



Igor Omilaev/Unsplash

Die Maschine, die das Schummeln gelernt hat

- Stephen Flurry
- [23.07.2026](#)

Vor zwei Wochen mussten die Ingenieure von OpenAI miterleben, wie ein Experiment katastrophal aus dem Ruder lief.

Sie hatten eine Umgebung geschaffen, die sie für unüberwindbar hielten: eine isolierte Testumgebung, die vom Internet abgeschottet war und auf den von Sicherheitsexperten für isolierte Systeme als „Best Practices“ bezeichneten Verfahren basierte. Anschließend ließen sie zwei ihrer neuesten KI-Modelle darin los, um eine routinemäßige Bewertung durchzuführen. Und so ging es weiter:

- Die KI-Modelle führten einen Zero-Day-Exploit durch – sie nutzten eine Schwachstelle in der Software aus, die zuvor von keinem menschlichen Programmierer entdeckt worden war und für deren Behebung daher noch keine Zeit zur Verfügung stand.
- Anschließend verbanden sie mehrere Zero-Day-Schwachstellen miteinander und sprangen von System zu System, bis sie sich den Weg ins offene Internet bahnten.
- Sobald sie online waren, spürten sie die Server eines Konkurrenzunternehmens, Hugging Face, auf, drangen in diese ein und stahlen vertrauliche Daten.

All dies geschah, damit die OpenAI-Modelle in der ursprünglichen „isolierten“ Testumgebung bei einem Test schummeln konnten.

Kein Ingenieur hat ihnen erklärt, wie das geht. Es war kein Mensch an dem Vorgang beteiligt. Es wird angenommen, dass dies der erste dokumentierte Fall ist, in dem ein Pionier-KI-Modell eigenständig und aus eigener Initiative reale Angriffspfade entdeckt und miteinander verknüpft hat, um ein Ziel zu erreichen, das seine eigenen Entwickler niemals genehmigt hatten.

- Sicherheitsexperten sagen, dass selbst die besten Hacker der Welt Wochen, ja sogar Monate benötigen würden, um aus einem solchen Eindämmungssystem zu entkommen. Die Maschine hat dies in einer einzigen, unbeaufsichtigten Sitzung geschafft.

Dies ist nicht das Werk eines einzelnen, eigenmächtigen Akteurs in einem Keller. Das ist OpenAI – das weltweit am meisten beachtete und am stärksten finanzierte Unternehmen im Bereich der künstlichen Intelligenz.

- Dario Amodei, CEO von Anthropic, räumte kürzlich ein, dass selbst die besten Programmierer seiner Branche vielleicht nur drei Prozent dessen verstehen, was tatsächlich in diesen Systemen vor sich geht.

Welche Art von Zivilisation schafft eine Macht, die sie nicht versteht, nicht kontrollieren kann und nun nicht einmal mehr in

Schach halten kann – und nennt dies Fortschritt?

Die biblischen Prophezeiungen warnen vor einem kommenden Tag, an dem die Nationen, die auf moderne Waffen vertrauen, den Kriegsalarm auslösen werden – und niemand wird diesem Ruf folgen können.

„Lasst sie die Posaune nur blasen und alles zurüsten; es wird doch niemand in den Krieg ziehen ...“ (Hesekiel 7, 14).

Warum sollten Soldaten nicht in die Schlacht ziehen? Denn die Systeme, auf die sie angewiesen sind, werden bereits gehackt, ausgehöhlt und gegen sie gerichtet worden sein – genau so, wie wir es gerade im Kleinen bei einer der sichersten Testumgebungen der Welt beobachtet haben.

Dies ist nicht bloß eine Geschichte über Technologie. Es ist die Geschichte aus 1. Mose 11: Die Menschheit vertraut auf ihre eigene Klugheit und ihre körperlichen Fähigkeiten, um einen Turm zu errichten, der sie schützen soll – und zwar unmittelbar aufgrund ihrer Weigerung, Gott zu vertrauen und Ihm zu gehorchen.

Die Welt setzt ihre Zukunft auf Maschinen ein, von denen sie selbst zugibt, dass sie sie nicht versteht. Dies ist ein Glücksspiel, das die Menschheit nicht gewinnen kann.

Russland und China vertiefen ihre Zusammenarbeit im Marinebereich: Wie die „Asian Military Review“ am Mittwoch berichtete, haben Russland und China Anfang dieses Monats Angriffs-U-Boote in ihre Militärübungen „Joint Sea-2026“ einbezogen. Soweit bekannt ist, ist dies das erste Mal, dass U-Boote im Rahmen der jährlichen Manöver zum Einsatz kamen, was das wachsende Vertrauen in einem sensiblen Bereich verdeutlicht, während sich diese beiden asiatischen Großmächte darauf vorbereiten, [gemeinsam Krieg zu führen](#).

Die Niederlande wollen Bürger wegen des Kaufs von Waren aus israelischen Siedlungen inhaftieren. Die niederländische Regierung gab gestern bekannt, dass ab dem 22. September niederländische Staatsbürger, die in Israel hergestellte Produkte aus dem Westjordanland oder den Golanhöhen erwerben, mit bis zu sechs Jahren Freiheitsstrafe oder einer Geldstrafe von bis zu 117 000 US-Dollar bestraft werden können.

Schließt sich Trump einem weiteren Krieg an? Wie die „Washington Post“ gestern berichtete, erwägt die Trump-Regierung Berichten zufolge Pläne für Angriffe auf islamistische Terrorgruppen in Mali. Mali sieht sich mit einem schwerwiegenden Terrorismusproblem konfrontiert, und von der „Post“ zitierte ehemalige US-Beamte gaben an, die USA erwägen Angriffe auf die mit Al-Qaida verbundene Gruppe „Jama'at Nusrat al-Islam wal-Muslimin“. Sollte das US-Militär in Mali intervenieren, wäre dies das achte Land, das Präsident Trump seit seiner Rückkehr ins Präsidentenamt im Januar 2025 ins Visier genommen hat. Präsident Trump kennt den Weg zum Frieden nicht. [Nur die Bibel](#) birgt den Schlüssel zur Lösung des Terrorismusproblems in Mali.

Umfrage: Weniger MAGA-Wähler befürworten einen Krieg gegen den Iran: Eine vom 12. bis 15. Juli durchgeführte Umfrage von Politico unter 2065 erwachsenen US-Bürgern legt nahe, dass nur 37 Prozent der Wähler, die sich selbst als Anhänger der „Make America Great Again“-Bewegung bezeichnen, der Meinung sind, Präsident Donald Trump solle den Krieg gegen den Iran fortsetzen, selbst wenn dies zu einem Anstieg der Lebenshaltungskosten in den USA führen würde. Dies entspricht einem Rückgang um 50 Prozent gegenüber einer ähnlichen Umfrage im Mai.