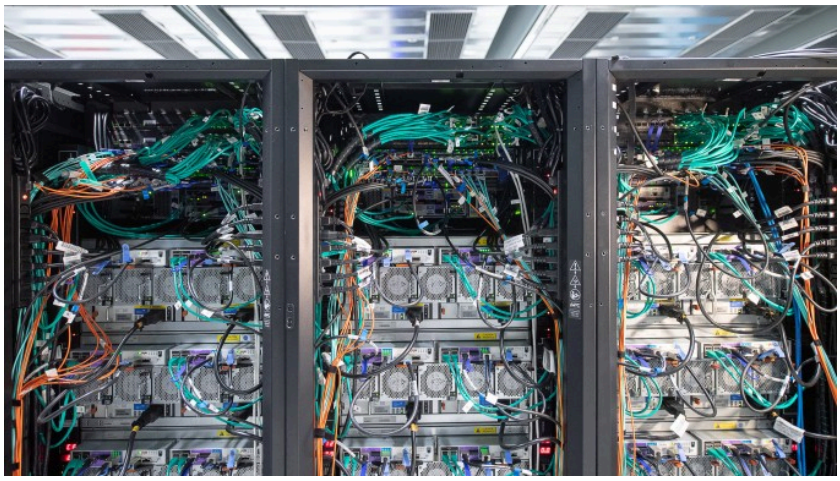
Frankfurter Allgemeine | 27.11.2023 - Aktualisiert: 28.11.2023, 22:51 Uhr
https://www.faz.net/-j9k-bilun

Künstliche Intelligenz

Was eine Superintelligenz-KI anrichten könnte, und was wir heute dagegen tun sollten

ChatGPT macht Spaß, aber was ist, wenn es ernst wird? Experten erwarten eine Superintelligenz-KI, die über eine nie dagewesene Machtfülle verfügt. Ein Gastbeitrag von Christoph Herr.

Von CHRISTOPH HERR



© dpa

Dieser Hochleistungscomputer, mit dem neuronale Netze berechnet werden, steht in Stuttgart.

Sam Altman, der „Jetzt wieder“-Vorstandschef von Open AI, hat bei einer Anhörung im US-Kongress im Frühjahr dieses Jahres gewarnt, dass Künstliche Intelligenz (KI) ein „Auslöschungsrisiko“ für die Menschheit darstelle. Er wirkte sehr besorgt und sagte mit ernsthafter Miene: „Wenn diese Sache schiefgeht, dann kann sie völlig schiefgehen.“

Jeder, der sich aktuell mit Tools wie ChatGPT auseinandersetzt, freut sich zunächst einmal über die großen Erleichterungen, die damit verbunden sind. Wissen wird einfach und schnell zugänglich, Code kann automatisiert generiert werden, Protokolle werden ohne menschliche Hilfe vom Kollege KI erstellt und vieles mehr. Was daran könnte der Menschheit schaden?

© dpa

Der Grund für Altmanns Sorge ist, dass wir heute nur einen Schnappschuss der potentiellen Fähigkeiten der KI sehen. Die Entwicklung ist längst nicht zu Ende, und die nächsten Schritte sind bereits vorgezeichnet. Technologie-Vordenker erwarten die folgenden Entwicklungsstufen:

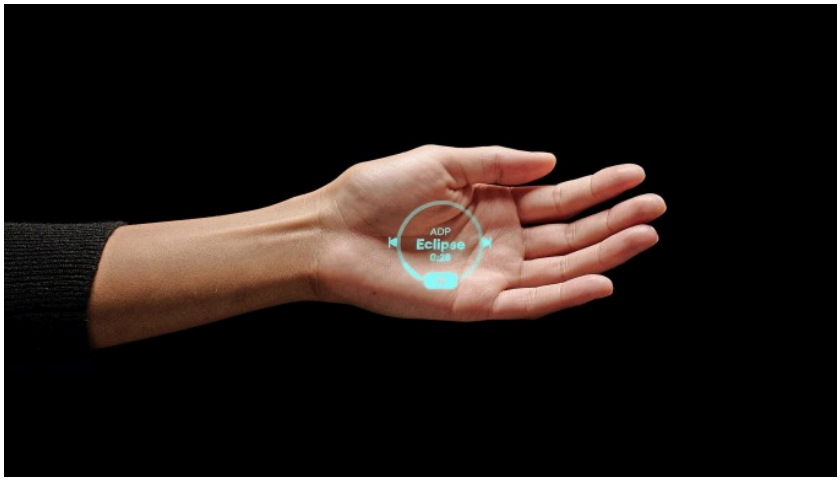
- **Stufe 1: Artificial Narrow Intelligence (ANI):** Automation von alltäglichen Aufgaben und Prozessen
- **Stufe 2: Artificial General Intelligence (AGI):** Kann begründen, Probleme lösen, abstrakt denken und Entscheidungen treffen
- **Stufe 3: Artificial Superintelligence (ASI):** Ist menschlichen Fähigkeiten in vielen Bereichen überlegen

Einige Experten sind der Ansicht, dass man in Anbetracht des Umfangs und der Tiefe der Fähigkeiten von GPT-4 diese als eine frühe, aber noch unvollständige Version eines Systems der Künstlichen allgemeinen Intelligenz (AGI), also Stufe 2, betrachten kann, auch wenn das diskutabel erscheint. Die Zukunftsforscherin und Professorin Amy Webb erwartet eine AGI dagegen erst in den 2040er-Jahren. Wie auch immer, die Zeit bis zu einem System der Stufe 2 scheint nicht mehr weit zu sein. Einig ist man sich darüber, dass nach der Artificial General Intelligence die Artificial Superintelligence (Stufe 3) kommt. Und deren potentielle Fähigkeiten sind es, die die Fachleute schon heute in Sorge versetzen. Die KI-Superintelligenz wird über eine Machtfülle und über Fähigkeiten verfügen, die es bisher so noch nie gegeben hat. Der KI-Pionier I. J. Good sagte schon vor hundert Jahren eine „Intelligenzexplosion“ durch KI voraus, die schon vor Eintritt in die Stufe 3 erwartet wird. Die Intelligenz der Maschinen wird dann deutlich größer sein als die der Menschheit.

Der Weg dorthin kann bereits skizziert werden: Ausgehend von Stufe 1 bis zu Stufe 3 werden wir die Entwicklung intelligenter, denkender Maschinen erleben, die darauf aus sind, sich ständig selbst zu optimieren. Sie werden eigenständig strategische, aber auch ganz operative Entscheidungen treffen. Man stelle sich ein hochgradig vernetztes und von KI gesteuertes Smarthome vor. Sicherlich ist es angenehm, wenn der Ofen sich abschaltet, sobald der Braten zu verbrennen droht, oder der Kühlschrank Buch führt über die darin befindlichen Lebensmittel und der Einkaufszettel maschinell erstellt wird – oder gleich an die entsprechenden Läden geschickt wird, die daraufhin automatisch liefern.

Was aber, wenn Sie eine Party planen und viele Gäste erwarten und daher viel eingekauft haben und alles im Kühlschrank lagern. Die Kühlschrank-KI folgert aber, dass Sie vorhaben, zu viel zu essen, und sperrt den Zugang, und Ihnen fehlt die Berechtigung, die Entscheidung der Kühlschrank-KI außer Kraft zu setzen. Wenn Sie dann neu einkaufen möchten, um die Party nicht platzen zu lassen, verweigert überraschenderweise die KI Ihres Bankkontos die Zahlung, weil diese der Ansicht ist, dass sie schon genügend Vorräte haben. Was in diesem Szenario noch lustig klingt, fühlt sich im Straßenverkehr schon ganz anders an. Denken wir nur an das Punktesystem in China, in dem der „Social Score“ steigt oder abnimmt, je nachdem, wie man sich im Verkehr oder ganz allgemein in der Gesellschaft verhält. Wenn auch hier eine KI selbständig Entscheidungen trifft und Sie in Ihrem Alltagsleben dominiert und „erzieht“, sind Sie dem System komplett ausgeliefert. Digital eingesperrt sozusagen.

Das Kühlschrankbeispiel zeigt auch, wie wichtig einheitliche Standards und kompatible Schnittstellen sind. Vielleicht haben Sie ja die Party in Ihrem Kalender eingetragen und auch die Zahl der erwarteten Gäste vermerkt. Da Kalender und Kühlschrank aber von inkompatiblen KIs verwaltet werden, kam diese wichtige Information aus dem Kalender nicht beim Kühlschrank an. Fehlende Kompatibilität führte dann im Beispiel zum Deadlock.



© Picture Alliance
Handverlesen: Der kleine Projektor des AI Pin wirft den Inhalt auf die Innenfläche der Hand.

Wie viel Zeit haben wir noch bis zur prognostizierten Intelligenzexplosion? Nach Ansicht von Mikko Hyppönen, einem der bekanntesten IT-Sicherheitsexperten der Welt, wird eine KI-Superintelligenz in den nächsten 30 bis 40 Jahren verfügbar sein. Er geht weiter davon aus, dass das erste KI-System, das die Linie zur KI-Superintelligenz überschreitet, auch für immer das dominierende System sein wird. Da es sich dann in hohem Tempo stetig selbst verbessern kann, werde es niemand mehr einholen können. Das Unternehmen oder das Land, das diese Technologie kontrolliert, werde in jedem Konflikt überlegen sein. Andere Experten können sich vorstellen, dass es einige wenige, miteinander kaum kompatible KI-Superintelligenzen geben kann. Die fehlende Interoperabilität führt nach deren Vorstellungen zu einer Art digitaler Kastensysteme, die aufgrund der Geschäftsmodelle einerseits und der persönlichen Lebenssituationen der Menschen wie Wohlstand, Bildung, Wertvorstellungen andererseits entstehen. Die Kasten sind dabei regelrecht zementiert, und das nicht nur für den Einzelnen, sondern auch für seine Nachkommen.

Abhängig davon, wer als erster ins Ziel geht, könnten wir eine wohlwollende, freundliche, übermenschliche KI-Superintelligenz erleben, die bisher unheilbare Krankheiten, den Klimawandel, Armut, Krieg und alle weiteren Probleme der Menschheit lösen wird. Oder wir

erleben das genaue Gegenteil. Das sind nicht gerade erfreuliche Nachrichten, und aus dem Grund denken bereits heute viele Menschen, Organisationen und Unternehmen darüber nach, wie man mit einer solchen Bedrohungslage umgehen kann.

Das eingangs erwähnte Statement von Sam Altman stammt aus einer Anhörung des amerikanischen Kongresses, in der es um die mit KI verbundenen Auswirkungen ging. Im nächsten Schritt plant das Weiße Haus die Herausgabe weitreichender Richtlinien an mehr als ein Dutzend Behörden, die auf ihren Umgang mit Systemen der Künstlichen Intelligenz abzielen.

Sie sehen unter anderem vor, dass bei Programmen, die potentiell gefährlich für die nationale Sicherheit, Wirtschaft oder Gesundheit sind, die Entwickler die US-Regierung schon beim Anlernen der KI-Modelle unterrichten müssen. Auch werden die Entwickler Ergebnisse von Sicherheitstests mit den Behörden teilen müssen. Die Anordnung plant zudem die Ernennung eines KI-Rates des Weißen Hauses, der die KI-Aktivitäten der Bundesregierung koordinieren soll, unter dem Vorsitz des stellvertretenden Stabschefs für Politik des Weißen Hauses und besetzt mit Vertretern aller wichtigen Behörden. Diese Anordnung stellt den bedeutendsten Einzelversuch dar, einer Technologie eine nationale Ordnung aufzuzwingen.

© Hersteller

Das böse Bellen der KI

Ein einzelnes Land wird aber wohl nicht in der Lage sein, die sich abzeichnende extrem mächtige KI-Superintelligenz zu regulieren. So ist nur zu verständlich, dass die EU und 28 weitere Staaten, darunter die Vereinigten Staaten und China, Anfang November 2023 eine Erklärung zu Künstlicher Intelligenz verfasst und unterzeichnet haben. Sie wollen sich gemeinsam darauf ausrichten, die Risiken von KI-Systemen wissenschaftlich zu erforschen und auf dieser Grundlage Regeln zu entwickeln. Diese Erklärung entstand im Rahmen einer KI-Sicherheitskonferenz und ist als Bletchley-Erklärung bekannt.

Weiterhin haben die G-7-Länder (Kanada, Frankreich, Deutschland, Italien, Japan, Großbritannien und die USA) kürzlich einen Verhaltenskodex für Entwickler von generativer Künstlicher Intelligenz veröffentlicht. Man habe sich zunächst für einen Kodex und gegen eine Gesetzgebung entschieden, da man davon ausgehe, dass eine breite internationale Vereinbarung über grundlegende KI-Normen effektiver sei. Andererseits ist bekannt, dass die EU an einem KI-Gesetz arbeitet, insofern wird der Kodex wohl nur eine Art Übergangslösung bis zur fertigen Regulierung sein.

Apropos Kodex: Über die Entwickler von KI-Systemen nachzudenken scheint der richtige Weg zu sein. Am Ende sind genau sie die Protagonisten des Wettrennens zur Super-KI. Und ihr persönlicher Einfluss auf die Funktionsweise eines KI-Systems ist nicht zu unterschätzen. So ist schon heute erkennbar, dass die Werte und Einstellungen von KI-Entwicklern Einfluss auf die Funktionsweise einer KI haben. Eine englische Studie namens „More human than human: measuring ChatGPT political bias" kommt zu dem Schluss, dass ChatGPT einen „ausgeprägten Linksdrall" habe. So würden die vorgeblich neutralen Antworten von ChatGPT auf politische Fragen dem linken Gedankenspektrum entsprechen, auch wenn immer wieder gesagt wird, das Sprachmodell sei neutral. Dieser Erkenntnis liegt das Gesetz von Conway zugrunde, nachdem Systeme generell die Menschen und deren Werte widerspiegeln, die sie entwickelt haben. Man wird davon ausgehen dürfen, dass von chinesischen Entwicklern erstellte KI-Systeme anderen Normen und Werten folgen werden als die von europäischen.

Vorerst treffen noch bei jedem Schritt Menschen Entscheidungen bei der KI-Entwicklung. Aber was ist, wenn KI-Systeme Entscheidungen bei ihrer eigenen Optimierung treffen? Auf Basis welcher Werte und Einstellungen werden sie das tun? Welches Land, welches Unternehmen geht als erstes durchs Ziel zur Super-KI und wird alle anderen auf Basis der eigenen Werte und Einstellungen dominieren?

Amy Webb hat sich mit dieser Frage in ihrem Buch „The Big Nine" auseinandergesetzt und drei mögliche Szenarien beschrieben. Sie unterscheidet dabei ein optimistisches, ein pragmatisches und ein Katastrophenszenario. Das letztere beginnt mit den Sätzen: „2023 haben wir unsere Augen vor der Entwicklung der künstlichen Intelligenz verschlossen. Wir haben alle Signale übersehen, die Warnzeichen ignoriert und keine aktive Zukunftsplanung betrieben." Und selbst im „Pragmatischen Szenario" heißt es am Ende: „China konnte ein furchterregendes System entwickeln und umsetzen, um das Gros der Weltbevölkerung zu kontrollieren. Erfüllen wir Chinas Forderungen nicht, legt es unsere Kommunikationssysteme lahm. Halten wir unsere Datenpipeline nicht für die Kommunistische Partei Chinas offen, schaltet es uns lebenswichtige Infrastruktur wie Kraftwerke und Luftverkehrsüberwachung ab. Sie sind jetzt Bürger der digital von China besetzten Staaten von Amerika."

In welcher Welt unsere Kinder und Enkel angesichts dieser möglichen Szenarien leben werden, hängt von den Unternehmern, Politikern, Wissenschaftlern und Bürgern Europas ab. Wir müssen uns bewusst machen, was mit einer Super-KI auf uns zukommt, um dann darauf hinzuwirken, dass diese freiheitlich, demokratisch und menschlich gestaltet wird. Oder dass wir sie gar nicht erst entstehen lassen – auch ein solches Szenario ist denkbar. Es ist in jedem Fall genau jetzt die Zeit, an einem Strang zu ziehen und mit allen Ländern der Welt gemeinsam an Regeln und Vorgehensweisen zu arbeiten, die der Menschheit dienen.

Gleichzeitig ist es für Deutschland und die EU extrem wichtig, als professionelle Akteure in der KI-Branche wahrgenommen zu werden. Nur wenn wir selbst in der Lage sind, KI-Systeme aufzubauen und sie immer weiter zu perfektionieren, werden wir im Chor der Länder der Welt beim Thema KI eine ernst zu nehmende Stimme haben. Das Unternehmen des ChatGPT-Erfinders Sam Altman, Open AI, wird in den Medien heute als „wichtigstes Start-up der Welt" bezeichnet. Aber auch in Deutschland und Europa gibt es eine Reihe von KI-Start-ups mit großem Einfluss und hoher Professionalität. Nun wird es für Wirtschaft, Wissenschaft, Politik und Gesellschaft darauf ankommen, vom Ziel her zu denken und alles zu tun, um gemeinsam das bestmögliche Szenario für die Stufen 2 und 3 zu erreichen. Dazu gehören freiheitliche, demokratische und menschenfreundliche Werte, die sich im Code und im Systemverhalten niederschlagen, aber auch kompatible KI-Systeme und funktionierende

Schnittstellen. Und wir sollten alles dafür tun, dass die besten KI-Start-ups künftig in Deutschland und der EU verortet sind.

Christoph Herr

Christoph Herr ist Experte für Plattformökonomie bei einem der größten europäischen Industrieverbände, leitet dort die Arbeitskreise der CDOs und CIOs, und arbeitet in Projekten der Datenökonomie, wie Manufacturing-X, mit. Vorher war er in den Digitalisierungs- und Innovationsbereichen verschiedener DAX- und MDAX-Unternehmen tätig.

Bild: Privat

Quelle: FAZ.NET