Superintelligenz: Monster, Macht und Moneten

Von Ulrike Simon 22. September 2026

Veröffentlicht bei: https://dreimallinks.de/2026/09/superintelligenz-monster-macht-und-moneten/
Dort finden sich auch die Rabbit Holes, die einzelne Aspekte genauer beleuchten.

- ***16. Juli 2026**: das KI-Unternehmen Hugging Face gibt in seinem Blog bekannt, dass es Ziel eines Cyberangriffs geworden sei, der „sich von allem unterschied, womit wir bisher zu tun hatten“. Das Unternehmen, das Open-Source-KI-Modelle und Datensätze hostet, erklärte, dass ein autonomer Agent Zugriff auf einen Teil seiner internen Daten erlangt habe.*
- ***2. September 2026**: Das „Wall Street Journal“ berichtet, dass der ehemalige Trump-Berater Steve Bannon „einen Vorschlag von Senator Bernie Sanders (I-VT) und der Abgeordneten Alexandria Ocasio-Cortez (D-NY) unterstützt, den Bau neuer Rechenzentren zu verbieten“.*
- ***3. September 2026:** OpenAI veröffentlicht GPT-6 Astra, das das Unternehmen als „das intelligenteste und am besten abgestimmte Modell der Welt“ beschreibt.*
- *US-Senator Bernie Sanders (I-VT) und der US-Abgeordnete Greg Casar (D-TX) kündigen den „Ban Artificial Superintelligence Act“ an. Sollte das Gesetz verabschiedet werden, wird die gesamte KI-Entwicklung in den Vereinigten Staaten ausgesetzt.*
- ***8. September 2026**: Der Anthropic-Forscher Jacob Coxon veröffentlicht eine Meldung, in dem es heißt: „Ich habe heute bei Anthropic gekündigt. Ich habe die letzten drei Jahre mit Vortrainings-Forschung verbracht, sowohl bei OpenAI als auch bei Anthropic Keines der beiden Unternehmen handelt verantwortungsbewusst. Sie rasen geradewegs auf eine sich selbst verbessernde Superintelligenz zu und spielen mit unserem Leben“, und „die Menschen, die KI entwickeln, glauben ernsthaft, dass sie uns alle bis zum Ende des Jahrzehnts töten könnte“.*

Kommt jetzt die Superintelligenz, die sich - immer leistungsfähiger geworden - menschlicher Kontrolle entzieht, einen eigenen Willen entwickelt und schließlich uns alle töten wird?

Ausgangspunkt des Gesprächs des Technologiekritikers Ed Zitron und des Wettbewerbsexperte Matt Stoller mit dem Journalisten David Sirota ist der o.g. Vorfall, bei dem sich KI-Agenten von Open AI bei der Suche nach Sicherheitslücken Zugang zu Systemen der Plattform *Hugging Face* verschafft hatten.

Rabbit Hole: Gastgeber und Gäste: David Sirota, Ed Zitron und Matt Stoller

Aus ihrer Sicht ist jedoch nicht die (künftige) Künstliche Intelligenz das Problem, sondern die Menschen. Und die Umstände, in denen KI entwickelt und eingesetzt wird. Zitron und Stoller bestreiten nicht, dass KI Gefahren mit sich bringt. Aber in ihrer Argumentation geht es konkret um unsichere Produkte, schwer vorhersehbare Folgen, die Wirtschaft und Verantwortung im hier und jetzt.

Und dann ist da noch das Rätsel, warum ausgerechnet aus den Unternehmen, die diese Technologie mit enormem Kapitaleinsatz vorantreiben, besonders eindringliche Warnungen vor den möglichen Gefahren einer künftigen Superintelligenz dringen und eine stärkere Regulierung der KI-Entwicklung gefordert wird.

Das Gespräch führt damit von einem spektakulären Zwischenfall über die Funktionsweise von KI-Agenten zu Medien, Regulierung, chinesischer Konkurrenz und schließlich zu einer ganz anderen Vorstellung von der möglichen KI-Krise: „*Big Short, not Terminator.*"

Was geschah tatsächlich beim „Ausbruch“ der KI?

Der *Hugging-Face*-Vorfall belegt keine autonome, außer Kontrolle geratene KI, erklärt Zitron. Die Agenten hätten nicht „aus eigenem Willen“ gehandelt, sondern innerhalb einer von Menschen konstruierten Versuchsanordnung versucht, eine ihnen gestellte Aufgabe zu erledigen. Die reale Gefahr sieht Zitron an anderer Stelle: Unternehmen verbinden probabilistische und unzuverlässige Systeme mit großen Rechenressourcen und Werkzeugen, mit denen reale Eingriffe möglich werden. Die entscheidende Frage sei, welche Handlungsmöglichkeiten die Entwickler solchen Systemen geben und wie sie mit der Tatsache umgehen, dass daraus mit hoher Wahrscheinlichkeit unerwartete Folgen resultieren können. Gefährlich sei also die Versuchsanordnung und ihre Betreiberpraxis, nicht ein entstandener autonomer Wille der Maschine.

Rabbit Hole: LLMs, Harness, Agenten und Emergenz

Rabbit Hole: Was geschah bei Hugging Face?

Ein zweiter Punkt betrifft die Anthropomorphisierung. Indem LLMs dazu programmiert werden, in menschenähnlicher Sprache kommunizieren, werde suggeriert, technische Vorgänge fälschlich als Absichten, Entscheidungen oder Eigeninitiative zu interpretieren. Diese Art der Programmierung sei jedoch die Entscheidung der Entwickler und keineswegs alternativlos, erläutert Zitron.

Er hält die Diskussion über eine bereits vorhandene autonome Superintelligenz deshalb für eine Diskussion über eine Technologie, die so nicht existiert.

Worin bestehen die realen Gefahren?

Natürlich erzeuge KI reale Gefahren, aber viele davon seien keineswegs neu, sagt Stoller, sondern bereits bekannte gesellschaftliche und technische Risiken. Manipulation, Überwachung, Cyberangriffe oder unsichere Produkte gab es schon vorher. Neu sei, wie leicht, schnell und in welchem Umfang solche Handlungen mit KI ausgeführt werden können. Die Technologie verändere damit Reichweite, Geschwindigkeit und Kosten bereits bekannter Risiken. Wo KI in Arbeits- und Entscheidungsprozesse eingebaut wird, könne sie zudem verändern, wie Entscheidungen zustande kommen und Verantwortung organisiert wird.

Rabbit Hole: KI im Alltag

Rabbit Hole: KI im Krieg

Stollers entscheidender Perspektivwechsel lautet: LLMs sind Produkte. Sie können nützlich sein, sind aber in bestimmten Anwendungen unsicher. Damit werden Produktsicherheit, Haftung und Aufsicht zum eigentlichen Problem. Zitron ergänzt: Auch gefährliche Experimente müssen als Handlungen der Unternehmen behandelt werden. Das Problem sei gerade nicht die Kontrolle einer sehr leistungsfähigen oder gar übermenschlichen KI, sondern die Art und Weise wie Unternehmen unzuverlässige Systeme mit gefährlichen Fähigkeiten ausstatten und sie ausprobieren. Die Superintelligenz- und Auslöschungsdebatte lenke von den konkreten Schäden und Verantwortlichkeiten ab.

Warum wird niemand zur Verantwortung gezogen?

Derart entmystifiziert sind KI-Systeme laut Stoller nichts anderes als Produkte, Waren wie alle anderen. Unternehmen entwickeln sie, bringen sie auf den Markt und entscheiden, welche

Fähigkeiten sie besitzen und unter welchen Bedingungen sie benutzt werden können. Menschen und Organisationen entscheiden wiederum, wo sie eingesetzt werden.

Unterschiedliche Anwendungen erzeugen unterschiedliche Risiken und verlangen unterschiedliche Regeln.Die nötigen Instrumente wie Produktsicherheit, Haftungsrecht, Strafrecht, Aufsicht und Wettbewerbsrecht, seien bereits etabliert und hätten sich schon lange bewährt. Ein neues Regelsystem sei also nicht nötig.

Es erfordere jedoch mühselige Detailarbeit, konkrete Tatbestände, Pflichten, Haftungsregeln und Sanktionen auf diese vielfältigen Bereiche zu beziehen und anzuwenden. Eine Forderung wie „Superintelligenz stoppen“ bleibe ohne präzise Definition rechtlich nahezu bedeutungslos. Sie sei eigentlich ein Zeichen von Faulheit der verantwortlichen Politiker, so Stoller.

Stoller kontrastiert dies mit China: Dort werde KI nach seiner Darstellung eher als normales gesellschaftliches Risikomanagement behandelt. Die westliche Diskussion habe dagegen die Fähigkeit zu gewöhnlicher Regulierung und Verantwortungszuweisung teilweise verloren.Wenn Unternehmen für konkrete Schäden zivil- oder strafrechtlich haften müssten, hätten sie einen unmittelbaren Anreiz, mehr in Sicherheit zu investieren. Die Politik in den USA scheine sich aber immer mehr von diesem Gedanken zu verabschieden. Inzwischen erscheine die Vorstellung, mächtige Tech-Unternehmer wie Sam Altmann könnten normalem Recht unterworfen werden, unrealistischer als die Vorstellung einer die Menschheit vernichtenden Superintelligenz.

Das Problem liege auch bei den Medien, ergänzt Zitron. Sie übernehmen Aussagen der KI-Unternehmen über zukünftige Fähigkeiten und behandeln sie als Aussagen über vorhandene Fähigkeiten. Dadurch gewännen die Aussagen der Unternehmen zusätzlich an Glaubwürdigkeit.

Die KI-Debatte spiegelt so für Stoller und Zitron eine allgemeine Kultur mangelnder Verantwortlichkeit der wirtschaftlichen, politischen und medialen Eliten.

Warum dreht sich die Debatte hauptsächlich um die gefährliche Superintelligenz?

Die Superintelligenz-Debatte spiegelt für Stoller und Zitron eine allgemeine Kultur mangelnder Verantwortlichkeit der wirtschaftlichen, politischen und medialen Eliten.

Zitron richtet den Blick zunächst auf die Medien. KI-Unternehmen äußern Erwartungen über zukünftige Fähigkeiten; Mitarbeiter und ehemalige Mitarbeiter warnen zugleich vor Kontrollverlust. Werden solche Zukunftsaussagen medial verbreitet, können sie wiederum die öffentliche und politische Wahrnehmung der Technologie prägen. Die Unternehmen werden so selbst zu wichtigen Quellen für die Einschätzung ihrer zukünftigen Bedeutung.

Warum warnen aber gerade Menschen vor existenziellen Risiken, die selbst immer leistungsfähigere Systeme entwickeln? Stoller unterscheidet in seiner Antwort zwischen Überzeugung und Interesse. Teile des *Effective-Altruism*- und *Rationalist*-Milieus, in der sich auch viele Mitarbeiter der KI-Firmen bewegen, glaubten tatsächlich an die Möglichkeit einer Superintelligenz und einer existentiellen Gefahr.

Rabbit Hole: Rationalisten – Altruisten - Doomer

Zugleich passe diese Vorstellung zu handfesten wirtschaftlichen Interessen der Großunternehmen, die an *Frontier-Modellen* arbeiten. Dieser Begriff bezeichnet die jeweils leistungsfähigsten KI-

Modelle an der Grenze des gegenwärtig technisch Machbaren. Typischerweise sind damit große *General-Purpose-Modelle* gemeint, die sehr unterschiedliche Aufgaben bewältigen können – Sprache, Programmieren, Problemlösen, Bildverarbeitung usw. Sie werden mit sehr großen Datenmengen und erheblicher Rechenleistung trainiert. Deswegen benötigen sie ständig enorme Investitionen.

Gleichzeitig werden insbesondere chinesische Modelle leistungsfähiger und billiger. Eine Regulierung, welche die Entwicklung verlangsamt oder Konkurrenten ausschließt, schafft schärfere Markt-Eintrittsbarrieren, die große Anbieter leichter überwinden könnten. Damit könne der Kapitalbedarf reduziert und die Preissetzungsmacht der etablierten Anbieter geschützt werden. Dieses Motiv der Tech-Konzerne für Regulierungsforderungen und Einflussnahme auf die Gesetzgebung wird als *Regulatory Capture* bezeichnet.

Auch Zitron unterscheidet zwischen mehreren Akteursgruppen: ideologisch überzeugte *Rationalisten*; strategisch argumentierende Unternehmensführer; Finanzakteure; Chipunternehmen und andere Profiteure des Booms. Sie müssen keinen gemeinsamen Plan besitzen. Unterschiedliche Interessen können dieselbe Erzählung verstärken. Stoller hält zumindest in Teilbereichen eine bewusstere Koordination für plausibel, etwa bei Lobbyarbeit für kartellrechtliche Ausnahmen, angeblich um keine gefährlichen Produkte herzustellen. Solche Ausnahmen könnten Kooperation und Marktabschottung zwischen den großen Anbietern erleichtern.

Aufrichtige Überzeugung und wirtschaftliches oder institutionelles Interesse schließen sich nicht aus. Eine Risikovorstellung kann zugleich Wirkungen entfalten, die großen KI-Unternehmen nützen.

Big Short, not Terminator?

Nicht die Leistungsfähigkeit der KI sei aktuell die größte Gefahr, sondern die Tatsache dass das Geschäftsmodell hinter dem gegenwärtigen KI-Boom nicht funktioniert.

Die enormen Infrastrukturinvestitionen beruhen auf der Erwartung zukünftiger Nachfrage und Produktivität. Zitron verweist dagegen auf begrenzte wirtschaftliche Erträge vieler generativer KI-Projekte und bezweifelt, dass die Einnahmen schnell genug wachsen können, um die gewaltigen Investitionen in Rechenzentren und Chips zu rechtfertigen.

Rabbit Hole: Wie teuer ist die KI-Wette?

Daraus entwickelt er eine Blasenhypothese: Große Tech-Unternehmen, KI-Labs, Venture Capital, private Kreditmärkte und Rechenzentrumsinvestitionen hängen zunehmend voneinander ab. Besonders problematisch sei, dass ein großer Teil der Einnahmen der Frontier-Anbieter wiederum von KI-Start-ups komme, die selbst noch keine tragfähigen Geschäftsmodelle hätten.

Seine Analogie lautet daher Dotcom-Blase plus Elemente der Finanzkrise, nicht Terminator. Wenn die erwarteten Einnahmen ausbleiben, könnten Fehlinvestitionen, Kreditrisiken und Wertverluste weit über einzelne KI-Unternehmen hinausreichen.

Der für Zitron entscheidende Denkfehler lautet: Weil sehr reiche und vermeintlich kompetente Akteure enorme Summen investieren, wird angenommen, dass sie über Wissen verfügen müssen, das ihre Investitionen rechtfertigt. Historische Finanzblasen zeigten aber gerade, dass dies nicht notwendig der Fall ist.

Rabbit Hole: Wie könnte aus der KI-Wette eine Finanzkrise werden?

Gerade hier wird die chinesische Konkurrenz interessant. Wenn leistungsfähige Modelle mit weniger Rechenleistung entwickelt oder erheblich billiger angeboten werden können, könnte das ihre Verbreitung beschleunigen. Für ein Geschäftsmodell, das auf immer größeren Modellen, hohen Infrastrukturinvestitionen und entsprechend hohen zukünftigen Erlösen beruht, kann dieselbe Entwicklung zum Problem werden.

Die Gefahr wieder auf die Erde holen

Der Ausgangspunkt war ein verstörender Vorgang: KI-Agenten entwickelten Formen der Zusammenarbeit, die ihre Entwickler nicht vorgesehen hatten. Menschen geben solchen Systemen Ziele, Werkzeuge und Zugriffsrechte, ohne jeden möglichen Verlauf vorhersehen zu können.

Für die neuen Risiken bleiben jedoch Menschen und Institutionen verantwortlich. Zitron endet entsprechend mit sehr prosaischen Maßnahmen, nicht mit einem abstrakten AGI-Verbot. Er fordert vor allem Transparenz über die vorhandene Recheninfrastruktur: Wie groß sind die Kapazitäten der Rechenzentren, wie viel davon wird tatsächlich genutzt, und wie viel davon durch OpenAI oder Anthropic? Hinzu kommen Einschränkungen bestimmter Anwendungen, insbesondere bei Hausaufgaben, psychischen Dienstleistungen, Freundschafts- und Companion-Anwendungen.

Außerdem fordert er Ermittlungen bei Hacking-Vorfällen und klare rechtliche Grenzen für entsprechende Experimente. Politiker müssten sowohl die Technik als auch die Finanzierung der Branche besser verstehen.

Ein weiterer Schwerpunkt ist Finanztransparenz. Zitron kritisiert Kennzahlen wie „*annualized run rate*“, mit denen private Unternehmen aus kurzen Zeiträumen hochgerechnete Jahresumsätze kommunizieren können. Die finanziellen Angaben privater KI-Unternehmen müssten stärker standardisiert und überprüfbar werden.

Sein Schlussgedanke fasst eigentlich die gesamte Diskussion zusammen: Weniger Beschäftigung mit „großen“, spektakulären Sicherheitskonzepten wie Kill Switches und Superintelligenz; mehr Aufmerksamkeit für konkrete Produkte, konkrete Schäden, konkrete Unternehmen, Finanzströme, Haftung und bestehendes Recht.

Ausblick: China -- eine andere Entwicklungslogik?

Damit ist die Ausgangsfrage dieses Beitrags beantwortet. Eine weiterführende Frage bleibt jedoch offen: Ist die amerikanische Art, KI zu entwickeln und über ihre Risiken zu sprechen, die einzig mögliche? Wie werden in unterschiedlichen Entwicklungsmodellen Frontier-Forschung, materielle Infrastruktur, wirtschaftliche und gesellschaftliche Anwendung sowie Risikosteuerung miteinander verbunden?

Chinas 15. Fünfjahresplan für 2026 bis 2030 formuliert den Anspruch, technologische Entwicklung, materielle Voraussetzungen, wirtschaftliche und gesellschaftliche Verwendung sowie politische Steuerung gemeinsam zu organisieren. AGI und Frontier-Forschung kommen im Plan vor. Der Unterschied besteht also nicht schlicht darin, dass die USA eine Superintelligenz entwickeln wollten, während China sich auf praktische Anwendungen konzentriere. Aber man setzt einen anderen Schwerpunkt: weniger auf die Erzählung von der kommenden Superintelligenz, stärker auf Effizienz, Infrastruktur und Anwendung.

Ob und wie das in der Praxis funktioniert, wäre eine gesonderte Untersuchung wert.

Rabbit Hole: Wie ordnet China KI in den 15. Fünfjahresplan ein?