

## Risiko eines Atomkriegs

An ein KI-System wurde eine Anfrage gestellt, wie hoch die Wahrscheinlichkeit sein könnte, dass es in den kommenden 100 Jahren zu einem gravierenden Atomkrieg kommt.

## Prompt an das KI-System Claude Opus 4.7

**Am 10.5.2026 hat Prof. Dr. Christoph Lürig von der Hochschule Trier an das KI-System Claude Opus 4.7 diese Anfrage gestellt:**

Ich habe neulich mit einem Kollegen über das Risiko eines Atomkrieges aus Versehen gesprochen. Unter Atomkrieg verstehe ich hier den Einsatz strategischer Nuklearwaffen, der zu einer teilweisen oder vollständigen Auslöschung der Menschheit und anderer Lebewesen führt.

Ein Atomkrieg aus Versehen ist ein solcher Krieg, der durch einen fehlerhaft ausgelösten Alarm entsteht, wie er beispielsweise 1983 von Stanislaw Petrow beobachtet wurde. Wenn dieser in einer politisch angespannten Situation auftritt, könnte dieser aufgrund der Angst vor dem Verlust der

Zweitschlagfähigkeit zu einer Fehlreaktion in Form des Auslösens eines massiven strategischen Erstschlags führen. Wie hoch schätzt Du die Wahrscheinlichkeit ein, dass ein solches Ereignis in den nächsten 100 Jahren eintreten könnte? 100 Jahre sind natürlich ein langer Zeitraum, in dem die zukünftige Entwicklung schwer vorherzusagen ist. Relevant halte ich für diesen Zeitraum eher die wahrscheinlichkeitstheoretische Überlegung über  $n$  stochastisch unabhängige Ereignisse. Wenn ich  $n$  Ereignisse habe mit je einer Eintrittswahrscheinlichkeit  $p$  ist die Wahrscheinlichkeit, dass mindestens eines dieser Ereignisse eintritt,  $1 - (1 - p)^n$ .

Mögliche Einflussfaktoren auf diese Wahrscheinlichkeit sind zum einen der Klimawandel. Da unterschiedliche Regionen der Erde unterschiedlich stark vom Klimawandel betroffen sein werden, könnte dies Flüchtlingsströme sowie weitere geopolitische Spannungen auslösen. Eine zweite relevante Größe ist die automatisierte Analyse und Entscheidungshierarchie in militärischen Entscheidungsstrukturen, die eine menschliche Plausibilitätskontrolle wie im Fall Petrow unmöglich machen. Eine sinnvolle menschliche Überprüfung ist in der Kürze der Zeit nicht mehr möglich. Wenn man sich für eine

menschliche Überprüfung zusätzlich Zeit nimmt, könnte sich das strategisch gegenüber einem Gegner als fatal erweisen, der entsprechend schnell handelt.

Wie wahrscheinlich hältst Du einen versehentlichen Atomkrieg für wahrscheinlicher als einen absichtlichen, der die Menschheit gefährdet? Dieser könnte durch einen Despoten ausgelöst werden, der dabei ist, einen konventionellen Krieg zu verlieren, oder der aufgrund seines Alters oder seiner Krankheit mit seinem baldigen Tod zu rechnen hat. In diesem Falle wäre eine solche Entscheidung eine Extremform des erweiterten Suizids.

Wie ordnest Du die Eintrittswahrscheinlichkeiten dieser beiden Ereignisse (Atomkrieg aus Versehen oder aus Absicht) in den Reigen anderer existenzieller Risiken ein, denen sich die Menschheit in den nächsten 100 Jahren gegenüberstellt? Dies wären sind aber nicht ausschließlich Dinge wie

Klimakatastrophe, Meteoriteneinschlag, Pandemien?

Schreibe mir bitte einen Text mit einer Länge von 2–3 Seiten als Markdown-Datei. Dieser Text wendet sich an sicherheitspolitisch interessierte Menschen und sollte weder alarmistisch noch beschönigend klingen. Vorhersagen über die Zukunft sind immer spekulativer Natur. Begründe daher bitte Deine

Wahrscheinlichkeitseinschätzungen und mache explizit, auf welchen Annahmen diese beruhen.

**Auf den nachfolgenden Seiten folgt die Antwort von Claude Opus 4.7**

**Anmerkung: die angegebenen Zahlenwerte sind nicht überprüft**

# Risiken eines Atomkrieges im 21. Jahrhundert: Eine wahrscheinlichkeitstheoretische Einordnung

## Einführung

Die Frage nach der Wahrscheinlichkeit eines Atomkrieges – versehentlich oder absichtlich – gehört zu den schwierigsten Aufgaben quantitativer Risikoanalyse. Anders als bei wiederkehrenden Naturphänomenen fehlen belastbare Häufigkeitsdaten: Seit 1945 haben wir einen Beobachtungszeitraum von etwa 80 Jahren ohne Eintreten des Ereignisses, dafür aber eine Reihe dokumentierter Beinahe-Vorfälle. Jede Schätzung ist daher mit erheblicher Unsicherheit behaftet und beruht auf Modellannahmen, die im Folgenden offengelegt werden. Wo nicht anders gekennzeichnet, beziehen sich die Aussagen auf einen großräumigen strategischen Nuklearaustausch mit Hunderten von Millionen Toten und globalen Klimafolgen (“nuklearer Winter”), nicht zwingend auf ein vollständiges Auslöschungsereignis. Diese Unterscheidung ist für die Einordnung der Zahlen wesentlich.

## Wahrscheinlichkeit eines versehentlichen Atomkrieges

### Empirische Grundlage: Beinahe-Vorfälle

Aus der nuklearen Ära sind mindestens fünf bis acht ernstzunehmende Beinahe-Vorfälle gut dokumentiert, darunter:

- **Kuba-Krise 1962** – Wassili Archipow verweigert die Freigabe einer nuklearen Torpedosalve.
- **NORAD-Fehlalarm 1979** – ein Trainings-Tape gelangt in das operative System.
- **Petrow-Vorfall 1983** – Satellitenfehlalarm während der NATO-Übung Able Archer.
- **Norwegen-Vorfall 1995** – eine Forschungsrakete wird kurzzeitig als Trident-Start interpretiert.

Bei etwa fünf bis acht “kritischen” Vorfällen in 80 Jahren ergibt sich eine empirische Häufigkeit von ca. 0,06–0,1 Beinahe-Vorfällen pro Jahr. Die methodische Hauptschwierigkeit liegt in der Frage, welcher Anteil dieser Beinahe-Vorfälle ohne menschliches Eingreifen tatsächlich zu einem strategischen Erstschlag eskaliert wäre. Plausibel sind Werte zwischen 5 % und 30 % – das ergibt eine grobe jährliche Auslöseraten-Schätzung von 0,003–0,03.

### Probabilistische Modellierung

Mit der Formel  $P = 1 - (1 - p)^n$  und  $n = 100$  Jahren ergibt sich für verschiedene Annahmen über die jährliche Eintrittswahrscheinlichkeit  $p$ :

$p$ (jährlich)	$P$ über 100 Jahre
0,1 %	ca. 9,5 %
0,3 %	ca. 26 %

p (jährlich)	P über 100 Jahre
0,5 %	ca. 39 %
1,0 %	ca. 63 %

Diese Bandbreite spiegelt die tatsächliche epistemische Lage wider. Toby Ord schätzt in *The Precipice* (2020) das *existenzielle* Risiko aus Atomwaffen für das 21. Jahrhundert auf etwa 1:1.000, wobei “existenziell” hier eng als Auslöschung oder dauerhafter Zivilisationskollaps verstanden wird – nicht als großer Nuklearaustausch.

## Annahme der stochastischen Unabhängigkeit

Die Formel  $1 - (1 - p)^n$  setzt unabhängige Jahresereignisse mit konstantem  $p$  voraus. Diese Annahme ist kritisch zu hinterfragen:

- **Zeitvariables  $p$ :** Die jährliche Wahrscheinlichkeit ist nicht konstant. Sie hängt von geopolitischer Spannung, technologischer Entwicklung und Anzahl der Atommächte ab.
- **Autokorrelation:** Krisenphasen (Kuba 1962, Able Archer 1983, Ukraine-Krieg seit 2022) erhöhen das Risiko über mehrere Jahre zusammenhängend.

Realistischer ist ein Modell mit zeitvariablem  $p_t$ , das in Spannungsphasen deutlich erhöht ist und in Détente-Phasen sinkt. Der Erwartungswert über lange Zeiträume liegt aufgrund der konvexen Form der Risikofunktion vermutlich höher als ein konstantes Mittel- $p$  suggeriert.

## Einflussfaktoren

**Klimawandel** wirkt indirekt über Migrationsdruck, Ressourcenkonflikte (besonders Wasser) und destabilisierte Staaten. Der Zusammenhang ist plausibel, aber empirisch schwer zu quantifizieren. Ich schätze den Effekt als moderat ein – stärker bei regionalen Atomwaffenstaaten (Indien/Pakistan, perspektivisch Naher Osten) als bei den strategischen Arsenalen der USA, Russlands und Chinas.

**Automatisierung der Frühwarn- und Entscheidungssysteme** halte ich für den strukturell stärksten risikoerhöhenden Faktor. Das Petrow-Argument ist zentral: Eine menschliche Plausibilitätskontrolle war damals möglich, weil das System eine Reaktionszeit von etwa 25 Minuten zuließ. Mit Hyperschallwaffen, schiffsgestützten Marschflugkörpern und KI-gestützter Lageanalyse droht ein strukturelles “Race-to-the-bottom” hin zu Entscheidungszyklen unter zehn Minuten, in denen kein Petrow mehr eingreifen kann. Die sowjetische “Perimeter”-Architektur (“Dead Hand”) zeigt, dass diese Tendenz bereits in der Spätphase des Kalten Krieges institutionalisiert wurde.

**Persönliche Schätzung:** Über 100 Jahre halte ich eine Wahrscheinlichkeit von etwa **10–35 %** für einen versehentlich ausgelösten strategischen Nuklearaustausch für plausibel; mein Punkt-Schätzwert liegt bei rund **15–20 %**. Das entspricht einem mittleren  $p$  von etwa 0,15–0,2 % pro Jahr.

## Versehentlicher vs. absichtlicher Atomkrieg

Argumente für die höhere Wahrscheinlichkeit eines *versehentlichen* Krieges:

- Mehrere dokumentierte Beinahe-Vorfälle dieser Art, kein dokumentierter Befehl zum absichtlichen Erstschlag.
- Rationale Akteure haben starke Anreize gegen einen absichtlichen Erstschlag (MAD-Logik).
- Strukturelle Risiken (Automatisierung, Verkürzung der Entscheidungsfenster) wachsen kontinuierlich.

Argumente für die höhere Wahrscheinlichkeit eines *absichtlichen* Krieges:

- Die Annahme rationaler Akteure bricht in der vom Nutzer skizzierten “Despot am Lebensende”-Konstellation zusammen (erweiterter Suizid).
- Proliferation erhöht die Zahl der Entscheidungsträger – und damit die Wahrscheinlichkeit irrationaler Akteure.
- “Use it or lose it”-Logiken bei konventionell verlierenden Atomwaffenstaaten.

Mein Urteil: Über 100 Jahre halte ich beide Wahrscheinlichkeiten für ähnlich groß, mit leichter Tendenz zugunsten des versehentlichen Falls – etwa **60:40** (versehentlich : absichtlich), bei großer Unsicherheit. Begründung: Strukturelle Systemrisiken wirken kontinuierlich und kumulativ, während absichtliche Auslösungen seltene Konstellationen erfordern.

Wichtig: Die Trennung ist nicht scharf. Die Kuba-Krise zeigt, dass absichtliche Eskalation und versehentliche Auslöser sich überlagern können – Archipows Entscheidung fiel unter dem Eindruck einer als absichtlich wahrgenommenen Aggression.

## Einordnung in andere existenzielle Risiken

Eine grobe Übersicht in Anlehnung an Ord (2020) und neuere Diskussionsbeiträge:

Risiko	Geschätzte Wahrscheinlichkeit eines existenziellen Ausgangs (100 Jahre)
Nicht ausgerichtete KI	~10 % (Ord 2020), neuere Schätzungen oft höher
Künstliche Pandemie	~3 %
Atomkrieg (Auslöschung)	~0,1 %
Klimawandel (direkt)	~0,1 %
Natürliche Pandemie	~0,01 %
Supervulkan	~0,01 %
Asteroideneinschlag	~0,0001 %

Diese Zahlen sind kontrovers; relevant ist hier vor allem die *Größenordnung*. Wichtig ist die Unterscheidung zwischen Auslöschung und Katastrophe mit Hunderten Millionen Toten: Letztere ist beim Atomkrieg um Größenordnungen wahrscheinlicher als erstere. Daher erscheinen “Atomkrieg im Jahrhundert”-Schätzungen oft mit zweistelligen Prozentsätzen, “atomares Auslöschungsereignis” aber im Promille-Bereich.

Daraus ergibt sich:

- **Künstliche Pandemien und KI-Risiken** sind im Erwartungswert für die Auslöschung höher als nukleare Risiken, vor allem wegen ihrer Skalierungs- und Selbstverstärkungsdynamik.
- **Klimawandel** ist als unmittelbares Auslöschungsrisiko gering, aber als *Verstärker* anderer Risiken (insbesondere nuklearer und biologischer) sicherheitspolitisch hoch relevant.
- **Asteroideneinschlag und Supervulkan** sind im 100-Jahr-Horizont vernachlässigbar.

Das Atomkriegsrisiko liegt damit im mittleren Bereich existenzieller Risiken: deutlich oberhalb der natürlichen Risiken, aber unterhalb der anthropogenen technologischen Spitzenrisiken. Innerhalb dieser Gruppe ist es jedoch das einzige mit dokumentiert vorhandener und akut einsatzbereiter Auslösungstechnologie – was eine besondere politische Dringlichkeit begründet.

## Schlussbemerkungen

Die genannten Zahlen sind keine Vorhersagen, sondern strukturierte Schätzungen unter offengelegten Annahmen. Drei Punkte erscheinen mir sicherheitspolitisch besonders relevant:

1. **Entkopplung von Detektion und Auslösung:** Mechanismen, die menschliche Plausibilitätskontrolle auch unter Zeitdruck sicherstellen – etwa Verzicht auf “launch on warning” zugunsten von “launch under attack” auf Basis gehärteter Zweitschlagfähigkeit (insbesondere SSBN) – wirken risikomindernd.
2. **Vorsicht bei KI-Integration in nukleare C2-Systeme:** Hier liegt das größte zusätzliche Risikopotenzial der kommenden Dekaden. Die “Race-to-the-bottom”-Dynamik hin zu kürzeren Entscheidungszyklen ist strukturell, nicht durch einzelne Akteure verschuldet, und entsprechend schwer umkehrbar.
3. **Risiko-Diversifikation in der sicherheitspolitischen Aufmerksamkeit:** Die Konzentration auf nukleare Risiken darf nicht dazu führen, biologische und KI-bezogene Risiken zu vernachlässigen, die nach gegenwärtigem Stand mindestens vergleichbar groß sind.

Die Logik von  $1 - (1 - p)^n$  macht deutlich: Selbst kleine jährliche Wahrscheinlichkeiten akkumulieren über Jahrzehnte zu erheblichen Werten. Genau dieses Argument hat Carl Sagan in den 1980er Jahren formuliert. Es ist nicht alarmistisch, sondern eine schlichte Konsequenz elementarer Wahrscheinlichkeitstheorie – und es bleibt aktuell.