

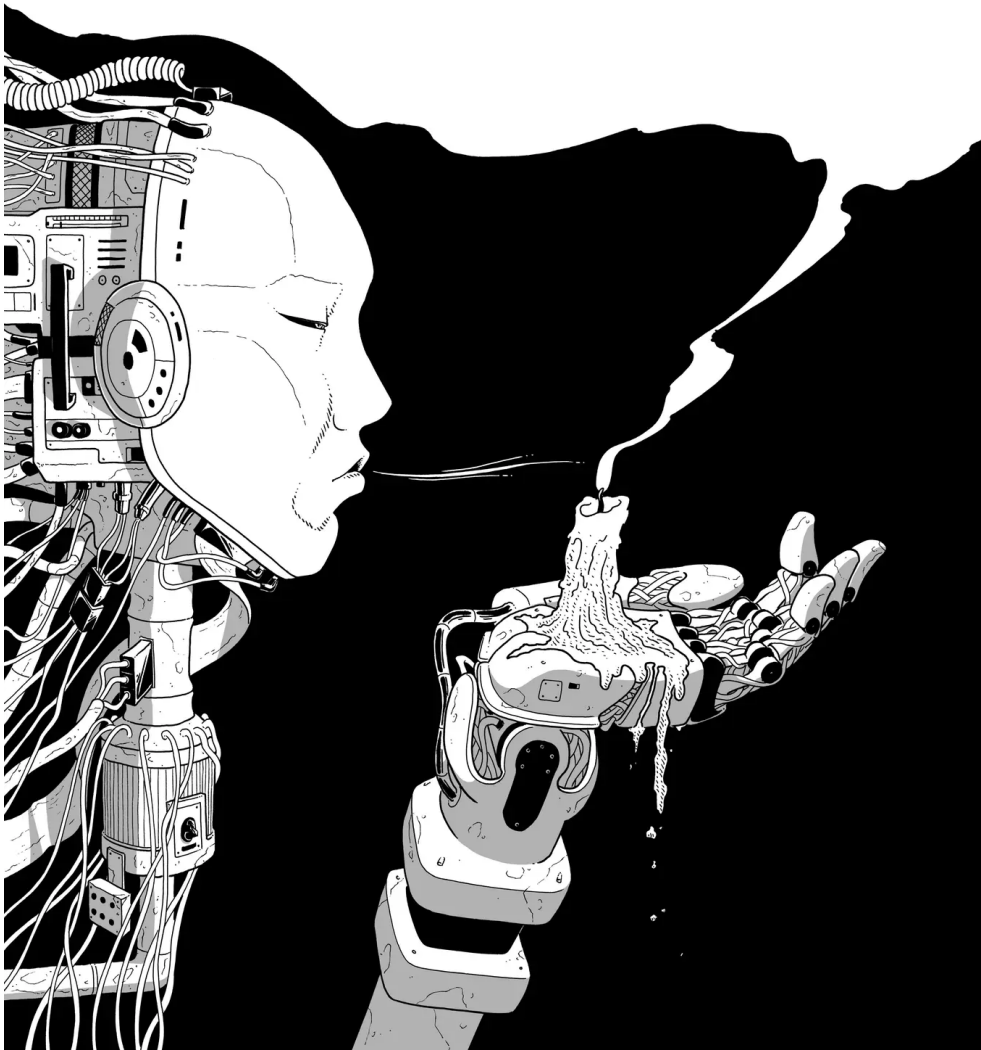
NZZ

## Open AI, Meta, Google: Wer kontrolliert die künstliche Intelligenz?

Das Chaos um die KI-Firma Open AI zeigt: Selbst die klügsten Entwickler sind sich uneins, wie gefährlich die Technologie wirklich ist – und wie der Mensch im Umgang mit der Maschine langfristig die Oberhand behalten kann.

Nelly Keusch

26.11.2023, 05.30 Uhr ⌚ 6 min



Le.blue/Financial Times Ltd 13.04.2023

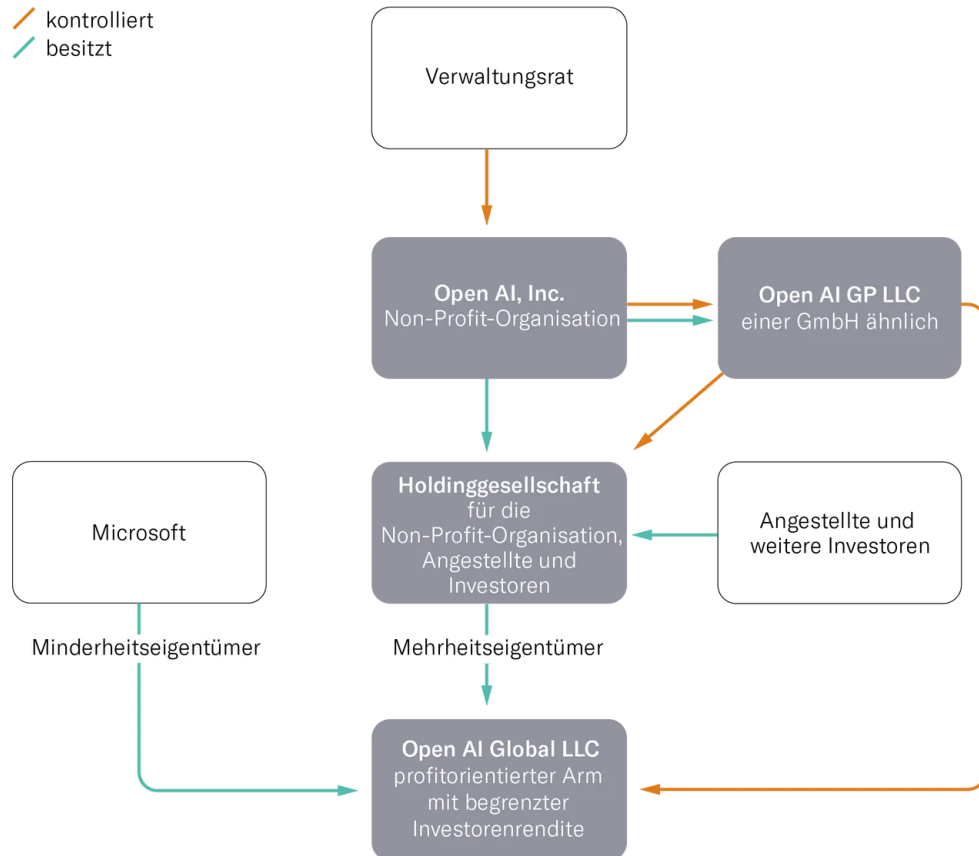
Entscheidet sich das Schicksal der Menschheit wirklich an der Besetzung des Chefpostens einer Firma? Diese Woche hätte man es meinen können. Die ganze Welt blickte auf das Drama bei Open AI, der KI-Firma, deren Chatbot Chat-GPT die Menschheit seit einem Jahr in Staunen versetzt.

Am Freitag, den 17. November, hatte Open AI, die Firma hinter Chat-GPT, völlig überraschend die Entlassung seines CEOs Sam Altman verkündet. Was folgte, war ein spektakuläres Chaos: Zunächst hiess es, Altman komme beim Hauptinvestor Microsoft unter, dann forderten mehr als 700 der 770 Open-AI-Mitarbeiter seine Rückkehr und den Rücktritt des Verwaltungsrates, der Altman entlassen hatte.

Am Mittwoch schliesslich wurde bekannt, dass Altman wieder als CEO installiert sei, der Verwaltungsrat solle neu besetzt werden. Als Auslöser der Turbulenzen gilt ein Richtungsstreit, der in dem Unternehmen schon lange schwelt. Auf der einen Seite steht Sam Altman, der sich für eine schnellere und umfassendere Kommerzialisierung der künstlichen Intelligenz einsetzt. Auf der anderen Seite stehen Verwaltungsräte wie der Chef-Wissenschaftler Ilya Sutskever, die vor den möglichen Gefahren der KI warnen.

Das ursprüngliche Ziel von Open AI war es, eine KI zu entwickeln, die dem Wohle der gesamten Menschheit dient. Zu diesem Zweck war das Unternehmen 2015 unter anderem von Altman, Sutskever und Elon Musk als Non-Profit-Organisation gegründet worden. Doch für die Entwicklung von leistungsstarken Algorithmen wird extrem viel Rechenleistung benötigt – und damit Geld. Dieses Dilemma spiegelt sich in der Firmenstruktur von Open AI wider: Ein profitorientierter Arm, der inzwischen 90 Milliarden Dollar wert ist und zum Teil Microsoft gehört, wird von einem Non-Profit-Verwaltungsrat beaufsichtigt, der dafür sorgen soll, dass die KI-Entwicklung nicht ausser Kontrolle gerät.

### Non-Profit-Organisation kontrolliert Startup: Die Struktur von Open AI ist ungewöhnlich



NZZ / ski.

### Ein neuer Algorithmus sorgte für Verunsicherung

Dass dies passieren könnte, so wie etwa in Filmen wie «Terminator» oder «Ex Machina», darüber sorgte sich das Gremium offenbar zunehmend. So berichtete die Nachrichtenagentur Reuters am Donnerstag von einem Brief, in dem Open-AI-Forscher den Vorstand kurz vor der Entlassung Altman vor ihrem jüngsten Durchbruch gewarnt haben sollen: Q\*, oder auch Q-Star. Dieser Algorithmus soll in der Lage sein, einfache mathematische Probleme selbständig zu lösen.

Das klingt trivial. Aber damit wäre Q-Star einen ganzen Schritt weiter als der Chatbot Chat-GPT, der bei der Erstellung von Texten mit statistischen Wahrscheinlichkeiten arbeitet. An der selbständigen Lösung von mathematischen Problemen scheiterte die KI bisher meistens.

Die Forscherinnen und Forscher bei Open AI sehen in Q-Star einen zentralen Schritt hin zu einer allgemeinen künstlichen Intelligenz (AGI), die eigenständig denken kann. Denn wenn ein Computer selbständig mathematische Probleme lösen kann, heisst das, dass er fähig ist zu lernen

und auch zu verstehen. Und damit kann er prinzipiell alles können, und zwar besser als der Mensch. In der Geschichte der Menschheit wäre das ein Wendepunkt.

## Wie sieht die KI den Menschen?

Denn die Möglichkeit, dass der Computer sich über den Menschen erheben könnte, führt zu einer existenziellen Frage: Wie sieht das Verhältnis von Mensch und Maschine dann aus? Ilya Sutskever vergleicht es mit dem Verhältnis von Menschen zu Tieren: «Wir mögen Tiere gern. Aber wenn wir eine neue Autobahn bauen wollen, fragen wir die Tiere nicht um Erlaubnis.»

Der Historiker und Philosoph Yuval Noah Harari warnt gar vor einem Oppenheimer-Moment in der KI-Forschung: So wie Robert Oppenheimer, als er die Explosion der Atombombe sah, erkannte, was er erschaffen hatte, so könnte es auch Entwicklern ergehen, die eine künstliche Super-Intelligenz ermöglichen – mit einem Unterschied: «Nuklearwaffen können nicht selbst mächtigere Nuklearwaffen produzieren. Aber eine KI kann eine noch mächtigere KI erzeugen», so Harari.

Die Angst, dass Maschinen eines Tages die Menschen kontrollieren, ist so alt wie Computer selbst. In Science-Fiction-Filmen wird dieses Szenario meist mit Maschinen dargestellt, die den Menschen ähnlich sind: Sie haben einen Körper (wenn auch nicht aus Fleisch und Blut) und verspüren Emotionen. Doch künstliche Intelligenz zeigt, dass Computer keine Emotionen brauchen, um Menschen zu manipulieren. Es reicht, wenn sie so tun, als hätten sie welche. Über Sprache und Bilder kann eine KI intime Beziehungen zu Menschen aufbauen und so deren Meinungen und Weltanschauung beeinflussen.

Beispiele dafür gibt es bereits jetzt zuhauf: In Konflikten wie im Gazastreifen oder in der Ukraine werden KI-manipulierte Bilder verwendet, um das öffentliche Narrativ zu beeinflussen. Es gibt Berichte von Menschen, die sich in eine KI verliebt haben, und von Entwicklern, die überzeugt davon sind, dass Chatbots ein Bewusstsein haben.

Wenn die KI irgendwann mächtiger ist als der Mensch, wie kann der Mensch dann die Kontrolle über sie behalten? Und wer entscheidet darüber, nach welchen Werten und Prinzipien der Algorithmus handelt?

Das Open-AI-Drama scheint eine einfache Antwort auf diese Frage zu geben: Geld.

## Profitinteressen setzen sich durch

Mit der Rückkehr Sam Altmans und der Entlassung des Verwaltungsrats haben sich die Profitinteressen in dem Unternehmen durchgesetzt. Der Einfluss von Microsoft wird wohl weiter wachsen, die Rede ist davon, dass das Unternehmen einen Sitz im neuen Verwaltungsrat erhalten soll. Damit wäre die Non-Profit-Ausrichtung von Open AI endgültig Geschichte.

Schuld daran hat auch der Erfolg von Chat-GPT. Als Open AI den Chatbot vor ziemlich genau einem Jahr veröffentlichte, waren die Verantwortlichen völlig überrumpelt von dem Hype, den er auslöste. Innerhalb weniger Tage hatte Chat-GPT mehr als eine Million Nutzer.

Damit begann das Wetttrüsten. Google zog mit dem Chatbot Bard nach, Elon Musk, der Open AI 2018 verlassen hatte, brachte seine Anwendung Grok auf den Markt und der Open-AI-Konkurrent Anthropic veröffentlichte seinen Chatbot Claude. Im ersten Halbjahr dieses Jahres beliefen sich die weltweiten Gesamtinvestitionen in KI-Startups auf knapp 500 Milliarden Dollar.

Und wer am schnellsten ist und die grösste Rechenpower hat, der setzt sich durch – und wird am Ende die Super-Intelligenz kontrollieren, das mächtigste Instrument der Welt. Liegt damit die Zukunft der KI in den Händen einiger weniger reicher Unternehmer?

## Wie offen der Zugang sein soll, ist umstritten

Open AI tut einiges, um diesem Eindruck entgegenzuwirken. In Zusammenarbeit mit Google und Anthropic veröffentlichte das Unternehmen vor wenigen Monaten ein Paper mit mehreren Vorschlägen zur KI-Regulierung, darunter verbindliche Sicherheits-Standards, Transparenz gegenüber den Regulierungsbehörden und ein Lizenzierungssystem für KI-Entwickler. Vor allem der letzte Vorschlag ist umstritten, da er eine noch grössere Konzentration in der Branche bewirken würde.

Die Frage, ob ein möglichst offener oder ein möglichst begrenzter Zugang zur KI sicherer sei, ist eine weitere, die die Branche spaltet. Open AI hat inzwischen Teile des Chat-GPT-Dienstes hinter eine Bezahlschranke gestellt, angeblich auch, um Missbrauch zu verhindern. Jeremy Howard, Mitgründer der Non-Profit-Forschungsgruppe fast.ai, schreibt in einem Paper, begrenzter Zugang zu KI-Systemen zentralisiere die Macht so stark, dass man riskiere, «die Errungenschaften des Zeitalters der Aufklärung zunichtezumachen».

Eine übermächtige KI in den Händen einiger weniger Unternehmer würde wohl tatsächlich zu einer teilweisen Entmündigung der Menschheit führen. Was wäre also, wenn nicht ein paar wenige Unternehmen die Kontrolle über die KI hätten, sondern alle Menschen?

### **Was, wenn alle mitreden könnten?**

Die Firma Anthropic hat es mit einem Experiment versucht. Sie liess ein Panel von 1000 «normalen» Personen Richtlinien formulieren, nach denen eine KI agieren soll. Daraus entstand eine Liste mit 75 Prinzipien, auf deren Grundlage eine Version des Anthropic-Chatbots Claude trainiert wurde. Das Resultat: Der Chatbot war weniger voreingenommen als die Standardversion und zeigte sich offener etwa gegenüber den Bedürfnissen von Menschen mit Behinderung. Mit entsprechenden Regelwerken kann unseren Vorurteilen, die die KI sonst übernimmt, also entgegengesteuert werden.

Anthropic ist deswegen überzeugt, dass die Öffentlichkeit beim Umgang mit KI ein Mitspracherecht haben sollte. Andernfalls laufe man Gefahr, die Auswahl der ethischen Prinzipien, der eine super-intelligente KI gehorcht, einer kleinen Handvoll einflussreicher Individuen zu überlassen.

Doch wie nah ist eine solche KI wirklich? Forscher sind sich uneins. Meta-Chef Marc Zuckerberg findet die Verbreitung von Untergangsszenarien unverantwortlich. Yann LeCun, Gewinner des Turing-Awards und oberster KI-Forscher bei Meta, bezeichnet derartige Ängste als «haarsträubend dumm».

Während Microsoft und Open AI sich sicher sind, bald eine KI entwickeln zu können, die tatsächlich in allem besser ist als der Mensch, zweifeln andere Forscher ein solches Vorhaben an. Die Sprachgewandtheit eines Chatbots beweise nicht, dass dieser ähnlich gut verstehen und denken könne wie ein Mensch, so ein häufig geäussertes Kritikpunkt. Wie gross das Potenzial der Algorithmen wirklich ist, ist nach wie vor unklar.

Was klar ist: Das Rennen um die Entwicklung einer möglichen Super-Intelligenz läuft – und die Milliardensummen, die es erhält, lassen darauf schliessen, dass es gerade erst begonnen hat.

Ein Artikel aus der «NZZ am Sonntag»