



Samstag, 10. August 2024, 15:57 Uhr
~6 Minuten Lesezeit

Riskanter Fortschritt

Experten warnen vermehrt vor globalen Gefahren durch KI.

von Werner Thiede
Foto: Ar_TH/Shutterstock.com

Dass künstliche Intelligenz (KI) faszinierende Möglichkeiten eröffnet und helfen kann, viele nutzbringende Dinge zu realisieren, ist unbestritten. Doch auf kaum einem anderen Gebiet zeigt sich die Ambivalenz des Fortschritts so deutlich wie hier. Warnungen von Hightech-Experten vor globalen Gefahren durch KI gibt es schon seit Jahren. In letzter Zeit aber sind solche Stimmen immer häufiger geworden. Und zwar nicht etwa als Äußerungen zweifelhafter Verschwörungstheoretiker oder schräger Kulturpessimisten, sondern ernst zu nehmender, seriöser KI-Fachleute.

Freilich hört in der Regel niemand gerne Unheilspropheten.

Dennoch sind die Warnungen mittlerweile unüberhörbar geworden, sodass sie sich nicht mehr ignorieren oder verdrängen lassen. Auch wenn immer noch eine Mehrheit von Fachleuten der Faszinationskraft von KI so sehr erliegt, dass sie entweder nichts von den massiven Zukunftsgefahren hören wollen oder dem Glauben anhängen, technische Intelligenz und politische Regelungen würden die Probleme schon irgendwie bewältigen können, wächst mittlerweile die Zahl kompetenter KI-Kritiker weltweit.

Beispielsweise warnte Ende März 2023 Sam Altman als OpenAI-Chef zusammen mit anderen führenden Fachleuten vor einem denkbaren Auslöschen der Menschheit durch KI. Gleichzeitig mahnte Eliezer Yudkowsky, ein führender KI-Forscher, in einem Kommentar des *Time Magazine* vom 29. März 2023:

„Viele Forscher, die sich mit diesen Fragen beschäftigen, darunter auch ich, gehen davon aus, dass das wahrscheinlichste Ergebnis der Entwicklung einer übermenschlich intelligenten KI unter den gegenwärtigen Umständen darin besteht, dass buchstäblich jeder auf der Erde sterben wird.“

Wenig später äußerten namhafte Persönlichkeiten um den Multimilliardär Elon Musk in einem Offenen Brief, alle Technologie-Labore sollten die Entwicklung von KI-Systemen sofort unterbrechen: Solange nicht einmal die Hersteller die Maschinen wirklich verstünden, seien die Risiken zu groß, dass die Systeme im Zuge erworbener Selbständigkeit eine inhumane Ausrichtung annehmen.

Auf dem KI-Sicherheitsgipfel 2024 in Seoul warnten dann 25 weltweit führende KI-Forscher, KI mache rasche Fortschritte in kritischen Bereichen und könne schon bald beispiellose

Kontrollprobleme aufwerfen. Mahnend äußerten sich dieses Jahr auch der deutsche Top-Manager Jan Leike sowie Ilya Sutskever, ein Mitbegründer von OpenAI, die enorme Macht einer „Superintelligenz“ könnte zur Entmachtung der Menschheit oder sogar zu ihrem Aussterben führen. Beide hatten deshalb bei OpenAI gekündigt. Sodann äußerte sich Karl von Wendt, der über KI promoviert und KI-relevante Start-ups gegründet hat, kürzlich in der Zeitschrift *Technikfolgenabschätzung in Theorie und Praxis*:

„Uns bleiben mit Glück noch höchstens zwei Jahrzehnte, wenn wir Pech haben nur noch wenige Jahre. Meine persönliche Erwartung ist angesichts der sich überschlagenden Ereignisse eher auf der kürzeren Seite.“

KI könne dafür sorgen, dass sie nicht mehr abzuschalten sei. Und der optimale Weltzustand, den KI auf Basis ihres Ziels anstrebe, sei „höchstwahrscheinlich nicht mit unseren Wünschen und oft auch nicht mit unserem Überleben vereinbar“.

Apokalyptisch anmutende KI-Risiken sind demnach nicht mehr von der Hand zu weisen. Hier hat man es offenbar nicht mit billigem oder gar abwegigem Alarmismus zu tun, sondern mit den Äußerungen kenntnisreicher und vorausschauender Verantwortungsträger. In der Welt *am Sonntag* vom 24. März 2024 hat Dirk Schümer darauf hingewiesen, dass viele führende Köpfe, die bereits Einblick in die Fähigkeiten ihrer „Zauberlehrlinge“ haben, tatsächlich pessimistisch sind:

„Wenn bald Systeme unser Zusammenleben mittels Gesichtserkennung und Kontodaten regeln, wenn sie im Internet beliebig künstliche Wahrheiten erzeugen, wenn sie das Finanzsystem stärker kontrollieren, wenn sie erst Viren entwickeln und Kriege führen, dann könnte für Menschen in dieser Wirklichkeit kein Platz mehr sein. Ehrliche Entwickler von Software gestehen, dass sie

aufgrund ihrer Erfindungen schlecht schlafen. Es gibt bereits spezielle Therapien gegen die Angst vor der digitalen Apokalypse.“

Für wie besorgniserregend manche Experten durch das Voranpreschen der KI-Entwicklung die Gesamtlage und namentlich die gar nicht mehr ferne Zukunft ansehen, hat am 14. Juni 2023 die ZDF-Sendung „Der gefährliche Goldrausch in Silicon Valley“ von Greta Buschhaus und Nicolas Wildschutz dargelegt. Bald darauf erschien ein *Spiegel*-Interview mit dem Oxford-Philosophen Nick Bostrom, der zu bedenken gab:

„Wenn wir jetzt Maschinen schaffen, die schlauer sind als wir, könnte das bedeuten, dass der Lauf der Dinge auf unserem Planeten künftig von diesen Maschinen bestimmt wird.“

Eine superintelligente KI „könne alle möglichen Wege finden, uns den Garaus zu machen.“

Weil die Wissenden auf dem Gebiet der KI-Entwicklung die drohenden Gefahren nicht einfach pauschal ignorieren können, haben sich seit 2023 die Bemühungen intensiviert, für KI brauchbare, meist ethisch motivierte Regelungen auf den Weg zu bringen.

Auf einem ersten globalen KI-Sicherheitsgipfel hatten sich im Herbst 2023 in London Staats- und Regierungschefs zu einem Minimalkonsens entschlossen, der in der Selbstverpflichtung zu einem verantwortungsvollen Umgang mit KI bestand. Im Vorfeld des zweiten KI-Sicherheitsgipfels im Mai 2024 in Seoul warnten weltweit führende KI-Forscher, es werde nicht genug getan, um die Menschheit vor den Risiken der Technologie zu schützen. Immerhin sagten dann zu Beginn des Gipfels führende Technologieunternehmen der Welt eine möglichst sichere KI-Entwicklung zu. Doch die Worte „möglichst sicher“ klingen sehr relativ; und was sind sie wert angesichts der gewichtigen

kommerziellen Bestrebungen, die in jenen Unternehmen vorherrschen? Vor allem aber: Wie effektiv sind angestrebte Regelungen tatsächlich? Warum hier leider Skepsis geboten ist, hat bereits ein Beitrag auf **Manova** (<https://www.manova.news/artikel/die-kapitulation-der-ethik>) verdeutlicht und muss hier nicht wiederholt werden.

Insbesondere stellen sich auch kritische Fragen mit Blick auf die Folgen von KI-Anwendungen auf militärischem Gebiet. Das hat eindrucksvoll die *arte*-Sendung „Killerroboter – KI im Krieg“ von Daniel Andrew Wunderer am 30. Januar 2024 dargelegt. Demnach gibt es zwar den Versuch eines Regulierungsprozesses der UN, aber der gilt bereits als gescheitert. Bis heute verbietet kein einziges Abkommen den Gebrauch autonomer Waffensysteme, die auf die Dauer das Töten doch nur immer weiter beschleunigen und maschinell präzisieren. Die erwähnten 25 führenden KI-Forscher warnten diesbezüglich in Seoul, autonome Systeme könnten die Kontrolle über Waffensysteme, darunter auch biologische Waffen, übernehmen; eine Auslöschung der Menschheit sei dann denkbar, wenn der Mensch die Kontrolle über solche Systeme verliere.

Hier ist zu bedenken, dass es wohl nur bei einer *Allgemeinen KI* ab einem bestimmten Punkt ihrer Leistungsfähigkeit unmöglich wird, vorauszusagen, ob und wann sie unkontrollierbar werden könnte. Karl von Wendt differenziert:

„Eine Gefahr geht vor allem von allgemeinen KIs aus, die nahezu beliebige Probleme lösen und langfristige Pläne schmieden können. Solche ‚Universalgenies‘ brauchen wir eigentlich gar nicht.“

Daher wäre es aus seiner Sicht das Klügste, vorläufig auf die Weiterentwicklung solch Allgemeiner KIs zu verzichten und Investitionen stattdessen in Spezialisierte KIs zu stecken. Indes – kennt man die Macher fortschrittlichster Technologien als zum Verzicht bereite und stets vorsichtig abwägende Leute oder nicht

viel eher als Zeitgenossen, die schon aus wirtschaftlichen Gründen und mit dem notwendigen Blick auf die Konkurrenz nach dem Motto handeln: „Was gekonnt wird, wird gemacht“? Gilt das nicht insbesondere auf militärischem Gebiet?

Angesichts solch düsterer Perspektiven ist es eigentlich recht kurzsichtig, ja fahrlässig, ungebremst in die Weiterentwicklung von KI zu investieren. Fortschritt um jeden Preis ist ein hochriskantes Programm geworden. Ob sich menschliche Intelligenz als intelligent genug erweisen wird, bei künstlicher Intelligenz entschieden auf die Bremse zu treten? Ulrike Guérot hat mit Recht gefragt:

„Was, wenn die Arbeit am Menschen und an der Menschlichkeit die einzige Chance ist, der Apokalypse zu entkommen, die künstliche Intelligenz aber genau das nicht gestattet?“

Es ist jetzt an der Zeit, nicht länger in mehr oder weniger bedenkenloser Weise den verlockenden Stimmen aus Technik und Naturwissenschaft zu folgen, sondern weisheitliche, spirituelle Überlegungen nach vorne zu bringen. Sollte das nicht global gelingen, dann dürften sehr bald wirklich apokalyptische Zeiten anbrechen – und zwar nicht nur wegen KI allein.

„Krieg, Klima, künstliche Intelligenz – die Angst vor der Apokalypse ist zurück“, war in der genannten Ausgabe der *Welt am Sonntag* zu lesen. Und Angst ist keineswegs immer irrational, sondern kann im Blick auf die Umstände sogar höchst rational, ja ein guter Lehrmeister sein. Was aber, wenn niemand mehr gute Lehren annehmen will, weil man sich angewöhnt hat, das Denken primär künstlicher Intelligenz zu überlassen?



Werner Thiede ist Pfarrer im Ruhestand, außerplanmäßiger Professor für Systematische Theologie an der Universität Erlangen-Nürnberg und Publizist. Seine neuesten Bücher sind „Himmlisch wohnen“, „Himmlische Freude. Vom tiefen Glück des Glaubens“ sowie „In Ängsten – und siehe, wir leben! Glaubenslieder“. Weitere Informationen unter **werner-thiede.de** (<https://www.werner-thiede.de/>).