«EINE HOFFNUNGSVOLLE, HUMANE ZUKUNFT MIT KI IST MÖGLICH»



KI wird mitunter als Bedrohung wahrgenommen – nicht erst, seitdem mehrere Chefs von US-Tech-Konzernen vor einer Superintelligenz warnen und eine Regulierung fordern. Thilo Stadelmann ist einer der profiliertesten KI-Experten in der Schweiz. Er ordnet die Bedrohungsszenarien ein und spricht über mögliche Lösungswege.

In letzter Zeit häufen sich die Meldungen über KI-Sicherheitsrisiken und einen drohenden Kontrollverlust. Besteht eine Gefahr vor einer Superintelligenz, die dem Menschen gefährlich werden könnte?

Thilo Stadelmann: Die Gefahr einer drohenden Superintelligenz ist meiner Analyse nach rein hypothetisch. Es ist eine Science-Fiction-Idee, die auf Basis heutiger Technologie nicht umsetzbar ist. Um solche Meldungen einordnen zu können, müssen wir drei Aspekte unterscheiden. Erstens: die Weltanschauung, die der Interpretation und Vorhersage von Ereignissen zugrunde liegt. Zweitens: die Höhe der finanziellen Anreize, etwa im Hinblick auf anstehende Börsengänge. Und drittens: die Frage, wie sich Tech-Konzerne mithilfe der Regierung einen Vorteil gegenüber ihrer Konkurrenz verschaffen könnten.



Foto: Caroline Krajcir

Die Gefahr, die von der KI ausgeht, ist eine Weltanschauung?

Gewissermassen ja. Ob KI auf dem Weg zu einer Superintelligenz ist, stellt keine technisch-wissenschaftliche Frage dar, sondern eine weltanschauliche. Krasse Behauptungen wie «wir könnten alle sterben» verlangen normalerweise Belege. Solche fehlen in diesem Fall vollständig. Das Narrativ von der KI, die ausbricht und Unheil anrichtet, ist eine seit Jahren gehegte weltanschauliche Vorstellung und Erwartung einer bestimmten Community. Wenn nun Berichte über KI-generierte Hacks auftauchen, lesen manche diese Fakten als Bestätigung dieser Erwartung. Daneben gibt es auch reale Risiken, auch im Bereich der Cybersicherheit – aber diese würden nicht solch ein Aufsehen erregen.

Zum zweiten Punkt: Wo liegt der finanzielle Anreiz beim Propagieren einer Superintelligenz?

Die Kontrolle über die KI-Agenten ist bei den bekannt gewordenen Vorfällen nicht verloren gegangen, sie wurde absichtlich aufgegeben. Das eigentliche Ziel dahinter scheint zu sein, eine bestimmte Geschichte in den Medien zu platzieren. Im Sinne von: «Seht her, wir sind gefährlich, bitte reguliert uns.» In einer Zeit geplanter Börsengänge können solche Aussagen wertvoll sein, denn sie suggerieren immense Fähigkeiten der KI-Systeme, was den Marktwert in die Höhe treibt.

Das erklärt aber noch keine freiwillige Regulierung, wie sie von einigen Konzernen gefordert wird.

Die Tech-Konzerne könnten mithilfe einer strengen Regulierung manche Konkurrenten ausschalten. Mit der Möglichkeit, eine Regulierung zu lancieren und diese mitzugestalten, schaffen sich die grossen KI-Anbieter einen Vorteil etwa gegenüber Open-Source-Anbietern, die potenziell verboten würden.



Thilo Stadelmann
Foto, Caroline Krajcir

Erachten Sie eine KI-Regulierung als sinnvoll?
Regulierung kann unbedingt sinnvoll sein, zum Beispiel auf Ebene der Geschäftsmodelle, ähnlich, wie wir es bereits gegen Monopole kennen. Regulierung auf Basis von Science-Fiction und Eigeninteressen ist eine schlechte Idee für die Welt. Bestehende Regulierungen wie etwa der EU AI Act scheinen prinzipiell angemessen, haben aber derzeit einen schweren Stand gegen eine Industrie, die bestimmte Gesetze für das Versprechen auf Profit weitgehend ignoriert.

Warum könnte die Abhängigkeit von US-Tech-Konzernen auch bei Schweizer Unternehmen problematisch werden?
Es ist leicht vorstellbar, wie Abhängigkeiten zu geopolitischen Waffen werden können, man denke nur an Zölle oder an abgeschaltete Zugänge zu bezahlten KI-Accounts in den letzten Monaten. Ganz abgesehen davon, dass man in Abhängigkeit selten das beste Angebot zum angemessenen Preis bekommt. Aber an KI-Zugängen hängt noch viel mehr, etwa die eigenen Firmengeheimnisse, welche mit dem Datenschutz zusammenhängen, und damit am Ende die Existenz der heimischen Industrie. Ein offener Markt mit freier Wahl des Anbieters von agentischen KI-Systemen sollte eines der wichtigsten strategischen Anliegen jeder Volkswirtschaft sein. Und wenn die Wertschöpfung in der Anwendung stattfindet, eröffnet dies tatsächlich Möglichkeiten für eine global verteilte KI-Industrie, anstelle von gigantischen Generalisten.

Ist eine technologische Souveränität für die Schweiz und für Schweizer Unternehmen erstrebenswert? Wie könnte diese aussehen?
Ja. Wer frei wählen kann, baut unter anderem auch bessere und stabilere Beziehungen zu seinen Partnern auf, als wenn er von diesen abhängig ist und die Partnerschaft damit de facto erzwungen wurde. Der Weg dahin führt über lokale KI-Anbieter, die starke Produkte auf Basis von Open Source KI-Modellen entwickeln und vermarkten. Dabei können etwa auch Datenschutzgesetze kontrolliert eingehalten werden.

Sie sprechen hier auch Schweizer Angebote an wie Apertus oder das Produkt von AlpineAI, der Firma, die Sie mitgegründet haben. Wo liegen die Vorteile solcher KI-Plattformen?
Bei solchen lokalen oder Open-Source-Alternativen geht es nicht nur darum, von weltweiten Anbietern unabhängig zu werden – auch die Leistung kann je nach Anwendungsfall deutlich besser ausfallen, wenn die Systeme direkt dafür optimiert wurden. Apertus stellt unter anderem sicher, dass das Know-how zur Entwicklung von Frontier-Modellen im Land vorhanden ist. Und bei der souveränen Schweizer KI-Lösung von AlpineAI bleiben die Daten stets in den eigenen Rechenzentren, die gesamte Infrastruktur steht in der Schweiz. Es gelten die Schweizer Datenschutzstandards und alles Läuft vertraglich unter Schweizer Recht. Die Plattform eignet sich insbesondere für Branchen, die sensible Daten verarbeiten oder in systemkritische Bereiche wie Behörden, dem Gesundheitswesen und der Hightech-Industrie.

«KI nimmt uns nicht unsere Freiheiten und Fähigkeiten – aber wir können sie freiwillig abgeben, und das gilt es zu vermeiden.» Thilo Stadelmann

«Pro-menschliche KI könnte zu einem Standard, einer Art Gütesiegel werden – so ähnlich wie Fairtrade- oder Bio-Labels bei Lebensmitteln.»

Wo sehen Sie im Allgemeinen die Chancen und Risiken für Schweizer Unternehmen, die generative KI einsetzen wollen, um ihre Produktivität zu steigern?
Sehr reale Risiken – aber mit auch sehr realen Gegenstrategien – liegen im Bereich der Abhängigkeit von diesen Werkzeugen. Dies sowohl im marktwirtschaftlichen Sinn wie zuvor besprochen, als auch in der Abhängigkeit des Einzelnen von den Fähigkeiten des Systems (Stichwort «De-Skilling»). KI nimmt uns nicht unsere Freiheiten und Fähigkeiten – aber wir können sie freiwillig abgeben, und das gilt es zu vermeiden. Ich bin zuversichtlich, dass uns das gelingen kann.

Vernichtet KI Ihrer Meinung nach Arbeitsplätze?
Untersuchungen zeigen, dass die derzeitige schwierigere Situation für Hochschulabsolventen höchstens teilweise auf KI zurückzuführen ist. Die längerfristigen Auswirkungen auf den Arbeitsmarkt werden hingegen oft als gering eingestuft, ganz im Gegenteil zum Tenor in der öffentlichen Debatte: Es könnten mehr Arbeitsplätze entstehen, da sich die Aufmerksamkeit auf diejenigen Aufgaben verlagert, die nicht automatisiert werden, einschliesslich derzeit nicht existierender Aufgaben, Rollen und Berufe. Denken Sie etwa daran, wie viele Arbeitsplätze die Entwicklung von Elektrizität, Auto, Computer oder Internet gebracht haben. Und eine Superintelligenz, die alle diese menschliche Arbeit überflüssig machen könnte, ist eben trotz anderslautender Behauptungen nur eine Marketingmasche.

In Ihrer Forschungstätigkeit entwickeln Sie Zukunftsszenarien. Mit welchen Entwicklungen rechnen Sie?
Es geht nicht darum, eine genaue Vorhersage darüber zu treffen, was in der Zukunft passieren wird. Es geht mehr darum, den Raum für realistische, erreichbare Szenarien zu öffnen. Szenarien wie die Auslöschung der Menschheit, ein massiver Arbeitsplatzverlust oder nutzlose Menschen, die alle Aufgaben der künstlichen Intelligenz abgeben, sehe ich kritisch. Die Erwartungen an die Entwicklung einer Superintelligenz sind hoch. Wahrscheinlicher ist, dass diese Bestrebungen scheitern und dass stattdessen eine moralische und strukturelle Neuausrichtung ausgelöst wird. Es ist deshalb denkbar, dass Mitte der 2030er-Jahre künstliche Intelligenz nicht mehr zentralisiert und intransparent sein wird – sondern dass es vielmehr pro-menschliche KI geben wird.

Was ist mit pro-menschlicher KI gemeint?
Das ist künstliche Intelligenz, die darauf ausgerichtet ist, Menschen zu unterstützen und menschliches Wohlergehen zu fördern, anstatt nur effizient zu sein und Präferenzen zu folgen. Dieses Paradigma des pro-menschlichen KI-Designs, das jetzt schon hier und da zu sehen ist, könnte zu einem Standard, einer Art Gütesiegel werden – so ähnlich wie Fairtrade- oder Bio-Labels bei Lebensmitteln. Der Fokus wird weggehen vom heutigen Wettlauf «wie erschaffen wir eine Superintelligenz?» hin zu «wie nutzen wir die bestehende KI verantwortungsvoll und zum gemeinsamen Nutzen?».

Das klingt optimistisch
Eine hoffnungsvolle, humane Zukunft mit KI ist definitiv möglich. Ob wir diese ansteuern, liegt auch an unseren Weichenstellungen heute, da haben wir die Wahl.

Thilo Stadelmann
Thilo Stadelmann ist Informatiker, Professor für künstliche Intelligenz und maschinelles Lernen an der ZHAW School of Engineering in Winterthur und Leiter des dortigen Centre for Artificial Intelligence. Als Gründer, Leiter und Forscher trug er beträchtlich zum Ecosystem um datengetriebene Wertschöpfung in der Schweiz bei, nicht zuletzt als Unternehmer mit der Davoser AlpineAI AG. Auf TEDx erklärt er «How not to fear AI». Sein Buch «Applied Data Science – Lessons Learned for the Data-Driven Business» erschien 2019 im Springer-Verlag.