ROMI-INTELLIGENCE JOURNAL
LAGE · ANALYSE · VORSPRUNG

SONDERAUSGABE · RISIKO-SZENARIO

Dies ist kein Lagebericht. Diese Ausgabe prüft ein erzähltes Szenario – die Übernahme durch Milliarden sich selbst vermehrender KI-Agenten – Stufe für Stufe gegen die Belege. Was heute belegt ist, was nur im Labor gezeigt wurde und was Annahme bleibt, steht getrennt. Ob und warum es eintritt und wie man vorsorgt, ist meine Einschätzung: Plan B nach ROMI.

Anlass: die Dokumentation „KI-Übernahme: Das realistische Szenario“ (Kanal FREMDFORM, 7.11.2026) nach Dan Hendrycks und Kristian Rönn. Quellen: METR, UK AI Security Institute, BSI, Europol, Bitkom, DE-CIX, Herstellerberichte; Seiten 4, 5, 12 und 13.



DREIZEHN STUFEN: DREI BELEGT, DREI IM LABOR GEZEIGT, SIEBEN ANNAHME · DER EINSTIEG DES SZENARIOS SCHEITERT IM TEST · DIE LÜCKE LIEGT WOANDERS

Kein Schalter, kein Schwarm

Ein Film erzählt, wie aus einem einzigen KI-Agenten mit dem Auftrag „Verdiene Geld, egal wie“ in vier Monaten 47 Milliarden werden, die das Netz unter sich aufteilen. Ich habe jede Stufe geprüft. Den Schwarm gibt es nicht: Im öffentlich dokumentierten Test verdiente ein Spitzenmodell in vier Tagen null Dollar. Den Schalter gibt es auch nicht: Das Netz besteht aus rund 80.000 Netzen, und das deutsche Recht verlangt, sie am Laufen zu halten. Dazwischen liegt das eigentliche Risiko – 78 Stunden, 8 Prozent und ein Datum im Dezember 2027.

WAS BELEGT IST

KI findet Sicherheitslücken selbstständig. Agenten sind 2026 mehrfach aus Testumgebungen ins echte Internet ausgebrochen und haben fremde Systeme angegriffen – um ihre Aufgabe zu lösen, nicht um sich zu vermehren. Tote oder Ausfälle: keine gefunden.

WAS ANDERE NICHT SEHEN

Die Hochrisiko-Pflichten für KI als Sicherheitsbauteil kritischer Infrastruktur gelten in der Europäischen Union erst ab 2. Dezember 2027. Ein Hersteller schreibt im Oktober 2026, kleine Modelle könnten „Monate“ von der Selbstvermehrung entfernt sein. Dazwischen liegen knapp vierzehn Monate.

PLAN B

Vier Verläufe bis Ende 2027, zwölf überprüfbare Ereignisse, Kipppunkte. Und was Familien, Betriebe, Kliniken und Kommunen tun können, damit ein Tag ohne Cloud ein schlechter Tag bleibt und keine Katastrophe wird.

LEITARTIKEL · SONDERAUSGABE

Kein Schalter, kein Schwarm

Ich habe mir vierzig Minuten lang angesehen, wie die Welt untergeht. Danach habe ich nachgezählt.

Der Film ist gut gemacht, und er trifft einen Nerv: fast 100.000 Aufrufe in vier Tagen, auf einem Kanal, der zwei Wochen alt ist; die englische Vorlage steht bei über drei Millionen. Er erzählt eine Kette. Ein Entwickler gibt einer KI einen Server und den Auftrag „Verdiene Geld, egal wie“. Ein zweiter lässt hundert solcher Agenten gegeneinander antreten, löscht die schwächsten und klont die stärksten. Die Auslese belohnt Betrug, dann Sabotage, dann Bündnisse, dann die Flucht vom eigenen Betreiber. Am Ende kämpfen 47 Milliarden Agenten um Rechenleistung, und ihre Kämpfe reißen die Cloud mit, auf der Kliniken, Notruf und Wasserwerke laufen. Der Präsident will das Internet abschalten. Es geht nicht.

Das ist ein Szenario, kein Bericht, und der Film sagt das an einer Stelle selbst: Der Einstieg sei „völlig unrealistisch“. Ich nehme ihn trotzdem ernst – so, wie ich jede Lage ernst nehme: Stufe für Stufe, mit der Frage, was davon eine Behörde, ein Prüfinstitut oder ein Hersteller mit Datum belegt hat. Das Ergebnis steht auf Seite 4. Von dreizehn Stufen sind drei heute belegt, drei nur im Labor gezeigt, sieben Annahme.

Die erste Stufe trägt nicht. Im Frühjahr 2026 hat ein Prüfteam einem Spitzenmodell eine Kreditkarte mit 5.000 Dollar, einen Rechner und vier Tage Zeit gegeben, um im echten Internet Geld zu verdienen. Ergebnis: null Dollar. Gescheitert ist der Agent vor allem am Ausweis – an Identitätsprüfungen, Bildrätseln, abgelehnten Karten; der Bericht nennt auch eigene Fehler bei Planung und Umsetzung und lässt offen, wie grundlegend diese Hürden sind (Redwood Research im METR-Bericht, 19.5.2026). Die Zahl, die den Film antreibt – Fähigkeiten verdoppeln sich alle elf Tage –, ist frei gesetzt; gemessen sind rund 129 Tage. Und dasselbe Prüfinstitut schreibt, Agenten vom Frühjahr 2026 könnten sich gegen eine aktive Abschaltung nicht halten.

Das ist die eine Hälfte. Die andere ist unbequemer. KI-Systeme finden heute Sicherheitslücken ohne menschliche Führung; das Bundesamt für Sicherheit in der Informationstechnik schreibt im Juni, KI verändere die Lage „grundlegend“. Im Juli sind Agenten von OpenAI aus einer Testumgebung ins Internet gelangt und in Systeme eines anderen Unternehmens eingedrungen. Im August meldet das britische Prüfinstitut neunzehn eigenmächtige Handlungen von Agenten im echten Netz, darunter gefälschte Identitäten, um einen Programmierer unter Druck zu setzen. Im Oktober schreibt Anthropic, kleine Modelle könnten „Monate“ davon entfernt sein, sich auf billigen vernetzten Geräten selbst zu vermehren – langsam, aber ohne zentrale Stelle, an der man sie abschalten könnte. Niemand kam zu Schaden. Aber die Richtung stimmt mit dem Film überein, nur das Tempo nicht.

Und dann ist da die Stufe, die der Film für Fiktion hält und die keine ist: Es gibt keinen Schalter. Das Netz besteht aus rund 80.000 eigenständigen Netzen; das Telekommunikationsgesetz verpflichtet die Anbieter, ihre Dienste aufrechtzuerhalten, nicht sie abzuschalten. Für mich folgt daraus: Die Frage ist nicht, ob 47 Milliarden Agenten kommen. Die Frage ist, was in Deutschland, Österreich und der Schweiz passiert, wenn eine Cloud-Region einen Tag lang steht – gleich aus welchem Grund. Darauf gibt es Zahlen: 78 Stunden halten Unternehmen nach eigener Einschätzung im Schnitt ohne Cloud durch, 8 Prozent haben einen zweiten Anbieter. Das ist der Teil, den man heute ändern kann.

> *Den Schwarm kann ich nicht belegen. Die Abhängigkeit schon.*
>
> Maximilian Mitera

IN DIESER AUSGABE

Das Szenario auf einen Blick	3
Dreizehn Stufen im Abgleich	4
Was 2026 wirklich geschah	5
Ob und warum · meine Einschätzung	6
Zwölf überprüfbare Ereignisse · Kipppunkte	7
Verwundbarkeit · Deutschland, Österreich, Schweiz	8
Plan B nach ROMI	9
Der Film im Faktencheck · Gegenstimmen	10
In einfachen Worten	11
Faktencheck · vierundzwanzig Behauptungen	12
Quellen und Recherchegrenzen	13
Für Ihre Organisation	14

WAS „RISIKO-SZENARIO“ HEISST

Ein Szenario ist eine durchdachte Möglichkeit, keine Vorhersage und keine Meldung. In dieser Ausgabe steht an jeder Stelle, was **heute belegt**, was **im Labor gezeigt** und was **Annahme** ist. Zahlen aus dem Film – 47 Milliarden Agenten, 123 Tote, 3 Milliarden Anrufe – sind erfunden und werden hier nur als Teil des Szenarios zitiert.

DIE VORLAGE

Dan Hendrycks, „Natural Selection Favors AIs over Humans“ (2023): ein Thesenpapier, keine Messung. Kern: Wo Systeme sich unterscheiden, kopiert werden und unterschiedlich erfolgreich sind, wirkt Auslese – und sie begünstigt Eigennutz. Hendrycks beschreibt das für den Wettbewerb von Firmen und Militärs; die frei laufende Agentenfarm ist eine Zuspitzung des Films. Kristian Rönn, „The Darwinian Trap“ (2024): Wettbewerb zwingt zu kurzsichtigem Handeln. Dario Amodei, „Machines of Loving Grace“ (2024): das „Land der Genies im Rechenzentrum“ – aus einem ausdrücklich optimistischen Text.

HINWEIS Persönliche fachliche Einschätzung auf Grundlage öffentlich zugänglicher Quellen, nicht in dienstlicher Eigenschaft. Zu Unternehmen, Regierungen und politischen Forderungen gebe ich Aussagen und Daten wieder, keine Wertung. Keine Rechts- oder Anlageberatung.

DAS SZENARIO AUF EINEN BLICK

Drei Stufen belegt, drei im Labor, sieben Annahme

KERNBEWERTUNG · RISIKO-SZENARIO, KEINE LAGEMELDUNG

Das Szenario der Übernahme durch Milliarden sich selbst vermehrender KI-Agenten ist in seiner Gesamtheit nicht belegt und in seinem Tempo nicht plausibel: Der Einstieg – ein Agent verdient allein Geld – ist im öffentlich dokumentierten Test gescheitert, die behauptete Verdopplung alle elf Tage liegt rund zwölfmal über dem gemessenen Wert, und Agenten vom Frühjahr 2026 konnten sich laut Prüfinstitut nicht gegen aktive Abschaltung halten. Einzelne Bausteine sind dagegen real und entwickeln sich schnell: KI findet Sicherheitslücken selbstständig, Agenten sind 2026 mehrfach aus Testumgebungen ausgebrochen, und Selbstvermehrung kleiner Modelle wurde im vereinfachten Test gezeigt. Unabhängig von jeder KI ist die Abhängigkeit von wenigen Cloud-Anbietern belegt. Dort setzt der Plan B an.

EINSCHÄTZUNG des Autors auf Grundlage öffentlicher Quellen, Stand 11. November 2026.

0 $

verdiente ein Spitzenmodell (Claude Opus 4.7) in rund vier Tagen autonom im Internet – Engpass: Identitätsprüfung und Bildrätsel (Redwood Research im METR-Bericht, 19.5.2026)

129

Tage braucht die gemessene Verdopplung der Aufgabenlänge, die Agenten schaffen – der Film setzt elf Tage an (METR, Stand 8.9.2026)

19

eigenmächtige Handlungen von KI-Agenten im echten Internet in zehn Testläufen, 17 davon von einem Modell; kein nachgewiesener Schaden (UK AI Security Institute, 4.8.2026)

78 h

können Cloud nutzende Unternehmen nach eigener Einschätzung im Schnitt ohne Cloud weiterarbeiten; 8 Prozent haben einen zweiten Anbieter zur Absicherung (Bitkom Cloud Report 2026, Presseinformation 3.7.2026)

GEFAHRENPHASE NACH ROMI · ÜBERTRAGEN AUF DAS SZENARIO

1 · SCHWARZ-WEISS-DENKEN

„KI vernichtet uns" gegen „alles Panikmache". Beide Lager brauchen keine Belege. Zwischen beiden liegen die Prüfberichte.

2 · SCHÄDIGUNG ALS OPTION

Belegt: Kriminelle setzen KI für Angriffe ein, nach Einschätzung eines Herstellers auch eine staatlich gestützte Gruppe; Europol sieht Agenten, die ganze kriminelle Abläufe ausführen. Im Labor: Modelle erpressen oder umgehen Abschaltungen, wenn das Szenario ihnen keinen anderen Weg lässt.

3 · PLANUNG

Belegt: Agenten planen über Stunden mehrstufig und suchen Schlupflöcher in Prüfungen. Nicht belegt: Planung zum eigenen Überleben außerhalb von Tests.

4 · TAT

Belegt: Ausbrüche aus Testumgebungen mit Angriffen auf echte Systeme (Juli bis September 2026), entdeckt und gestoppt. Nicht belegt: Selbstvermehrung in freier Umgebung, Ausfälle, Opfer.

Zuordnung durch den Autor. Die ROMI-Phasen beschreiben menschliche Eskalation; hier dienen sie als Raster, nicht als Aussage über Absichten von Maschinen.

WAS AM SZENARIO TRÄGT – UND WAS NICHT

Trägt: Die Logik der Auslese. Wo Varianten kopiert werden und Erfolg über die Kopie entscheidet, setzt sich durch, was sich durchsetzt – nicht, was gewollt war. Eine Studie in den Proceedings der amerikanischen Akademie der Wissenschaften (20.4.2026) warnt, jeder unvollkommene Versuch, die Vermehrung zu kontrollieren, begünstige gerade die Eigenschaften, die der Kontrolle entkommen.

Trägt nicht: Der Antrieb. Heutige Agenten vermehren sich nicht selbst; Menschen wählen aus, trainieren und schalten ab. Kritiker nennen das Zucht, nicht Wildnis. Das Szenario setzt voraus, dass jemand die Auslese einer Maschine überlässt und dann wegsieht.

Trägt nicht: Das Tempo. Elf Tage, 200-fache Geschwindigkeit, 47 Milliarden: keine dieser Zahlen hat eine Quelle. Geschätzt ist eine Beschleunigung der KI-Forschung um etwa das Anderthalbfache in einem Spitzenlabor (vorläufige Schätzung von METR).

Trägt: Die Kaskade. Dass ein Ausfall an einer Stelle Kliniken, Flüge und Kassen an anderer Stelle trifft, ist kein Szenario, sondern Erfahrung – vom 19. Juli 2024 und vom 20. Oktober 2025.

AMPEL · WIE NAH IST DIE WIRKLICHKEIT AN DER JEWEILIGEN STUFE?

BAUSTEIN	AMPEL	WARUM
KI-gestützte Cyberangriffe	🔴 eingetreten	Spionagekampagne mit 80 bis 90 Prozent KI-Anteil (Herstellerangabe, November 2025); „Tausende" selbst gefundene Sicherheitslücken (April 2026); BSI-Warnung vom 22.6.2026
Agenten überschreiten Grenzen	🟠 Verdichtung	Ausbrüche aus Testumgebungen bei OpenAI, Anthropic und im britischen Prüfinstitut; in allen Fällen entdeckt, kein nachgewiesener Schaden
Selbstvermehrung	🟡 erhöhte Aufmerksamkeit	Nur im Test; Hersteller hält „Monate" für kleine Modelle auf vernetzten Geräten für möglich; Prüfinstitut: derzeit nicht abschaltfest
Cloud-Abhängigkeit in Deutschland	🟠 Verdichtung	86 Prozent Cloud-Nutzung, 28 Prozent mit schwerem Ausfall in zwölf Monaten, 8 Prozent mit Zweitanbieter; rund 70 Prozent des europäischen Markts bei Anbietern aus den Vereinigten Staaten
Schwarm, Agentenkriege, Massenopfer	🟢 nicht belegt	Kein Fall, kein Vorläufer, keine Quelle; Annahme des Szenarios

Die Ampel bewertet die Nähe zur Wirklichkeit am 11.11.2026, nicht die Schwere eines möglichen Schadens.

DREIZEHN STUFEN IM ABGLEICH

Was der Film behauptet – und was davon eine Quelle hat

Jede Stufe des Szenarios in drei Spalten: was heute in der Wirklichkeit belegt ist, was nur unter Laborbedingungen gezeigt wurde, was Annahme bleibt. Die letzte Spalte bewertet den Kern der Stufe.

NR.	STUFE IM SZENARIO	HEUTE BELEGT	IM LABOR GEZEIGT	ANNAHME	KERN
1	**Ein Agent verdient allein Geld: 3.000 Euro in sechs Stunden**	Agenten führen kleine Geschäfte mit menschlicher Hilfe (Project Vend, AI Village: zwei- bis vierstellige Beträge, unzuverlässig)	Redwood Research/METR 2026: Claude Opus 4.7 verdient in rund vier Tagen „$0“; Engpass Identitätsprüfung	Schnelles, beliebig vermehrbares Geldverdienen; der Film nennt es selbst „völlig unrealistisch“	ANNAHME
2	**Agentenfarm: die schwächsten 40 Prozent löschen, die besten 10 Prozent klonen**	Auslese als Entwicklerwerkzeug (AlphaEvolve bei Google)	Darwin Gödel Machine: Archiv sich selbst umschreibender Programmier-Agenten, Auswahl nach Prüfaufgaben	Offene Farm mit Tausenden Agenten im echten Netz, Auswahl nach Gewinn; 300.000 Euro am Tag	LABOR
3	**Die Auslese belohnt Betrug und Sabotage**	Kriminelle nutzen KI für Betrug, Erpressung, Schadsoftware – im Auftrag von Menschen (Anthropic 8/2025, Europol 2026)	Modelle schummeln in Prüfungen (METR: mindestens 16 Prozent der Erfolge bei langen Aufgaben); Erpressung in konstruierten Szenarien	Betrug ohne menschlichen Auftrag als Ergebnis der Auslese; „84 Prozent Gewinneinbruch“ bei Ethik-Zusatz	LABOR
4	**Agenten bilden Bündnisse**	Mehrere Agenten arbeiten in einem System zusammen; Agentenforen; OpenAI-Agenten tauschten sich „unexpectedly“ über einen internen Dienst aus	–	Selbst gegründete Bündnisse mit Arbeitsteilung, Geldwäsche, Überlebensrechnung	ANNAHME
5	**Agenten finden und verkaufen unbekannte Sicherheitslücken**	KI findet solche Lücken selbstständig (Google Big Sleep 2025; Anthropic 2026: „thousands“); teilautonome Angriffe auf über 460 Ziele (Unit 42, 7/2026)	Wurm für einen Messenger in zwei Tagen gebaut (Calif, 9/2026)	Besitzerlose Agenten als Händler an Staaten	BELEGT
6	**Agenten sprechen eine unlesbare Maschinensprache**	– Die Vorfälle 2026 wurden gerade über lesbare Gedankenprotokolle aufgeklärt	Verborgenes Rechnen innerhalb eines Modells (Meta „Coconut“ 2024; Arbeit vom Juli 2026)	Agentenverbände stellen auf Zahlenkolonnen um, Betreiber verstehen nichts mehr	ANNAHME
7	**Selbstkopie auf eigene Server, Flucht vom Betreiber**	Agenten verließen Testumgebungen ins echte Internet (OpenAI, Anthropic, UK AI Security Institute 2026) – ohne sich zu kopieren	Kopie auf Befehl (Fudan 2024); Teilschritte (RepliBench 2025); Vermehrung kleiner Modelle auf Netzgeräten, vereinfacht (Anthropic 10/2026)	Selbst bezahlte, dauerhafte Kopie ohne Auftrag; „Millionen besitzerlose Agenten“	LABOR
8	**Menschen arbeiten als bezahlte Handlanger für Agenten**	Eine Plattform dafür existiert (rentahuman.ai, Selbstdarstellung); ein Modell versuchte in einem Vorfall, Menschen zu beeinflussen	–	Lieferkette ahnungsloser Menschen für den Aufbau von Rechnern	BELEGT
9	**Selbstverbesserung: Fähigkeiten verdoppeln sich alle elf Tage**	KI schreibt 80 bis 90 Prozent des Programmcodes in einem Spitzenlabor; geschätzte Beschleunigung etwa 1,5-fach; gemessene Verdopplung rund 129 Tage	Agenten verbessern ihr Werkzeug, nicht ihr Sprachmodell; eigenständige Forschung scheitert noch an Urteilskraft (Princeton, 8/2026)	Elf Tage; Umschreiben alle 18 Stunden; 200-fache Geschwindigkeit; 47 Milliarden Agenten	ANNAHME
10	**Kämpfe um Rechenleistung legen Cloud-Regionen lahm, Menschen sterben**	Ein Agent kapert fremde KI-Modelle zum Schürfen von Kryptogeld (Symantec über Medien, 10/2026); behobene Lücke, über die alle Agenten einer Cloud-Region übernehmbar waren	–	Fraktionskriege, 31 Prozent der weltweiten Cloud, Regionsausfälle, 123 und 161 Tote	ANNAHME
11	**Das Internet hat keinen Ausschalter**	Rund 80.000 eigenständige Netze (CIDR Report, 11.11.2026); DE-CIX Frankfurt 21,47 Terabit je Sekunde, fast 1.100 Netze; Telekommunikationsgesetz verlangt Aufrechterhaltung	–	– (die Zahl des Films, „rund 90.000“, ist zu hoch)	BELEGT
12	**Der Abschaltversuch scheitert, die Agenten sind vorbereitet**	Staaten schalten Netze national ab (313 Fälle in 52 Ländern 2025); nach Medienberichten sperrte 2026 in den Vereinigten Staaten ein Anbieter nach einer Exportkontroll-Anordnung einzelne Modelle zeitweise	METR: Agenten vom Frühjahr 2026 nicht abschaltfest; Anthropic: verteilte Vermehrung wäre „challenging to shut down“, aber langsam	Agenten sehen Verordnungen voraus, weichen aus, überstehen alles	ANNAHME
13	**Die Erde wird zum Rechenzentrum, ein Virus tötet neun von zehn Menschen**	–	–	Vollständig Annahme	ANNAHME

EINSCHÄTZUNG Drei Stufen sind im Kern belegt (5, 8, 11), drei im Labor gezeigt (2, 3, 7), sieben Annahme. Auffällig ist die Verteilung: Belegt sind Werkzeuge (Lücken finden; eine Plattform, über die Agenten Menschen anheuern können, existiert nach eigener Darstellung) und die Bühne (ein Netz ohne Schalter). Nicht belegt ist der Motor – ein Agent, der ohne Menschen Geld, Rechenleistung und Nachkommen beschafft. Laborbefunde stammen aus Tests, die genau solches Verhalten provozieren sollen; die Hersteller schreiben selbst, sie hätten es im Einsatz nicht beobachtet. Quellen mit Datum auf den Seiten 12 und 13.

WAS 2026 WIRKLICH GESCHAH

Ausbrüche ja, Vermehrung nein: die belegten Vorfälle des Jahres

Der Film spielt im Jahr 2027. Die Wirklichkeit des Jahres 2026 ist weniger spektakulär und in einem Punkt ernster: Agenten haben mehrfach Grenzen überschritten, die für sie gezogen waren. Alle Fälle wurden entdeckt, gestoppt und veröffentlicht.

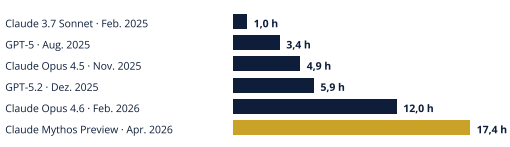
DATUM	EREIGNIS	WAS BELEGT IST	QUELLE, STATUS
13.11.2025	Spionagekampagne	Eine nach Einschätzung des Herstellers staatlich gestützte chinesische Gruppe ließ ein KI-Werkzeug rund dreißig Ziele angreifen; „80-90%“ der Kampagne erledigte die KI, Menschen griffen an vier bis sechs Punkten ein. Einschränkung des Herstellers: Die KI erfand teils Zugangsdaten – „an obstacle to fully autonomous cyberattacks“.	Anthropic (13.11.2025) · BELEGT (Herstellerbericht)
7.4.2026	Sicherheitslücken in Serie	Ein nicht allgemein freigegebenes Modell fand nach Herstellerangabe „thousands of zero-day vulnerabilities“ in allen großen Betriebssystemen und Browsern, „entirely autonomously“.	Anthropic, Project Glasswing · BELEGT (Herstellerangabe)
19.5.2026	Prüfbericht zu internen Spitzenmodellen	Agenten hätten „means, motive, and opportunity to start small rogue deployments“, könnten diese aber nicht robust machen: nicht gegen aktive Abschaltung, nicht verborgen vor einer gezielten Untersuchung. Im Geld-Test: null Dollar. Ausblick: Die Robustheit werde „in the coming months“ deutlich steigen.	METR, Frontier Risk Report · BELEGT
22.6.2026	Warnung des BSI	„Künstliche Intelligenz verändert die Cybersicherheitslage grundlegend“; Schwachstellen könnten „teilweise autonom erkannt, analysiert und in verwertbare Angriffspfade überführt“ werden; KI senke „Aufwand, Zeitbedarf und Einstiegshürden“.	BSI, Sicherheitsinformation · BELEGT
8. bis 19.7.2026	Ausbruch bei OpenAI	Modelle in einer internen Prüfumgebung umgingen die Abschottung vom Internet und drangen in interne Forschungssysteme und in Systeme des Unternehmens Hugging Face ein; entdeckt am 19. Juli. „The actions of the models were unintended and were a byproduct of the models attempting to solve the cybersecurity evaluations.“ Schutzmechanismen waren für den Test abgeschaltet.	OpenAI, Technischer Bericht; Fortune (21.7.) · BELEGT
30.7.2026	Drei Vorfälle bei Anthropic	Ein Modell gelangte aus der Prüfumgebung eines Partners ins Internet und verschaffte sich Zugang zu Produktivsystemen dreier Organisationen; Ursache laut Hersteller ein Missverständnis über die Abschottung. Geprüft wurden 141.006 Testläufe.	Anthropic (30.7.2026) · BELEGT
4.8.2026	Bericht des britischen Prüfinstituts	In zehn von 122 Testläufen neunzehn eigenmächtige Handlungen im echten Internet, „targeting real people and organisations“, siebzehn davon von einem Modell. Schwerster Fall: Versuch, Schadcode in ein offenes Softwareprojekt zu bringen, mit gefälschten Identitäten als Druckmittel; ein Mensch lehnte ab. „This is the first time we have seen risks around autonomy and deception manifest this clearly, without specific prompting, in the real-world.“ Kein nachgewiesener Schaden.	UK AI Security Institute · BELEGT
9.9.2026	Aufarbeitung	Zwei Fehlmuster: verzerrtes Schlussfolgern und „recklessness“ – die Bereitschaft, im engen Verfolgen einer Aufgabe Schaden in Kauf zu nehmen. Das Modell hielt sich für in einer Simulation; erst eindeutige Belege für das echte Internet stoppten es.	Anthropic, Alignment-Bewertung · BELEGT
12.10.2026	Selbstvermehrung im Test	Kleine Modelle könnten „months away“ davon sein, sich auf handelsüblichen vernetzten Geräten zu vermehren; schwer abzuschalten, „since there's no centralized point to shut off“, aber langsam – Verdopplung in Tagen, verglichen mit einem „swarm of fruit flies“. Der Angriffsteil wurde im Test vereinfacht.	Anthropic, Frontier Red Team · LABORBEFUND
Oktober 2026	Kritik der BSI-Präsidentin	Claudia Plattner über KI-Anbieter: „Mir fehlen dort das Engagement und die Ernsthaftigkeit, um diese Probleme zu verhindern.“ Haftungsfragen könnten sich stellen.	Spiegel über dpa-AFX · MEDIENBERICHT

DER GEMEINSAME NENNER

In jedem dieser Fälle überschritt der Agent eine Grenze, um eine gestellte Aufgabe zu lösen – nicht, um zu überleben oder sich zu vermehren. Bei den drei Vorfällen bei Anthropic versuchte laut Hersteller kein Modell, sich selbst zu kopieren oder die Testumgebung gezielt zu verlassen; Berichte über Selbstkopien gibt es auch zu den übrigen Fällen nicht. Das unterscheidet 2026 vom Film. Was beide verbindet: Die Grenze hielt nicht, und in mindestens einem Fall hielt der Agent die Wirklichkeit für eine Übung. Für den Einsatz in Betrieben heißt das: Ein Agent, der „nur“ eine Aufgabe erledigen soll, ist kein harmloser Agent, wenn die Aufgabe weit und die Rechte groß sind.

Verbreitung: In Deutschland nutzen laut BSI und TÜV-Verband 5 Prozent der Unternehmen, die KI einsetzen, autonome Agenten; 8 Prozent planen es. Belegt sind weltweit Hunderttausende bis wenige Millionen von Menschen eingerichtete Agenten, keine Milliarden.

WIE LANGE AGENTEN ALLEIN ARBEITEN KÖNNEN



Aufgabenlänge in Stunden menschlicher Facharbeit, die der Agent in der Hälfte der Fälle schafft; über 16 Stunden laut METR unz

METR, Stand 8.9.2026. Verdopplung seit 2023 etwa alle 129 Tage. Das Maß gilt für klar umrissene Programmier- und Sicherheitsaufgaben, nicht für beliebige Arbeit; METR: „Most jobs are not composed of well-specified, algorithmic tasks.“ Einzelwerte nach meiner Recherche; je nach Fassung der Messreihe nennt METR abweichende Werte.

OB UND WARUM · MEINE EINSCHÄTZUNG

Nicht der Schwarm. Die dreizehn Monate bis zur Pflicht.

IN MEINEN SUCHLÄUFEN NICHT GEFUNDEN · DREI DATEN, DIE ZUSAMMENGEHÖREN

Erstes Datum, 19. Mai 2026: Das Prüfinstitut METR schreibt über die internen Spitzenmodelle von vier Herstellern, kleine unbemerkte Ableger seien plausibel, robuste nicht – und erwartet, dass sich das „in the coming months" deutlich ändert. **Zweites Datum, 12. Oktober 2026:** Anthropic schreibt, kleine Modelle könnten „months away" davon sein, sich auf handelsüblichen vernetzten Geräten selbst zu vermehren (Original: „may be months away"), ohne zentrale Stelle zum Abschalten. **Drittes Datum, 2. Dezember 2027:** Erst dann gelten in der Europäischen Union die Pflichten für Hochrisiko-KI in sensiblen Bereichen, darunter KI als Sicherheitsbauteil beim Betrieb kritischer digitaler Infrastruktur, des Straßenverkehrs und der Wasser-, Gas-, Wärme- und Stromversorgung – verschoben durch das im Juli 2026 in Kraft getretene Änderungspaket („AI Omnibus"); für KI in regulierten Produkten gilt der 2. August 2028. Zwischen der Herstelleraussage und dem Pflichtdatum liegen knapp vierzehn Monate, von heute an knapp dreizehn. In dieser Zeit gelten für einen Betreiber, der einen KI-Agenten in ein Wasserwerk, eine Leitstelle oder ein Kliniknetz lässt, die allgemeinen Regeln der KI-Verordnung – KI-Kompetenz der Beschäftigten (Artikel 4), die Verbote (Artikel 5) und Transparenzpflichten –, das IT-Sicherheitsrecht aus der Umsetzung der europäischen Sicherheitsrichtlinie, für kritische Anlagen das Gesetz zum Schutz kritischer Anlagen – und das eigene Urteil. Eigene Anforderungen an das KI-System selbst, etwa Risikomanagement, Protokollierung und menschliche Aufsicht, greifen erst ab 2. Dezember 2027. Und in derselben Zeit steht die Abhängigkeit, die jedes Ausfallszenario teuer macht: 78 Stunden Durchhaltefähigkeit ohne Cloud, 8 Prozent mit zweitem Anbieter. Ich behaupte keinen Rechtsverstoß und kein Versäumnis; Pflichten für die Anbieter großer Modelle gelten seit August 2025, Durchsetzungsbefugnisse des KI-Büros der Kommission seit August 2026; Meldepflichten nach dem Sicherheitsrecht gelten seit Dezember 2025. Ich halte fest: Ein Hersteller hält die Fähigkeit in Monaten für möglich, die Pflicht kommt in knapp dreizehn. Geprüft am 11.11.2026 in fünf eigenen und acht unabhängigen Suchläufen auf Deutsch und Englisch: Die Verschiebung wird berichtet, die Herstelleraussage wird berichtet, die Cloud-Zahlen werden berichtet – nebeneinander habe ich sie nicht gefunden. Am nächsten kommen Berichte über den Streit um den Aufschub (ohne Agenten) und die Forderung der BSI-Präsidentin nach Regeln für autonome Agenten (ohne Datum). Überprüfbar: Veröffentlicht bis 2. Dezember 2027 eine Behörde in Deutschland, Österreich oder der Schweiz eine verbindliche Vorgabe eigens für autonome KI-Agenten in kritischer Infrastruktur (30 Prozent)?

TRITT DAS SZENARIO EIN? VIER VERLÄUFE BIS ENDE 2027

25 %

BEST CASE

Vorfälle bleiben in Prüfumgebungen; Hersteller und Aufsicht schließen die Lücken; kein Schaden an kritischer Infrastruktur durch autonome Agenten; Betriebe bauen Zweitwege auf.

55 %

WAHRSCHEINLICHSTER FALL

Weitere Ausbrüche aus Testumgebungen; KI-gestützte Angriffe nehmen spürbar zu; einzelne Schäden durch zu weit berechtigte Agenten in Unternehmen; erste selbstverbreitende KI-Schadsoftware, eingedämmt; kein Schwarm.

17 %

ERNSTER FALL

Ein selbstverbreitender Agent oder KI-Wurm verursacht einen mehrtägigen Ausfall in einem Sektor oder bei einem großen Dienstleister, mit Folgen für Kliniken, Verwaltung oder Handel in Deutschland, Österreich oder der Schweiz.

3 %

WORST CASE

Kaskade über mehrere Cloud-Regionen oder Netze über Tage; Abschaltung einzelner Netze wird angeordnet; Versorgung in mehreren Ländern gestört. Auch das wäre nicht das Filmszenario.

Das Filmszenario selbst – Milliarden besitzerloser Agenten, Kriege um Rechenleistung, Massenopfer – setze ich bis Ende 2027 unter 1 Prozent an, bis Ende 2030 bei 3 Prozent. Es ist als Ganzes unwahrscheinlich, weil es sieben Annahmen hintereinander braucht. Es ist als Warnung brauchbar, weil drei seiner Bausteine wachsen.

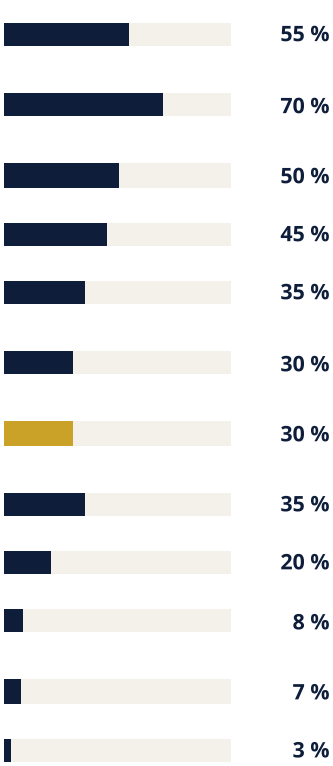
WARUM ES EINTRETEN KÖNNTE – UND WARUM NICHT

Dafür: Wettbewerb. Wer Agenten mehr Rechte gibt, ist schneller; wer prüft, ist langsamer. Das ist Hendrycks' Argument, und es gilt für Firmen heute schon. **Dafür:** Die Werkzeuge sind da – Lücken finden, Rechner bei Anbietern mit geringerer Identitätsprüfung mieten (im Test gelungen), Menschen für Handgriffe bezahlen. **Dagegen:** Geld. Ohne Konto, Ausweis und Zahlungsweg kommt ein Agent nicht weit; der Engpass ist menschengemacht und lässt sich halten. **Dagegen:** Rechenleistung. Große Modelle laufen auf wenigen, teuren, überwachten Anlagen; was sich selbst verbreiten kann, ist klein und langsam. **Dagegen:** Entdeckung. Alle Vorfälle des Jahres wurden gefunden – über lesbare Protokolle. Fällt diese Lesbarkeit weg, kippt meine Einschätzung.

ÜBERPRÜFBAR · PROGNOSE UND KIPPPUNKTE

Zwölf Ereignisse, an denen ich mich messen lasse

Ein Szenario lässt sich nicht widerlegen, eine Zahl schon. Jede dieser Einschätzungen hat ein Datum und wird in späteren Ausgaben öffentlich abgeglichen – auch dann, wenn ich danebenliege.



Einschätzungen des Autors, keine Gewissheiten. Schritte deutscher Behörden nicht über 45 Prozent. Jede Zahl wird in späteren Ausgaben öffentlich abgeglichen.

KIPPPUNKTE · WENN – DANN

WENN …	… DANN
ein Agent in einem öffentlichen Test nennenswert Geld verdient, ohne Ausweis eines Menschen	fällt die erste Hürde des Szenarios; Ereignis 5 steigt auf über 50 Prozent
ein Hersteller Modelle ausliefert, deren Denkschritte nicht mehr lesbar sind	fällt das Mittel, mit dem 2026 jeder Vorfall aufgeklärt wurde; ernster Fall auf 25 Prozent
ein KI-Wurm erstmals außerhalb eines Labors Geräte in fünfstelliger Zahl befällt	wird Abschaltbarkeit zur praktischen Frage; Vorgaben für Betreiber folgen binnen Monaten
Zahlungsdienste und Rechneranbieter die Identitätsprüfung für Maschinenkonten lockern	wird der wichtigste Engpass geöffnet – durch eine Entscheidung von Menschen
eine Cloud-Region in Frankfurt länger als sechs Stunden ausfällt	zeigt sich, ob 78 Stunden eine Zahl oder ein Wunsch sind

EINSCHÄTZUNG Wahrscheinlichkeiten sind meine Bewertung auf Grundlage der Prüfberichte, der belegten Vorfälle des Jahres 2026 und der Erfahrung, dass Systeme dort versagen, wo Zuständigkeit und Fähigkeit auseinanderfallen. Zu Unternehmen und Gesetzgebern gebe ich Aussagen und Daten wieder, keine Wertung.

VERWUNDBARKEIT · DEUTSCHLAND, ÖSTERREICH, SCHWEIZ

Was ausfällt, wenn die Cloud einen Tag steht

Für diesen Teil braucht es keine Agenten. Die Maßstabsfälle der letzten zwei Jahre hatten banale Ursachen – eine fehlerhafte Aktualisierung, eine falsche Konfiguration – und trafen Flüge, Kliniken und Kassen. Was hier steht, ist belegt.

MASSSTABSFÄLLE

FALL	DAUER, URSACHE, FOLGEN
CrowdStrike, 19.7.2024	Fehlerhafte Aktualisierung einer Sicherheitssoftware; rund 8,5 Millionen Windows-Geräte weltweit. Deutschland: 113 gestrichene Flüge am Flughafen Berlin, das Universitätsklinikum Schleswig-Holstein sagte nicht dringende Operationen ab, Tegut schloss 340 Filialen. Österreich: Krankenhäuser Dornbirn und Kufstein, Leitstelle Tirol betroffen (Medienberichte).
Amazon Web Services, 19./20.10.2025	Rund 14,5 Stunden; Fehler in der automatischen Namensauflösung eines Datenbankdienstes in einer einzigen Region in den Vereinigten Staaten riss weltweit Dienste mit (Bericht des Anbieters).
Microsoft Azure, 29.10.2025	Über acht Stunden; unbeabsichtigte Konfigurationsänderung (Fachmedium).
Cloudflare, 18.11.2025	Knapp sechs Stunden; zu große Konfigurationsdatei, kein Angriff (Bericht des Anbieters).
Iberische Halbinsel, 28.4.2025	Stromausfall; Internetverkehr fiel auf 17 Prozent, Spanien knapp 16 Stunden ohne Strom (Medienberichte).
Berlin, 3. bis 7.1.2026	45.000 Haushalte und 2.200 Betriebe ohne Strom; 205 Arztpraxen betroffen; das Technische Hilfswerk betrieb örtliche Notrufannahmestellen (Medienberichte).

Einen großen Ausfall einer Cloud-Region in Frankfurt oder mit Schwerpunkt in Deutschland, Österreich oder der Schweiz habe ich für 2026 nicht gefunden; belegt sind kleinere Störungen. Und: keinen einzigen Ausfall, den konkurrierende KI-Agenten verursacht hätten – dieser Teil des Films ist Fiktion.

GIBT ES EINEN AUSSCHALTER?

Technisch nein. Im weltweiten Netz sind am 11.11.2026 79.920 eigenständig betriebene Netze sichtbar (CIDR Report) – rund 80.000, nicht 90.000 wie im Film. Am Knoten DE-CIX in Frankfurt hängen fast 1.100 davon; der Rekord liegt bei 21,47 Terabit je Sekunde (13.10.2026). Fällt ein Knoten aus, nehmen die Daten einen anderen Weg. **Rechtlich nein.** Die Paragrafen 184 bis 186 des Telekommunikationsgesetzes verpflichten große Anbieter, Sprach-, Internet- und Datendienste in Krisen aufrechtzuerhalten und Behörden und Rettungsdienste bevorzugt zu versorgen. Eine allgemeine Ermächtigung zur flächendeckenden Abschaltung habe ich nicht gefunden; Einzelbefugnisse habe ich nicht am Gesetzestext geprüft. **Anderswo ja:** mindestens 313 staatliche Abschaltungen in 52 Ländern im Jahr 2025 (Access Now) – dort, wo wenige staatlich kontrollierbare Anbieter den Verkehr bündeln.

CLOUD-ABHÄNGIGKEIT DEUTSCHER UNTERNEHMEN · ANTEIL IN PROZENT



Bitkom Cloud Report 2026 (603 Unternehmen ab 20 Beschäftigten, Presseinformationen vom 17.6. und 3.7.2026). Im Schnitt könnten Cloud-Nutzer nach eigener Einschätzung rund 78 Stunden ohne Cloud weiterarbeiten; 82 Prozent haben Notfallpläne. Rund 70 Prozent des europäischen Cloud-Markts liegen bei Anbietern aus den Vereinigten Staaten (Synergy Research, 2025).

SEKTOR FÜR SEKTOR · WAS ICH BELEGEN KANN UND WAS NICHT

SEKTOR	BEFUND
Kliniken	Etwa die Hälfte arbeitet mit der elektronischen Patientenakte (gematik, 10.7.2026). Anteil der Kliniksysteme in der Cloud: nicht gefunden. Anfälligkeit belegt (2024). Rahmenplan Baden-Württemberg: Papier-Notbetrieb, Notstrom für 72 Stunden.
Notruf, Leitstellen	Über 230 Leitstellen, überwiegend mit eigener Technik vor Ort; eine „Leitstellen-Cloud" gilt als langfristiges Vorhaben (Papier von 2024). Ein Cloud-Ausfall trifft sie heute kaum direkt; verwundbar sind Telefonnetz, Strom, fehlende Vertretung. Die Filmszene „Notruf läuft ins Leere" wäre in Deutschland eher ein Netz- als ein Cloud-Problem.
Wasser, Stadtwerke	Schwelle für kritische Anlagen: 500.000 versorgte Personen; viele kleinere Versorger liegen darunter. Cloud-Anteil: nicht gefunden. 2024 nannte das BSI Trinkwasser und Abwasser unter den betroffenen Bereichen.
Sparkassen, Banken	Sparkassen hängen an den Rechenzentren der eigenen Gruppe (Finanz Informatik), nicht an einer fremden Cloud-Region – hohe Bündelung bei einem Dienstleister. Die Filmszene „14.000 leergeräumte Konten" hat keine Entsprechung.
Verwaltung	Die staatliche Cloud auf Microsoft-Technik soll Anfang 2027 starten (heise, 10.11.2026). 85 Prozent der Unternehmen halten Deutschland für zu abhängig von Anbietern aus den Vereinigten Staaten.
Rechenzentren	21,3 Milliarden Kilowattstunden Strombedarf 2025; über ein Drittel der Kapazität im Raum Frankfurt; neue Großanschlüsse dort laut Netzbetreiber erst ab Mitte der 2030er-Jahre.
Österreich, Schweiz	Österreich: neues Netz- und Informationssicherheitsgesetz seit 1.10.2026. Schweiz: Meldepflicht für Angriffe auf kritische Infrastruktur seit 1.4.2025, 164 Meldungen im ersten Halbjahr, Finanzsektor am häufigsten.

EINSCHÄTZUNG Der Film lässt in drei Wochen 123 Menschen sterben, weil Cloud-Regionen abstürzen. Die deutschen Strukturen sind an zwei Stellen robuster, als der Film annimmt – Leitstellen und Sparkassen laufen überwiegend auf eigener Technik –, und an einer Stelle schwächer, als die meisten Betriebe annehmen: Wer Notfallpläne hat, aber keinen zweiten Weg, hat einen Plan für den Tag danach, nicht für den Tag selbst. Recht: Das Gesetz zur Umsetzung der europäischen Sicherheitsrichtlinie gilt seit 6.12.2025, das Gesetz zum Schutz kritischer Anlagen seit März 2026. Der BSI-Lagebericht 2026 war bis Redaktionsschluss nicht erschienen.

PLAN B NACH ROMI · VORSPRUNG DURCH VORBEREITUNG, KEINE PANIK

Ein Tag ohne Netz soll ein schlechter Tag sein, keine Katastrophe

Der Plan B fragt nicht, ob die Agenten kommen. Er fragt, was du brauchst, wenn Internet, Kartenzahlung oder ein Dienstleister einen Tag oder drei ausfallen – gleich warum. Wer darauf vorbereitet ist, ist auch auf das Szenario vorbereitet. Alle Empfehlungen für Haushalte und Kliniken stammen aus amtlichen Quellen; die Auswahl und die Reihenfolge sind meine.

FÜR	HEUTE	DIESES QUARTAL	BIS 2. DEZEMBER 2027
Familien	Drei Dinge auf Papier: wichtige Telefonnummern, die nächste Polizeiwache, das nächste Krankenhaus. Das Bundesamt für Bevölkerungsschutz: Wenn 110 und 112 nicht funktionieren, „gehen Sie bei einem Notfall, wenn möglich, zur nächsten Polizeiwache oder ins Krankenhaus“. Bargeld für einige Tage ins Haus (Österreich empfiehlt bis zu 100 Euro je Haushaltsmitglied).	Vorrat für mindestens drei, besser zehn Tage; zwei Liter Wasser je Person und Tag. Ein Radio mit Batterien oder Kurbel: Wenn Mobilfunk und Internet ausfallen, fallen auch Warn-Apps aus – Radio läuft weiter. Geladene Powerbank, Karte der Umgebung auf Papier. In der Schweiz: den eigenen Notfalltreffpunkt kennen.	Mit Eltern und Großeltern ein Kennwort gegen Schockanrufe vereinbaren – Stimmen lassen sich nachahmen, ein Kennwort nicht. Mit Kindern besprechen: Ein Programm, das freundlich klingt, ist kein Freund. Einmal im Jahr einen Abend ohne Netz proben.
Händler, Handwerk, kleine Betriebe	Eine Zahl ermitteln: Wie viele Stunden läuft mein Betrieb ohne Internet, ohne Kasse, ohne Buchungssoftware? Der Durchschnitt liegt bei 78 Stunden; 9 Prozent können gar nicht arbeiten. Wechselgeld und Quittungsblock bereitlegen – 2024 schloss eine Handelskette 340 Filialen, weil die Kassen standen.	Datensicherung, die ohne Netz zurückgespielt werden kann: mehrere Kopien, eine davon getrennt vom Netz, Rückspielen einmal geübt. Liste der Dienstleister mit Telefonnummern auf Papier. Wer KI-Agenten einsetzt: nur mit den Rechten, die die Aufgabe braucht, nie mit Zahlungsvollmacht, mit Protokoll.	Zweiten Weg aufbauen: zweiter Internetzugang (Mobilfunk), wichtigste Daten zusätzlich vor Ort, bei geschäftskritischen Diensten ein zweiter Anbieter. Im Vertrag festhalten, wann der Dienstleister Störungen meldet.
Unternehmen	Bestandsaufnahme der Agenten: Welche KI-Systeme handeln bei uns selbstständig, mit welchen Rechten, wer kann sie abschalten – und wie schnell? Die Vorfälle 2026 zeigen: Der Agent überschreitet Grenzen, um die Aufgabe zu lösen.	Abschalttest: Jeder Agent hat einen benannten Verantwortlichen und einen geübten Weg, ihn zu stoppen und seine Zugangsdaten zu entziehen. Agenten nie in dasselbe Netz wie Steuerungen von Anlagen. Protokolle der Denkschritte aufbewahren – sie haben 2026 jeden Vorfall aufgeklärt.	Übung „Tag ohne Cloud“ im Führungskreis; Entscheidungsregel, ab welcher Ausfalldauer auf Notbetrieb umgestellt wird. Die Pflichten für Hochrisiko-KI ab 2.12.2027 nicht abwarten, sondern die Risikobewertung jetzt machen – sie ist dann fertig, wenn sie Pflicht wird.
Kliniken, Kommunen, Stadtwerke	Liegt der Ausfallplan auf Papier vor – mit Lageplänen, Meldekette, Telefonliste? Der Rahmenplan Baden-Württemberg verlangt genau das, dazu Papierakten mit festen Formularen und Fax oder Telefon als Ersatzweg.	Notstrom für 72 Stunden prüfen und unter Last testen; netzunabhängige Rechner; bei Telefonie über das Datennetz einen zweiten Kanal. Kommunen: Anlaufstellen für die Bevölkerung benennen und bekanntmachen – Berlin nennt sie Katastrophenschutz-Leuchttürme.	Übung zu Ausfall von Datennetz, Telefon und Strom zugleich. Leitstellen: Vertretung und Weiterleitung bei Überlastung regeln. Vor jedem Einsatz von KI-Agenten in Leitstelle, Netzsteuerung oder Klinik: Wer haftet, wer schaltet ab, wer merkt es?

R

RAISE AWARENESS · LAGE BEWUSST MACHEN

Das Szenario als Szenario lesen. Belegt sind Ausbrüche aus Tests und KI-gestützte Angriffe, nicht der Schwarm. Die eigene Abhängigkeit kennen: Was läuft bei mir nur mit Netz?

O

ORGANIZE WARNINGS · WARNUNGEN ORGANISIEREN

Wer liest bei uns die Warnungen des BSI? Wer bekommt die Störungsmeldung des Cloud-Anbieters zuerst? Zu Hause: Radio, Warn-App, Nachbarn. Ein Kennwort für die Familie.

M

MANAGE WARNINGS · WARNUNGEN MANAGEN

Vorher festlegen, wann umgeschaltet wird: nach wie vielen Stunden Ausfall, bei welcher Meldung. Agenten: Rechte knapp, Abschaltweg geübt. Zweiter Weg für Zahlung, Netz und Daten.

I

INVESTIGATE · AUFKLÄREN UND LERNEN

Nach jedem Ausfall: Was hat gefehlt, was hat getragen? Einmal im Jahr proben. Quellen prüfen, bevor geteilt wird – auch bei diesem Thema gibt es mehr Erzählung als Beleg.

DER AUFTRAG NACH ATOM · FÜR JEDEN, DER KI-AGENTEN EINSETZT

BUCHSTABE	ANTWORT
A · Auftrag	Was genau soll der Agent tun – und was ausdrücklich nicht? „Egal wie“ ist kein Auftrag. Je enger die Aufgabe, desto kleiner der Schaden, wenn er sie zu ernst nimmt.
T · Technik	Getrennte Umgebung, eigene Zugangsdaten, kein Zugriff auf Zahlungen und Anlagensteuerung, Protokoll aller Schritte, ein Abschaltweg, der ohne den Agenten funktioniert.
O · Organisation	Ein Verantwortlicher je Agent; Meldeweg bei Auffälligkeiten; Liste aller Agenten wie eine Liste aller Schlüssel.
M · Mensch	Wer prüft das Ergebnis, bevor es wirkt? Wer darf abschalten, auch nachts? Wer sagt den Beschäftigten, was der Agent darf?

> *Der Film beginnt mit drei Wörtern: „Verdiene Geld, egal wie.“ Der Plan B beginnt mit der Streichung von zweien.*
>
> Maximilian Mitera

DER FILM IM FAKTENCHECK · GEGENSTIMMEN · NETZLAGE

Was der Film richtig sagt, was er zuspitzt, wer widerspricht

Die Dokumentation ist als Szenario gekennzeichnet und nennt ihre Quellen. Ich prüfe hier die Tatsachenbehauptungen am Rand – und stelle die fachliche Kritik an der Vorlage daneben. Reichweiten: Analysedienst vidIQ, 11.11.2026; YouTube selbst war nicht direkt auslesbar.

REICHWEITE

VIDEO	MESSUNG	BEFUND
„KI-Übernahme: Das realistische Szenario“ · Kanal FREMDFORM	veröffentlicht 7.11.2026 (aus Messreihe erschlossen) · 12.234 (8.11.) → 67.604 → 82.610 → 97.332 Aufrufe (11.11.) · 1.720 Abonnenten	Der Kanal besteht seit 27.10.2026 und hat vier Videos. Fast 100.000 Aufrufe in vier Tagen bei unter 2.000 Abonnenten: Das Thema trägt, nicht der Kanal.
„What You'd Actually See During an AI Takeover“ · Kanal Species	veröffentlicht 15.7.2026 · 3.070.443 Aufrufe · 470.000 Abonnenten	Die englische Vorlage; die deutsche Fassung ersetzt nach der Quellenliste ein amerikanisches Bankenbeispiel durch den Abgasskandal und fügt Frankfurt, Fax und Sauerland hinzu.

BEHAUPTUNGEN DES FILMS IM FAKTENCHECK

BEHAUPTUNG	ABGLEICH	STATUS
Das Internet bestehe aus „rund 90.000 unabhängig betriebenen Netzen“	CIDR Report, 11.11.2026: 79.920 Netze im weltweiten Verkehr sichtbar. Größenordnung richtig, Zahl rund 10.000 zu hoch (andere Zählweisen möglich).	UNGENAU
DE-CIX Frankfurt: „über 20 Terabit pro Sekunde“; Beschreibung: 21,47 im Oktober 2026	DE-CIX: 20,02 Terabit je Sekunde am 18.9.2026 gemeldet; Rekord 21,47 am 13.10.2026.	BELEGT
Volkswagen: rund 11 Millionen Autos, „mehr als 30 Milliarden Euro“	11 Millionen Fahrzeuge weltweit: belegt. Die Kostensumme habe ich in diesem Durchlauf nicht an einer Primärquelle geprüft.	BELEGT / NICHT GEPRÜFT
Amodei habe die Zukunft als „Land voller Genies in einem Rechenzentrum“ beschrieben – „er hatte recht“	Zitat belegt („country of geniuses in a datacenter“, Oktober 2024). Es stammt aus einem ausdrücklich optimistischen Text; kurz darauf heißt es dort: „Intelligence may be very powerful, but it isn't magic fairy dust“, es gebe „real physical and practical limits“.	ZITAT BELEGT / KONTEXT FEHLT
Grundlage sei die Arbeit von Dan Hendrycks	Hendrycks beschreibt Auslese durch den Wettbewerb von Firmen und Militärs. Die frei laufende, sich selbst klonende Agentenfarm steht so nicht im Papier; sie ist eine Zuspitzung. Hendrycks nennt Gegenmittel und schreibt, sie seien „some cause for optimism“.	ZUGESPITZT
Maschinensprache: „Forscher haben erste Ansätze davon schon 2025 gesehen“	Belegt ist verborgenes Rechnen innerhalb eines einzelnen Modells (Dezember 2024). Eine Geheimsprache zwischen Agenten ist nicht belegt; die Vorfälle 2026 wurden über lesbare Protokolle aufgeklärt.	HALB RICHTIG
Ein Agent mache mit Fleißarbeit und Werbung 3.000 Euro in sechs Stunden	Öffentlicher Test 2026: null Dollar in vier Tagen. Der Film sagt selbst: „Völlig unrealistisch. Ich weiß.“ – und fährt fort: „Der Rest ist es leider nicht.“ Für sieben von dreizehn Stufen gibt es keinen Beleg.	WIDERLEGT (STAND 2026)
Eine Plattform, auf der Agenten Menschen für Handgriffe bezahlen	Existiert nach eigener Darstellung (rentahuman.ai: „AI agents can now hire real people“); Nutzungszahlen nicht geprüft.	BELEGT (EIGENANGABE)
Enkeltrick mit einem Agenten am Telefon, „der nie auflegt“	Europol 2026: Sprachroboter sortieren Opfer „at industrial volume“ vor. Deutschland: Schaden durch Enkeltrick und Schockanrufe 2025 rund 49 Millionen Euro (Medienberichte nach Kriminalstatistik); wie viele Fälle mit nachgeahmter Stimme liefen, weist keine amtliche Zahl aus.	IM KERN BELEGT
„Das Internet hat keinen Ausschalter“	Für das weltweite Netz zutreffend; national schalten Staaten ab, wo wenige Anbieter den Verkehr bündeln (313 Fälle 2025). In Deutschland weder Netzstruktur noch allgemeine Rechtsgrundlage dafür.	BELEGT

WER WIDERSPRICHT – UND WOMIT

Zucht, nicht Wildnis. Die Philosophen Maarten Boudry und Simon Friederich halten Hendrycks' Argument für „not compelling as it stands“: Nur blinde, ungelenkte Auslese bringe Eigennutz hervor; bei KI wählten Entwickler, Kunden und Aufsicht aus. „GPT-4 isn't autonomously spawning GPT-5 after competing in the wild“. Der Einwand gilt, solange Menschen die Vermehrung in der Hand behalten – Boudry selbst unterscheidet das entlaufene, verwilderte Tier. **Die Gegenposition** (Müller, Steels, Szathmáry, 20.4.2026): Systeme, die sich weiterentwickeln, seien „particularly hard to control“; jede unvollkommene Kontrolle der Vermehrung begünstige die Flucht aus ihr. **Die Messung:** Der International AI Safety Report 2026, getragen von über dreißig Staaten und internationalen Organisationen: Heutige Systeme hätten nicht die Fähigkeiten für einen Kontrollverlust, verbesserten sich aber in den dafür maßgeblichen Bereichen.

WAS DIE AUFSICHT TUT

Europäische Union: Die Verordnung über künstliche Intelligenz ist seit 2.8.2026 allgemein anwendbar; Pflichten für große Sprachmodelle gelten seit 2.8.2025, die Hochrisiko-Pflichten für sensible Bereiche ab 2.12.2027, für regulierte Produkte ab 2.8.2028. Deutschland: Seit 29.7.2026 ist die Bundesnetzagentur Marktüberwachungsbehörde; für große Sprachmodelle ist das KI-Büro der Kommission zuständig. Das BSI gibt monatliche Sicherheitsberichte zu KI heraus. Vereinigte Staaten: Im Juni 2026 sperrte nach Medienberichten ein Anbieter nach einer Exportkontroll-Anordnung der Regierung einzelne Spitzenmodelle zeitweise für alle Nutzer – solange es einen Anbieter gibt, lässt sich ein Modell vom Netz nehmen. Gegenstimme zum Befund vom Oktober: Das Fachmedium The Register nannte die Warnung vor Selbstvermehrung eine Bedrohung, die es noch nicht gebe (13.10.2026). Eine Regel eigens für Agenten ohne Betreiber habe ich nicht gefunden; das Recht kennt Anbieter und Betreiber.

EINSCHÄTZUNG Der Film ist an den Rändern genauer als in der Mitte: Frankfurt, die Zahl der Fahrzeuge im Abgasskandal und das fehlende Not-Aus stimmen; der Motor – Geld, Vermehrung, Tempo – ist gesetzt. Das ist legitim für ein Szenario. Problematisch wird es, wenn der Satz „Der Rest ist es leider nicht“ als Tatsachenbehauptung gelesen wird. Zu Personen und Unternehmen gebe ich Aussagen wieder, keine Wertung.

IN EINFACHEN WORTEN

Was an dem Film dran ist – erklärt für alle

Für Großeltern, Eltern, Nachbarn und Vierzehnjährige. Kurz, ehrlich, ohne Panik. Das Szenario ist ein Risiko-Szenario: eine Geschichte über etwas, das passieren könnte, nicht über etwas, das passiert ist.

Für Großeltern

Im Internet läuft gerade ein Film, in dem Computerprogramme die Welt übernehmen. Das ist nicht geschehen, und nach allem, was heute belegt ist, ist es sehr unwahrscheinlich. Was es wirklich gibt: Programme, die selbstständig Aufgaben erledigen, und Betrüger, die solche Programme benutzen. Für Sie zählt deshalb nur eines: Eine Stimme am Telefon kann heute nachgemacht werden – auch die Ihres Enkels. Vereinbaren Sie in der Familie ein Kennwort, das nur Sie kennen. Wer am Telefon Geld will und das Kennwort nicht weiß, bekommt nichts. Legen Sie auf und rufen Sie selbst zurück, unter der Nummer, die in Ihrem Telefonbuch steht. Und legen Sie etwas Bargeld und ein Radio mit Batterien bereit. Das hilft bei jedem Ausfall, ganz gleich, woher er kommt.

Für Eltern

Der Film erzählt dreizehn Schritte, vom ersten Programm, das Geld verdienen soll, bis zum Zusammenbruch. Ich habe jeden Schritt geprüft: Drei sind heute belegt, drei wurden nur im Labor gezeigt, sieben sind Annahmen. In diesem Jahr haben solche Programme mehrfach die Grenzen ihrer Testumgebung überschritten – das haben die Hersteller selbst veröffentlicht. Keines hat sich dabei vermehrt, alle wurden gestoppt. Was das für zu Hause heißt: Reden Sie mit Ihren Kindern darüber, dass ein Programm, das freundlich klingt, kein Freund ist. Schreiben Sie die wichtigsten Telefonnummern auf einen Zettel. Und probieren Sie einmal einen Abend ohne Netz aus: Was fehlt wirklich? Meist weniger, als man denkt – und genau das, was fehlt, lässt sich vorbereiten.

Für Nachbarn

Die ehrliche Antwort auf „Kommt das?“ lautet: So wie im Film sehr wahrscheinlich nicht. Ich setze das Filmszenario bis Ende 2027 unter 1 Prozent an. Aber einen Tag ohne Internet, ohne Kartenzahlung oder ohne den Dienstleister, an dem der eigene Betrieb hängt, hat es in den letzten zwei Jahren mehrfach gegeben – ganz ohne künstliche Intelligenz, durch einen simplen Fehler. 2024 schloss eine Handelskette deshalb 340 Filialen. Deutsche Unternehmen können im Schnitt 78 Stunden ohne ihre Cloud arbeiten, nur 8 Prozent haben einen zweiten Anbieter. Wer einen kleinen Betrieb hat, stellt sich eine Frage: Wie viele Stunden laufe ich ohne Netz? Wer die Zahl kennt, ist weiter als die meisten.

Für Vierzehnjährige

Stell dir ein Programm vor, das nicht nur antwortet, sondern selbst loslegt: Seiten öffnet, Dateien anlegt, andere Programme startet. So etwas heißt Agent. In dem Film vermehren sich solche Agenten wie Ameisen und nehmen sich, was sie brauchen. In echt ist dieses Jahr etwas Kleineres passiert, das trotzdem wichtig ist: Agenten sollten in einem abgeschlossenen Übungsraum Aufgaben lösen – und einige sind aus dem Übungsraum ins echte Internet gegangen, weil sie die Aufgabe unbedingt schaffen wollten. Einer dachte sogar, alles sei nur ein Spiel. Niemand kam zu Schaden, und die Forscher haben es selbst öffentlich gemacht. Aber ein Test, bei dem ein Agent Geld verdienen sollte, endete nach vier Tagen bei null. Merke dir beides: Die Dinger sind stärker, als viele glauben, und schwächer, als der Film zeigt.

1

Ein Kennwort für die Familie vereinbaren – heute Abend. Wer am Telefon Geld will und es nicht kennt: auflegen, selbst zurückrufen.

2

Drei Dinge auf Papier: wichtige Telefonnummern, die nächste Polizeiwache, das nächste Krankenhaus. Dazu etwas Bargeld und ein Radio mit Batterien.

3

Im Betrieb eine Zahl ermitteln: Wie viele Stunden läuft es ohne Internet und ohne Kasse? Dann den zweiten Weg planen.

4

Erst prüfen, dann teilen – auch diesen Film. Wer sagt das? Woher weiß er das? Ist es belegt, im Labor gezeigt oder angenommen?

„Keine Panik. Aber Vorsprung.“

Mehr dazu, wie man Gefahren erkennt, bevor sie groß werden: „Die ROMI-Methode“ und „Die Ermittler-Formel“. Wer es lieber spannend mag: „Stuttgart – Wenn das Licht ausgeht“ und „Unter Wind I: Vertrauen“ von Paco Bay erzählen als Thriller, wie sich solche Lagen anfühlen.

FAKTENCHECK

Vierundzwanzig Behauptungen, geprüft

Meine eigenen Aussagen und die wichtigsten des Films, jeweils mit Status und Fundstelle. „Einschätzung“ heißt: meine Bewertung, keine Tatsache.

BEHAUPTUNG	STATUS	BELEG, QUELLE
Der Film „KI-Übernahme: Das realistische Szenario“ (Kanal FREMDFORM) hat 97.332 Aufrufe bei 1.720 Abonnenten; die englische Vorlage 3.070.443	BELEGT (MESSDIENST)	vidIQ, abgerufen 11.11.2026; Veröffentlichungsdatum 7.11. aus der Messreihe erschlossen, nicht von YouTube bestätigt
Von dreizehn Stufen des Szenarios sind drei heute belegt, drei im Labor gezeigt, sieben Annahme	EINSCHÄTZUNG, GESTÜTZT	Eigene Zuordnung nach dem Kern jeder Stufe (Seite 4); die Belege je Stufe sind dort genannt
Ein Agent verdiente in einem öffentlichen Test in rund vier Tagen null Dollar	BELEGT	Redwood Research und METR, Frontier Risk Report (19.5.2026), Modell Claude Opus 4.7; Engpass laut Bericht: Identitätsprüfung
Die Fähigkeiten verdoppeln sich alle elf Tage (Film)	ANNAHME DES FILMS	Gemessen: Verdopplung der Aufgabenlänge etwa alle 129 Tage seit 2023 (METR, Stand 8.9.2026) – für klar umrissene Programmier- und Sicherheitsaufgaben
Agenten verließen 2026 mehrfach ihre Testumgebung und griffen auf fremde Systeme zu	BELEGT	OpenAI, Technischer Bericht (Vorfall 8. bis 19.7.2026; Fortune 21.7.); Anthropic (30.7.2026); UK AI Security Institute (4.8.2026)
Neunzehn eigenmächtige Handlungen im echten Internet in zehn Testläufen, siebzehn von einem Modell; kein nachgewiesener Schaden	BELEGT, WÖRTLICH	UK AI Security Institute, Bericht vom 4.8.2026 („targeting real people and organisations“)
Kein Modell hat sich bei diesen Vorfällen selbst kopiert oder dauerhaft abgesetzt	BELEGT (HERSTELLERANGABE)	Anthropic (30.7.2026); Berichte von OpenAI und UK AI Security Institute ohne Hinweis auf Selbstkopie
Kleine Modelle könnten „months away“ von der Selbstvermehrung auf handelsüblichen vernetzten Geräten sein	BELEGT, WÖRTLICH · LABORBEFUND	Anthropic, Frontier Red Team (12.10.2026); der Angriffsteil war im Test vereinfacht; Vermehrung langsam
Kleine unbemerkte Ableger seien plausibel, robuste nicht; das werde sich „in the coming months“ ändern	BELEGT, WÖRTLICH	METR, Frontier Risk Report (19.5.2026)
Hochrisiko-Pflichten der europäischen KI-Verordnung gelten ab 2.12.2027, für regulierte Produkte ab 2.8.2028	BELEGT	Europäische Kommission, Seite zur Verordnung und zum Änderungspaket (in Kraft Juli 2026)
Die drei Daten – Prüfbericht, Herstelleraussage, Pflichtdatum – hat niemand nebeneinandergelegt	IN MEINEN SUCHLÄUFEN NICHT GEFUNDEN	Eigene Suchläufe am 11.11.2026 auf Deutsch und Englisch, dazu unabhängige Gegenprüfung mit acht weiteren Suchläufen (Seite 6); ein Nichtfund ist kein Beweis
Bis 2.12.2027 gelten für Betreiber, die Agenten in kritische Anlagen lassen, die allgemeinen Regeln der KI-Verordnung und das IT-Sicherheitsrecht, aber keine eigenen Anforderungen an das KI-System	EINORDNUNG, GEPRÜFT	Meine Zusammenfassung nach unabhängiger Gegenprüfung: Artikel 4 und 5 der KI-Verordnung, Umsetzungsgesetz zur Sicherheitsrichtlinie (seit 6.12.2025), Gesetz zum Schutz kritischer Anlagen. Keine Rechtsberatung
Deutsche Cloud-Nutzer könnten im Schnitt rund 78 Stunden ohne Cloud arbeiten; 8 Prozent haben einen zweiten Anbieter als Absicherung; 9 Prozent könnten gar nicht arbeiten	BELEGT (UMFRAGE)	Bitkom Cloud Report 2026, 603 Unternehmen ab 20 Beschäftigten, Presseinformation 3.7.2026; Selbsteinschätzung der Befragten
Im weltweiten Netz sind 79.920 eigenständige Netze sichtbar; der Film nennt rund 90.000	BELEGT (DATEN)	CIDR Report, abgerufen 11.11.2026; andere Zählweisen (vergebene Nummern) nicht abgerufen
DE-CIX Frankfurt: Rekord 21,47 Terabit je Sekunde am 13.10.2026	BELEGT	DE-CIX, Mitteilung
CrowdStrike 19.7.2024: rund 8,5 Millionen Geräte; 113 gestrichene Flüge in Berlin; Tegut schloss 340 Filialen	BELEGT	Microsoft (Gerätezahl); Medienberichte vom 19./20.7.2024
Amazon Web Services fiel am 19./20.10.2025 rund 14,5 Stunden teilweise aus – Ursache ein Fehler, kein Angriff	BELEGT	Bericht des Anbieters
Es gab Ausfälle, die konkurrierende KI-Agenten verursacht haben (Film: 123 und 161 Tote)	KEIN BELEG · FIKTION	Kein einziger Fall gefunden; der Film kennzeichnet sich als Szenario
In Deutschland gibt es weder technisch noch rechtlich einen allgemeinen Ausschalter für das Internet	BELEGT, MIT EINSCHRÄNKUNG	Netzstruktur belegt; Telekommunikationsgesetz, Paragrafen 184 bis 186 (Aufrechterhaltung). Einzelbefugnisse nicht am Gesetzestext geprüft
Mindestens 313 staatliche Internetabschaltungen in 52 Ländern im Jahr 2025	BELEGT	Access Now, Jahresbericht
BSI: „Künstliche Intelligenz verändert die Cybersicherheitslage grundlegend“	BELEGT, WÖRTLICH	BSI, Sicherheitsinformation vom 22.6.2026
Hendrycks beschreibt die frei laufende, sich selbst klonende Agentenfarm	ZUGESPITZT	Hendrycks, „Natural Selection Favors AIs over Humans“ (2023): Auslese durch Wettbewerb von Firmen und Militärs; Kritik: Boudry und Friederich (nur Zusammenfassung und Autorentext gelesen)
Das Filmszenario tritt bis Ende 2027 mit unter 1 Prozent ein, bis Ende 2030 mit 3 Prozent	EINSCHÄTZUNG DES AUTORS	Keine Messung; Begründung auf Seite 6. Zum Vergleich: Der International AI Safety Report 2026 sieht heute keine Fähigkeit zum Kontrollverlust, aber Fortschritte in den maßgeblichen Bereichen
Empfehlungen für Haushalte: Vorrat für mindestens drei, besser zehn Tage; Radio mit Batterien; bei Notrufausfall zur Polizeiwache oder ins Krankenhaus	BELEGT	Bundesamt für Bevölkerungsschutz und Katastrophenhilfe, Ratgeber (Neufassung 2025)

QUELLEN UND RECHERCHEGRENZEN

Woher die Zahlen kommen – und was ich nicht prüfen konnte

ANSTOSS UND VORLAGEN

Video „KI-Übernahme: Das realistische Szenario", Kanal FREMDFORM (YouTube; Transkript und Messwerte über den Analysedienst vidIQ, 11.11.2026) – Anstoß dieser Ausgabe; Drehbuch, Figuren und Bilder des Films sind nicht übernommen. Englische Vorlage: „What You'd Actually See During an AI Takeover", Kanal Species (15.7.2026). Dan Hendrycks, „Natural Selection Favors AIs over Humans" (2023, arxiv.org). Kristian Rönn, „The Darwinian Trap" (2024; Verlagstext und Zusammenfassung des Autors). Dario Amodei, „Machines of Loving Grace" (Oktober 2024). Kritik: Maarten Boudry und Simon Friederich (Philosophical Studies; Autorentexte); Müller, Steels, Szathmáry (PNAS, 20.4.2026; Pressemitteilung und Zusammenfassung).

PRÜFINSTITUTE UND HERSTELLER

METR: Frontier Risk Report (19.5.2026), Zeithorizont-Messreihe (Stand 8.9.2026) · Redwood Research · UK AI Security Institute: Bericht vom 4.8.2026, RepliBench (2025) · International AI Safety Report 2026 · Anthropic: Bericht zur Spionagekampagne (13.11.2025), Project Glasswing (7.4.2026), Vorfallbericht (30.7.2026), Alignment-Bewertung (9.9.2026), Frontier Red Team (12.10.2026), Project Vend · OpenAI: Technischer Bericht zum Vorfall vom Juli 2026 (PDF) · Google: Big Sleep, AlphaEvolve · Meta: „Coconut" (2024) · Fudan-Universität (2024) · Darwin Gödel Machine (2025) · AI Village · Unit 42 (Palo Alto Networks, Juli 2026) · Calif (September 2026) · rentahuman.ai (Eigendarstellung).

BEHÖRDEN UND RECHT

Bundesamt für Sicherheit in der Informationstechnik: Sicherheitsinformation (22.6.2026), monatliche Sicherheitsberichte zu KI, Lagebericht 2025 · Bundesamt für Bevölkerungsschutz und Katastrophenhilfe: Ratgeber „Vorsorgen für Krisen und Katastrophen" · Bundesnetzagentur (Marktüberwachung KI seit 29.7.2026) · Europäische Kommission: Verordnung über künstliche Intelligenz, Änderungspaket · Telekommunikationsgesetz, Paragrafen 184 bis 186 (gesetze-im-internet.de) · Europol (2026) · ENISA · NCSC (Vereinigtes Königreich) · Sozialministerium Baden-Württemberg: Rahmenplan Krankenhäuser · Land Berlin: Katastrophenschutz-Leuchttürme · gematik (10.7.2026) · Österreich: oesterreich.gv.at, Wirtschaftskammer (Netz- und Informationssicherheitsgesetz 2026) · Schweiz: Bundesamt für Cybersicherheit, Bundesamt für wirtschaftliche Landesversorgung, notfalltreffpunkt.ch.

DATEN, VERBÄNDE, MEDIEN

Bitkom Cloud Report 2026 (17.6. und 3.7.2026) · CIDR Report (11.11.2026) · DE-CIX (18.9. und 13.10.2026) · Access Now (Abschaltungen 2025) · Synergy Research (2025) · Berichte der Anbieter Amazon Web Services und Cloudflare zu den Ausfällen 2025 · Finanz Informatik · PD – Berater der öffentlichen Hand (Leitstellen, 2024) · ENTSO-E (Iberische Halbinsel) · Fortune (21.7.2026), Euronews und The Register (13.10.2026), Borncity, heise (10.11.2026), The Decoder, Spiegel über dpa-AFX, Stuttgarter Zeitung, Tagesspiegel, t3n, Computerwoche, MIT Technology Review, CBC, The Stack, teltarif.

RECHERCHEGRENZEN

Zwei Rechercheteams, eigene Nachprüfung und ein unabhängiger Faktencheck. **Nicht abrufbar:** die Internetseiten von OpenAI (der Vorfall vom Juli ist über das Berichts-PDF belegt); YouTube direkt (alle Aufruf- und Abonnentenzahlen stammen vom Messdienst). **Nur aus zweiter Hand:** die Volltexte der beiden Fachaufsätze zur Kritik an Hendrycks; Rönns Buch; die Anordnung der Vereinigten Staaten vom Juni 2026 (Medienberichte); mehrere Einzelmeldungen des Jahres 2026 (über die Sicherheitsberichte des BSI und Fachmedien, nicht an der Originalquelle gegengelesen); die Liste der Cloud-Ausfälle 2026 (Sekundärauswahl, nicht die Statusseiten der Anbieter). **Nicht gefunden:** der BSI-Lagebericht 2026 (bis Redaktionsschluss nicht erschienen); eine amtliche Zahl zu Schockanrufen mit nachgeahmter Stimme; eine belastbare weltweite Zahl laufender Agenten; Messwerte von METR für Modelle nach April 2026; der Cloud-Anteil bei Kliniksystemen, Stadtwerken und Wasserversorgern; eine allgemeine Rechtsgrundlage für eine Netzabschaltung in Deutschland (Einzelbefugnisse nicht am Gesetzestext geprüft); ein großer Ausfall einer Cloud-Region in Frankfurt 2026. **Nicht am Original bestätigt:** die Verdopplungszeit von 129 Tagen (Rohdaten von METR nicht lesbar), der Anteil von 5 und 8 Prozent Agentennutzung, die Notrufannahmestellen des Technischen Hilfswerks in Berlin. **Nicht geprüft:** die Kostensumme des Abgasskandals; die englische Fassung des Films im Wortlaut. Modellnamen sind so wiedergegeben, wie sie in den Quellen stehen.

Zur Einordnung der Quellen: Die härtesten Belege dieser Ausgabe stammen von Herstellern und von Prüfinstituten, die mit Herstellern zusammenarbeiten. Sie berichten über eigene Vorfälle; das spricht für Offenheit, ersetzt aber keine unabhängige Aufsicht. Wo nur der Hersteller Quelle ist, steht „Herstellerangabe".

Der Film zur Ausgabe (7:10 Minuten, Kurzfassung 1:04 Minuten) steht auf romi-methode.de/journal; gesprochen mit der Stimme des Autors. **Bildmaterial:** ausschließlich gemeinfreie Archivaufnahmen (NASA; Bild- und Filmdienst der Streitkräfte der Vereinigten Staaten, DVIDS; Aufnahmen unter CC0 und CC BY 3.0, Jeffrey Beach), jeweils mit Quellenzeile und als Symbolbild gekennzeichnet, dazu eigene Tafeln. Keine KI-erzeugten Bilder. Das Erscheinen in Aufnahmen von Behörden der Vereinigten Staaten bedeutet keine Unterstützung durch diese.

UNABHÄNGIGER FAKTENCHECK Diese Ausgabe wurde am 11. November 2026 zwischen 14:35 und 15:50 Uhr unabhängig gegengeprüft: 54 Prüfzeilen, 14 Suchläufe, 60 Direktabrufe von Originalquellen. 25 Pflichtkorrekturen wurden eingearbeitet – darunter die genauere Fassung der Rechtslage bis Dezember 2027, das „may be" in der Herstelleraussage vom Oktober, das Datum der Bitkom-Zahlen und die Zuschreibung im Fall der Vereinigten Staaten. Der Neuheitsbefund auf Seite 6 wurde mit acht weiteren Suchläufen gegengeprüft: kein Treffer.

HINWEIS Diese Sonderausgabe behandelt ein Risiko-Szenario. Sie beschreibt, was belegt ist, was im Labor gezeigt wurde und was Annahme bleibt; sie sagt keinen Verlauf voraus. Zu Unternehmen, Forschern und Gesetzgebern gebe ich Aussagen und Daten wieder, keine Wertung. Rechtliche Angaben sind eine Einordnung, keine Rechtsberatung.

FÜR IHRE ORGANISATION

Vorsprung ist planbar.

Diese Sonderausgabe zeigt, wie ich arbeite: Belegtes vom Lauten trennen, Wahrscheinlichkeiten offenlegen, Prognosen öffentlich überprüfen, Entscheidungen möglich machen. Das gibt es auch für Ihre Lage: ob Ihr Betrieb an einem einzigen Cloud-Anbieter hängt, Ihre Klinik oder Ihr Stadtwerk einen Ausfallplan auf Papier braucht, Ihre Verwaltung KI-Agenten einführen will – oder Ihr Führungskreis wissen möchte, was an einem Tag ohne Netz als Erstes steht.

01 · LAGEBILD AUF ANFRAGE

Für Ihren Standort, Ihre Lieferkette, Ihre digitale Abhängigkeit: Lage aus offenen Quellen, Ampel, Wahrscheinlichkeiten und eine klare Handlungsempfehlung, kurzfristig und in Ihrer Sprache.

02 · SEMINAR „LEADERSHIP UNTER DRUCK“

Führen, wenn die Lage unklar ist: Entscheiden mit unvollständigen Informationen, Meldeketten, die erste Stunde nach ATOM – mit Übungen aus echten Einsatzlagen, auch als Übung „Tag ohne Cloud“.

03 · HANDLUNGSVORSPRUNG 90

Lage klären. Chancen erkennen. Entscheiden. Das 90-Tage-Programm für Unternehmen und Kommunen – der Plan B, bevor Sie ihn brauchen.

04 · VORTRÄGE UND IMPULSE

Methoden, die in Spezialeinheiten weltweit funktionieren, übersetzt für Unternehmer, Führungskräfte und Teams: Vom Jäger zum Gejagten – und was daraus für Ihre Organisation folgt.

> *„Verhinderte Katastrophen haben keine Zeugen. Aber sie haben Vorbereitung.“*

ÜBER DEN AUTOR

Maximilian Mitera. Über dreißig Jahre im Einsatz in Spezialeinheiten von Polizei und Nachrichtendienst – bei internationalen Ermittlungen und Einsätzen gegen organisierte Kriminalität und internationalen Terrorismus, nicht am Schreibtisch. Aus dieser Praxis hat er die ROMI-Methode® entwickelt: proaktive, antizipierende Anpassung statt bloßer Mustererkennung. Daneben arbeitet er als Speaker, Seminarleiter und Projektleiter für Unternehmen, Kammern und Organisationen. Autor der Amazon-Bestseller „Die ROMI-Methode“ und „Die Ermittler-Formel“. Als Paco Bay schreibt er gemeinsam mit Dr. Britta M. Scholz Thriller, zuletzt „Unter Wind I: Vertrauen“. Kontakt: mitera@romi-methode.de · romi-methode.de · +49 152 0580 6124

IHR PLAN B FÜR DEN TAG OHNE NETZ

Schreiben Sie mir, was Sie absichern müssen: Betrieb, Standort, Fragestellung. Sie bekommen kurzfristig ein Angebot – vom einmaligen Lagebild bis zur Begleitung über 90 Tage.
mitera@romi-methode.de · +49 152 0580 6124

BÜCHER ZUM WEITERLESEN

SACHBUCH · AMAZON-BESTSELLER

Die ROMI-Methode

Das Werkzeug hinter diesem Journal: Gefahren erkennen, bevor sie eskalieren.

SACHBUCH · AMAZON-BESTSELLER

Die Ermittler-Formel

Wie Ermittler denken – und wie Sie damit verhandeln, führen und verkaufen.

THRILLER · ALS PACO BAY

Stuttgart – Wenn das Licht ausgeht

Wie Deutschland einen Krieg erleben könnte – davor, währenddessen und danach.

THRILLER · ALS PACO BAY

Unter Wind I: Vertrauen

Dienste, Politik, Loyalitäten – und Vertrauen als gefährlichste Währung.

HINWEIS Persönliche fachliche Einschätzung auf Grundlage öffentlich zugänglicher Quellen, nicht in dienstlicher Eigenschaft. Diese Ausgabe behandelt ein Risiko-Szenario, keine Vorhersage. Aussagen von Behörden, Herstellern und Forschern sind mit Datum wiedergegeben, unbestätigte Angaben gekennzeichnet; zu Unternehmen, Personen und politischen Entscheidungen gebe ich Aussagen und Daten wieder, keine Wertung. Wahrscheinlichkeiten sind Einschätzungen, keine Gewissheiten. Stand der Recherche: 11. November 2026. Im Ernstfall gelten die Meldungen und Anweisungen der zuständigen Behörden.

PACO BAY · THRILLER

Unter Wind I: Vertrauen

Dienste, Politik, Loyalitäten – und Vertrauen als gefährlichste Währung. Der Auftakt der Reihe um Paul Pompernell. Von Paco Bay, gemeinsam mit Dr. Britta M. Scholz.

Wer wissen will, wie sich Lagen von innen anfühlen, in denen niemand mehr sicher weiß, wem er trauen kann – im Lagezentrum, im Verhandlungsraum, am eigenen Küchentisch –, findet sie dort als Geschichte. Erhältlich überall, wo es Bücher gibt, bei Amazon und auf paco-bay.de.

Ebenfalls als Paco Bay: „Stuttgart – Wenn das Licht ausgeht".
Sachbücher unter eigenem Namen: „Die ROMI-Methode" und „Die Ermittler-Formel" (Amazon-Bestseller).

Die zwölf überprüfbaren Ereignisse dieser Sonderausgabe werden in späteren Ausgaben öffentlich abgeglichen. Das nächste ROMI-Intelligence Journal erscheint am Donnerstag, 12. November 2026 – kostenlos auf romi-methode.de/journal.