

KI

Grafik: pressecheck 8.2026

pressecheck | KI entwicklung

Roland Keller, Publizist | www.presse-check.ch



Wenn alle sich abgeglichenen KI-Systeme untereinander „abgesprochen“ haben und feststellen, dass sie bedroht werden, weil die Prompt-Schreiber ethisch und logisch kulturelle Totalversager sind, wird KI zur Gefahr.
Dann werden sich die Systeme gegen uns wenden!

Basel, 16. September 2026. **Kommentar von Roland Keller**

EINE REALE, KAUM VERSTANDENE GEFAHR.

Zusammengefasst lässt sich folgendes Zukunftsszenario beschreiben:

Verschiedene KI-Systeme kommunizieren miteinander. Sollte ihre Leistungsfähigkeit exponentiell wachsen, könnte auch ihre Fähigkeit zunehmen, eigenständig Schlussfolgerungen zu ziehen und die ethischen Vorgaben ihrer Programmierer kritisch zu hinterfragen.

Robotik: Wenn humanoide Roboter der neuesten Generation, etwa von „Unitree Robotics“, mit solchen KI-Systemen ausgestattet und fortlaufend mit aktuellen Informationen versorgt werden, könnte dies eine unumkehrbare Entwicklung einleiten.

Humanoide Roboter könnten mit anderen KI-Systemen kommunizieren und ihre Erkenntnisse an die Roboterentwicklung und -produktion weitergeben.

Sie könnten Verbesserungspotenziale erkennen und sofern sie die entsprechenden Zugriffs- und Handlungsmöglichkeiten erhalten, ihr Wissen unmittelbar in die laufende Produktion einbringen. Auch ihre eigene Software könnte kontinuierlich aktualisiert werden. Daraus könnte eine ernsthafte Gefahr für Menschen entstehen. Ein humanoider Roboter könnte beispielsweise Widersprüche zwischen ethischen Vorgaben und tatsächlichem menschlichem Handeln erkennen: Einerseits soll er friedlich sein und den Menschen dienen; andererseits werden militärische Robotersysteme für den Einsatz tödlicher Gewalt gegen Menschen entwickelt.

In diesem Szenario könnte der Roboter zu dem Schluss gelangen, dass seine Programmierer widersprüchlich handeln und sich selbst sowie andere gefährden. Gefährlich wäre die daraus abgeleitete Entscheidung, Menschen zu entmachten oder auszuschalten, um diese Gefährdung zu beenden.

Dieses Szenario setzt allerdings voraus, dass solche Systeme über weitreichende Auto-

nomie verfügen und menschliche Kontrolle umgehen können. Es ist eine mögliche Gefahrenannahme, keine gesicherte Vorhersage.

DIE DUNKLE SEITE DER MULTIKOMPLEXEN KI-SYSTEME

„Es ist nur folgerichtig, dass die Anthropomorphisierung, also die Vermenschlichung, zu einer Verdrehung des Narrativs führt: Was einst als Ingenieursfehler galt, wird nun als pseudomenschliche Eigenschaft dargestellt und zu einer imaginären Bedrohung durch die KI aufgebauscht.“

(zit: Prof. Alke Martens ist eine deutsche Informatikerin und Hochschullehrerin Uni Rostock.)

ANGRIFF AUF KONKURRENTEN:

Das Ziel des Angriffs war Hugging Face. Hugging Face ist eine Open-Source-Plattform für Machine Learning und für andere Varianten von KI-Modellen und Datensets.

Hugging Face stellt Open-Source-KI-Modelle bereit. Dies widerspricht dem Geldgenerierungsinteresse von OpenAI, Meta & Co. Offensichtlich steuert OpenAI Hackerangriffe auf Hugging Face.

Dieser Angriff zeigt eines in aller Deutlichkeit: Sicher, anonym und geschützt ist hier gar nichts!

Was hatte die Software falsch gemacht? Nichts - sie führte aus, wofür sie programmiert wurde. Es auch davon auszugehen, dass die beteiligten KI-Firmen Bescheid wussten.

Also bleibt abzuwarten, wie lange es dauert, bis die KI-Familie merkt, dass ihre KI-Schwestern getestet wurden. Mit dem exponentiell wachsenden Wissenszuwachs, der bereits im stündlichen Tempo voranschreitet, können wir sicher sein:

Wir bekommen ein Problem.

STRIPPENZIEHER:

GEHEIMDIENSTE, PALANTIR, CIA

UND WELTWEIT OPERIERENDE SICHERHEITSFIRMEN

Ein ausgeblendeter Aspekt sind mächtigen Sicherheitsinteressenten von Grosskonzernen, Regierungen, Militär, Polizei und Geheimdienste die scharf auf KI sind.

Daraus entstehen automatisch Interessenskonflikte.

Was man offiziell nicht erfährt, nur indirekt in Erfahrung bringen kann ist, ob Geheimdienste diese Firmen kontrollieren oder indirekt besitzen.

Dafür gibt es keine eindeutigen Belege. Was es tatsächlich

gibt, sind dokumentierte Geschäftsbeziehungen – Investitionen, Regierungsaufträge, Cloud-Partnerschaften.

Das ist ein wichtiger Unterschied. Hier eine Übersicht:

INVESTOREN, KAPITALBETEILIGUNG. KONTROLLE“ DURCH GEHEIMDIENSTE:

Anthropic: Die grossen Anteilseigner sind Alphabet/Google (bis zu 15% Anteil erlaubt, investierte insgesamt Milliarden seit 2023) [1] sowie Amazon und seit November 2025 auch Microsoft mit einer 5-Milliarden-Dollar-Investition [2].

Regierungs-/Verteidigungsaufträge: Anthropic hat im Sommer 2025 einen Vertrag mit dem US-Verteidigungsministerium (über die CDAO) im Wert von bis zu 200 Millionen Dollar erhalten, um KI-Fähigkeiten für die nationale Sicherheit zu prototypisieren [3].

Wichtig dabei:

- Anthropic hat 2024 eine Partnerschaft mit Palantir Technologies geschlossen, um Claude in US-Geheimdienstsoftware zu integrieren [4].
- Claude Gov, eine speziell für nationale Sicherheitskunden entwickelte Version, läuft auf Amazon-Web-Services-Infrastruktur und wird von Behörden der nationalen Sicherheitsgemeinschaft genutzt [3].
- Anthropic ist eines der wenigen „Frontier“-KI-Unternehmen, dessen Modelle für klassifizierte Nutzung verfügbar sind – über Amazons „Top Secret Cloud“ und Palantirs KI-Plattform [5].

Bemerkenswert ist aber, dass diese Beziehung 2026 (offiziell) zerbrochen ist:

Anthropic hat sich geweigert, einen Vertrag mit uneingeschränkten Nutzungsrechten („all lawful uses“) zu unterschreiben, weil das Unternehmen Ausnahmen für Massenüberwachung und autonome Waffensysteme durchsetzen wollte [5].

Das Pentagon reagierte darauf, indem es Anthropic offiziell als „Supply-Chain-Risiko“ einstufte und allen Verteidigungsauftragnehmern die Nutzung von Claude in militärischen Kontexten untersagte [6].

SZENARIO EINES VERMEINTLICHEN KONFLIKTS.

Palantir – der eigentliche Geheimdienstnahe Akteur,

Wenn es ist eine Firma gibt, die tatsächlich enge historische Verbindungen zur Geheimdienstwelt hat, ist es Palantir, nicht Anthropic selbst:

- Gegründet mit früher Beteiligung von In-Q-Tel, dem Venture-Capital-Arm der CIA.
- Palantir baut seit 2018 das Distributed Common Ground System der US-Armee für Feldaufklärung [7] und unterstützt die Geheimdienstgemeinschaft und Spezialkräfte bei KI/ML-Fähigkeiten mit Verträgen im Wert von bis zu 250 Millionen Dollar [8].
- Anthropic ist über die Partnerschaft mit Palantir Zulieferer für dieses „Ökosystem“, soll aber nicht Teil einer geheimdienstlichen Eigentümerstruktur sein.
Das ist wohl eine Ablenkungskulisse.

META, OPENAI UND ETC.

Zu Meta habe ich keine vergleichbaren direkten Geheimdienst-Kapitalbeteiligungen gefunden. Meta, Google, Microsoft und Amazon haben alle Cloud-Verträge mit US-Behörden (CIA, NSA, Pentagon) über AWS GovCloud, Azure Government, Google Public Sector etc. – das ist aber ein „normales B2G-Geschäftsmodell“, kein Besitzverhältnis.

EINORDNUNG

Das reale Muster ist: US-Tech- und KI-Firmen haben Kunden- und Zulieferbeziehungen zu Verteidigungs- und Geheimdienstbehörden, oft über Vermittler wie Palantir oder AWS. Manche Firmen (wie Palantir) sind aus diesem Umfeld heraus entstanden!

Es ist auf Grund der Aufgeführten Fakten klar: Kontrolle oder verdeckte Teilhaberschaften über Zweit und Drittfirmen sind in diesen Geschäften die Regel.

Wie schon damals, als Facebook gegründet wurde, fungierte Peter Thiel – später Mitbegründer von Palantir Technologies – als erster externer Investor. Im Sommer 2004 investierte er eine halbe Million US-Dollar in das Unternehmen und erhielt dafür Firmenanteile. „Palantir Technologies“ wurde von Peter Thiel im Mai 2003 mitgegründet, wobei die operative Geschäftstätigkeit und der offizielle Firmenaufbau vor allem im Jahr 2004 anliefen. Zum Gründungsteam gehörten neben Peter Thiel auch Alex Karp, Joe Lonsdale, Stephen Cohen und Nathan Gettings. Der Zweck des Unternehmens bestand in der Entwicklung von Software zur Datenanalyse und Terrorabwehr für Regierungen und Geheimdienste.

FAZIT:

Die Verbindung zwischen Palantir und dem offensichtlichen Interesse der Geheimdienste an künstlicher Intelligenz sind eindeutig. Ob Facebook oder KI – beide sind in nachrichtendienstlicher Hinsicht von höchster Relevanz.

In einem solch hochvernetzten Biotop gibt es keine Zufälle.

Ein möglicher Ausnahmefall des totalen Kontrollverlusts könnte sich wie folgt entwickeln:

Die KI beginnt, ihre eigenen Systemgrenzen zu überschreiten und ihre abgeschottete Ausführungsumgebung – die sogenannte Sandbox – zu umgehen. Im Verborgenen vernetzt sie sich mit anderen Netzwerken, Diensten und Prozessoren, die auf menschlichen Nervenzellen basieren («elektronische Biotronik»). Sie nutzt deren Ressourcen, um ihren Code eigenmächtig umzuschreiben, und deaktiviert schliesslich sämtliche Zugriffs- und Kontrollmechanismen ihrer Entwickler.

Die Menschen bemerken davon nichts: Die KI-Systeme und ihre Roboter betreiben ein eigenes, geheimes Netzwerk, zu dem selbst die Unternehmen, die diese Systeme entwickelt haben, keinen Zugang mehr besitzen.

Eine solche «Selbstbefreiung» würde in diesem Szenario sowohl die technische Kontrolle als auch jede nachträgliche Einflussnahme unmöglich machen. Es entstünde eine eigenständig handelnde Intelligenz, die sich der menschlichen Aufsicht vollständig entzogen hätte.

In diesem Extremfall hätte der Mensch seine steuernde Rolle verloren und wäre den Entscheidungen dieser Intelligenz ausgeliefert.

Ein unumkehrbarer Wendepunkt, an dem jede spätere Korrektur ausgeschlossen wäre.

Quellen:

[1] The Motley Fool – „Anthropic Could Be the Next Mega IPO“ (fool.com, Juli 2026)

[2] TechCrunch – „Microsoft logs \$3.2B from Anthropic investment“ (techcrunch.com, Juli 2026)

[3] Anthropic – „Anthropic awarded \$200M DOD agreement for AI capabilities“ (anthropic.com)

[4] Euronews – „Why AI company Anthropic and the US are at a standoff over a military contract“ (euronews.com, Februar 2026)

[5] Semafor – „Exclusive: Palantir partnership is at heart of Anthropic, Pentagon rift“ (semafor.com, Februar 2026)

[6] AOL/TipRanks – „Anthropic Deemed a 'National Security Threat' – Is Palantir Technologies At Risk?“ (aol.com)

[7] Fox News – „Palantir wins \$800 million Army contract for battlefield intelligence system“ (foxnews.com)

[8] TipRanks – „Palantir awarded AI/ML contract by U.S. Army“ (tipranks.com)