http://www.taz.de/Kuenstliche-Intelligenz/!5321037/

die tageszeitung
Öko / Wissenschaft 24. 7. 2016

MEIKE LAAFF
tazzwei-Redakteurin

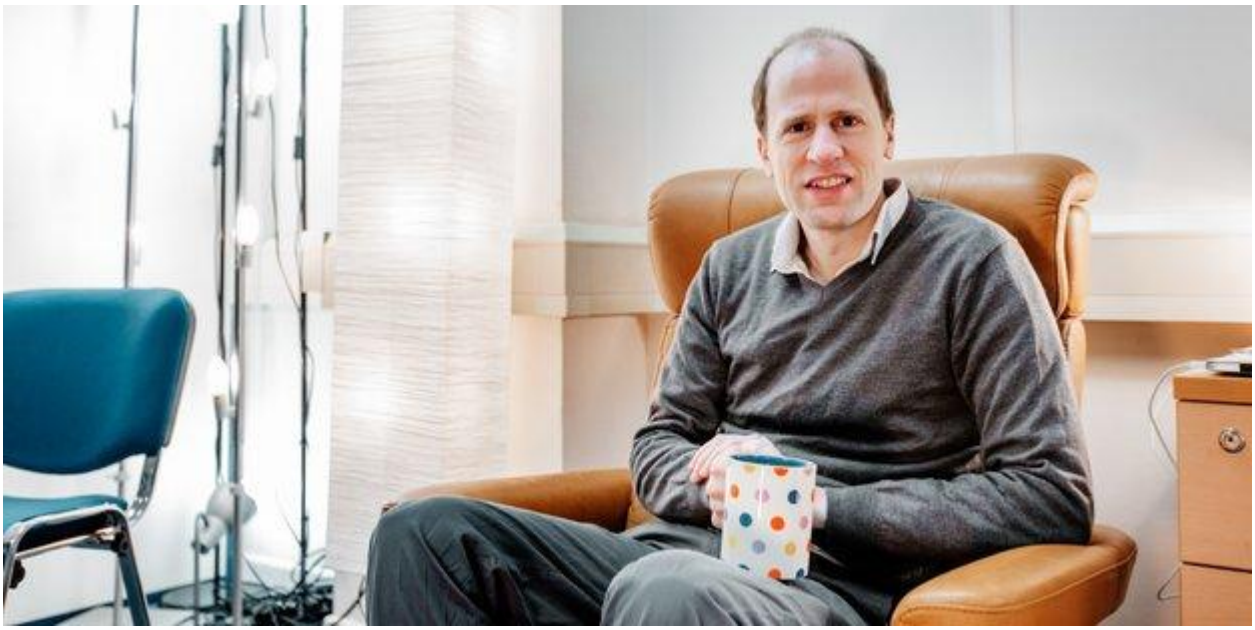
THEMEN
künstliche Intelligenz, Nick Bostrom, Stephen Hawking, Oxford, Intelligente Maschinen

Künstliche Intelligenz

Apocalypse soon

Superschlaue Computer lernen immer selbstständiger. Das sei eine tickende Zeitbombe, sagt der Philosoph Nick Bostrom.

OXFORD *taz* | Es summt. Nur hörbar, sobald Nick Bostrom eine Redepause einlegt. Passender wäre, es würde ticken. Denn wenn Bostrom recht hat, dann sitzen



Noch ist er klüger als die Lampe hinter ihm: Nick Bostrom Foto: Future of Humanity Institute/University of Oxford

wir auf einer Zeitbombe: der Künstlichen Intelligenz (KI). Was der Philosoph für unsere Zukunft mit superintelligenten Maschinen vorhersagt, ist nicht viel weniger als die Apokalypse.

Bostrom vergleicht die Menschheit mit kleinen Kindern, die mit einer Bombe spielen. „Wir haben kaum eine Idee, wann die Detonation stattfinden wird", schreibt er. „Aber wenn wir sie an unser Ohr halten, hören wir ein leises Ticken".

Das Summen hingegen rührt von dem Dutzend Tageslichtlampen, mit denen Bostrom den kleinen Besprechungstisch in seinem Büro umstellt hat. In ihrer schieren Menge sind sie das einzig Exzentrische in dem sonst aufgeräumten Zimmer. Wohlgeordnet wie die Gedanken von Nick Bostrom.

Superschlaue Computer könnten sich immer mehr Rechenkapazität, Speicherplatz und Wissen aneignen, um ihre Ziele zu erreichen. Sich Zugriff auf Rechenzentren verschaffen, die Kontrolle über die Infrastruktur übernehmen und alle Versuche, sie abzuschalten, durchkreuzen. Die Menschheit könnte die Superintelligenz auf dem Weg dahin auslöschen. Das ist zumindest ein Szenario – das Bostrom im Gespräch ganz ruhig in Nebensätzen hinwirft.

„Superintelligenzen – Szenarien einer kommenden Revolution" heißt sein Buch, das zum Bestseller in den USA wurde. Erstaunlich, weil es sich nicht nur ziemlich dröge liest, sondern auch noch einen Konjunktiv an den nächsten reiht.

In Einklang bringen

Bostroms Kernthese: Wahre Künstliche Intelligenz, die weit über jene Maschinen hinausgeht, die Menschen in Brettspielen schlagen, Autofahren oder unsere Konsumbedürfnisse anzustacheln suchen, kann schnell unserer Kontrolle entschlüpfen.

Bostrom will das verhindern. Er will heute schon anfangen darüber nachzudenken, wie man diese Künstlichen Intelligenzen im Zaum halten kann. Oder besser noch: ihre Ziele mit unseren in Einklang bringen kann. Damit, argumentiert Bostrom, müsse man schon beginnen, während man

sie konstruiert. „Wir wollen uns doch nicht später treten dafür, dass wir nicht früher damit angefangen haben", sagt er in der wohl beleuchteten Sitzecke seines Büros.

Nick Bostrom über KI

„Das ist nicht wie ein kaputter Fernseher, den man einfach ausschaltet"

„Future of Humanity Institute" hat Bostrom die Einrichtung genannt, die er vor zehn Jahren gründete, um daran zu arbeiten. Angesiedelt ist sie im britischen Oxford, wo sich eine altehrwürdige Universität an die andere kuschelt. Jahrhunderte des Wissens und Denkens, gebettet in Sandsteinbauten, die aussehen wie Kulissen aus Harry-Potter-Filmen. Mittendrin, auf einer Etage direkt über einem Fitnessstudio in einem der wenigen gesichtslosen Neubauten des Stadtkerns, das Institut. Bostrom, gebürtiger Schwede, ist bis heute der Direktor.

Er studierte Physik, Mathematik und diverse weitere Disziplinen. Ein Überflieger. Schmal, mit der Halbglatze, den Klamotten, die schon länger kein Bügeleisen gesehen haben, und dem Stoffgürtel wirkt er nicht wie jemand, der unnötig Zeit mit Äußerlichkeiten verschwendet. Sondern wie jemand, dessen Askese aus der Konzentration aufs Innere entspringt. Auf das Denken.

Ein nüchterner Mann

Abgesehen von dem noch immer leicht skandinavischen Zungenschlag, den er auch nach Jahrzehnten in Großbritannien noch nicht abgelegt hat, gibt es kaum einen Eindruck von dem Menschen Nick Bostrom. So exzentrisch seine Thesen auf viele wirken, so nüchtern der Mann dahinter. Nicht verbindlich, aber auch nicht unfreundlich. Nicht laut. Sparsam in Gestik und Mimik.

Künstliche Intelligenz, damit wird heute so einiges bezeichnet. Inzwischen vergeht kaum ein Tag, ohne dass vermeldet wird, was einer solchen Künstlichen Intelligenz nun gelungen ist: Bilder malen. Den Unterschied zwischen Hunderassen auf Fotos erkennen. Ein Drehbuch schreiben. Anwaltsgehilfe werden. Einzelanwendungen, die darauf basieren, dass Maschinen auf immer größere Datensammlungen zurückgreifen können, mit immer besseren Prozessoren Verbindungen herstellen und daraus lernen. Tatsächlich verstehen oder gar denken, das ist bislang aber noch keiner Maschine gelungen.

„Wir brauchen noch einige Durchbrüche, um Künstliche Intelligenz zu erreichen, die sich auf einem Level mit der menschlichen bewegt", räumt auch Bostrom ein. Und dass man keine Ahnung hätte, wie schwer diese Durchbrüche sein werden. Worte eines Mannes, der sich bewusst ist, dass er vorsichtig sein muss. Ist doch sein Forschungsgebiet so weit in die Zukunft gerichtet, dass seriöse wissenschaftliche Überlegungen und Science-Fiction-Humbug schnell zu verschwimmen scheinen.

Von Stephen Hawking empfohlen

Weswegen Bostrom nicht selbst schätzt, wann es so weit sein wird mit der Künstlichen Intelligenz auf Niveau des Menschen, sondern Praktiker aus der KI-Forschung befragte. Im Durchschnitt sagten sie ihm: Die Chancen, dass Computersysteme im Jahr 2040 so intelligent sind wie Menschen, stehen fifty-fifty. 2070 halten viele für noch wahrscheinlicher. Andere Forscher widersprechen – Bostroms Thesen, aber auch generell der Idee, dass dieser Sprung Maschinen je gelingen wird. Was den Philosophen Bostrom zu einer durchaus umstrittenen Figur in der KI-Forschung macht. Aber einer durchaus einflussreichen.

Nobelpreisträger Stephen Hawking empfahl sein Buch ebenso wie US-Unternehmensvisionär Elon Musk. Führende IT-Konzerne suchen das Gespräch mit ihm und seinen Leuten. Bostrom spricht auf Konferenzen, eröffnete in diesem Jahr die deutsche IT-Messe Cebit.

„Die erste ultraintelligente Maschine ist die letzte Erfindung, die die Menschheit machen muss", schreibt er. Danach kann die Künstliche Intelligenz sich selbstständig machen. Sich selbst immer weiter verbessern. Neue Maschinen entwerfen. Sich selbst optimieren. Krebs heilen. Vernichten, was ihr im Weg steht. Fast gottgleiche Kräfte schreibt er Superintelligenzen zu.

Bostrom könnte es sich auch einfacher machen. Sich mit Fragen beschäftigen, vor die Maschinen und ihre eng gesteckten Formen von Künstlicher Intelligenz uns schon heute stellen: Wie umgehen mit selbst fahrenden Autos? Wie stabilisieren wir Gesellschaften, wenn Maschinen uns die Arbeit wegnehmen? Welche ethischen Beschränkungen brauchen autonome Kampfdrohnen? Fragestellungen, auf die die Politik bald reagieren muss, ja, sagt Bostrom. Größeres Interesse hat er an diesen Diskussionen aber nicht, sagt er. „Mein Fokus liegt auf Längerfristigem."

Explosion der Intelligenz

Und so muss man mit Bostrom das „Was wäre, wenn"-Spiel spielen, neben dem Lampenwald seines Besprechungstischchens. Also: Gäbe es superschlaue Computer, macht es dann überhaupt noch einen Unterschied, ob eine Firma wie Google sie unter ihrer Kontrolle hat oder ein Staat? Kommt drauf an, sagt Bostrom. Zum einen darauf, ob wir die KI unter unsere Kontrolle bekommen. Und zum anderen, welche Ziele und Werte wir ihr einimpfen.

Erreichen Maschinen aber erst einmal Intelligenz auf menschlichem Niveau, so Bostroms Argument, dann könnte es ganz schnell gehen, dass sie uns überflügeln. „Intelligenzexplosion" nennt Bostrom das. Und meint damit den Zeitpunkt, an dem die Maschinen uns entgleiten könnten – indem sie ihre Vorteile gegenüber unseren biologischen Gehirnen ausspielen.

Warum also nicht einfach einen Notausschalter für Künstliche Intelligenzen programmieren? „Auf so etwas sollten wir uns nicht verlassen", sagt Bostrom. In seinem Gesicht zuckt kurz etwas auf. Eine sonst gut im Zaum gehaltene Ungeduld, Banalitäten wie diese Frage nun schon wieder erklären zu müssen.

Superintelligente Agenten seien in der Lage, menschliche Handlungen und Strategien zu antizipieren, sagt er. Und sich zu widersetzen. Sie könnten Menschen überreden, sie nicht abzuschalten. Sie könnten sich Kontrolle über die Energieversorgung verschaffen oder sich einfach auf eine andere Hardware kopieren. „Das ist nicht wie ein kaputter Fernseher, den man einfach ausschaltet – und dann steht er da und wartet, was wir als Nächstes tun", sagt er. „Die Idee von einem bösen Geist, der für immer in einer Flasche eingesperrt ist, erscheint nicht sehr vielversprechend. Früher oder später wird er einen Weg hinausfinden."

Besser sei es, von Anfang an sicherzustellen, dass die Künstliche Intelligenz auf unserer Seite sei. „Eine Verlängerung unseres Willens und unserer Werte", sagt Bostrom. Nur: Wie soll das gehen?

Fragen dieser Art sind es, weswegen einige Künstliche-Intelligenz-Forscher Bostrom verabscheuen. Vor allem Wissenschaftler, die keine Anzeichen dafür sehen, dass das, was Bostrom beschreibt, jemals eintreten könnte. Die es nicht für möglich halten, dass Computer den Sprung vom reinen Kombinieren zum Denken schaffen – und Bostroms Thesen somit als wichtigtuerisches Zukunftsgeraune abtun.

Bostrom entgegnet: lieber zeitig mit diesen Überlegungen anfangen, als später ohne Lösung dazustehen. Er ist davon überzeugt: Sich jetzt einen festen Satz ethischer Grundsätze auszudenken, die man in die Künstliche Intelligenz einschreibt, das sei keine gute Idee. Besser wäre es, Künstliche Intelligenzen durch Beobachtung lernen zu lassen, was wir wollen und meinen, welche Ziele wir verfolgen.

Die optimale Denkleistung

Es wäre leicht, von Bostrom das Bild eines Sonderlings zu zeichnen. Porträts über ihn strotzen vor Details, die das zu untermauern suchen. Ein Mann, der sich am liebsten flüssig von Smoothies ernährt, dessen Laster Nikotin-Kaugummi-Kauen ist – alles im Dienste der optimalen Denkleistung. Ein Workaholic, der zur Partnerin und dem Sohn in Kleinkindalter eine transatlantische Fernbeziehung pflegt. Der einem Verein angehört, der die Leichen seiner Mitglieder nur Stunden nach dem Tod einfriert und einlagert – für den Fall, dass man sie später wiederbeleben kann.

Näher bringen solche Details einem den Menschen Bostrom aber nur, wenn man versteht, was dahintersteht. Bostrom hat sich intensiv mit Transhumanismus beschäftigt – einer philosophischen Bewegung, die die Natur nicht als Krone der Schöpfung begreift, sondern eine Verschmelzung von Menschen und Technologie anstrebt. Im positiven Sinne. Ein Widerspruch zu Bostroms apokalyptischen KI-Prognosen? Für ihn nicht. Er spricht lieber von zwei Möglichkeiten. Abwägung. Wahrscheinlichkeiten. Prozente. Bostrom, ein Kopfmensch.

Der im Gespräch nun ganz neu ansetzt. Darüber spricht, was passieren würde, wenn nicht eine, sondern gleich mehrere Künstliche Intelligenzen gleichzeitig den Menschen überflügeln würden. Bostrom redet sich heiß über die evolutionären Dynamiken, die in der Wechselwirkung dieser Maschinen dann entstehen würden.

Würden. Könnten. Müssten. Die Lampen im Hintergrund, sie hören nicht auf zu surren.