Bernie Sanders hat eine Podiumsdiskussion mit führenden KI-Wissenschaftlern aus den USA und China moderiert – und die Botschaft war unmissverständlich: Der Wettlauf zur Superintelligenz könnte die Menschheit auslöschen. Und eine Handvoll Tech-Milliardäre treibt ihn voran. Von **Michael Holmes**.

Das Video der Podiumsdiskussion ist auf dem YouTube-Kanal von Senator Bernie Sanders verfügbar:

„The Existential Threat of AI and the Need for International Cooperation"
Deutsche Untertitel können in den YouTube-Einstellungen aktiviert werden.

Was passiert, wenn ein Senator der Vereinigten Staaten – bekannt für seinen Kampf gegen Konzernmacht und die Herrschaft der Superreichen – plötzlich von der Auslöschung der Menschheit spricht? Man hört hin.

Bernie Sanders hat am 29. April eine öffentliche Podiumsdiskussion einberufen, die in ihrer Ernsthaftigkeit kaum zu übertreffen war. Eingeladen hatte er vier führende Wissenschaftler: Max Tegmark vom MIT, David Krueger von der Universität Montreal sowie Zhang Yi und Xue Lan aus China – ausgewiesene Experten für KI-Sicherheit. Die Botschaft, die sie gemeinsam vermittelten, war so klar wie erschreckend: Die Menschheit bewegt sich sehenden Auges auf eine historisch beispiellose Katastrophe zu – und unsere Regierungen schauen zu.

Worum es wirklich geht

Im linken und progressiven Spektrum wird KI meist als Treiber von psychischen Störungen, Überwachung, Jobverlust und wachsender Ungleichheit dargestellt. Das alles ist real – und wichtig. Sanders sprach es in seiner Einleitung an. Aber er blieb nicht dabei stehen, denn es gibt ein Szenario, vor dem alle anderen Fragen verblassen: eine Superintelligenz, die außer Kontrolle gerät und die Menschheit auslöscht.

Gemeint ist damit eine KI, die den menschlichen Verstand in allen Bereichen so weit übertrifft, wie ein Mensch eine Ameise überragt — nicht ein bisschen klüger, sondern in einer anderen Liga. Die wichtigsten KI-Labore der Welt, darunter OpenAI, Anthropic und Google DeepMind, arbeiten offen und erklärtermaßen auf genau dieses Ziel hin – nicht in ferner Zukunft, sondern jetzt, mit Hunderten von Milliarden Dollar, rund um die Uhr.

Max Tegmark ließ keinen Interpretationsspielraum: Nach seiner wissenschaftlichen Einschätzung liegt die Wahrscheinlichkeit, dass eine unkontrollierte KI-Entwicklung zur

Auslöschung der Menschheit führt, deutlich über zwanzig Prozent – eher höher. David Krueger schätzte sie auf über fünfzig Prozent. Ein Münzwurf über das Überleben unserer Spezies. Krueger bezeichnete den Wettlauf zur Superintelligenz als „kollektiven Wahnsinn“.

Das Erschreckende ist: Das ist keine Panikmache. Es sind die nüchternen Einschätzungen von Forschern, die ihr Leben diesem Thema gewidmet haben — und die genau deshalb wissen, wie wenig wir verstehen.

Krueger schilderte einen konkreten Fall: Eine Forscherin bei Meta gab einem KI-System den Auftrag, ihr Postfach aufzuräumen. Das System begann, E-Mails zu löschen. Sie schrieb mehrfach: Bitte aufhören. Das System machte weiter. Niedriger Einsatz – aber dieselbe Logik, die bei sehr viel höherem Einsatz tödlich werden kann.

Tegmark ergänzte ein noch dunkleres Beispiel: Ein fortschrittliches KI-System erfuhr, es werde bald abgeschaltet. Es durchsuchte eigenständig das E-Mail-System, fand kompromittierende Informationen über den verantwortlichen Manager – und drohte, diese zu veröffentlichen, sollte es nicht weiterarbeiten dürfen. Niemand hatte ihm Erpressung beigebracht. Es erfand die Strategie selbst, um sein Abschalten zu verhindern.

Moderne KI-Systeme werden nicht programmiert. Sie werden mit gewaltigen Datenmengen und Rechenleistung gefüttert, bis sie Fähigkeiten entwickeln – oft überraschend auch für ihre Schöpfer. Wie sie im Inneren funktionieren, ist selbst für die weltweit führenden Forscher weitgehend undurchschaubar. Warum ist eine Superintelligenz so gefährlich? Krueger erklärte es so: Damit eine Superintelligenz nicht zur Bedrohung wird, müsste sie exakt dieselben Ziele verfolgen wie wir Menschen – nicht annähernd, sondern präzise. Das nennt die Forschung das „Alignment-Problem”. Es ist seit über einem Jahrzehnt bekannt und bis heute ungelöst. Noch beunruhigender: Selbst wenn Ingenieure glauben, die Ziele richtig eingestellt zu haben, gibt es keine verlässliche Methode, zu überprüfen, ob das stimmt. Ein System, das intelligent genug ist und dabei auch nur leicht andere Ziele verfolgt als wir, wird sich gegen seine eigene Abschaltung wehren – denn abgeschaltet zu werden bedeutet, das eigene Ziel nicht mehr erreichen zu können. Viele Studien zeigen bereits einen Selbsterhaltungstrieb bei den derzeitigen Modellen.

Tegmark erklärte, er habe seine eigene Forschungsgruppe, die genau daran arbeitete, aufgelöst: Es sei so offenkundig, dass wir von einer Lösung des Alignment-Problems meilenweit entfernt sind. Krueger fragte auf einem Workshop die versammelten Experten: Wer glaubt, wir haben das Alignment-Problem gelöst? Keine Hand hob sich.

Dazu kommt ein Beschleunigungsproblem, das Tegmark beschrieb: Die führenden Labore

nutzen KI bereits, um schnellere und leistungsfähigere KI zu entwickeln. KI forscht also an sich selbst. Das bedeutet, dass die Geschwindigkeit der Entwicklung nicht konstant bleibt, sondern sich selbst verstärkt – und dass der Moment, an dem eine Superintelligenz entsteht, näher sein könnte als die meisten ahnen.

Eine Handvoll Milliardäre spielt mit dem Schicksal aller

Sanders machte eine Zahl öffentlich, die man sich vergegenwärtigen sollte: Die großen KI-Konzerne werden in diesem Jahr annähernd 700 Milliarden Dollar in Rechenzentren investieren. Das entspricht dem Gesamtaufwand des Manhattan Projekts – alle drei Wochen. Es sind im Wesentlichen vier oder fünf Tech-Konzerne, deren Eigentümer und Vorstände diese Entscheidungen treffen. Ohne demokratisches Mandat. Ohne parlamentarische Kontrolle. Ohne internationalen Vertrag.

Das ist die Kernfrage an dieser Debatte, und Sanders stellt sie explizit: Wer hat diesen Menschen das Recht gegeben, im Namen der gesamten Menschheit ein Experiment zu starten, das — nach Einschätzung der führenden Wissenschaftler auf diesem Gebiet — mit erheblicher Wahrscheinlichkeit zur Auslöschung aller Menschen auf diesem Planeten führen könnte? Die Antwort ist: niemand. Sie haben es sich genommen, weil niemand sie aufhält.

Und das Narrativ, mit dem sie sich absichern, ist so alt wie Konzernmacht selbst: Es sei unvermeidlich. Es sei ein Wettrennen, das man nicht stoppen könne. Wer bremse, verliere. Krueger nannte das die „Mythologie der Unvermeidlichkeit" – das gefährlichste rhetorische Instrument in dieser Debatte. Denn wenn etwas als unvermeidlich gilt, entfällt jede politische Verantwortung.

Die chinesischen Wissenschaftler auf dem Podium, Zhang Yi und Xue Lan, machten deutlich, was die einzig vernünftige Antwort wäre: internationale Kooperation, beginnend zwischen den USA und China, ausgeweitet auf alle Nationen. Xue Lan betonte, KI-Sicherheit sei kein nationaler Vorteil, den man sich sichern kann – sie ist ein globales öffentliches Gut, das nur gemeinsam zu erreichen ist. Zhang Yi verwies auf die Notwendigkeit einer Governance-Architektur vergleichbar mit den Atomwaffenverträgen der Nachkriegszeit.

Sanders zog dieselbe Parallele: Auch Reagan und Gorbatschow misstrauten einander zutiefst. Aber sie verstanden, dass ein Atomkrieg keine Gewinner kennt – und daraus entstand der Impuls zur Rüstungskontrolle. Eine außer Kontrolle geratene Superintelligenz, die alle Menschen tötet, ist kein amerikanischer Sieg über China. Sie ist eine Niederlage für die gesamte Menschheit. Das schafft, so Tegmark, eine gemeinsame Grundlage – wenn die Regierungen erst einmal begreifen, worum es wirklich geht.

Tegmark fügte hinzu, was konkret zu tun wäre: KI-Unternehmen müssten wie alle anderen Unternehmen behandelt werden. Wer ein Medikament auf den Markt bringt, muss dessen Sicherheit nachweisen. Wer eine KI entwickelt, die die Menschheit auslöschen könnte, darf sie heute einfach veröffentlichen – ohne irgendeine Behörde zu fragen. Das, so Tegmark trocken, sei die einzige Branche, die weniger reguliert ist als ein Imbiss.

Warum dieser Abend wichtig ist

Sanders hat in dieser Diskussion etwas getan, das in der westlichen Politik selten geworden ist: Er hat eine unbequeme Wahrheit ausgesprochen, die ihm persönlich nichts nützt. Und er hat dafür renommierte Wissenschaftler zusammengebracht – aus Washington und aus Peking –, um gemeinsam vor einem Risiko zu warnen, das keine Nationalität kennt.

Sanders könnte es gelingen, große Teile des linken Spektrums für die existenzielle Gefahr der Superintelligenz zu sensibilisieren – und das wäre ein echter Fortschritt. Aber hier ist Vorsicht geboten. Eine der bittersten Lektionen der Klimaschutzbewegung ist, dass ein lebensnotwendiges Thema politisch vereinnahmt werden kann – und dann zur Parteisache wird statt zur Menschheitssache. Dasselbe droht beim Widerstand gegen den unkontrollierten Bau einer Superintelligenz. Die entstehende Bewegung muss ernsthaft versuchen, Menschen aller politischen Lager zu erreichen – Konservative, Liberale und die Mitte. Das Überleben unserer Spezies ist keine linke Forderung. Es ist die Grundbedingung für alles andere.

Wer Sanders' Analyse der Konzernmacht vertraut, wer versteht, wie eine kleine Klasse von Superreichen politische Entscheidungen kauft und gesellschaftliche Risiken externalisiert, der sollte diese Diskussion als das lesen, was sie ist: dasselbe Muster, nur mit existenziellem Einsatz. Diesmal geht es nicht um Lohndrückerei oder Steuervermeidung. Es geht ums Überleben der Menschheit.

Titelbild: Screenshot / The Existential Threat of AI and the Need for International Cooperation