

Künstliche Intelligenz – Politik und Sicherheit

Korrespondenzadresse

Prof. Dr. Dr. Manfred Spitzer
Universität Ulm
Abteilung für Psychiatrie
Leimgrubenweg 12–14
89075 Ulm, Deutschland

Bibliografie

Nervenheilkunde 2023; 42: 680–684
DOI 10.1055/a-2106-0106
ISSN 0722-1541
© 2023. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Kaum eine Woche vergeht ohne neue Nachrichten über erstaunliche Fortschritte durch Künstliche Intelligenz (KI) in den Natur- und den Geisteswissenschaften sowie in so unterschiedlichen Bereichen wie Medizin, Militär, Klima, Verbrechensbekämpfung, Politik und Wirtschaft [7]. KI wird unsere Welt in einem Ausmaß verändern wie die Erfindung der Schrift oder des Buchdrucks. Einen Eindruck davon vermitteln die US-amerikanischen Autoren *Ian Bremmer* und *Mustafa Suleyma*¹ in der September/Oktober-Ausgabe des internationalen Politik-Fachblatts *Foreign Affairs*. Dort beschreiben sie in einem Artikel zum Verhältnis von Regierungen und KI die Situation der Welt in gut 10 Jahren.

KI in 10 Jahren

„Wir schreiben das Jahr 2035, und KI ist allgegenwärtig. KI-Systeme leiten Krankenhäuser, betreiben Fluggesellschaften und bekämpfen sich gegenseitig vor Gericht. Die Produktivität ist auf ein noch nie dagewesenes Niveau gestiegen, und zahllose, zuvor unvorstellbare Unternehmen haben sich in rasanter Geschwindigkeit vergrößert, was zu immensum Wohlstandswachstum geführt hat. Weil Wissenschaft und Technik auf Hochtouren laufen, kommen täglich neue Produkte, Medikamente und Innovationen auf den Markt. Und doch wird die Welt immer unberechenbarer und zerbrechlicher, weil Terroristen neue Wege finden, die Gesellschaft mit intelligenten, sich selbst weiterentwickelnden Cyberwaffen zu bedrohen, und zugleich vor allem Anwälte, Ärzte und höhere Angestellte massenhaft ihren Arbeitsplatz verlieren“ [1].

Noch vor weniger als einem Jahr hätte man Leute ungläubig ausgelacht, die so etwas prognostizierten. Heute dagegen erscheint das geschilderte Szenario durchaus plausibel – und keiner lacht! Die Veröffentlichung der generativen KI *ChatGPT* wird von Bremmer und Suleyman als „Big Bang“ (Urknall) bezeichnet, der „den Beginn einer weltverändernden technologischen Revolution, die Politik, Wirtschaft und Gesellschaft neugestalten wird,“ markiert.

In ihrem Artikel sprechen sie vom „KI-Macht Paradoxon“ (AI power paradox), welches darin besteht, dass KI nicht nur die Politik im Hinblick auf die Notwendigkeit der Regulierung herausfordert, sondern aufgrund ihrer „hyper-evolutionären Natur“ zugleich immer schwerer zu regulieren ist, je länger man damit wartet.

Sie führen zu Recht die sehr rasche Entwicklung von KI an, verglichen mit der Leistungsfähigkeit von Computern. Für letztere gilt seit Jahrzehnten das *Mooresche Gesetz*, dem zufolge sich die Leistungsfähigkeit von Computern alle 18–24 Monate verdoppelt. Für KI wurde dagegen während der letzten 10 Jahre eine Verzehnfachung der Leistungsfähigkeit pro Jahr beobachtet: Prozesse, die vor 10 Jahren noch Monate brauchten, geschehen heute in Sekunden. Zudem erfolgt die Verbreitung neuer KI-Modelle oft innerhalb von Tagen. Diese Modelle sind nichts weiter als riesige Zahlenmengen und für diese besteht – im Unterschied zu Sachen – kein „Transportproblem“. Sie lassen sich daher auch leicht „stehlen“ und illegal verbreiten. „Keine dermaßen leistungsfähige Technologie war jemals so leicht, so breit und so schnell für jedermann verfügbar“, bemerken die Autoren. „KI-Systeme sind nicht wie Autos oder Flugzeuge, die auf einer Hardware basieren, die sich schrittweise verbessern lässt und deren kostspieligste Ausfälle in Form von einzelnen Unfällen auftreten. Sie sind nicht mit chemischen oder nuklearen Waffen vergleichbar, die schwierig und teuer zu entwickeln und zu lagern sind, geschweige denn heimlich weitergegeben oder eingesetzt werden können. Da ihre enormen Vorteile offensichtlich werden, werden KI-Systeme immer größer, besser, billiger und allgegenwärtiger werden“ [1].

Die meisten Länder der Welt verfügen nicht über die Ressourcen, KI an der vordersten Front des Kenntnisstandes mitzuentwickeln, weswegen die großen KI-Firmen den Fortschritt bestimmen. Viele Länder hängen daher sowohl von den Firmen als auch von einigen anderen Ländern ab, die den Fortschritt tragen. So entwickeln große Staaten wie die USA und China KI und konkurrieren um die Vorherrschaft – und um die hierzu notwendigen Ressourcen. Weil KI auf einem ganz bestimmten Typ Hardware besonders schnell läuft – den sogenannten Grafikprozessoren (GPU), die vor allem für Computerspiele entwickelt wurden – haben die USA ein Exportverbot für bestimmte GPU nach China ausgesprochen, um den Konkurrenten hardwareseitig zu behindern. Keines der beiden Länder hält sich an das geforderte Moratorium der Weiterent-

1 Bremmer ist ein US-amerikanischer Politikwissenschaftler mit den Spezialgebieten Außenpolitik, Transformation von Staaten und globale politische Risiken, Gründer und Präsident des international tätigen Beratungsunternehmens *Eurasia Group* mit Hauptsitz in New York, Initiator einer gleichnamigen Stiftung sowie des *Global Political Risk Index* (GPRI) an der Wall Street. Der KI-Wissenschaftler und Unternehmer Suleyman ist Mitbegründer und war Chef der KI-Firma *DeepMind*.

wicklung von KI, weil eine solche Entwicklungspause „einer leichtsinnigen und lächerlichen einseitigen Abrüstung gleichkäme“ [1].

Sieht man einmal von den beiden genannten Staaten ab, läßt sich festhalten dass all dies bedeutet, dass die Entwicklung der KI zumindest in den nächsten Jahren weitgehend von den Entscheidungen einer Handvoll privater Unternehmen bestimmt werden wird, unabhängig davon, was die politischen Entscheidungsträger in Peking, Washington oder Brüssel tun. Mit anderen Worten: „Nicht Politiker oder Bürokraten, sondern Technologen werden die Macht über eine Kraft ausüben, die sowohl die Macht der Nationalstaaten als auch ihre Beziehungen zueinander tiefgreifend verändern könnte. Das macht die Beherrschung und Verwaltung von KI zu einer Herausforderung, mit der Regierungen noch nie zuvor konfrontiert waren, und zu einem regulatorischen Balanceakt, der heikler ist und bei dem mehr auf dem Spiel steht als alles, was politische Entscheidungsträger bisher zu regulieren hatten“, so die ebenso beunruhigenden wie bedenkenswerten Überlegungen von Bremmer und Suleyman [1], die sich für einen weisen Umgang mit KI aussprechen – sie nennen ihren Ansatz „Technoprudentialism“ („Technikweisheit“), bei dem KI weder kaputtreguliert bzw. in Verwaltung erstickt werden dürfe, noch unreguliert oder durch die Falschen reguliert auf die Menschheit losgelassen werden darf. Das ist keineswegs einfach, aber man sollte erstens alle Beteiligten mit an den Tisch holen, und sich zweitens vom Vorgehen bei anderen großen Herausforderungen (man spricht heute ja nicht mehr von Problemen) inspirieren lassen – „nämlich Klimawandel, globale Abrüstung und die Stabilität unseres Finanzsystems“.

„Ein starkes KI-Governance-System würde sowohl die von KI ausgehenden, gesellschaftlichen Risiken mindern als auch die Spannungen zwischen China und den USA abbauen, indem es das Ausmaß verringert, in dem KI ein Schauplatz – und ein Instrument – des geopolitischen Wettbewerbs ist. Ein solches System hätte zudem noch grundsätzlichere und dauerhaftere Auswirkungen, denn es wäre ein Modell für den Umgang mit anderen, neu aufkommenden disruptiven Technologien“, und sie meinen dies beispielsweise in Bezug auf Quantencomputer, Biotechnologie, Nanotechnologie und Robotik (man könnte ergänzen: Raumfahrt, neue Energiequellen, neue Formen des nachhaltigen Wirtschaftens mit nachhaltigen Ressourcen). Eine erfolgreiche Regulierung der KI würde der Menschheit dabei helfen, diese Technologien erfolgreich zu beherrschen. Sie schließen ihre Überlegungen wie folgt ab: „Das 21. Jahrhundert wird nur wenige so große Herausforderungen und so vielversprechende Möglichkeiten bieten wie die KI. Im letzten Jahrhundert begannen die politischen Entscheidungsträger mit dem Aufbau einer globalen Governance-Architektur, von der sie hofften, dass sie den Aufgaben des Zeitalters gewachsen sein würde. Jetzt müssen sie eine neue Governance-Architektur aufbauen, um die gewaltigste und potenziell bestimmende Kraft dieser Ära einzudämmen und nutzbar zu machen. Das Jahr 2035 steht vor der Tür. Es gibt keine Zeit zu verlieren“ [1].

Cyber-Sicherheit

KI wird von Menschen produziert und angewendet – was leider auch das Potenzial zu ihrem Missbrauch mit sich bringt. Ende März 2023 erschreckten führende Köpfe im KI-Bereich die Welt, weil sie in einem offenen Brief den sofortigen Stopp der Entwicklung und

des Einsatzes von KI sowie eine 6-monatige Pause (Moratorium) der Weiterentwicklung bzw. des Trainings von KI gefordert hatten. Der Brief wurde vom US-amerikanischen *Future of Life Institute (FLI)*² publiziert und von etwa 1300 Experten und Wissenschaftlern aus der KI- und Tech-Branche unterzeichnet – darunter Elon Musk und Steve Wozniak, der Mitbegründer von Apple. Ende Mai 2023 warnte eine Gruppe von KI-Spezialisten und KI-Firmenchefs erneut davor, dass KI eine existenzielle Bedrohung für die Menschheit darstellen könnte und ihr gesellschaftliches Risiko auf einer Stufe mit Pandemien und Atomkriegen steht. Ihre (englischsprachige) Erklärung bestand aus nur 22 Wörtern, war von mehr als 350 Führungskräften, Forschern und Ingenieuren im KI-Bereich – u. a. dem Chef der Firma *OpenAI*, Sam Altman, dem Chef der Firma *DeepMind*, Demis Hassabis, und dem zuweilen als „Pate der KI“ bezeichneten Informatiker Geoffrey Hinton – unterzeichnet worden. Sie wurde vom in San Francisco ansässigen Zentrum für KI-Sicherheit (*Center for AI Safety, CAIS*) und danach sofort von allen großen Zeitungen der Welt verbreitet. Der Text der Erklärung lautete: „Die Verminderung des Risikos der Auslöschung [der Menschheit] durch KI sollte neben anderen gesellschaftlichen Risiken wie Pandemien und Atomkriegen eine globale Priorität sein“ (CAIS 2023). Kurz: KI ist sehr gefährlich.

Wie kommen KI-Experten darauf? *ChatGPT* ist nichts weiter als ein recht unzuverlässiger Plauderroboter, von dem man nie weiß, ob er gerade halluziniert (wie man das mittlerweile ohne allzu großen Bezug zur Psychiatrie nennt) oder die Wahrheit spricht. Und warum sollte KI, die das Wetter besser vorhersagt oder in der Medizin Diagnose und Therapie optimiert, zu einem Weltkrieg oder einer Pandemie führen? – Alles nur Hype und Science-Fiction?

Aus meiner Sicht sollten wir uns über KI weniger Sorgen machen als darüber, wozu Menschen KI verwenden könnten. „Es ist einfach, keine KI in ein unternehmenskritisches System einzubauen“, stellt hierzu Sandra Wachter [6], Professorin für *Technology and Regulation* am *Oxford Internet Institute* der *University of Oxford*, lapidar fest. Sie fügt hinzu, dass es durchaus genügen sollte, dafür zu sorgen, dass keiner KI die Möglichkeit gegeben wird, beispielsweise einen Atomkrieg auszulösen. Die Anweisung „Den roten Knopf nicht mit KI direkt verbinden“ sollte ihrer Ansicht nach für intelligente Ingenieure einfach genug zu realisieren sein.

Viele Informatiker machen sich allerdings durchaus Sorgen über die Probleme, die *Menschen* mit KI schaffen könnten, einschließlich der aktuellen generativen KI-Systeme. So sorgt sich Wachter beispielsweise über die Umweltauswirkungen der vielen energieintensiven Rechenzentren, die für den Betrieb von KI erforderlich sind, und über die Bedrohung bestimmter Arbeitsplätze, die KI mit sich bringt. Hinzu kommen vor allem Gefahren für unsere Online-Sicherheit, denn generative KI kann die Fähigkeit von Kriminellen und anderen böartigen Akteuren zur Durchführung von Betrügereien und Cyberangriffen durchaus verbessern. Sofern Chatbots wie *ChatGPT* beim Programmieren helfen können, so das Argument, helfen sie eben auch Mächtgern-Cyberangreifern, die nicht programmieren können, beim Schreiben von Schadsoftware.

Hinzu kommt, dass Chatbots sehr rasch eine Lawine von Falschnachrichten oder auf bestimmte Gruppen zugeschnittene Des-

2 Das FLI wurde im Jahr 2014 in Cambridge, MA/USA, von Wissenschaftlern gegründet. Die gemeinnützige Institution hat das Ziel, die globalen Risiken durch fortgeschrittene KI zu verringern.

informationen aussenden können. Grafik-KI wie *Dall-E-2* oder *Mid-journey* kann gefälschte Bilder und Videos erstellen (man denke nur an den Papst in Daunenjacke und Turnschuhen oder die Verhaftungen von Putin und Trump), ohne dass man hierzu Fachkenntnisse braucht: Zuruf (man nennt es „prompt“) genügt. Weil solche fotografisch wirkenden Bilder eine sehr große Überzeugungskraft besitzen – man „sieht ja, was los ist“ – lassen sich mit ihnen Menschen leicht manipulieren.

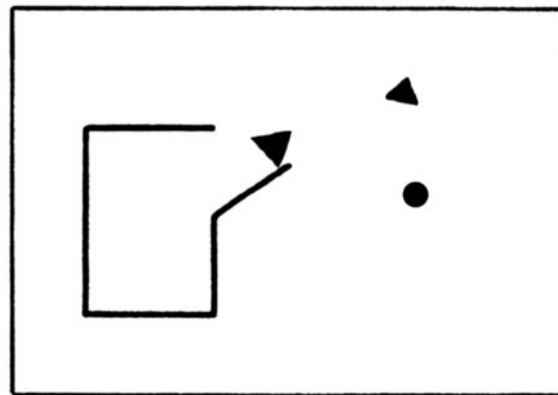
„Menschlich“, „sachlich-objektiv“, voller Vorurteile und daher gefährlich

Vor fast 80 Jahren führte der Sozialpsychologie Fritz Heider (1896–1988) experimentelle Studien durch, mit denen nachgewiesen werden konnte, wie leicht es Menschen fällt, auch den einfachsten „Agenten“ die Eigenschaft *Menschlichkeit* zuzuschreiben [7]. Er führte Versuchspersonen einen kurzen Trickfilm vor, in dem sich Symbole bewegten – ein großes und ein kleines Dreieck sowie ein Kreis (► **Abb. 1**). Danach wurden die Teilnehmer gefragt, was sie gesehen hatten. Allein aufgrund der Art der Bewegung der 3 Symbole sprachen die meisten Versuchspersonen davon, dass eine *Frau* (Kreis) von einem *Mann* (großes Dreieck) verfolgt bzw. bedroht wird und ein anderer freundlicher *Mann* (kleines Dreieck) sie schützen möchte [2].

Noch weiter gingen die Studien des Stanford-Professors für Mensch-Maschine-Interaktion Clifford Nass (1958–2013), dem Direktor des dortigen *Communication between Humans and Interactive Media (CHIME) Lab*. Er ließ beispielsweise Menschen die Leistung eines Desktop-Computers bewerten, wobei sich zeigte, dass sie den Computer positiver bewerteten, wenn sie ihre Antwort in den zu bewertenden Computer selbst eingaben, anstatt sie entweder (Kontrollgruppe 1) mit Bleistift auf ein Blatt Papier zu schreiben oder (Kontrollgruppe 2) in einen anderen Computer einzugeben. Sie verhielten sich dem Computer gegenüber also ebenso wie gegenüber einem Menschen, den man ja auch weniger kritisch beurteilt, wenn man ihm ins Gesicht schaut [5].

Nass konnte weiterhin zeigen, dass Menschen nicht nur Sozialverhalten, wie z. B. Höflichkeit und Gegenseitigkeit gegenüber Computern an den Tag legen, sondern sogar menschliche soziale Kategorien (geschlechtsspezifische oder ethnische Vorurteile sowie Persönlichkeitseigenschaften) auf Computer übertragen [4]. Solches Verhalten gegenüber Sachen ist völlig sinnlos, aber Menschen tun dies eben trotzdem. Daher muss man sich nicht darüber wundern, dass Sprachassistenten wie *Siri* (Apple, 2011), *Alexa* (Amazon, 2015) oder *Cortana* (Microsoft, 2014) trotz ihrer begrenzten Fähigkeiten (sie reagieren auf Fragen oder Befehle wie „Alexa, wie wird das Wetter heute?“ oder „Hey Siri, stell meinen Wecker auf 7 Uhr“) wie schon in den 1960er-Jahren *Eliza* (die vom Informatiker Joseph Weizenbaum programmierte „Psychotherapeutin“) sehr „menschlich“ behandelt werden: Jeden Tag sagen Hunderttausende von Menschen zu *Alexa* „Guten Morgen“, eine halbe Million Menschen haben ihr ihre Liebe gestanden, und mehr als 250 000 Menschen haben ihr einen Antrag gemacht. Eines der häufigsten zu *Alexa* gesprochenen Wörter ist „Danke“ [9].

Es ist genau diese Eigenschaft des Menschen, aus wenigen Informationen sozial anmutende Agenten zu konstruieren, was uns so anfällig gegenüber KI macht, mit der wir nicht nur zunehmend



► **Abb. 1** Bild aus dem Film von Heider. Man glaubt kaum, dass Menschen die Bewegungen von Dreiecken und einem Kreis als „handelnde Menschen“ interpretieren, aber genau dies war das wesentliche Ergebnis der sozialpsychologischen Experimente (nach Daten aus [2]).

interagieren, sondern die wir oft auch wie verlässliche Experten behandeln. Dies ist das Argument von Celeste Kidd, einer Psychologin an der *University of California* in Berkeley, und *Abeba Birhane*, einer Informatikerin bei der *Mozilla Foundation* in San Francisco. „Menschen bilden stärkere und länger anhaltende Überzeugungen, wenn sie Informationen von Personen erhalten, die sie als vertrauenswürdig und sachkundig einschätzen“, schrieben sie im Fachblatt *Science* [3] und wiesen darauf hin, dass generative KI immer mehr Material mit Verzerrungen und Fälschungen produziert. Und Millionen von Menschen fallen darauf herein. „Menschen unterstellen generativen KI-Modellen häufig Intentionalität, menschliche Intelligenz und das Auftreten von Empfindungen, obwohl diese Attribute [von KI] nicht belegt sind. Diese Bereitschaft, generative KI-Modelle als kenntnisreiche, intentionale Agenten wahrzunehmen, geht mit der Bereitschaft einher, die von ihnen gelieferten Informationen schneller und mit größerer Sicherheit zu übernehmen. Diese Tendenz kann noch verstärkt werden, weil sie [die Nutzer] mit den Modellen in mehreren Sinnesmodalitäten umgehen können, die es den Nutzern ermöglichen, die Modelle aufzufordern, Handlungen wie „sehen“, „zeichnen“ oder „sprechen“ auszuführen, die mit Intentionalität assoziiert sind. Der potenzielle Einfluss der problematischen Ergebnisse von Modellen auf menschliche Überzeugungen übersteigt somit den Einfluss, der typischerweise für andere Formen algorithmischer Inhaltsvorschläge wie beispielsweise Suche [im Internet] beobachtet wird“, fügen die Autoren hinzu [3]. Der Umgang mit KI ist aus den genannten Gründen also manipulativer als der Umgang mit digitaler Informationstechnik in Form von herkömmlichen Computern oder Suchmaschinen.

Dies sei im Folgenden noch etwas näher ausgeführt. Wenn Menschen eine KI wie *ChatGPT* etwas fragen, dann sind sie definitionsgemäß im Hinblick auf die Antwort unwissend und zugleich neugierig. Sobald sie dann eine Antwort erhalten haben, nehmen – ebenfalls definitionsgemäß – Ungewissheit und Neugier ab. Die Antwort wird Teil ihres Vorwissens, und – gemäß dem unvermeidbaren hermeneutischen Zirkel bei jeglichem Erkenntnisprozess –, beeinflusst das weitere Vorgehen. Auch das ist ganz grundsätzlich so und kann

gar nicht anders funktionieren [8, 10]. Nachfolgende Erfahrungen werden daher unweigerlich „mit der Brille“ der Vorerfahrungen gemacht, was dem Erkenntnisfortschritt einerseits förderlich sein kann, aber andererseits bedeutet, dass nachfolgende Erfahrungen nicht mehr in dem Maße berücksichtigt bzw. bewertet werden wie die ersten Erfahrungen. Geht es um Meinungen und Überzeugungen, so folgt daraus, dass Menschen in der frühen Phase ihrer Meinungsbildung besonders beeinflussbar sind. Wenn also Menschen bei Unwissenheit grundlegend „erstmal im Internet suchen“, und wenn in wenigen Monaten alle größeren Suchmaschinen durch KI „im Hinblick auf ihr User-Interface verbessert“, sprich: durch KI gefiltert werden (wie von allen großen KI-Firmen geplant), dann folgt daraus, dass KI die Meinung von Menschen in Zukunft stärker beeinflussen wird. Aus der Sicht des Nutzers bekommt er von der KI eine Antwort auf seine Frage, also von einer Maschine, die sehr viel weiß und – weil sie ja „nur“ eine Maschine ist – vermeintlich ohne jegliche Vorurteile und Wertungen. Faktisch jedoch hat die KI schon die Vorurteile, die sie selbst enthält, und (sofern die Suche personenbezogen funktioniert) die Vorurteile des Nutzers in ihre Antwort mit einbezogen. Schon der Beginn jeder Suche nach Erkenntnis im KI-basierten Internet ist damit gleich mehrfach problematisch: Die KI liefert – ebenso unbemerkt wie unvermeidbar – ihre eigenen Vorurteile und die des Nutzers gleich mit, und ihre Nutzung am Anfang geht notwendig mit einer geringeren Wahrscheinlichkeit dafür einher, dass noch eine Änderung der Meinung erfolgt, von einem Erkennen der Vorurteile gar nicht zu reden.

Der Leser kann sich ausmalen, was es bedeutet, wenn weltweit in den fortschrittlichen Gesellschaften die Nutzung der Informationsbeschaffung mit Hilfe von KI erfolgt. Es sei hier noch einmal betont, dass der beschriebene Mechanismus nicht allein auf die KI zurückzuführen ist, sondern auf einer – ungunstigen – Wechselwirkung zwischen künstlicher und menschlicher Intelligenz beruht.

Die Autoren der *Science*-Arbeit drücken dies wie folgt aus: „Dieser Aspekt der menschlichen Neugier hat langfristige Auswirkungen darauf, wie KI-Systeme menschliche Überzeugungen beeinflussen. Er hat zur Folge, dass Informationen, die von großen Sprachmodellen [Large Language Models, LLM] an eine unsichere Person übermittelt werden, im Nachhinein nur schwer zu aktualisieren sind, da die vom Modell gelieferten Informationen die Unsicherheit der Person auflösen, selbst wenn sie falsch sind“ [3]. So ergibt sich ein Mechanismus, der – von vielen Menschen einer Gesellschaft benutzt – zur Aufrechterhaltung von falschen Überzeugungen führt. Kurz: Das Entstehen und vor allem der Verbleib von Verschwörungstheorien bis zum Sankt Nimmerleinstag werden durch KI begünstigt.

Ein weiteres Problem besteht darin, dass von KI produziertes Material mittlerweile das Internet überflutet und damit zu einem Teil der Trainingsdaten für die nächste Generation von KI wird. Dadurch werden „systematische Verzerrungen in einer kontinuierlichen Rückkopplungsschleife in die Zukunft hineinprojiziert, verstärkt

und verfestigt. Je schneller solche Systeme genutzt und angenommen werden und je mehr sie in überall verwendete KI-Systeme eingebaut werden, desto mehr Einfluss haben diese Systeme auf die menschlichen Überzeugungen“ [3].

Weil wir im Grunde genommen erst am Anfang dieser Entwicklung stehen, sollte es – noch – möglich sein, ihre Auswirkungen mit wissenschaftlichen Methoden zu untersuchen. Wie eingangs jedoch bereits deutlich wurde, wird das Zeitfenster hierfür sehr rasch kleiner. Wird KI erst einmal von vielen Menschen oft verwendet, ist es zu spät. Aus meiner Sicht heben die Autoren auch den folgenden Aspekt zu Recht hervor: „Die Erforschung der Auswirkungen generativer KI-Modelle auf die Überzeugungen von Kindern ist von besonderer Priorität. Kinder sind anfälliger für Glaubensverzerrungen, da sie stärker dazu neigen, Technologie zu vermenschlichen, und weil ihr Wissensstand noch nicht so weit entwickelt und [daher besonders] beeinflussbar ist“ [3]. Aus dieser Sicht erscheint es mir besonders problematisch, dass viele Pädagogen es gar nicht abwarten können, große Sprachmodelle, also KI wie *ChatGPT*, an Schulen einzusetzen. Im Fachblatt *Science* steht, wie gefährlich das ist. – Nicht allein für die Kinder, sondern für unsere gesamte Gesellschaft.

Literatur

- [1] Bremmer I, Suleyman M. The AI Power Paradox. *Foreign Affairs* 2023; 9/10. <https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox>; abgerufen am 11.9.2023
- [2] Heider F, Simmel M. An experimental study of apparent behavior. *American Journal of Psychology* 1944; 57: 243–259
- [3] Kidd C, Birhane A. How AI can distort human beliefs. *Science* 2023; 380: 1222–1223
- [4] Moon Y, Nass C. How “real” are computer Personalities? *Communication Research* 1996; 23: 651–674
- [5] Nass C, Moon Y, Carney P. Are People Polite to Computers? Responses to Computer-Based Interviewing Systems. *Journal of Applied Social Psychology* 1999; 29: 1093–1109
- [6] Russon MA. Forget human extinction – these are the real risks posed by AI today. *New Scientist*, 25.7.2023. <https://www.newscientist.com/article/2384063-forget-human-extinction-these-are-the-real-risks-posed-by-ai-today/>; abgerufen am 25.8.2023
- [7] Spitzer M. Künstliche Intelligenz. München: Droemer, 2023
- [8] Spitzer M. Keilschrift, KI und Hermeneutik. *Nervenheilkunde* 2023; 42: 237–242
- [9] Turk V. How we fell in love with our voice-activated home assistants. *New Scientist* 2016; 3104
- [10] Wade S, Kidd C. The role of prior knowledge and curiosity in learning. *Psychon Bull Rev* 2019; 26: 1377–1387