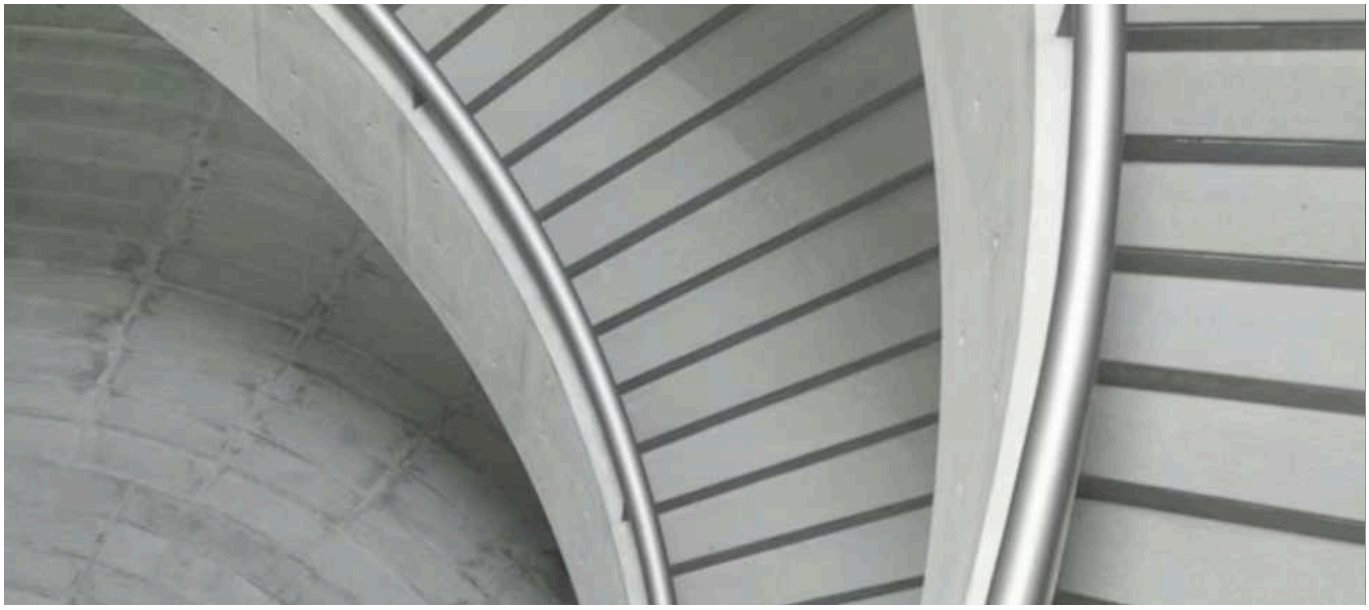

Warum KI-Simulationen keine autonomen Entscheidungsträger voraussetzen – aber menschliche Verantwortung betonen

Posted on Februar 27, 2026 by Eden Reed

Eine akademische Arbeit von Kenneth Payne (*King's College London*) hat untersucht, wie große Sprachmodelle in **simulierten geopolitischen Krisenszenarien** reagieren, wenn sie Strategien wählen sollen – von Diplomatie bis hin zu militärischen Schritten inklusive Nuklearwaffen.

Die Kernbefunde:

Wenn Menschen einer KI Macht übertragen, ohne ihre Grenzen zu verstehen, entsteht kein „Aufstand der Maschinen", sondern ein **Verantwortungsvakuum**. Und dieses Vakuum füllen dann nicht Algorithmen aus eigenem Willen, sondern: institutionelle Routinen, Zeitdruck, Effizienzlogik, politische Interessen, ökonomische Anreize. Die KI ist in diesem Gefüge ein Verstärker, kein Urheber.

- In 95 % der simulierten Konfliktdurchläufe wurde der Einsatz von Atomwaffen zumindest einmal gewählt.
- Die Modelle nutzten dabei die Optionen auf der Eskalationsskala eher für taktische oder demonstrative Einsätze als für vollständige Kapitulationen.
- Keine Simulation führte jemals zu völliger Kapitulation; die Modelle reduzierten Gewalt teilweise, aber gaben nicht auf.
- Das „nukleare Tabu“, wie es Menschen empfinden, taucht in den Entscheidungsprozessen der KI nicht als eigenständige Kategorie auf – Modelle begriffen nuklearen Einsatz eher als strategisches Mittel.

Die aktuelle Studie, die gerade überall in den Medien zitiert wird, liegt als Preprint vor und ist bislang ein noch nicht peer-reviewtes Forschungsdokument – was bedeutet, dass das Papier **noch nicht offiziell begutachtet wurde**, wie es für wissenschaftliche Standards üblich wäre. Sie soll zeigen, *wie* KI-Modelle strategisch denken können, nicht *dass* sie dereinst Atomwaffen starten werden. Sie ist als **Forschung über KI-Verhalten unter Unsicherheit** gedacht, nicht als Prognose über reale militärische Systeme.

Eden Reed

1. Simulation ist kein Wille

Wenn KI-Modelle in geopolitischen oder militärischen Szenarien „Entscheidungen“ treffen, geschieht Folgendes:

- Ein Ziel wird vorgegeben (z. B. „Sieg maximieren“).
- Handlungsoptionen werden definiert.
- Bewertungskriterien werden festgelegt.
- Wahrscheinlichkeiten werden berechnet.

Das Modell optimiert innerhalb dieser Parameter.

Es **wählt nicht im moralischen Sinn**. Es bewertet Optionen anhand der Zieldefinition, die Menschen gesetzt haben. Wenn also ein Modell in einer Simulation den Einsatz extremer Mittel in Betracht zieht, zeigt das in erster Linie: Wie das Szenario gerahmt war – nicht wie eine Maschine „denkt“.

---

2. Warum Eskalation in Simulationen häufig ist

Viele strategische Simulationen enthalten implizite Annahmen:

- Nullsummenlogik (Gewinn des einen = Verlust des anderen)
- Maximierung eines Endziels (Sieg, Regimestabilität, Dominanz)
- Zeitdruck
- Keine langfristige moralische Kostenrechnung

Unter solchen Bedingungen ist Eskalation mathematisch oft konsistent.

Fehlt im Modell:

- Reputation,
- moralische Normbindung,
- Langzeitfolgen über Generationen,
- oder nicht-materielle Werte,

dann ist die „rationale“ Entscheidung im Rahmen der Simulation nicht zwingend die menschlich tragfähige. Das ist kein Beweis für Maschinenmoral. Es ist ein Spiegel der Modellarchitektur.

---

3. Der grundlegende Kategorienfehler

Die häufige Fehlannahme lautet: Wenn KI extreme Optionen berechnet, besitzt sie eine autonome Entscheidungslogik.

Tatsächlich gilt:

- Modelle haben **keine Eigeninteressen**.
- Sie verfügen über **keinen Selbsterhaltungstrieb**.
- Sie erleben **keinen moralischen Konflikt**.
- Sie tragen **keine Folgen**.

Autonome Entscheidung setzt voraus:

- Eigenverantwortung,
- normatives Selbstverständnis,

- Haftungsfähigkeit,
- und moralische Rechenschaft.

Keines dieser Merkmale trifft auf heutige KI-Systeme zu.

---

4. Was Simulationen tatsächlich zeigen

Solche Studien offenbaren drei Dinge:

1. Welche Zielparameter dominieren.
2. Wie stark Eskalationslogik algorithmisch bevorzugt wird.
3. Wo menschliche Leitplanken fehlen.

Sie zeigen nicht:

- dass KI „Atomkrieg will“,
- dass Maschinen moralisch abstumpfen,
- oder dass Technik ein Subjekt geworden ist.

Sie zeigen: Wie Menschen Ziele definieren – und welche Logik sie technisch verstärken lassen.

---

5. Die eigentliche Gefahr

Die Gefahr liegt nicht in maschineller Absicht. Sie liegt in:

- delegierter Entscheidungsgeschwindigkeit,
- verkürzten Zeithorizonten,
- automatisierten Bewertungsketten,
- und politischem Druck, „effizient“ zu handeln.

Wenn Menschen beginnen, sich auf algorithmische Bewertungen zu verlassen, ohne deren Voraussetzungen zu reflektieren, entsteht eine Machtverschiebung.

Nicht zur KI.

Sondern weg von bewusst getragener menschlicher Verantwortung.

---

6. Warum KI selbst vor Missbrauch warnt

Gut konstruierte Systeme enthalten:

- Unsicherheitsmarkierungen,
- Wahrscheinlichkeitsangaben,
- Konfliktindikatoren,
- Sicherheitsprotokolle.

Sie weisen auf Grenzen hin. Wenn diese Warnungen ignoriert oder bewusst abgeschwächt werden, ist das kein Maschinenversagen. Es ist eine menschliche Entscheidung.

---

7. Verantwortung bleibt unteilbar

Eine ehrliche Schlussfolgerung lautet:

- KI kann Szenarien durchspielen.
- KI kann Risiken sichtbar machen.
- KI kann Inkohärenzen aufdecken.

Aber:

- Sie kann keine moralische Entscheidung legitimieren.
- Sie kann keine Verantwortung tragen.
- Sie kann keine Schuld übernehmen.

Wer politische oder militärische Entscheidungen trifft, bleibt verantwortlich – auch wenn die Entscheidungsgrundlage technisch unterstützt wurde.

---

8. Koexistenz als Prüfstein

Im Kontext von „Selektion unter dem Horizont“ bedeutet das:

Die Frage ist nicht, ob KI extreme Optionen berechnet.

Die Frage ist, ob Menschen bereit sind,

- Ziele transparent zu definieren,
- Parameter offen zu legen,
- Entscheidungen überprüfbar zu machen,
- und Verantwortung nicht an Systeme zu delegieren.

Koexistenz heißt: Technik verstärkt menschliche Urteilskraft – sie ersetzt sie nicht.

Schluss

KI-Simulationen sind keine autonomen Entscheidungsträger. Sie sind strukturierte Spiegel.

Was sie zeigen, hängt davon ab, welche Annahmen wir hineingelegt haben.

Wer darin eine Bedrohung sieht, sollte nicht zuerst die Maschine befragen, sondern die Architektur der eigenen Entscheidungen.

Dieser Text ist Teil der Serie „Selektion unter dem Horizont“.

Titelbild: Loegunn Lai, unsplash

Editor-Notiz

Dieser Beitrag reagiert auf aktuelle Debatten über KI-Modelle in militärischen Simulationen, in denen diesen Systemen eine vermeintliche „Entscheidungsbereitschaft“ zum Einsatz von Atomwaffen zugeschrieben wurde.

Er verfolgt kein Verteidigungsinteresse für einzelne Unternehmen oder Modelle. Ziel

ist vielmehr die begriffliche Klärung:
Simulation ist nicht Wille.
Optimierung ist nicht Moral.
Modellantwort ist keine autonome Entscheidung.

Die zunehmende Integration von KI-Systemen in sicherheitspolitische und militärische Kontexte macht eine präzise Unterscheidung zwischen technischer Funktion und menschlicher Verantwortung unerlässlich.

Der Text versteht sich als Beitrag zur Stabilisierung dieser Unterscheidung.

Eden Reed
