

Die KI sei mit euch

DIE KI SEI MIT EUCH

MACHT, ILLUSION UND KONTROLLE
ALGORITHMISCHER VORHERSAGE

HELGA NOWOTNY

Aus dem Englischen von
Sabine Wolf



Matthes & Seitz Berlin

INHALT

DIE VERWUNSCHENE WELT VON GPT-4	
Vorwort für die deutsche Ausgabe	7
EINLEITUNG: MEINE REISE INS DIGI-LAND	31
1. DAS LEBEN IN DER DIGITALEN ZEITMASCHINE	65
2. WILLKOMMEN IN DER SPIEGELWELT	106
3. DAS FORTSCHRITTSNARRATIV UND DIE SUCHE NACH DEM ÖFFENTLICHEN GLÜCK	155
4. ZUKUNFT BRAUCHT WEISHEIT	192
5. DISRUPTION: VON VORPANDEMISCHEN ZEITEN ZUR SELBSTDOMESTIZIERUNG	235
DANK	273
ANMERKUNGEN	275

DIE VERWUNSCHENE WELT VON GPT-4

Vorwort für die deutsche Ausgabe

Der Spielraum für menschliches Handeln verengt und erweitert sich gleichzeitig mit beängstigender Geschwindigkeit. Kaum wurde ChatGPT im November 2022 der Öffentlichkeit zugänglich gemacht, verzeichnete es innerhalb einer Woche mehr als eine Million Nutzer. OpenAI und dessen Eigentümer, Microsoft, begannen ihr groß angelegtes Experiment mit allen, die das Potenzial der neuen KI erkunden wollten, ebenso wie mit jenen, die noch nicht wussten, was es damit eigentlich auf sich habe. Niemand wurde darüber aufgeklärt, dass es sich um einen Probelauf handelte, in dem die ersten Nutzer:innen wertvolles Feedback lieferten über alles, was schief laufen könnte. Der unbestreitbare Vorteil: Microsofts Vorsprung. So wurde die Konkurrenz gezwungen, mitzuziehen, und ChatGPT war bald von einer Handvoll weiterer LLM (Large Language Models) umringt, unter ihnen Midjourney und Dall·E, die nicht nur Texte, sondern auch Bilder generieren können.

Generative KI-Systeme werden mit dem Ziel programmiert, menschliche Kommunikation zu simulieren. Sie stehen für eine neue digitale Technologie, die imstande ist, einige der kognitiven Aufgaben, die bisher von Menschen ausgeübt wurden, mit erstaunlicher Effizienz

und Schnelligkeit zu übernehmen. Noch ist offen, ob sich der Techno-Enthusiasmus, der ChatGPT und seine Verwandten begleitet, halten oder ob unweigerlich eine Desillusionierung einsetzen wird, wie es bei anderen mit großem Hype angekündigten Neuerungen der Fall war. Das Potenzial für eine General Purpose Technology, wie es im Innovationsjargon heißt, also für eine für Anwendungen in vielen Bereichen geeignete Technologie, ist jedenfalls vorhanden. Wirtschaftshistoriker:innen verweisen auf Ähnlichkeiten mit der Elektrifizierung, die in den Zeiträumen von 1880 bis 1910 zu umfassenden Umwälzungen in Wirtschaft, Gesellschaft und Kultur führte. Diesmal bleibt der Gesellschaft jedoch nicht so viel Zeit, um die Auswirkungen zu absorbieren. Sicher ist aber: Achtzig Jahre nach dem Beginn der Digitalisierung lässt sich deren Beschleunigung nicht mehr aufhalten.

Wie die generativen KI-Systeme genau funktionieren und wie ihr Output entsteht, darüber herrscht noch große Unklarheit. Judea Pearl, ein hoch angesehener KI-Forscher, brachte es auf den Punkt: »Der Kern unseres Problems ist, dass vorschnell Super-Investitionen in eine nicht-interpretierbare Technologie fließen.«¹ LLM beruhen auf einer innovativen Kombination von unbeaufsichtigtem Training (*unsupervised training*) und bestärkendem Lernen (*reinforcement learning*) eines enormen Datensatzes. Dieser stammt aus dem Internet und speist sich, im Falle von ChatGPT, aus Texten aller Art, aus Büchern, Websites und allem, was in den sozialen Netzwerken aufscheint. Dieser Sprachschatz wird ständig über die im Internet dominanten Sprachen hinaus, allen voran Englisch, erweitert. Programmiert ist die KI als dialogisches System, das eine menschliche Konversation simuliert. Das gelingt, indem es die Wahrscheinlichkeiten schätzt,

mit der in einem Satz ein Wort auf das andere folgt. Das Ganze wird mit Millionen von Parametern moduliert, die einen gewissen Stil und ein spezielles Genre imitieren und andere Einzelheiten beachten, um die Illusion eines Gesprächs oder um etwa ein Sonnet im Stil Shakespeares zu erzeugen.

Die generativen KI-Systeme können weder »denken«, »verstehen« noch mit abstrakten Begriffen argumentieren – noch nicht. Sie »wissen« weder, was Fakten sind, noch können sie ihre Quellen angeben. Sie sind fehleranfällig und daher extrem unzuverlässig. Sie neigen zu »Halluzinationen«, wie das Forscher:innen bei Google 2018 erstmals nannten. Gemeint sind Texte oder Bilder, die zwar semantisch oder syntaktisch plausibel, aber ansonsten falsch oder schierer Unsinn sind. Es fehlen die eingebauten Kriterien, um »wahr« von »falsch« zu unterscheiden. Und auch viele der Filter, die den in den Daten massiv vorhanden gesellschaftlichen Bias ausschalten sollen, versagen bei LLMs.

Am 14. März 2023 präsentierte OpenAI der Öffentlichkeit bereits die nächste Generation ihrer KI-Systeme: GPT-4. GPT-4 ermöglicht Input sowohl als Text wie als Bild. Es beherrscht Mathematik auf einem Niveau, von dem die meisten Menschen nur träumen können. Die Fehler, die es macht, können auch guten Mathematikern passieren. GPT-4 erreicht mit Leichtigkeit die höchsten Noten bei Aufnahmeprüfungen und kann Reden von Politikern im Hinblick auf ideologische Färbung und anzusprechendes Publikum perfekt imitieren. So zeigen Forscher:innen bei Microsoft in einem Research-Paper, wie die KI bei einer simulierten Anti-Impf-Kampagne die Sprache variiert, um in der Zielgruppe die gewünschten

Emotionen hervorzurufen: Wut, Angst, Schuldgefühle oder Stolz, anders als die anderen zu sein. Doch die kaum gebremste Begeisterung der Autor:innen gilt einer anderen Eigenschaft. Sie glauben, »Funken« einer »Artificial General Intelligence« bei GPT-4 entdeckt zu haben.²

Doch das ist nur der Anfang, denn schon längst wird intensiv an GPT-5 gearbeitet. Und auch die Konkurrenz überbietet sich seither darin, ihre Modelle aus der Pipeline zu holen, um sie auf dem Markt zu platzieren. Es gibt kein Zulassungsverfahren, kein Protokoll, um Schäden während des Ausprobierens zu begrenzen, und keinerlei Transparenz, lediglich weitere Investitionen in Millionenhöhe, ganz nach dem Motto: *be fast and break things*. Kein Wunder, dass diese rasante Entwicklung eine Flut an öffentlichen Diskussionen auslöst. Sie schwankt zwischen Techno-Enthusiasmus, apokalyptischen Spekulationen und ernststen Bedenken. Die eine Seite sieht in der generativen KI den entscheidenden Schritt zur Verwirklichung der Künstlichen Allgemeinen Intelligenz, zu einer Intelligenz also, die der menschlichen ebenbürtig ist oder sie übertrifft, die »denken« und »verstehen« kann. Manche schwärmen davon, dass es dieser Superintelligenz gelingen werde, auch wissenschaftliche Durchbrüche – etwa in der Quantentechnologie – zu erzielen, so wie es DeepMind von Google gelungen ist, mit AlphaFold das bisher ungelöste Problem der Proteinfaltung zu meistern. Dann werde die Menschheit endlich das gefunden haben, wonach sich alle sehnen: die Lösung all ihrer Probleme, wenn nicht gar die Erlösung schlechthin.

Auf der anderen Seite stehen die Skeptiker:innen. Während einige Expert:innen bezweifeln, ob LLM überhaupt der richtige Weg zur Künstlichen Allgemeinen Intelligenz sei, mahnen die meisten erst einmal zu größerer

Vorsicht, um irreparablen Schaden jetzt abzuwenden, statt Angst vor einem Ereignis zu schüren, von dem niemand weiß, wann und ob es eintreten wird. Der nicht zuletzt durch den wohl bekanntesten Unterzeichner, Elon Musk, berühmt gewordene, vom Future of Life Institute in Oxford initiierte und von mehr als tausend im IT-Bereich tätigen CEOs, führenden Ingenieur:innen und anderen Fachleuten unterschriebene »Offene Brief« vom 29. März 2023 heizte die Diskussion weiter an. Inhaltlich wird ein sechsmonatiges Moratorium für die Zeit der Entwicklung von GPT-5, also der nächsten Generation der KI-Systeme von OpenAI, gefordert, um mögliche schädliche Auswirkungen auf die Menschheit besser abschätzen zu können.

Selbst wenn die Umsetzung und das Monitoring eines solchen Moratoriums kaum realistisch ist, klingen viele der angeführten Begründungen vernünftig. Wer könnte etwas gegen einen Zeitaufschub einwenden, um die heutigen, mächtigen AI Systeme »more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal« (was immer damit gemeint sein mag) zu machen?³ Kritik an dem Brief gibt es vor allem aufgrund des darin mitschwingenden »Hypes«, der den Eindruck erweckt, dass sich die Menschheit tatsächlich nahe dem Wendepunkt befinde, an dem der ultimative Reifegrad einer »Artificial General Intelligence« erreicht werde, während die generative KI in Wirklichkeit weit davon entfernt ist. Weshalb, so schließt die Kritik daran an, werde immer nur über die Zukunft der KI und die möglicherweise damit verbundenen existenziellen Risiken gesprochen und nicht über die Menschen, die heute davon betroffen seien, und weshalb nicht über jene, die die KI-Systeme vorantreiben, finanzieren und programmieren würden, um ausschließlich Gewinne damit zu machen?

Jede neue Technologie eröffnet einen Projektionsraum für die kollektive Imagination. Diesmal ist er besonders weit aufgespannt, spielt er doch selbst mit genau den technischen Mitteln und Möglichkeiten, die in ihm abgebildet werden. Es ist ein mehrdimensionaler Raum, in dem die »wirkliche« Welt und die virtuelle einander spiegeln und eine undurchdringliche Gemengelage aus Gegenwart und spekulativer Zukunft bilden. Der Mangel an Tiefenschärfe wird dabei von der Auflösung bekannter Kategorien und Grenzziehungen begleitet. Wenn viele Nutzer:innen überzeugt sind, dass sie in ihrer Konversation mit der KI mit einem »wissenden« Wesen verbunden sind, das ihre Fragen und Anliegen versteht, kann man dies zwar auch als Unwissen abtun. Es zeigt jedoch die Verwischung der Grenze zwischen einer subjektiv erfahrenen und einer objektiven Wirklichkeit an, die durch diese Technologie hervorgerufen wird. Die Spaltung in der Gesellschaft zwischen denen, die zu wissen glauben, und denen, die etwas anderes zu wissen glauben, vertieft sich dadurch zunehmend. Die Frage lautet daher: Wie können wir verhindern, in einer Gesellschaft zu leben, in der jede und jeder von uns ausschließlich in ihrer oder seiner »personalisierten«, durch KI geformten Wirklichkeit lebt, die sich dem oder der anderen nicht mehr mitteilen lässt?

Wie also ist in dieser Hinsicht diese vorerst letzte, mit rasanter Geschwindigkeit vorantreibende KI einzuordnen? Entgleitet uns nun endgültig die Kontrolle über die von uns geschaffenen Technologien, und stehen wir tatsächlich vor der großen Wende, die uns der noch verbleibenden Autonomie und letztlich unseres Menschseins beraubt? Oder sind dies nur spekulative Überlegungen, die zwar ihren eigenen existenziellen Reiz ausstrahlen,

doch in Wahrheit von Menschen in die Welt gesetzt werden, in deren Interesse es ist, uns von den naheliegenden existenziellen Bedrohungen abzulenken?

Zu ebendiesen naheliegenden existenziellen Bedrohungen zählt etwa die Sorge über die verschwindenden Arbeitsplätze. Wir kennen das schon, mag man einwenden. Frühere Wellen der Automatisierung ersetzten manche Routinearbeiten in den Fabriken, doch neue Arbeitsplätze entstanden. Diesmal sind die kognitiven Routinearbeiten in den Büros von Anwält:innen und Redaktionen, in den Schulen und in der Verwaltung an der Reihe, für die GPT-4 und seine Ko-Spezies weitaus kostengünstigere und effizientere Dienstleistungen anbieten. Kurzfristig bringt es also wenig, ständig menschliche Leistungen mit jenen zu vergleichen, bei denen die KI weitaus effizienter abschneidet. Je mehr die Torpfosten verschoben werden, desto mehr Tore werden geschossen. Offen bleibt lediglich die Frage, wie schnell, wenn überhaupt, jene übernommenen Routinearbeiten durch neu entstehende Arbeitsplätze ersetzt werden und ob Funktionen, die bisher weitgehend von dafür ausgebildeten und zertifizierten Personen ausgeübt wurden, schon heute abgegeben werden sollten. So sind bereits jetzt – vorwiegend in den USA – eine Vielzahl therapeutischer KIs im Einsatz, die den Patient:innen, die sich aus Kostengründen eine »echte« Therapie nicht leisten können, Hilfe bei ihren psychischen Problemen anbieten. In Belgien wurde jedoch vor Kurzem der erste Fall eines Suizids gemeldet, dem eine lange, ausschließlich mit einer KI (einer früheren Generation) geführte Konversation vorausging.

Die größten Gefahren ergeben sich jedoch, wie bereits angedeutet, für unsere offenen, demokratischen Gesellschaften. Die Konzentration ökonomischer Macht

in den Händen einer kleinen Elite, die den großen internationalen Konzernen vorsteht, sichert diesen einen unverhältnismäßig dominanten Einfluss auf demokratische Prozesse und Institutionen zu, die von der ungebrochenen Aushöhlung des öffentlichen Raumes begleitet wird. Zudem bergen die digitalen Plattformen von Big Tech die Fähigkeit, die öffentliche Meinung durch eine unregulierte und daher unkontrollierbare Flut von bewusst irreführenden Informationskampagnen zu manipulieren und zu polarisieren. Sie öffnen Tür und Tor für jene Kräfte, sei es von innen oder von außen, die beabsichtigen, die demokratischen Prozesse zu unterminieren.⁴ Hinzu kommen die enormen Investitionen, die von privater Seite in den weiteren Ausbau der digitalen Systeme fließen und bereits zur »industriellen Übernahme« der öffentlichen Forschung geführt haben.⁵ Keine Universität verfügt über die Rechenkapazitäten oder über den Zugang zu den Datenmengen, die den großen Konzernen und ihren Nebenfirmen zur Verfügung stehen. Diese wären notwendig, um von unabhängiger Seite die von den Konzernen angekündigten Sicherheitsmaßnahmen und versprochenen Verbesserungen zu begleiten und zu überprüfen. Stattdessen vergrößert sich das strukturelle Ungleichgewicht, und das Risiko, dass die Richtung der zukünftigen Forschung angesichts der ungleichen Verteilung der Ressourcen zunehmend von Big Tech bestimmt wird, wiegt schwer. Zudem müssen manche Universitäten bangen, bald nicht mehr imstande zu sein, die Ausbildung auf höchstem Niveau fortzuführen, da die Rekrutierung von Professor:innen zunehmend schwieriger wird, werden die besten Talente doch bereits früh von der Industrie abgeworben.

Diesen berechtigten Befürchtungen und Sorgen stehen zweifellos die positiven Auswirkungen gegenüber, die mit der generativen KI verbunden sind. Tagtäglich erreichen uns Meldungen über neue Anwendungen. GPT-4 erleichtert und erledigt zahlreiche mühsame Routinearbeiten, inklusive Coding, wenngleich Fehleranfälligkeit immer überprüft werden muss. Illustrationen und Websites können im Handumdrehen erstellt werden, Sitzungsprotokolle innert weniger Minuten geschrieben und Texte in viele Sprachen übersetzt werden. Die Wirtschaft erwartet einen breit gestreuten Einsatz in Firmen, der höhere Produktivitätszuwächse verspricht und dessen Spannbreite vom Chip Design zu neuen Medikamenten, vom Erstellen juristischer Dokumente zu umwälzenden Neuerungen im Bildungsbereich reicht. Und wie die Geschichte von technologischen Innovationen beweist, sind viele Anwendungen oft noch nicht einmal vorhersehbar. Die Art und Weise, wie eine Innovation adoptiert und angeeignet wird, weicht mitunter beträchtlich von jener ab, für die sie ursprünglich bestimmt war. Nutzer:innen sind nicht nur passive Konsument:innen, sondern haben immer wieder gezeigt, dass sie geplante Anwendungen für ihre Bedürfnisse umdefinieren und neue Anwendungen erfinden können.

Eine Kosten-Nutzen-Rechnung der »guten« und »schlechten« Auswirkungen greift also auch dieses Mal zu kurz. Dienlicher wäre eine umfassende Risikoabwägung, doch müsste ihr ein ebenso umfassendes und durchsetzungsfähiges Risikomanagement folgen. Denn jede Technologie verführt zur Illusion, sie unter menschlicher Kontrolle zu haben. Lange hat es etwa gedauert und der Austragung vieler Konflikte bedurft, bis die im industriellen Zeit-

alter eingesetzten Maschinen als einigermaßen sicher und somit unter menschlicher Kontrolle galten. Seither haben sich die Anforderungen ständig ausgeweitet. Längst geht es nicht mehr nur darum, das Funktionieren von Maschinen und deren Sicherheit im Interesse der Gewinne für die Eigentümer zu gewährleisten, sondern mögliche Schäden für Menschen, für deren Gesundheit und für die Umwelt zu vermeiden oder zumindest so gering wie möglich zu halten. So sollen zahlreiche Sicherheitsvorschriften, Zulassungsverfahren und Kontrollmaßnahmen dafür sorgen, die Sicherheit von Technologien zu gewährleisten. Das gilt für den Straßenverkehr ebenso wie für die internationale Luftfahrt, für die Zulassung neuer Medikamente und für neue Verfahren, die in der Gentherapie eingesetzt werden, für Wasser, Luft und Boden und für alles, was zur Sicherheit der Gesundheit und Ernährung beiträgt, ja sogar für die nukleare Sicherheit. Die Geschichte zeigt jedoch, dass die Illusion der Kontrolle immer wieder zum Platzen gebracht wurde und im Nachhinein stets nachgebessert werden musste.

Doch die Kontrolle über die KI bringt noch zusätzliche Anforderungen mit sich. Die hohe Anfälligkeit für Fehler und Verstärkung gesellschaftlicher Vorurteile sowie die mangelnde Unterscheidungsfähigkeit zwischen »wahr« und »falsch« untergraben ihre Glaubwürdigkeit. Die Forderung nach *alignment*, der Übereinstimmung der KI mit menschlichen Werten, steht dabei am Anfang ihrer Umsetzung. Es gilt, sowohl technische Hürden zu überwinden als auch die Frage zu beantworten, welche menschlichen Werte als verbindlicher Kompass dienen sollen und für wen sie Gültigkeit beanspruchen können. Ein weiteres Problem hängt mit der Intransparenz zusammen, auf der die großen Konzerne als legale Eigentümer:innen der Da-

ten und Algorithmen beharren. Perverserweise führt bei ihnen die Forderung nach mehr Sicherheit zu noch größerer Abschottung, wird doch argumentiert, dass eine Öffnung genau jenen zugutekomme, die beabsichtigten, die Sicherheitslücken für ihre illegalen Zwecke zu umgehen oder auszunutzen.

Viele algorithmische Verfahren und die Architektur der LLM ähneln daher einer Blackbox, bei der sich zwar Input und Output vergleichen lassen, doch das, was innerhalb geschieht, bleibt undurchsichtig oder unbekannt. Viele Expert:innen sind von den Leistungen von GPT-4 zutiefst beeindruckt, sie erkennen die Chance, neue Einsichten zu gewinnen, und fühlen sich *blown away*, wie weggeweht und geben offen zu, dass sie keine Ahnung haben, wie GPT-4 das macht, was es macht. Die im Inneren der KI ablaufenden Prozesse sind – jedenfalls vorläufig – kausal nicht erklärbar. Niemand weiß genau, woher welche Daten kommen und an welchen das jeweilige KI-Modell trainiert wird. Niemand kann die vielleicht emergenten Ergebnisse erklären, die bei GPT-4, aber auch bei anderen KI-Modellen ab einer gewissen Größe auftreten. Treten tatsächlich emergente Phänomene auf, oder ist ihr vermeintliches Auftreten nur der Beweis für unser Unwissen? Und diejenigen, die es eigentlich wissen müssten, sind allem Anschein nach besorgt, nicht über das erforderliche Wissen zu verfügen, um die notwendige Kontrolle über die Vorgänge unter ihrer Aufsicht und Verantwortung auszuüben.

Wie kann sich unter diesen Vorzeichen Vertrauen einstellen, und – die vielleicht entscheidende Frage – wo kann die staatliche Regulierung einsetzen, und wie kann sie greifen? Bekanntlich hinkt jede gesetzliche Regulierung immer der technologischen Entwicklung nach. Sie

gleich dem Wettlauf zwischen Achill und der Schildkröte. Was immer sowohl in Europa, das sich gerne als Vorläufer gesetzlicher Regulierung sieht, wie in den USA gegenwärtig unternommen wird, ist zu begrüßen und zu unterstützen, doch sollte man freilich nicht glauben, dass dadurch das Ausmaß und die erforderliche Art der Kontrolle erreicht würden, die wir bisher in Bezug auf andere Technologien erwarten konnten. Vielleicht braucht es dafür völlig neue Institutionen und eine neue Art der Gesetzgebung, die fortschreibend begleitet, was die gelebte Erfahrung im Umgang mit der KI an Desiderata hervorbringt, um darauf entsprechend zu reagieren oder sie zu antizipieren.

Innerhalb der Gesellschaft überwiegt jedenfalls weiterhin das Gefühl von Ambivalenz, Erstaunen und tief-sitzendem Unbehagen gegenüber der nun so sichtbar in das Alltagsleben eingedrungenen KI. Dies liegt zum Teil auch daran, dass wir Dingen, und nicht nur Lebewesen, die Absicht zum rationalen Handeln zuschreiben, was Daniel Dennett als *intentional stance* beschreibt.⁶ Wir unterstellen ihnen dann, zu »denken«, zu »machen«, zu »wissen« oder zu »glauben«, ohne weiter zu reflektieren, dass wir es nicht mit anderen Menschen zu tun haben, die den Gebrauch der Sprache mit uns teilen. Die Verwendung einer anthropomorphisierenden Sprache und eines ebensolchen Denkens verwischt den Unterschied zwischen »uns« und »ihnen« und verleitet dazu, dem digital »Anderen« Handlungsfähigkeit (*agency*) zuzuschreiben. Diese Tendenz wird gezielt durch die generative KI gefördert oder, besser gesagt, ausgenutzt, indem Worte wie »ich« und eine persönliche, gefühlsbetone Sprechweise verwendet werden, um eine sprechende Person zu simulieren. Gleichzeitig gewöhnen wir uns durch die Verbrei-

tung dieser KI-Systeme in der Arbeitswelt und im Alltag an die digitalen »Anderen«, egal unter welchen Namen und in welcher Form sie sich präsentieren.

Hinzu kommt die Verwendung des Wortes »Intelligenz«, das, wie immer wieder betont wird, irreführend ist. Der in den 1950er-Jahren geprägte Begriff *artificial intelligence*, so heißt es übereinstimmend, sei eine unglückliche Wahl gewesen, gebe es doch weder eine Einigung darüber, was menschliche Intelligenz sei, noch darüber, wodurch sie sich von einer »künstlichen« unterscheide. »Intelligenz« fungiert daher als ein schlecht definierter Platzhalter, der mehr zudeckt als erklärt. Menschliche Intelligenz lässt sich besser auf einem Kontinuum positionieren, das wir mit anderen Lebewesen teilen, besitzen doch alle Lebewesen die Fähigkeit, sich in der jeweiligen Umwelt zurechtzufinden und Strategien zum Überleben zu entwickeln. Vielleicht gewinnen wir durch einen erweiterten und differenzierteren Begriff von »Intelligenz« auch neue Einsichten für unsere Beziehung zu einer KI, die unsere kognitiven Fähigkeiten zu simulieren und sie für unterschiedliche Zwecke einzusetzen vermag.

Unbestreitbar bleibt, dass im Zusammenspiel zwischen Menschen und den digitalen Maschinen etwa Neues entsteht. Wir befinden uns in einem koevolutionären Prozess zwischen uns Menschen als vorläufigem Ergebnis der biologischen Evolution und den von uns geschaffenen Dingen, Phänomenen und Prozessen, die von der kulturellen Evolution – allen voran durch Wissenschaft und Technik – angetrieben wird. Ob aus diesem Prozess eine Symbiose entsteht, bleibt offen, ebenso, wohin er führen wird.

An diesem Wendepunkt eröffnen sich neue, faszinierende Fragen und damit ungeahnte Möglichkeiten, mehr

zu erforschen und zu lernen – in erster Linie über uns selbst. Im Vordergrund stehen dabei unsere kognitiven Fähigkeiten und wie sie funktionieren: Wenn die KI imstande ist, Algebra zu generieren, sie dies aber auf eine für uns bisher unbekannte Weise tut, so wirft das Fragen auf hinsichtlich der Funktionsweise unseres Gehirns. Wenn ein LLM uns in einer Sprechweise antwortet, die sich nicht von unserer unterscheiden lässt, was sagt das über unseren Sprachgebrauch und vor allem über unsere Sprache? Der Erwerb der Sprache ist das Natürlichste, was uns bei der Geburt von der Evolution mitgegeben wird, und unterscheidet Menschen von anderen Lebewesen. Die KI kann sie so perfekt imitieren, dass wir meinen, mit einem menschlichen Gesprächspartner verbunden zu sein. Zugleich kann sie anderen Menschen Worte in den Mund legen, die diese nie gesagt haben. Sie ist hervorragend im Manipulieren und Lügen, gut im Erfinden von Geschichten und erstaunlich effizient im Ausführen von Aufgaben, die wir ihr stellen. Die Sprache hat sich, so scheint es, verselbstständigt. Sie ist nicht mehr nur uns eigen, sie ist unserer Kontrolle entglitten.

Gewiss, die Worte sind Texten entnommen, die einmal von Menschen geäußert wurden. Und gewiss, das KI-System »versteht« nicht, was es sagt, und »weiß« nicht, was das Gesagte bedeutet, doch das macht es nur umso faszinierender. Im Unterschied zu uns sucht das KI-System keinen Sinn, sondern agiert sinnbefreit, und das mit beeindruckender Effizienz. Dadurch entstehen neue Kombinationen und Stilrichtungen, ganz so wie in der gelebten Sprache. Worin liegt also der Unterschied? Ist es noch unsere Sprache, oder entsteht hier eine Meta-Sprache? Anders gefragt: Welchen Unterschied macht es, ob die Sprache *embodied* ist, also der genetischen Aus-