KI könnte die Menschen auslöschen: Was ist dran an der Warnung?

Am 22. März 2023 veröffentlichte das gemeinnützige „Future of Life Institute", eine Denkfabrik mit Sitz in Cambridge, Massachusetts, einen offenen Brief mit dem Titel „Pause Giant AI Experiments: An Open Letter", der unter anderem von Elon Musk, George Dyson, Steve Wozniak und Yuval Noah Harari unterschrieben war: „Stoppt die gigantischen Experimente mit Künstlicher Intelligenz!".

Es drohe nicht weniger als der Untergang der Welt, wenn Künstliche Intelligenz ohne Regulierung einfach so weiterentwickelt werde. Deswegen brauche es eine Pause von mindestens sechs Monaten. Ausgelöst wurde der Alarm wohl von ChatGPT, diesem bahnbrechenden Large Language Model, das durch seine Eloquenz, sein Allgemeinwissen und sein Unwissen beeindrucken kann. Jeder, der mit ChatGPT gechattet hat, weiß: Das wird groß, das wird die Welt verändern.

Ist Eloquenz aber gleichbedeutend mit Intelligenz? Man sollte sich vergegenwärtigen, wie ein Large Language Model arbeitet: Ein künstliches neuronales Netz wird trainiert, im Fall von ChatGPT mit sehr viel Text, auch aus dem Internet; welche Texte genau dafür ausgewählt wurden, ist nicht ganz klar.

ChatGPT lernt dann, mit welcher Wahrscheinlichkeit ein Wort auf das andere folgt und gibt auf jede Frage die Wörter entsprechend wieder raus. Wenn ChatGPT also gefragt wird, was Berlin ist, spuckt es aus, dass Berlin die Hauptstadt von Deutschland ist. Nicht weil es eine Vorstellung davon hätte, was Berlin ist, was eine Stadt ist oder wo Deutschland liegt, sondern weil es die statistisch wahrscheinlichste Antwort ist. Würde ich ein neuronales Netz auf Text trainieren, der behauptet, Berlin wäre der Name eines chinesischen Nudelgerichts, würde es mir auf Nachfrage diese Antwort geben.

Das Internet ist jetzt schon voll von Propaganda

Erfahrungsgemäß können auch Menschen, an deren Intelligenz man zweifeln darf, Schaden anrichten. Welche Schäden sind es aber, die Musk und seine Mitunterzeichner im Fall der Künstlichen Intelligenz befürchten? Im offenen Brief werden sie als Fragen formuliert. Die erste, nämlich ob wir die Maschinen unsere Informationskanäle mit Propaganda und Unwahrheit fluten lassen wollen, entbehrt nicht der Komik – immerhin war es Elon Musk, der so gut wie alle Menschen bei Twitter, die sich mit der Moderation beschäftigten, entlassen hat.

Sam Altman: „Wenn mit dieser Technologie etwas schiefgeht, dann kann es ziemlich schiefgehen." Bild: dpa

Darüber hinaus kündigte Musk erst letzte Woche den freiwilligen Verhaltenskodex gegen Desinformation der EU auf, durch den sich Twitter unter anderem verpflichtete, das Verbreiten von Desinformationen über Onlinewerbung zu verhindern. Das Internet ist jetzt schon voll mit Propaganda und Unwahrheit, das Ende, was sich durch KI verändern könnte ist, dass es damit mehr Propaganda und Unwahrheit geben wird. Das Problem löst man nicht technisch, sondern gesellschaftlich.

Sollen wir, fragt der offene Brief weiter, alle Jobs automatisieren, auch die, die Erfüllung und Befriedigung versprechen? Schon diese Unterscheidung zwischen Jobs, die erfüllen und solchen, die das nicht tun, sollte jedem ein Hinweis sein, welche Jobs auf jeden Fall automatisiert werden sollten. Mit welcher Rechtfertigung sollten Menschen stumpfe Arbeiten für wenig Geld ausführen, wenn es auch eine Maschine tun könnte, die darüber deshalb nicht depressiv wird, weil sie einfach nicht depressiv werden kann?

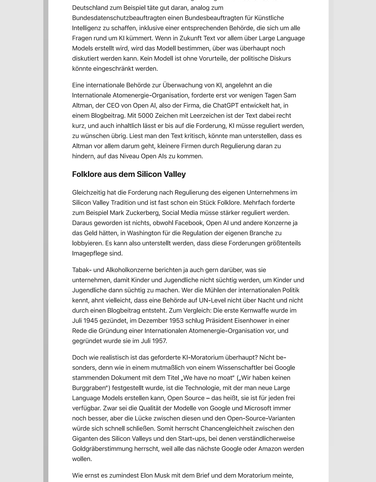
In der Geschichte des Kapitalismus und der Technologie wurde jede, wirklich jede Technologie früher oder später dazu benutzt, Jobs wegzurationalisieren. Das ging so weit, dass zum Beispiel in der metallverarbeitenden Industrie die Facharbeiter ersetzt wurden durch Maschinen, die schlechterte Qualität lieferten, dafür aber keinen Lohn haben wollten und auch nicht streiken. Was uns zeigt, dass nicht die Technologie das Problem ist. Frau Müller hätte mit Sicherheit kein Problem damit, dass ihr Job bei einer Kundenservicehotline zukünftig von einem Large Language Model gemacht wird, wenn sie nur weiter ihren Lohn erhalten würde. Naturgemäß wird das nicht passieren. Ist das aber die Schuld des Large Language Models?

Wird der Mensch obsolet?

Die letzten beiden Fragen des offenen Briefs haben es dann in sich. Zunächst geht es darum, ob wir Intelligenzen entwickeln sollten, die klüger und zahlreicher sind als der Mensch und die uns deshalb obsolet machen und auslöschen werden.

Die Frage wird jeder mit Nein beantworten. Ist es theoretisch möglich, dass eine KI die Menschheit auslöscht? Ja. Gleichzeitig kann man aber auch die Gegenfrage stellen: Warum sollte die KI das tun? Um einen kybernetischen Ödipuskomplex zu überwinden?

Von einem Large Language Model bis zu jenem „Skynet", das im Film „Terminator" die Menschheit knechtet, ist es noch ein weiter Weg. Umso weltfremder ist es aber, darauf hinzuweisen, dass uns eine KI auslöschen könnte – zu einem Zeitpunkt, da die Menschheit grade auf dem besten Weg ist, die Zivilisation, wie wir sie kennen, zusammenbrechen zu lassen, vielleicht sogar das Ende unserer Spezies herbeizuführen. Die Menschen vernichten mit atemberaubender Geschwindigkeit ihre eigenen Lebensgrundlagen, ganz ohne dass es ChatGPT oder Künstliche Intelligenz dafür brauchte.



So sieht die Künstliche Intelligenz ein Wohnhaus der Zukunft Bild: Midjourney / Spehr

Ist Ödipus eine Software?

Woher kommt also die Idee, dass eine KI diametrale Interessen zu denen der Menschheit hätte und diese deswegen auslöschen wollte? Ist das eine Projektion? Würde Elon Musk das tun, wenn er die Mittel dafür hätte? Wird die Vorstellung eines strafenden Gottes auf KI übertragen, inklusive der Angst, für die eigenen Sünden büßen zu müssen? Warum nicht einfach mal folgendes Gedankenexperiment? Eine KI, die wirklich in der Lage wäre, autonom auf einem Level zu denken, das sich kein Mensch mehr vorstellen kann: Wäre es für die nicht viel einfacher, sich den Menschen durch intelligente Anreize gefügig zu machen?

Immerhin lieben wir unsere Geräte! Wir tragen unsere Smartphones immer am Körper oder in der Nähe, wir schützen sie vor Schmutz und Schaden, wir haben ein sozioökonomisches System geschaffen, das es quasi unmöglich macht, dass irgendwann keine Smartphones mehr hergestellt werden. Warum um Himmels willen sollte uns eine KI auslöschen, wenn sie uns noch gebrauchen könnte, für Arbeiten zum Beispiel, auf die sie keine Lust hat?

Mehr zum Thema

Bei der letzten Frage drängt sich der Verdacht auf, dass einige der Unterzeichner, die ja teilweise tatsächlich sehr klug sind, den unterzeichneten offenen Brief gar nicht gelesen haben. Die Frage lautet, ob wir das Risiko eingehen sollen, die Kontrolle über unsere Zivilisation zu verlieren. Die Frage wäre allerdings, wann wir jemals die Kontrolle darüber hatten. Hätten wir sie, würden wir doch in einem Utopia ohne Leid und Elend wohnen.

Wir leben aber in einer Welt, in der es auf deutschen Autobahnen keine Geschwindigkeitsbegrenzung gibt und in republikanisch regierten US-Bundesstaaten keine Regulierung von Schusswaffen. Zivilisation ist ein Kampf diametral entgegenstehender Interessen, der zivilisierteste Ausgleich dieser findet über demokratische Wahlen statt. Aber es grenzt an Verschwörungsglauben zu denken, eine Instanz, egal wie hoch entwickelt, wäre in der Lage, bestimmen zu können, wie sich acht Milliarden Menschen entwickeln werden.

Was will Altman wirklich?

Die industrielle Revolution automatisierte die Arbeit, die uns bevorstehende Revolution wird das Denken automatisieren.

Höchstwahrscheinlich wird kein Stein auf dem anderen bleiben. Der Punkt des Briefes, den man gelten lassen muss, ist der, dass sich Regierungen und Parlamente schnell Gedanken über die Regulierung von KI machen sollten. Deutschland zum Beispiel täte gut daran, analog zum Bundesdatenschutzbeauftragten einen Bundesbeauftragten für Künstliche Intelligenz zu schaffen, inklusive einer entsprechenden Behörde, die sich um alle Fragen rund um KI kümmert. Wenn in Zukunft Text vor allem über Large Language Models erstellt wird, wird das Modell bestimmen, über was überhaupt noch diskutiert werden kann. Kein Modell ist ohne Vorurteile, der politische Diskurs könnte eingeschränkt werden.

Eine internationale Behörde zur Überwachung von KI, angelehnt an die Internationale Atomenergie-Organisation, forderte erst vor wenigen Tagen Sam Altman, der CEO von Open AI, also der Firma, die ChatGPT entwickelt hat, in einem Blogbeitrag. Mit 5000 Zeichen mit Leerzeichen ist der Text dabei recht kurz, und auch inhaltlich lässt er bis auf die Forderung, KI müsse reguliert werden, zu wünschen übrig. Liest man den Text kritisch, könnte man unterstellen, dass es Altman vor allem darum geht, kleinere Firmen durch Regulierung daran zu hindern, auf das Niveau Open Als zu kommen.

Folklore aus dem Silicon Valley

Gleichzeitig hat die Forderung nach Regulierung des eigenen Unternehmens im Silicon Valley Tradition und ist fast schon ein Stück Folklore. Mehrfach forderte zum Beispiel Mark Zuckerberg, Social Media müsse stärker reguliert werden. Daraus geworden ist nichts, obwohl Facebook, Open AI und andere Konzerne ja das Geld hätten, in Washington für die Regulation der eigenen Branche zu lobbyieren. Es kann also unterstellt werden, dass diese Forderungen größtenteils Imagepflege sind.

Tabak- und Alkoholkonzerne berichten ja auch gern darüber, was sie unternehmen, damit Kinder und Jugendliche nicht süchtig werden, um Kinder und Jugendliche dann süchtig zu machen. Wer die Mühlen der internationalen Politik kennt, ahnt vielleicht, dass eine Behörde auf UN-Level nicht über Nacht und nicht durch einen Blogbeitrag entsteht. Zum Vergleich: Die erste Kernwaffe wurde im Juli 1945 gezündet, im Dezember 1953 schlug Präsident Eisenhower in einer Rede die Gründung einer Internationalen Atomenergie-Organisation vor, und gegründet wurde sie im Juli 1957.

Doch wie realistisch ist das geforderte KI-Moratorium überhaupt? Nicht besonders, denn wie in einem mutmaßlich von einem Wissenschaftler bei Google stammenden Dokument mit dem Titel „We have no moat" („Wir haben keinen Burggraben") festgestellt wurde, ist die Technologie, mit der man neue Large Language Models erstellen kann, Open Source – das heißt, sie ist für jeden frei verfügbar. Zwar sei die Qualität der Modelle von Google und Microsoft immer noch besser, aber die Lücke zwischen diesen und den Open-Source-Varianten würde sich schnell schließen. Somit herrscht Chancengleichheit zwischen den Giganten des Silicon Valleys und den Start-ups, bei denen verständlicherweise Goldgräberstimmung herrscht, weil alle das nächste Google oder Amazon werden wollen.

Wie ernst es zumindest Elon Musk mit dem Brief und dem Moratorium meinte, wurde einen Monat später, am 17. April, klar, als er in einem Interview mit Fox News bekannt gab, dass er seine eigene KI-Firma gegründet habe und mit „TruthGPT" ein Large Language Model schaffen wolle, das nicht so politisch korrekt wie das von Open AI sei, was immer das bedeuten soll.

Wenn ich ChatGPT frage, wie man Crystal Meth herstellt, weist mich die KI darauf hin, dass es mir sowas nicht erklären möchte, weil Crystal Meth gefährlich ist. Will Musk also eine KI, die Menschen erklärt, wie man Meth herstellt oder Autobomben baut? Die also eine Gefahr für die Menschheit wäre? Wir werden sehen.