Jetzt klingt seine Warnung noch lauter

Der Physik-Nobelpreis geht an zwei Vordenker der künstlichen Intelligenz. Einer von ihnen mahnt inzwischen gegen seine eigene Schöpfung VON ULRICH SCHNABEL

Es ist in zweierlei Hinsicht bemerkenswert, dass die Königlich Schwedische Akademie der Wissenschaften in diesem Jahr zwei Pioniere der künstlichen Intelligenz (KI) auszeichnet. Zum einen ist deren Arbeit gar keine Physik im engeren Sinne; zum anderen gehört der geehrte Geoffrey Hinton heute zu den vehementesten Mahnern gegen die Gefahren von KI. Das ist in etwa so, als hätte das Nobelkomitee in den 1950er-Jahren Robert Oppenheimer den Preis verliehen, dem »Vater der Atombombe« – der später zum Warner vor Kernwaffen wurde.

Die Parallele zwischen KI und Bombe hat Hinton selbst gezogen. »Das Risiko des Aussterbens durch KI sollte neben anderen Risiken von gesellschaftlichem Ausmaß wie Pandemien und Atomkrieg eine globale Priorität sein«, forderte er im Mai 2023 als Erstunterzeichner in einem offenen Brief mit Hunderten anderen Forschern und Unternehmern.

Nun verleihen die Stockholmer Juroren solchen Warnungen noch mehr Gewicht. Wenn der britische Informatiker künftig vor KI warnt, tut er dies fortan mit der Aura des Nobelpreisträgers. Ebenso, wenn er erklärt, dass die KI nicht etwa die menschliche Intelligenz simuliere, sondern »eine andere Form von Intelligenz« hervorbringe, »eine neue und bessere Form von Intelligenz«, wie er 2023 der *Technology Review* sagte. Das sei etwa so, »als wären Außerirdische gelandet und wir hätten es nicht bemerkt, weil sie sehr gut Englisch sprechen«.

Ob das Nobelkomitee diese Ansichten teilt, ließ es bei der Verkündung des Preises nicht erkennen. Immerhin aber verwies es mehrfach auf die Verantwortung und die Risiken, die mit der prämierten Forschung einhergingen.

Vor zwei Jahren beschlichen ihn düstere Ahnungen

Eigentlich ist der diesjährige Physik-Preisträger von Haus aus Psychologe und Informatiker. Eher dem klassischen Nobel-Schema entspricht der zweite Laureat, John Hopfield. Dem Physiker und Molekularbiologen gelang 1982 die Entwicklung einer frühen Form eines neuronalen Netzes – so genannt, weil es das Zusammenwirken der menschlichen Hirnzellen (Neuronen) in simpler Form imitiert. Dieses »Hopfield-Netz« entwickelte Hinton dann weiter zu künstlichen Netzwerken, die in der Lage sind, in einem vorliegenden Datensatz selbstständig Muster zu erkennen.

Während es um den 1933 geborenen Hopfield in den letzten Jahren ruhig geworden ist, sorgt der 76-jährige Hinton bis heute für Aufregung in der KI-Szene. Der Mann, der gern als »Godfather« (Pate) der künstlichen Intelligenz bezeichnet wird und lange für Google arbeitete, zog im vergangenen Jahr einen aufsehenerregenden Schlussstrich: Er kündigte seinen Konzernjob, um freier über die Gefahren der KI sprechen zu können. Gleichzeitig äußerte Hinton Bedauern über seine Arbeit, deren Folgen er nicht vorhergesehen habe.

Jahrzehntelang habe er geglaubt, dass die künstlichen neuronalen Netze niemals mit dem menschlichen Gehirn mithalten könnten. Schließlich gebe es im Gehirn rund 100 Billionen neuronale Verbindungen, die Modelle der KI kämen nur auf einen Bruchteil davon, höchstens eine Billion. Doch als im Jahr 2022 das Sprachmodell ChatGPT veröffentlicht wurde und den weltweiten KI-Hype befeuerte, beschlichen Hinton Zweifel. Vor Kurzem blickte er in einem *Spiegel*-Interview zurück und sagte, er habe »gemerkt, wie schnell sich das alles entwickelt. Wie katastrophal es womöglich wird.«

Was Hinton vor allem umtreibt, ist die erstaunliche Lernfähigkeit der modernen KI-Programme. Denn die neuronalen Netze lernen nicht nach einem fest vorgegebenen Schema, sondern finden selbstständig den richtigen Weg. Dazu müssen sie zunächst trainiert werden und immer wieder eine Rückmeldung bekommen, ob ihr Ergebnis stimmt. Allmählich strukturiert sich dann das neuronale Netz so um, dass es immer bessere Ergebnisse erzeugt und irgendwann in der Lage ist, auch neue Aufgaben zu lösen – etwa ein Gedicht im Stile Shakespeares über Donald Trump zu schreiben. Wie das System das genau macht und was im Inneren der Netzwerke abläuft, lässt sich von außen im Einzelnen gar nicht nachverfolgen – ähnlich wie niemand sagen kann, was im Kopf eines Kindes geschieht, das gerade lesen oder schreiben lernt.

FOTO: JOHN SIBLEY/REUTERS

Ob eine KI geschrieben hat, erkennt nur noch der Computer

Wenigstens weiß man beim Kind aus Erfahrung, wie lange es braucht von krakeligen Versuchen zum flüssigen Schreiben. Die KI hingegen verblüffte viele Fachleute mit dem Tempo ihres Fortschritts. Konnte man sich etwa Anfang 2023 noch damit trösten, dass ChatGPT weit davon entfernt sei, das deutsche Abitur zu bestehen, nahm der Chatbot diese Hürde wenige Monate später. Heute gestehen selbst Uniprofessoren ein, dass KI-generierte Texte oft besser sind als die Erzeugnisse ihrer Studierenden; wer künstliche von menschengeschriebenen Texten unterscheiden will, schafft das ironischerweise oft nur noch mit speziellen Programmen.

Auch die Herstellerszene verändert sich rasant. Das Unternehmen OpenAI etwa, das ChatGPT hervorbrachte, wurde einst als gemeinnützige Organisation gegründet, um KI auf verantwortungsbewusste Weise »zum Wohle der Menschheit« zu entwickeln. Das war 2015, also gemessen an KI-Maßstäben in grauer Vorzeit. Inzwischen sind viele der Gründer und Vordenker von Bord gegangen. Übrig geblieben ist CEO Sam Altman, der erst vor wenigen Tagen einen enormen Deal verkündete: insgesamt 6,6 Milliarden Dollar hat er eingesammelt von Investoren wie Microsoft, dem Chiphersteller Nvidia, SoftBank und der Investmentfirma MGX aus Abu Dhabi. Ziel: Die einstige Non-Profit-Organisation soll zur Gewinnmaschine werden.

Kein Wunder, wenn Fachleute wie Hinton ein mulmiges Gefühl beschleicht. Der Mann, der noch in den 1980er-Jahren dafür verlacht wurde, welches Potenzial er der künstlichen Intelligenz prophezeite, steht heute als Zauberlehrling da. Mit Staunen sieht er, wie sich seine Schöpfung selbstständig macht. Dabei kann man ihn weder als Apokalyptiker noch als Wirrkopf abstempeln wie vielleicht den Google-Entwickler Blake Lemoine, der vor zwei Jahren ein Computerprogramm des Konzerns als bewusstes Wesen bezeichnete und daraufhin seinen Job verlor. Nein, Geoffrey Hintons Argumente sind viel differenzierter – und gerade deshalb ernster zu nehmen.

Üblicherweise wogt die Debatte um die Gefahren der KI zwischen zwei Polen hin und her: Entweder die Maschinen werden uns bald ebenbürtig bis überlegen. Oder sie werden niemals menschliches Niveau erreichen. Vertreter der zweiten Position argumentieren gerne, KI simuliere nur mithilfe statistischer Wahrscheinlichkeiten Verständnis. Von einem echtem Welterfassen im menschlichen Sinne sei sie weit entfernt.

Doch aus Hintons Sicht verfehlen beide Positionen den eigentlichen Punkt: Es geht nicht darum, ob KI-Programme irgendwann *genauso* denken können wie der Mensch; sondern darum, dass sie ganz andere Fähigkeiten entwickeln. Diese sind dem menschlichen Denken gleichermaßen unter- wie überlegen. So scheitern sie einerseits an simplen Knobelaufgaben, die kleine Kinder spielend bewältigen; andererseits können sie tausendmal mehr Wissen verarbeiten als selbst der klügste Homo sapiens. Dadurch können sie etwa Muster in Klima-, Medizin- oder Wirtschaftsdaten entdecken, die kein menschlicher Experte sähe.

Zugleich können sie erstaunliche Analogien herstellen. So ließ sich Hinton etwa von ChatGPT erklären, warum eine Atombombe funktioniert wie ein Komposthaufen: Beide basieren auf Kettenreaktionen. Wenn ein Komposthaufen heißer wird, erzeugt er immer schneller Hitze. Wenn eine Atombombe mehr Neutronen produziert, erzeugt sie immer schneller Neutronen. »Die zugrunde liegende Logik ist dieselbe, eine Kettenreaktion. ChatGPT hatte das verstanden«, sagte Hinton dem *Spiegel*. Im Internet sei diese Analogie nirgendwo zu finden gewesen, das Programm müsse sie selbst gezogen haben.

Es sollte uns also nicht etwa die Sorge vor einer allzu menschlichen KI umtreiben, sondern eher jene vor einer maschinellen Lernfähigkeit, die uns irgendwann ebenso fremd wie überlegen sein könnte – eben wie das Denken von Außerirdischen.

Wurde er früher gefragt, wieso er an einer Technologie arbeite, die potenziell so gefährlich sei, habe er gern Robert Oppenheimer zitiert, verriet Hinton einmal der *New York Times*. »Wenn man etwas sieht, das technologisch attraktiv ist, dann versucht man es zu erreichen«, habe der Atomforscher gesagt. Hinton sagt das heute nicht mehr.

QUELLE: ZEIT 4/20243, Seite 35

Maschinelles Lernen

Die Technik hinter den jüngsten Fortschritten bei **künstlicher Intelligenz** (KI) erlaubt es Maschinen, **Muster** in Daten wie Texten und Bildern zu finden. Nach einigem Training können sie selbst Inhalte **generieren.** Zentral ist die Belohnung korrekter Lernschritte (**»backpropagation«**) in **neuronalen Netzen.** – Für ihre grundlegenden Arbeiten zu diesen Themen erhalten **Geoffrey Hinton** und **John Hopfield** den Nobelpreis