsprojekte und Wettbewerbe in diese Richtung – etwa der weltbekannte Loebner-Preis, bei dem eine Version des Turing-Tests zum Einsatz kommt –, doch die Ergebnisse blieben bislang stets ernüchternd. Zudem haben sich die breit diskutierten AGI-Zielmarken, also welche technischen Probleme dafür gelöst werden müssen, regelmäßig verschoben. Ehemals galt es eine Partie gegen den Schachweltmeister zu gewinnen, inzwischen ist automatische Texterzeugung der Maßstab. Viele Fachleute zweifeln sogar grundsätzlich an der Möglichkeit einer allgemeinen künstlichen Intelligenz in Form eines Digitalcomputers.[17] Dass dennoch oft über diese Art von KI diskutiert wird, hat mit geschäftlichen Motiven wirtschaftlicher Akteure zu tun,[18] aber auch mit der zentralen Rolle solcher Systeme in Science-Fiction-Filmen, siehe etwa Samantha in ‘HER’, Data in ‘Star Trek’, HAL 9000 in ‘2001’, Ava in ‘Ex Machina’, C3PO in ‘Star Wars’, Bishop in ‘Aliens’, T-800 in ‘Terminator’ oder auch dem Maschinenmenschen in ‘Metropolis’.[19]

Um die schwammige Verwendung von künstlicher Intelligenz im gesellschaftlichen Diskurs zu charakterisieren, schlage ich den Begriff der *Zeitgeist-KI* vor.[20] In dem Zusammenhang kann der Begriff KI dann von Big Data und Statistik über Software, IT, Digitalisierung, Algorithmen, Roboter, Apps und IKT bis hin zum Internet in etwa alles bedeuten. Auch im

politischen Diskurs wird pauschal von „künstlicher Intelligenz“ gesprochen, egal, ob es um selbstfahrende Autos, Roboterhunde, automatisierte Entscheidungssysteme, Klimamodelle, automatisierte Arbeitsmarktvermittlungssysteme, Tischreservierungssysteme oder smarte Verkehrsleitsysteme geht; beizeiten werden auch traditionelle Informatikprodukte mit dem Label versehen. Alles ist „KI“, obwohl in alledem wenig bis keine künstliche Intelligenz steckt. In einem aktuellen Bericht zum Stand von KI in der öffentlichen Verwaltung heißt es etwa, „dass oftmals Projekte als KI-basiert bezeichnet würden, jedoch de facto konventionelle IKT-Anwendungen nutzen“.[21] Aber auch in Wissenschaft und Wirtschaft ist jenes schwammige Verständnis anzutreffen.[22] Um eine reflektierte gesellschaftliche Auseinandersetzung mit künstlicher Intelligenz zu ermöglichen, muss dieses Verständnis expliziert, mithin aufgelöst werden.[23]

Besprechung aktueller KI-Narrative

Vor diesem Hintergrund können wir nun gängige Narrative zu künstlicher Intelligenz umreißen und diskutieren – auf welchen Annahmen beruhen sie, inwieweit entsprechen diese dem aktuellen Stand der Forschung, welche gesellschaftspolitischen Implikationen haben sie? Oftmals überschneiden und vermischen sich diese Erzählungen, sodass die folgende Darstellung nach Stichworten eine analytische Trennung vornimmt, die eher einer schlaglichtartigen Analyse von Diskursen zum Thema KI als einer systematischen Darstellung ihres Gegenstands dient. Die Ausführungen stützen sich auf Erkenntnisse aus der kritischen Informatik sowie den *critical data and algorithm studies*, der Datenschutztheorie und ferner der Philosophie des Geistes und der Semiotik.

Autonomie/Agency – Oft wird KI-Systemen die Fähigkeit zugesprochen, eigenständig „zu agieren“ oder etwas „zu entscheiden“.[24] Solche Diskussionen gehen implizit von der Annahme aus, KI-Systeme seien zu selbständigem Handeln (im Gegensatz etwa zu bloßem Verhalten) befähigt[25] und nicht bloß (komplexe) Werkzeuge, die einen extern gesetzten Zweck umsetzen. Handlungsfähigkeit impliziert jedoch eigene Intention, innere Vorstellung der Sachlage, einen potenziell alternativen Handlungsausgang, überdies ein Moment der bewussten Entscheidung und dann letztlich auch Verantwortlichkeit. KI-Systeme verhalten sich jedoch deterministisch – grundsätzlich erfolgt bei gleicher Eingabe, wie bei anderen Maschinen auch, die gleiche Ausgabe.[26] Ist dies nicht der Fall, wie etwa bei ChatGPT, so liegt das an absichtlich eingebauten Zufallsparametern und somit an Entscheidungen bei der Entwicklung, nicht an einer vermeintlich eigenen Handlungsfähigkeit der KI. Intention und eine innere Vorstellung sind gar nicht vorhanden. Solange es keine AGI gibt, sind die Ergebnisse aller KI-Systeme primär das Ergebnis von Designentscheidungen,

gegebenenfalls inklusive Zufall oder Fehlern. Wenn die eingesetzten Systeme ihre Zwecke nicht erfüllen, ändern Hersteller und Betreiber sie, justieren nach und korrigieren, was den Werkzeugcharakter solcher Systeme weiter unterstreicht. Dass Anwendungen künstlicher Intelligenz eher großtechnischen Anlagen denn einfachen Werkzeugen gleichen, ändert nichts daran, dass ihre Charakterisierung als eigenständige Akteure von den tatsächlich verantwortlichen Organisationen (Firmen, Behörden etc.) ablenkt, die diese Systeme für ihre eigenen Zwecke programmieren oder kaufen und einsetzen. Autonomie – in einem bedeutsamen Sinn – haben diese Systeme daher keine.[27]

Wahrhaftigkeitsanspruch – Aktuelle KI-Systeme können weder lügen noch hochstapeln. Selbstverständlich können die Aussagen falsch sein (und das sind sie auch oft[28]), aber eine Lüge impliziert das Wissen um die Wahrheit und eine absichtliche Abkehr davon. Gleichermaßen impliziert Hochstapelei das Wissen um die eigene Identität und ein absichtliches Vorspielen einer anderen. KI-Systeme können keine Wahrheitsansprüche erheben, weil sie genau das ausgeben, was Modellarchitektur und Daten vorgeben. Sie können daher auch nicht achtlos Unsinn reden, strategisch manipulieren oder bewusst ablenken, denn all dies würde mindestens einen inneren Zustand, eine Motivation, ein Geistesmodell und ein Modell des Anderen voraussetzen. Wie oben beschrieben, bestehen KI-Systeme jedoch aus zwar komplexen, aber rein formalen Abläufen. Die einsetzenden Organisationen hingegen haben sehr wohl Interessen und potenziell auch das Wissen um die Wahrheit und die eigene Identität. Entsprechende Ansprüche müssen gegenüber den Entwicklerfirmen und -organisationen geltend gemacht werden.

Wissen und Wahrheit – Aktuelle textbasierte KI-Anwendungen wie ChatGPT beruhen auf Sprachmodellen, die gemäß ihrer Architektur aus den Unmengen an Eingabetexten mehrdimensionale Vektorräume errechnen, in denen der „Abstand“ von Worten und Wortarten abgespeichert wird; das ist das Sprachmodell. Daraus werden in der normalen Nutzung Texte generiert, indem das Modell auf Basis einer User-Eingabe („Prompt“) das wahrscheinlichste nächste Wort ermittelt.[29] Dann sind Prompt plus erstes Ergebniswort wiederum die interne Eingabe für das nächste Wort. Dies wird mehrfach wiederholt. So setzt sich die Ausgabe („Antwort“) des Chatbots zusammen. Die Ausgaben sind also formal-mathematisch begründete Aneinanderreihungen von Zeichenketten, die statistisch rekombiniert aus den Eingabetexten abgeleitet werden und deren Bedeutung – anders als bei den weniger leistungsfähigen symbolbasierten Wissensmodellen – für das System technisch bedingt gar keine Rolle spielen kann.[30] Dementsprechend sind Wahrheit oder Richtigkeit keine relevanten Kriterien für die Antworten und können es auch nicht ohne Weiteres werden. Die Ergebnisse solcher KI-Systeme (re)produzieren die formalen

Verhältnisse von Worten zueinander in den Ausgangstexten, genau das ist ihre Funktion. Die Ergebnisse sind teilweise beeindruckend, aber umso obskurer sind auch die Fehler, etwa wenn nichtssagende Allgemeinplätze oder reine „Fantasiefakten“ ausgegeben werden. Streng genommen können Sprachmodelle jedoch keine „Fehler“ machen, die Wortreihen in den Ausgangstexten sind eben, wie sie sind.[31] Gerade der Stil – also die Form – der generierten Texte ist in der Regel tadellos, da auch die Texte, an denen das Modell trainiert wurde, in der Regel stilistisch sehr gut sind. Insofern ergibt es wenig Sinn, zu behaupten, dass ChatGPT „sogar“ Texte im Stil von Gedichten, akademischen Artikeln, Zeitungsmeldungen oder Handbüchern produzieren könne, denn es ist ja gerade die Form, auf die Sprachmodelle optimiert sind und etwas anderes ist mit dieser Technologie auch gar nicht möglich.[32] Mathematisch gesprochen sind die Ergebnisse also Variationen eines statistischen Durchschnittstexts innerhalb des Modells und zwar relativ zur jeweils konkreten Abfrage. Als passende Analogie bietet sich ein stochastischer Papagei an, der gemäß statistischer Regeln Wortfolgen reproduziert.[33] Mit echter Sprachkompetenz und Verständnis haben wir es weder in dem einen noch dem anderen Fall zu tun.

Vorhersagen – KI-Systeme können formale Muster und Abhängigkeiten von Variablen in vorhandenen Datensätzen detektieren, seien es Wetter-, Geschäfts- oder Verhaltensdaten. Diese Muster beziehen sich notwendigerweise immer auf die Vergangenheit, können aber statistisch ausgewertet und mathematisch in die Zukunft projiziert werden. Inwiefern das dann aber als Vorhersage im eigentlichen Sinne taugt, hängt vom Gegenstandsbereich ab. Sind es Daten und Abhängigkeiten, die physikalischen Gesetzen unterliegen, etwa Wetterdaten, kann das funktionieren. Sind es hingegen soziale Daten, erzeugen solche Vorberechnungen zwar ein Ergebnis, welches aber nur dann die Zukunft vorhersagt, wenn diese genau so strukturiert ist, wie es die (Daten-)Vergangenheit war – eine nicht nur aus wissenschaftstheoretischer und sozialwissenschaftlicher Sicht heikle Annahme, selbst wenn die Daten der Vergangenheit vollständig wären.[34] Der Aufstieg und Niedergang von Predictive Policing (zu Deutsch „vorrausschauende Polizeiarbeit“) illustriert das Problem gut, denn dort wurde angenommen, dass sich Kriminalität auf ähnliche Weise wie Erdbeben geografisch vorhersagen lässt. Die jeweilig zugrunde liegenden Prozesse sind jedoch sehr verschieden – gesellschaftliche Prozesse folgen im Gegensatz zu physikalischen Vorgängen gerade keinen feststehenden Gesetzen. In der Praxis waren die Ergebnisse von Predictive Policing-Softwares kaum zu gebrauchen – teilweise trafen die Voraussagen in weniger als ein Prozent der Fälle zu und in anderen Fällen reproduzierten sie sogar Rassismus, indem etwa Straftaten übermäßig in schwarzen Nachbarschaften angezeigt wurden.[35]

Neutralität und Objektivität – Datenbasierten Systemen wird häufig Neutralität und Objektivität attestiert, so ist es auch mit KI-Systemen. Dabei gestalten bestimmte Akteure die System- und Modellarchitektur für einen bestimmten Zweck. Diesem Zweck entsprechend wählen sie auch die Datengrundlage des Systems aus, die wiederum entscheidend für die Ergebnisse des Systems ist. Prinzipiell gibt es keine objektiven oder neutralen Daten, sondern nur jeweils passende Daten in Relation zu einem bestimmten Einsatzzweck.[36] Beispielsweise ist es möglich, Feinstaubsensoren in einer Stadt anzubringen, um die Luftqualität zu messen, aber in welcher Höhe und an welchen Orten (Parks, Straßen, Kindergärten etc.) dies geschieht, ist eine Entscheidung, die von Akteuren und Zwecken abhängt. Und selbst wenn es – rein hypothetisch – möglich wäre, eine „totale" Datenbasis in perfekter Qualität zu haben, müssten dennoch unzählige Entscheidungen für die konkrete Verarbeitung getroffen werden, um praktisch nutzbare Ergebnisse zu liefern. Wäre etwa bei der vorausschauenden Polizeiarbeit die Schadenshöhe vergangener Verbrechen miteinbezogen worden, hätten die Systeme überwiegend die Hauptgeschäftsviertel und Finanzhandelsplätze von Städten als Kriminalitätsschwerpunkte angezeigt, doch das entsprach nicht dem Zweck. Diese Art Entscheidungen braucht es bei allen datenbasierten Systemen – inklusive KI, weshalb keines davon als „neutral" oder „objektiv" gelten kann.

Selbst Ansprüche und Kritik an KI-Systemen unterstellen jedoch oft deren (mögliche) Neutralität beziehungsweise Objektivität, etwa wenn sie Fairness einfordern oder Diskriminierung durch diese Systeme bemängeln. Diese Ansätze greifen mitunter zu kurz, weil es sich bei Fairness nur teilweise um eine technische Eigenschaft der Systeme handelt und gerade die Frage, was als nicht-diskriminierend und fair gilt, primär von gesellschaftlichen Aushandlungsprozessen abhängt.[37] Ein automatisches Kreditvergabesystem basierend auf dem Einkommen einer Person beispielsweise kann aktuell nur korrekt oder fair sein: entweder es ist korrekt, aber reproduziert dann das geschlechtsspezifische Lohngefälle, oder es ist fair, aber mathematisch gesehen inkorrekt.[38]

(A)politische Optimierung – Beizeiten werden gesellschaftliche und politische Fragen als im Wesentlichen durch KI lösbare Aufgaben diskutiert.[39] Dazu zählen etwa das Abwenden des Klimawandels,[40] die Einführung eines automatisierten Grundeinkommens oder die Beendigung des Welthungers.[41] Dem Narrativ nach erscheint das als möglich, mithin vielversprechend, weil KI apolitische und „unideologische" Lösungen für gesellschaftliche Probleme erzeugen könne. Diese Erzählung ist aus mehreren Gründen problematisch. Sie verkennt etwa, dass bevor KI-Systeme Muster detektieren oder Prozesse optimieren können, zunächst einmal festgelegt werden muss, was genau gesucht und was in einem

bestimmten Sachbereich als Optimum gelten soll. Erst daraus ergibt sich eine sinnvolle Definition der technischen Zielfunktion des Systems. Auch diese Festlegung ist prinzipiell eine politische Frage, die gesellschaftlicher Aushandlung bedarf. Denn selbst wenn Einigkeit bezüglich eines Ziels herrscht, gibt es in der Regel unzählige Strategien, um es zu erreichen. Auch die Wahl der Zwischenschritte und die Gewichtung der relevanten Faktoren sind kategorisch keine technischen Fragen, sondern politische. Insofern besteht der Widerspruch des Narrativs in der Annahme, dass KI-Systeme ihre eigene gesellschaftliche Ausgangsbedingung als technisches Ergebnis liefern könnten. Das lässt sich an zwei Beispielen verdeutlichen. Eine KI kann den Energieverbrauch eines Rechenzentrums reduzieren (Ziel), indem es anhand von Lastdaten energieeffiziente Kühlzyklen berechnet (Strategie).[42] Dies ist möglich, wenn Ziel und Strategie unstrittig sind. Im Bereich der Stadtplanung wiederum kann eine KI bei Mobilitätsfragen helfen und beispielsweise dazu beitragen, die Auslastung von Autostraßen und Parkplätzen im Hinblick auf Verkehrsaufkommen zu optimieren (Stichwort „smart cities“). Die Frage aber, wie eine Stadt lebenswerter gestaltet werden kann, und ob dabei nicht eher eine Fahrradinfrastruktur mit erweiterten Fußgängerzonen zielführend wäre, ist jedoch keine technische Frage. Generalisiert gilt das auch für die Bekämpfung des Welthungers. Sollte man bestimmten Ländern Schulden erlassen, internationale Wirtschaftsverträge ändern, die Nahrungsverteilung überdenken, bestimmte Technologien global zugänglich machen oder ganz andere Maßnahmen ergreifen? Und wer sollte sie finanzieren? All das sind politische Fragen, die KI-Systeme nicht lösen können. Sie können bestenfalls helfen, Beispielszenarien zu errechnen und gegebenenfalls die Umsetzung zu erleichtern. Selbst bei der Auswahl und dem Modellieren der relevanten Rahmenbedingungen handelt es sich jedoch um Aufgaben, die kollektive, also politische Entscheidungen erfordern.

Demokratisierung – Auch wenn KI-Systeme komplexe Werkzeuge sind, die aktuell von mächtigen Konzernen entwickelt werden, wird oft gemutmaßt, dass diese Werkzeuge über kurz oder lang durch gesteigerte Effizienz und andere technische Verbesserungen bald allen zur Verfügung stehen könnten und somit eine Demokratisierung von KI ermöglichen würden.[43] Was zunächst wie ein erstrebenswertes politisches Anliegen klingen mag, entpuppt sich bei genauerem Hinsehen als Wirtschaftsagenda. „Demokratisierung“ meint in diesem Fall nämlich nicht die Kontrolle oder Mitgestaltung solcher Technologien durch selbstorganisierte Gemeinschaften, sondern lediglich den breiten Nutzungszugriff für Unternehmen.[44] Die Kontrolle solcher Systeme kann auch schwerlich demokratisiert werden. Während es im Hinblick auf kleine, hochspezialisierte KI-Systeme mit einer vergleichsweise kleinen Datenbasis noch möglich sein kann, ist es bei großen KI-Systemen wie etwa LLMs hingegen illusorisch, da der technische, organisatorische (und energetische)

Aufwand für Gestaltung und Herstellung schlechthin immens ist. Ihr Entwicklungsprozess schließt die Datenerhebung, ihre Qualitätssicherung, Klassifikation und Speicherung ein, betrifft weiter das inkrementelle Modelldesign und das Vorkonfigurieren („Training“) der Modelle und hört auch bei der Moderation und Verfeinerung der KI-Systeme durch menschliche Arbeitskräfte nicht auf (Stichwort *Reinforcement Learning from Human Feedback*), die größtenteils in den globalen Süden ausgelagert wird.[45] Dementsprechend steckt hinter großen KI-Systemen eine komplexe globale Lieferkette.[46] Zudem müssen die Systeme regelmäßig aktualisiert, sämtliche genannten Schritte also regelmäßig wiederholt werden. Daten und Systeme altern schließlich auch. Es ist also kein Zufall, dass nur finanzstarke Player wie OpenAI, Meta, BAAI oder Google regelmäßig Durchbrüche bei LLMs vermelden können, denn nur sie verfügen über das nötige Kapital, um solche Projekte zu stemmen. Man kann behaupten, dass diese Art KI quasi die Atomkraft des Digitalen ist,[47] also grundsätzlich nur durch mächtige Akteure kontrolliert und zentralisiert betrieben werden kann. Der Zugriff wird dann vermietet, aber die Kontrolle bleibt beim Eigentümer. Daran ändern auch die freien und offenen KI/ML-Programmbibliotheken von Google und Co. oder kleine LLMs zum Selbstbetreiben[48] nichts, weil stets die restlichen Zutaten zur eigenen Erstellung fehlen.

Arbeitsplatzverlust – Abschließend soll noch kurz das Narrativ angerissen werden, dem zufolge KI ein „Jobkiller“ sei. Obwohl diese grundlegende Debatte so alt ist, wie die Automatisierung selbst, wird sie auch im KI-Kontext verkürzt geführt – bereits die Ludditen oder andere Maschinenstürmer waren ja keineswegs technikfeindlich. Zugespitzt besagt das Narrativ, dass durch den Einsatz von KI massenweise Arbeitsplätze verschwinden würden, manche Studien sprechen gar von knapp der Hälfte aller Jobs in den USA,[49] sodass auf absehbare Zeit eine Massenarbeitslosigkeit drohe. Dieses Narrativ hat mindestens drei fragwürdige Aspekte: Erstens ist es ja zunächst begrüßenswert, wenn Maschinen und KI-Systeme Menschen die Arbeit abnehmen, besonders wenn diese repetitiv, beschwerlich oder gar gefährlich ist.[50] Die Frage ist also, welche Jobs wegfallen, wer die Betroffenengruppen sind und welche Rolle Lohnarbeit überhaupt gesellschaftlich künftig spielen soll. Zweitens zeigen Studien, dass Arbeitsplätze so gut wie nie plötzlich wegfallen, sondern sich eher langsam verändern. Statt technologiebedingter Massenarbeitslosigkeit sehen wir einen Strukturwandel des Arbeitsmarktes.[51] Dabei besteht die Veränderung darin, dass sich moderat und zunehmend auch gut bezahlte Arbeitsplätze in eine überschaubare Anzahl lukrativer Stellen („high skilled labour“) wandeln, während der Großteil schlechter bezahlten oder gar prekären Beschäftigungsformen („low skilled labour“) weicht.[52] An der Arbeitslosenquote ändert das wenig, politischer Handlungsbedarf besteht dennoch. Drittens verschleiert das Narrativ wie so oft die tatsächlichen

Zusammenhänge und verantwortlichen Akteure. Schließlich nimmt künstliche Intelligenz den Menschen die Arbeit nicht einfach weg und Arbeitsplätze verschwinden auch nicht von Zauberhand. Es sind Entscheider:innen in Konzernen, die beschließen, dass es wohl profitabler wäre, Stellen abzubauen, Arbeitsprozesse zu automatisieren und/oder die Produktion zu verlagern. Diese Entwicklung hängt primär von rechtlichen und wirtschaftlichen Rahmenbedingungen ab, von unternehmerischen Strategien und nicht zuletzt politischen Entscheidungen bezüglich der Rolle von KI in der gesamtgesellschaftlichen Transformation.[53] Solange die Entwicklung künstlicher Intelligenz als Naturgewalt betrachtet und von Angstszenarien wie einer drohenden Massenarbeitslosigkeit begleitet wird, profitieren von ihr vor allem Unternehmen, die KI-Produkte herstellen oder vertreiben und darüber hinaus Konzerne, denen angstinduzierte Niedriglöhne zupass kommen. Für gesellschaftliches Gestaltungspotenzial bleibt dann kaum Raum.

Fazit und Ausblick

Bei der Analyse dieser KI-Narrative fällt auf, dass sie meist gar nicht auf den technischen Eigenschaften dieser Systeme fußen, oft Fehlannahmen über ihre Einsatzdynamik enthalten und somit nicht zu einer fruchtbaren und produktiven Debatte beitragen, sondern ihr teilweise sogar entgegenstehen. Dass sie in aktuellen Diskursen trotzdem prominent vorkommen, mag daran liegen, dass die Technologie missverstanden wird, gesellschaftliche Kontexte und Dynamiken ausgeblendet werden oder auch daran, dass solche Erzählungen oft einem kommerziellen Zweck dienen und entsprechend verbreitet werden.

KI wird gemäß den Zwecken gestaltet und verwendet, die Unternehmen, Regierungen und andere Organisationen für sie vorsehen, wenn auch manchmal mit nicht intendierten Effekten. Oberflächliche Zeitgeist-KI-Debatten oder dramatisch zugespitzte Gegenüberstellungen à la „Mensch gegen Maschine“ sind jedoch fehl am Platze, verschleiern relevante Machtfragen[54] und gehören in den Bereich der Science-Fiction.

Gleichwohl müssen die Komplexität und Beherrschbarkeit dieser Systeme im Auge behalten werden. Dafür gibt es längst methodische Ansätze und regulatorische Maßnahmen, die getroffen werden könnten,[55] wenn denn der politische Wille dafür vorhanden wäre. Dass nicht entsprechend gehandelt wird, hängt nicht zuletzt mit neoliberalen Gesellschaftsbildern, kokettierenden „Neuland“-Haltungen, wirkmächtigem Lobbyismus, Innovation um ihrer selbst willen oder auch mit naiver Technikgläubigkeit

zusammen. In dieser Gemengelage sind Fragen ethischer und vertrauenswürdiger KI zwar technisch und philosophisch interessant, hängen aber weit hinter dem Stand der Diskussion in den technikregulatorischen Sozialwissenschaften[56] und der Datenschutztheorie zurück.[57] Denn die Fragen, die sich stellen, wenn Organisationen KI-Systeme einsetzen, sind zwar neu, aber nicht fundamental anders, als die, die sich stellen, wenn Organisationen für ihre Zwecke Beton, Differentialgleichungen oder Spürhunde einsetzen. Ein hohes Maß an Entpolitisierung sowie eine Selbststilisierung sind jedoch wesentliche Aspekte der Selbsterzählung von KI – und auch von Digitalisierung allgemein.

Der Einsatz von KI-Technologien verändert tatsächlich unsere Gesellschaften, aber nicht so, wie oft – wahlweise utopistisch[58] oder dystopisch – argumentiert wird. Die Rettung der Menschheit durch eine leistungsstarke, mithin gute KI steht ebenso wenig zu erwarten wie die drohende Auslöschung der Menschheit durch eine unkontrollierbare, mithin böse Superintelligenz.[59] Tatsächliche Folgen und Gefahren dieser Technologien, die Gegenstand gesellschaftlicher Debatten sein müssen, sind etwa die Strukturveränderung des Arbeitsmarktes, die Ausweitung ausbeuterischer Lieferketten bei der Datenerhebung und auswertung, insbesondere im globalen Süden,[60] die einseitige Macht- und Produktivitätssteigerung weniger Firmen, die Ausweitung personalisierter Überwachung, die unendliche Welle von KI-generiertem Spam und Betrug,[61] die automatisierte Aneignung und Kommerzialisierung von Online-Kulturwerken, die Herausbildung globaler Abhängigkeiten durch Oligopoltendenzen in der Industrie, die Implikationen militärischer KI-Nutzung,[62] der exponentiell ansteigende Energieverbrauch durch KI-Systeme und nicht zuletzt der unreflektierte und nur scheinbar unpolitische Einsatz solcher Systeme im Wohlfahrtsstaat und in weiteren sensiblen Gesellschaftsbereichen (Kreditvergabe, Bewerbungsprozesse, Strafverfahren etc.).[63] Eine weitere drängende politische Frage betrifft die kollektive Gestaltung der geistigen Arbeit im Kontext der Maschinisierung unter den aktuell gegebenen Macht- und Besitzverhältnissen. Technikgestaltung ist immer auch die Gestaltung von (technisch geprägten) sozialen Praktiken, aber gleichzeitig ist sie selbst das Ergebnis sozialer Praktiken – und künstliche Intelligenz ist seit jeher ein diskursiv umkämpftes Terrain.

Doch die Zukunft ist offen und all diese Fragen sind zu diskutieren, sodass wir als Gesellschaft diesen interessanten und mächtigen Werkzeugkasten kreativ und reflektiert nutzen können. Auch wenn schon viele theoretische und praktische Vorarbeiten existieren, sind die Einsatzmöglichkeiten von KI und anderen digitalen Technologien jenseits von Profit und Markt noch nicht einmal ansatzweise erdacht und entwickelt worden – von kritischer Informatik[64] und anarchistischer Softwaregestaltung[65] über das digitale

Commoning[66] und herrschaftsfreie Informationssysteme[67] bis hin zur Unterstützung der Nachhaltigkeitstransformation[68] und der Rückeroberung der Computationsmittel.[69] Wenn wir die gesellschaftliche Gestaltbarkeit künstlicher Intelligenz und ihre vielfältigen Anwendungsmöglichkeiten im Blick behalten wollen, müssen wir nicht nur die beizeiten wiederkehrenden ELIZA-Momente durchschauen, sondern auch überlegen, inwiefern der Fokus auf riesenhaft-überkomplexe KI-Systeme dabei überhaupt hilfreich ist, oder ob er nicht eher ablenkt von den eigentlichen Zutaten für ein gutes Leben für alle.

Förderhinweis: Diese Arbeit wurde durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert (Förderkennzeichen: 16DII131 – „Deutsches Internet-Institut“).

Endnoten

1. John McCarthy u.a., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, Stanford University, https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (06.11.2023).

2. Alan Turing, Computing Machinery and Intelligence, in: Mind 59 (1950), 236, S. 433–460 (https://doi.org/10.1093/mind/LIX.236.433).

3. McCarthy u.a., Proposal on Artificial Intelligence.

4. Wolfgang Coy / Lena Bonsiepen, Erfahrung und Berechnung. Kritik der Expertensystemtechnik, Berlin / Heidelberg 1989.

5. Vgl. etwa Martin Popel u.a., Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals, in: Nature Communications 11 (2020), doi.org/10.1038/s41467-020-18073-9.

6. Joseph Weizenbaum, ELIZA – A Computer Program for the Study of Natural Language Communication Between Man And Machine, in: Communications of the ACM 9 (1966), 1, S. 36–45.

7. Joseph Weizenbaum, Die Macht der Computer und die Ohnmacht der Vernunft [1978], übers. von Udo Rennert, Frankfurt am Main 2000; vgl. aktuell Emily M. Bender u.a., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in: Association for Computing Machinery (Hg.), FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York 2021, S. 610–623.

8. Hubert L. Dreyfus, What Computers Can’t Do. A Critique of Artificial Reason, New York 1972.

9. Vgl. etwa Carl Benedikt Frey / Michael A. Osborne, The Future of Employment: How Susceptible Are Jobs to Computerisation?, Oxford Martin School, https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf (6.11.23); oder Joseph Briggs / Devesh Kodnani, The Potentially Large Effects of Artificial Intelligence on Economic Growth, Goldman Sachs Publishing, https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45

d7-967b-d7be35fabd16.html (6.11.23).

10. Vgl. Ray Kurzweil, The Singularity is Near: When Humans Transcend Biology, New York 2005.

11. Vgl. Elon Musk u.a., Pause Giant AI Experiments: An Open Letter, Future of Life Institute, https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (6.11.23); oder Geoffrey Hinton u.a., Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war, Center for AI Safety, https://www.safe.ai/statement-on-ai-risk (6.11.23).

12. Vgl. Tom Mitchell, Machine Learning, New York 1997.

13. Rainer Rehak, The Language Labyrinth: Constructive Critique on the Terminology Used in the AI Discourse, in: Pieter Verdegem (Hg.), AI for Everyone? Critical Perspectives, S. 87–102, London 2021.

14. Vgl. Turing, Computing Machinery and Intelligence; siehe auch Martin Holland, Hat Chatbot LaMDA ein Bewusstsein entwickelt? Google beurlaubt Angestellten, in: heise online, 13.6.23, https://www.heise.de/news/Hat-Chatbot-LaMDA-ein-Bewusstein-entwickelt-Google-beurlaubt-Angestellten-7138314.html (6.11.23).

15. Vgl. Kurzweil, The Singularity is Near; Musk u.a., Pause Giant AI Experiments.

16. Mariana Lenharo, Consciousness theory slammed as ‘pseudoscience’ — sparking uproar, in: Nature, 20.09.23, https://www.nature.com/articles/d41586-023-02971-1 (6.11.23).

17. Dreyfus, What Computers Can’t Do; Ragnar Fjelland, Why general artificial intelligence will not be realized, in: Humanities and Social Sciences Communications 7 (2020), 10, https://www.nature.com/articles/s41599-020-0494-4; Andrea Roli / Johannes Jaeger / Stuart A. Kauffman, How Organisms Come to Know the World: Fundamental Limits on Artificial General Intelligence, in: Frontiers in Ecology and Evolution 9 (2021), https://doi.org/10.3389/fevo.2021.806283; Rainer Rehak, The Language Labyrinth.

18. O.A., X.AI: Elon Musk gründet KI-Unternehmen trotz Brief mit Moratorium, in: Frankfurter Allgemeine Zeitung, 16.4.23,

https://www.faz.net/aktuell/wirtschaft/unternehmen/x-ai-elon-musk-gruendet-ki-unternehmen-trotz-brief-mit-moratorium-18825457.html.

19. Isabella Hermann, Künstliche Intelligenz in der Science-Fiction: Mehr Magie als Technik, in: Helen Ahner / Max Metzger / Mathis Nolte (Hg.), Von Menschen und Maschinen: Interdisziplinäre Perspektiven auf das Verhältnis von Gesellschaft und Technik in Vergangenheit, Gegenwart und Zukunft. Proceedings der 3.Tagung des Nachwuchsnetzwerks "INSIST", 05.-07. Oktober 2018, https://www.ssoar.info/ssoar/handle/document/67663.

20. Rainer Rehak, Artificial Intelligence for Real Sustainability?, in: Patricia Jankowski u.a. (Hg.), Shaping Digital Transformation for a Sustainable Society. Contributions from Bits & Bäume, Technische Universität Berlin, https://doi.org/10.14279/depositonce-17526 (6.11.23).

21. Deutscher Bundestag, Bericht zum Stand von KI in der öffentlichen Verwaltung, Bildung, Forschung und Technikfolgenabschätzung — Bericht — hib 520/2022, https://www.bundestag.de/presse/hib/kurzmeldungen-914308 (6.11.23).

22. Im Hinblick auf die Wissenschaft vgl. stellvertretend Daniel Innerarity, The epistemic impossibility of an artificial intelligence take-over of democracy, AI & Society, https://doi.org/10.1007/s00146-023-01632-1 (6.11.23); für die Wirtschaft vgl. James Vincent, Forty percent of ‘AI startups’ in Europe don’t actually use AI, claims report, in: The Verge, 5.3.2019, https://www.theverge.com/2019/3/5/18251326/ai-startups-europe-fake-40-percent-mmc-report (6.11.23).

23. Ein positives Beispiel liefert dazu der japanische Rat für soziale Grundsätze einer humanzentrierten künstlichen Intelligenz, vgl. Council for Social Principles of Human-Centric AI, Social Principles of Human-Centric AI, https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf (6.11.23).

24. Vgl. etwa Emma Farge, Robots say they won't steal jobs, rebel against humans, in: Reuters, 7.7.2023, https://www.reuters.com/technology/robots-say-they-wont-steal-jobs-rebel-against-humans-2023-07-07/.

25. Vgl. Florian Butollo u.a. (Hg.), Lecture Series "Autonomous Systems & Self-

Determination", Weizenbaum Institute for the Networked Society, https://doi.org/10.34669/wi/2 (6.11.23).

26. Stephen Wolfram, What Is ChatGPT Doing … and Why Does It Work?, Stephen Wolfram Writings, https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/ (6.11.23).

27. Florian Butollo u.a. (Hg.), Lecture Series "Autonomous Systems & Self-Determination".

28. Christiane Floyd, From Joseph Weizenbaum to ChatGPT: Critical Encounters with Dazzling AI Technology, in: Weizenbaum Journal of the Digital Society 3 (2023), 3, https://doi.org/10.34669/WI.WJDS/3.3.3 (6.11.23).

29. Stephen Wolfram, What Is ChatGPT Doing.

30. Emily Bender u.a., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in: Association for Computing Machinery (Hg.), FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York 2021, S. 610–623.

31. Das zeigt sich etwa an den allgemeinen Antworten eines Chatbots, die für bestimmte Zwecke unpassend bis regelrecht schädlich sind: Lauren Aratani, US eating disorder helpline takes down AI chatbot over harmful advice, The Guardian, 31.7.2023, https://www.theguardian.com/technology/2023/may/31/eating-disorder-hotline-union-ai-chatbot-harm.

32. Paola Lopez, ChatGPT und der Unterschied zwischen Form und Inhalt, in: Merkur 77 (2023), 891, S. 15–27, https://www.merkur-zeitschrift.de/artikel/chatgpt-und-der-unterschied-zwischen-form-und-inhalt-a-mr-77-8-15/ (6.11.23).

33. Bender u.a., On the Dangers of Stochastic Parrots.

34. Florian Eyert / Paola Lopez, Rethinking Transparency as a Communicative Constellation, in: FAccT ’23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, New York 2023, S. 444–454.

35. Aaron Sankin / Surya Mattu, Predictive Policing Software Terrible at Predicting Crimes,

in: Wired, 2.10.2023, https://www.wired.com/story/plainfield-geolitica-crime-predictions (6.11.23); und Will Douglas Heaven, Predictive policing algorithms are racist. They need to be dismantled, in: MIT Technology Review, 17.7.2020, https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/ (6.11.2023).

36. Rob Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences, Thousand Oaks, CA 2014.

37. Eyert / Lopez, Rethinking Transparency.

38. Sayash Kapoor / Arvind Narayanan, Quantifying ChatGPT's gender bias, in: AI Snake Oil, 26.4.2023, https://aisnakeoil.substack.com/p/quantifying-chatgpts-gender-bias (6.11.23).

39. Vgl. etwa Michael Chui u.a., Applying Artificial Intelligence for Social Good, McKinsey & Company, https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good (6.11.23).

40. Vgl. etwa o.A., Das Potenzial von KI für den Klimaschutz, in: UmweltDialog, 28.12.2020, https://www.umweltdialog.de/de/umwelt/klimawandel/2020/Das-Potenzial-von-KI-fuer-den-Klimaschutz.php (6.11.23).

41. Eileen Guo / Adi Renaldi, Worldcoin: Tausche Kryptowährung gegen Augen-Scan, in: MIT Technology Review, 3.7.2023, https://www.heise.de/hintergrund/Worldcoin-Tausche-Kryptowaehrung-gegen-Augen-Scan-7097714.html (6.11.23).

42. Richard Evans / Jim Gao, DeepMind AI Reduces Google Data Centre Cooling Bill by 40%, in: Google DeepMind, 20.7.2016, https://www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40 (6.11.23).

43. Vgl. Stefan Betschon, Die Demokratisierung der künstlichen Intelligenz, in: Neue Zürcher Zeitung, 15.12.2016, https://www.nzz.ch/digital/aktuelle-themen/machine-learning-die-demokratisierung-der-kuenstlichen-intelligenz-ld.135034 (6.11.23).

44. Vgl. Lisa Mayerhofer, „Künstliche Intelligenz wird demokratisiert", in: Werben & Verkaufen, 31.1.2020,

https://www.wuv.de/Archiv/%22K%C3%BCnstliche-Intelligenz-wird-demokratisiert%22 (6.11.23).

45. Rainer Mühlhoff, Human-Aided Artificial Intelligence: Or, How to Run Large Computations in Human Brains? Towards a Media Sociology of Machine Learning, in: New Media & Society 10 (2019), 10, S. 1868–1884.

46. Kate Crawford, Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence, New Haven, CT 2021.

47. Rehak, Artificial Intelligence for Real Sustainability?

48. Craig G. Smith, The Long and Mostly Short of China's Newest GPT, in: IEEE Spectrum, 28.7.2023, https://spectrum.ieee.org/china-chatgpt-wu-dao (6.11.23).

49. Frey / Osborne, The Future of Employment.

50. Constanze Kurz / Frank Rieger, Arbeitsfrei: Eine Entdeckungsreise zu den Maschinen, die uns ersetzen, München 2013.

51. Florian Butollo, Automatisierungsdividende und gesellschaftliche Teilhabe, in: regierungsforschung.de, 24.5.2018, https://regierungsforschung.de/automatisierungsdividende-und-gesellschaftliche-teilhabe/ (6.11.23).

52. Ebd.

53. Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen (Hg.), Unsere gemeinsame digitale Zukunft, Berlin 2019, http://www.wbgu.de/hg2019 (6.11.23).

54. Jeanette Hofmann, Demokratie und Künstliche Intelligenz, Digitales Deutschland, https://digid.jff.de/demokratie-und-ki/ (6.11.23).

55. Martin Rost, Künstliche Intelligenz. Normative und operative Anforderungen des Datenschutzes, in: DuD – Datenschutz und Datensicherheit – DuD 42 (2018), 9, S. 558–565, https://www.maroki.de/pub/privacy/2018-09_DuD-KI.pdf (6.11.23).

56. Vgl. etwa Florian Eyert / Florian Irgmaier / Lena Ulbricht, Extending the framework of algorithmic regulation. The Uber case, in: Regulation & Governance 16 (2022), 1, S. 23–44, https://doi.org/10.1111/rego.12371 (6.11.23).

57. Rost, Künstliche Intelligenz.

58. Vgl. etwa Marc Andreessen, Why AI Will Save the World, in: Andreessen Horowitz, 6.6.2023, https://a16z.com/ai-will-save-the-world/ (6.11.2023).

59. Vgl. Musk u.a., Pause Giant AI Experiments; oder Hinton u.a., Mitigating the risk of extinction.

60. Mühlhoff, Human-Aided Artificial Intelligence.

61. Amy Castor / David Gerard, Pivot to AI: Pay no attention to the man behind the curtain, in: Amy Castor, 12.9.2023, https://amycastor.com/2023/09/12/pivot-to-ai-pay-no-attention-to-the-man-behind-the-curtain/ (6.11.2023).

62. Vgl. allgemein Hans-Jörg Kreowski / Aaron Lye, Künstliche Intelligenz im Dienst des Militärs, FIfF-Kommunikation 4/23, S. 19–23; und konkret: Yuval Abraham, 'A mass assassination factory': Inside Israel's calculated bombing of Gaza, in: +972 Magazine, 30.11.2023, https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/.

63. Rainer Mühlhoff, „Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitische Wende des Digitalen Kapitalismus", in: Deutsche Zeitschrift für Philosophie 68 (2020), 6, S. 867–890, https://doi.org/10.1515/dzph-2020-0059.

64. Wolfgang Coy, Für eine Theorie der Informatik!, in: ders. u.a. (Hg.), Sichtweisen der Informatik, Braunschweig/Wiesbaden 1992.

65. Ralf Klischewski, Anarchie – Ein Leitbild für die Informatik: Von den Grundlagen der Beherrschbarkeit zur selbstbestimmten Systementwicklung, Bern 1996.

66. Silke Helfrich (Hg.), Wem gehört das Wissen der Welt? Zur Wiederentdeckung der Gemeingüter, München 2009.

67. Christian Ricardo Kühne, Zu Bedingungen und Möglichkeiten der Gestaltung herrschaftsfreier Informationssysteme, in: Rainer Fischbach / Klaus Lenk / Jörg Pohle (Hg.), Der Weg in die »Digitalisierung« der Gesellschaft, 2. Auflage. Marburg (im Erscheinen); Christian Ricardo Kühne, GNUnet und Informationsmacht: Analyse einer P2P-Technologie und ihrer sozialen Wirkung, Humboldt-Universität zu Berlin, https://doi.org/10.18452/14270 (6.11.23).

68. Tilman Santarius / Josephin Wagner, Digitalization and sustainability: A systematic literature analysis of ICT for Sustainability research, in: GAIA - Ecological Perspectives for Science and Society, 23 (2023), S1.

69. Cory Doctorow, Seizing the means of computation – how popular movements can topple Big Tech monopolies, Transnational Institute, https://www.tni.org/en/article/seizing-the-means-of-computation (6.11.23).

Rainer Rehak
Rainer Rehak ist wissenschaftlicher Mitarbeiter in den Forschungsgruppen „Digitalisierung, Nachhaltigkeit und Teilhabe“ und „Technik, Macht und Herrschaft“ am Weizenbaum-Institut für die vernetzte Gesellschaft sowie assoziierter Mitarbeiter am Wissenschaftszentrum Berlin für Sozialforschung (WZB). Er promoviert aktuell an der TU Berlin zu systemischer IT-Sicherheit und gesellschaftlichem Datenschutz. Seine Forschungsfelder sind Datenschutz, IT-Sicherheit, staatliches Hacking, Informatik und Ethik, Technikfiktionen, Digitalisierung und Nachhaltigkeit, konviviale und demokratische Digitaltechnik sowie die Implikationen und Grenzen von Automatisierung durch KI-Systeme. Er studierte Informatik und Philosophie in Berlin und Hong Kong und beschäftigt sich seit über 15 Jahren mit den Implikationen der Computerisierung der Gesellschaft. Er ist Sachverständiger für Parlamente (z. B. den Deutschen Bundestag) und Gerichte (z. B. das Bundesverfassungsgericht) und publiziert zudem regelmäßig auch in nicht-wissenschaftlichen Medien.

Dieser Beitrag wurde redaktionell betreut von Karsten Malowitz, Nikolas Kill.
Artikel auf soziopolis.de:
https://www.soziopolis.de/zwischen-macht-und-mythos.html