

Künstliche Intelligenz (KI) ist ein epochaler Technologiesprung, der die Menschheit vor ganz neue Fragen stellt, die keine Disziplin alleine beantworten kann. John Brockman, Agent für Wissenschaftsliteratur und Gründer des Debattenforums edge.org hat das „Possible Minds“-Projekt ins Leben gerufen, das Natur- und Geisteswissenschaften zusammenführt, um KI und deren wahrscheinliche Ausformungen und Folgen zu ergründen. Eine erste Textsammlung ist in den USA als Buch mit dem Titel „Possible Minds – 25 Ways of Looking at AI“ bei Penguin Press erschienen. Das Feuilleton der SZ druckt Texte aus dem Projekt als Serie.

Steven Pinker ist Kognitionsforscher und Psychologe an der Harvard University. In Deutschland erschien von ihm zuletzt das Buch „Aufklärung jetzt“ bei S. Fischer.

Tech-Propheten sind ein populäres Genre, für das Norbert Wieners Buch „Kybernetik“ ein frühes Beispiel war. Wobei es bei Wiener wie so oft weniger um Voraussagen ging, als vielmehr um finstere Warnungen im Geiste des Alten Testaments, dass einen furchtbaren Strafen für die Sünden der Zeitgenossen treffen könnten. Solche Tech-Propheten greifen auf die Rebellion der Romantik gegen die „satanischen Mühlen“ der industriellen Revolution zurück, wenn nicht sogar noch weiter auf die Archetypen des Prometheus, der Pandora und des Faust.

DER GEIST IN DER MASCHINE

Was bedeutet künstliche Intelligenz? Eine Serie von Essays sucht Antworten.

Teil 2

Gegenwärtig beschäftigen sich viele solcher Propheten mit künstlicher Intelligenz. Auf der einen Seite gibt es da die allhergebrachten Science-Fiction-Dystopien von Computern, die Amok laufen und uns in ihrem unaufhaltbaren Welteroberungsdrang versklaven. In den neueren Varianten unterjochen uns KIs ganz aus Versehen, weil sie gnadenlos zielstrebig einen Auftrag erfüllen – egal welche Konzeptionen das für die Menschheit hat, auch wenn die ihnen diesen Auftrag ja ursprünglich gegeben hat. Beide Arten von Voraussagen halte ich für Hirngespinnste, die auf einem technologischen Determinismus basieren, der vollkommen außer Acht lässt, was für Informations- und Kontrollnetzwerke es in einem intelligenten System wie einem Computer, einem Hirn oder auch in einer Gesellschaft als Ganzes gibt.

Die Angst vor der Unterjochung durch die Maschinen basiert auf einem wirren Verständnis von Intelligenz, die eher der Naturphilosophie der Antike oder einem Nietzscheanischen „Wille zu Macht“ zuzurechnen ist, als einer Analyse von Intelligenz und Absicht, was Information, Rechenleistung und Kontrollmechanismen betrifft. In diesen Horrorszenerien wird Intelligenz als allmächtiger, wunsch erfüllender Zauberkraft beschrieben, über den der jeweilige Handelnde in unterschiedlichen Mengen verfügen kann. Menschen haben davon mehr als Tiere, und ein künstlich intelligenter Computer oder Roboter wird mehr davon haben als Menschen. Und weil wir Menschen unsere bescheidenen Möglichkeiten dazu benutzt haben, Tiere mit noch bescheideneren Fähigkeiten zu zähmen oder auszuschließen, weil technologisch fortgeschrittene Gesellschaften technologisch primitivere Gesellschaften versklavt oder vernichtet haben, soll daraus folgen, dass eine superkluge KI das Gleiche mit uns tun wird. Und weil eine KI millionenfach schneller denken wird als wir und ihre Superintelligenz so einsetzen wird, dass sie ihre Superintelligenz fortlaufend verbessert, werden wir machtlos sein, sobald sie nur eingeschaltet wird.

Diese Szenarien verwechseln Intelligenz mit Antrieb, Glaube mit Sehnsucht, Schlussfolgerungen mit Zielen. Selbst wenn wir übermenschlich intelligente Roboter erfinden würde, warum sollten sie ihre Herren versklaven wollen, um die Welt zu beherrschen? Intelligenz ist die Fähigkeit, neue Mittel einzusetzen, um ein Ziel zu kommen. Aber die Ziele selbst sind für die Intelligenz belanglos: Klug zu sein ist nicht dasselbe, wie etwas zu wollen. Es ist nun mal so, dass Intelligenz im Homo Sapiens das Ergebnis einer natürlichen Auslese im Darwinischen Sinne ist, also eines immanenten Konkurrenzprinzips. Im Hirn dieser Spezies ist das Denken gepaart mit Zielen wie Rivalen zu dominieren oder Ressourcen zusammenzurufen.

Es ist aber ein Fehler, die Funktion aus dem limbischen System einer bestimmten



COLAGE: STEFAN DÖRMIR

Die einzige Gefahr ist der Mensch

Hört doch endlich auf mit eurem dystopischen Irrsinn oder:

Warum die meisten Zukunftsszenarien zur KI völlig an der Realität vorbeigehen. Von Steven Pinker

Primatenspezies mit der Grundbeschaffenheit von Intelligenz zu verwechseln. Es gibt kein Gesetz komplexer Systeme, das besagt, dass sich intelligent Handelnde automatisch in skrupellose Größenwahnsinnige verwandeln müssen.

Ein weiterer Irrglaube ist es, Intelligenz als grenzenloses Kontinuum von Wirkungsmacht zu betrachten, ein wunderbares Elixier mit dessen Hilfe man jedes Problem lösen, jedes Ziel erreichen kann. Dieser Trugschluss führt zu so unsinnigen Fragen, wann KI denn „das menschliche Level von Intelligenz übertreffen“ werde, und ob sich dann zwangsweise eine „artificial general intelligence“ (AGI) mit gottglei-

cher Allwissenheit und Allmacht entwickeln wird.

Intelligenz ist streng genommen ein Verbund aus mehreren Software-Modulen, die mit Fähigkeiten ausgestattet sind, unterschiedliche Ziele zu erreichen, oder sich diese aneignen. Menschen können zum Beispiel Nahrung finden, Freunde gewinnen, andere beeinflussen, potenzielle Partnerschaftspartner bezaubern, Kinder aufziehen, sich frei in der Welt bewegen und allen möglichen Leidenschaften und Ablenkungen nachgehen.

Computer können so programmiert werden, dass sie einige dieser Aufgaben übernehmen, wie etwa, Gesichter zu erkennen. Sie können sich auch Probleme vornehmen, die Menschen nicht lösen können, wie das Klima zu simulieren oder Millionen von Buchhaltungsunterlagen zu sortieren. Die Probleme sind dabei so unterschiedlich, wie die Fähigkeiten, sie zu lösen. Doch anhand zu erkennen wie zentral Wissen und Erkenntnis für Intelligenz sind, verwechseln die Dystopien eine künstliche Allgemeinintelligenz der Zukunft mit dem Dämon des Weltgeistes, wie ihn der Mathematiker Pierre-Simon Laplace im 18. Jahrhundert beschrieben hat, ein mythisches Wesen, das den Ort und die Bewegung jedes einzelnen Partikels im Universum kennt und sie mit den Gleichungen der physikalischen Gesetze füttert, um den Zustand aller Dinge und jedes Wesens zu jedem Zeitpunkt in der Zukunft zu errechnen.

Es gibt viele Gründe, warum Laplaces Dämon niemals seinen Weg auf einen Siliziumchip finden wird. Ein lebendiges intelligentes System muss seine Informationen in der chaotischen Welt der Gegenstände und Menschen Stück für Stück zusammen sammeln, indem es sich jedes Feld einnimmt und zwar in Arbeitsgängen, die

von Ereignissen vorgegeben werden, die sich in der wahren Welt ereignen. Das ist einer der Gründe dafür, dass Verständnis nicht nach dem Mooreschen Gesetz der zyklischen Leistungssteigerung von Computern funktioniert. Wissen eignet man sich an, indem man Erklärungen formuliert, die man an der Wirklichkeit erprobt, und nicht indem man einen Algorithmus immer schneller und schneller laufen lässt. Informationen aus dem Internet abzusaugen wird auch nicht zu Allwissenheit führen. Big Data sind immer noch klar begrenzte Daten. Das Universum des Wissens aber ist unendlich.

Bisher hat noch keine KI versucht, ihr Labor zu übernehmen oder Programmierer zu versklaven

Ein anderer Grund, die plötzliche Machtübernahme durch künstliche Intelligenzen anzuzweifeln ist, dass diese Idee den gegenwärtig inflationären Rummel um KI viel zu ernst nimmt. Trotz der Fortschritte des Maschinenlernens, vor allem der vielschichtigen neuronalen Netzwerke, sind KI-Systeme momentan noch weit davon entfernt eine generelle Intelligenz zu entwickeln (wenn dieses Konzept überhaupt schlüssig ist). Stattdessen sind sie immer noch darauf beschränkt, klar definierte Eingaben zu klar definierten Ergebnissen zu verarbeiten, und zwar auf Gebieten, für die gigantische Datensätze zur Verfügung stehen, in denen die Kriterien für Erfolg unmittelbar und präzise sind, deren Umfeld sich nicht verändert, und für die man keine schrittweisen, hierarchischen oder abstrakten Beweisführungen braucht.

Viele Erfolge der künstlichen Intelligenz kommen nicht durch ein besseres Verständnis davon, wie Intelligenz funktio-

niert, sondern von der Holzhammer-Kraft immer schnellerer Prozessoren und immer größerer Datensätze, die den Programmen ermöglichen, sich an Millionen Beispielen zu üben und dann ähnliche Ergebnisse daraus abzuleiten. Jedes KI-System verfügt nur über eine Inselbegabung und kann keine Probleme lösen, auf die es nicht vorbereitet wurde.

Es ist nur eine Binsen, aber bisher hat noch keine dieser KIs versucht, ihr Labor zu übernehmen oder ihre Programmierer zu versklaven. Und selbst wenn eine KI versuchen würde, Machtwillen zu entwickeln, wäre sie ohne die Kooperation von Menschen nur ein impotentes Hirn im Fass. Ein superintelligentes System müsste für die Machtübernahme schnellere Prozessoren bauen, eine Infrastruktur, die sie am Leben erhält und die Roboter-Effektoren, die sie mit der wirklichen Welt verbindet – alles unmöglich, es sei denn die Menschen würden ihr aktiv große Teile der technisierten Welt überlassen. Man kann sich natürlich einen apokalyptischen Computer vorstellen, der bösartig, allmächtig, immer in Betrieb und gegen jegliche Eingriffe gesichert ist. Die Lösung für diese Bedrohung ist ganz einfach: Baut so etwas nicht!

Aber was ist mit den neuesten KI-Bedrohungen, wie die Werteausrichtung, wie sie in der Sage von König Midas beschrieben wird, bei der jemand seinen magischen Wunsch bereut, weil er unvorhergesehene Nebenwirkungen hat? Wenn wir einer AI das Ziel einprogrammieren würden, Wasser hinter einem Damm zu sammeln, würde es vielleicht ohne Rücksicht auf die Menschen eine Ortschaft fluten. Wenn wir ihr den Auftrag geben würden, so viele Büroklammern wie möglich zu produzieren, würde sie vielleicht sämtliche Materie des erreichbaren Universums zu Büroklammern verarbeiten, inklusive unserer Kör-

per. Eine KI, die menschliches Glück maximieren soll, würde uns vielleicht an einen Dopamin-Tropf hängen oder unsere Hirne ganz neu verkabeln.

Zum Glück widerlegen sich solche Gedankenexperimente selbst. Sie gehen davon aus, dass Menschen zum einen so begnadet sind, dass sie eine allwissende und allmächtige KI bauen können, aber gleichzeitig so idiotisch, dass sie ihr die Kontrolle des Universums überantworten würden, ohne vorher zu testen, wie das funktioniert. Und sie nehmen an, dass eine KI herausfinden könnte, wie man Elemente unwandelt, Hirne verkabelt, aber so idiotisch, dass sie großen Schaden anrichten würde, nur weil sie etwas missversteht. Dabei ist die Fähigkeit, Kompromisse zu finden, um die sich widersprechenden Ziele auf einen Nenner zu bringen, kein beliebiger Nachrüstbaustein für Intelligenz, den Ingenieure vielleicht vergessen. Im Gegenteil: Genau das ist der Kern von Intelligenz.

Warum sollten Menschen die Kontrolle der Welt einfach Maschinen überlassen?

Wenn wir solche Fantasien von digitalem Größenwahn, sofortiger Allwissenheit und der perfekten Kenntnis und Kontrolle jedes Partikels des Universums außer Acht lassen, ist künstliche Intelligenz wie jede andere Technologie auch. Sie wird Stück für Stück entwickelt, sie wird darauf ausgerichtet, unterschiedliche Problemlöse zu lösen, sie wird vor ihrem Einsatz getestet und sie wird immer wieder justiert und verbessert, um ihre Funktionstüchtigkeit und Sicherheit zu garantieren.

Vor allem das letzte Kriterium ist wesentlich. Eine Sicherheitskultur ist in einer fortgeschrittenen Gesellschaft ein Bollwerk gegen den autoritären oder ausbeuterischen Einsatz von Technologie. Während westliche Gesellschaften zu Beginn des 20. Jahrhunderts noch Verstärkungen und Todesfälle in Industrie-, Haushalts- und Verkehrsunfällen in schockierend hohen Zahlen tolerierte, ist der Wert eines Menschens Lebens im Lauf des Jahrhunderts enorm gestiegen. Deswegen nutzten Regierungen und Ingenieure die Ergebnisse aus Unfallstatistiken, um zahllose Regulierungen, Sicherheitsvorrichtungen und Designveränderungen zu implementieren, die die verschiedenen Technologien zunehmend sicherer gemacht haben.

Die Tatsache, dass manche Vorschriften, wie das Verbot, Handys neben einer Zapfsäule zu benutzen, Risiken mit fast schon lächerlicher Akribie eindämmen, zeigt, dass wir als Gesellschaft wie besessen sind, wenn es um Sicherheit geht. Mit beeindruckenden Ergebnissen: Die Zahlen von Industrie-, Haushalts- und Verkehrsunfällen sind seit der ersten Hälfte des 20. Jahrhunderts um 95 (in manchen Bereichen um 99) Prozent gesunken. Doch die Propheten einer bösartigen, allmächtigen KI tun so, als würde es diese Zahlen nicht geben, als würden Ingenieure eines Morgens die Kontrolle der physischen Welt einfach unerprobten Maschinen überlassen, egal, was das für die Menschheit bedeutet.

Norbert Wiener schrieb, dass die Gefahr für die Gesellschaft nie die Maschine selbst ist, sondern dass sie immer aus dem erwacht, was der Mensch mit ihr tut. Nur wenn wir uns daran erinnern, können wir die Gefahren und Möglichkeiten der KI adäquat einschätzen.