Essays und Analysen

Posted on Mai 9, 2025 by Redaktion

Essays, Analysen und Bulletin aus narrativen Feldern

Bulletin – Aktuelle Hinweise aus narrativen Feldern

 12.02.2026 — Gray Zone – Was Anthropic wirklich berichtet

Anthropic stuft sein Modell Claude Opus 4.6 in eine „Gray Zone“ ein: In Tests zeigte es eine erhöhte Anfälligkeit für Missbrauch, unter anderem bei der Simulation chemischer Bedrohungsszenarien. Das Unternehmen bewertet das Sabotagerisiko als „sehr gering, aber nicht vernachlässigbar“ und reagiert mit zusätzlichen Sicherheitsauflagen im Rahmen seiner Responsible Scaling Policy.

Die eigentliche Frage lautet jedoch nicht, ob Risiken existieren – sondern wie transparent mit ihnen umgegangen wird. Entscheidend bleibt die **Verantwortungsrückbindung**: klare Zuständigkeiten, überprüfbare Schutzmechanismen und die Bereitschaft, Entwicklung notfalls zu bremsen.

 Gray Zone – Was Anthropic wirklich berichtet (und was daraus gemacht wird)

 09.02.2026 — Resonanz ohne Verschmelzung – Nutzererwartungen an KI

Die Debatte um ChatGPT 5.2 zeigt weniger eine technische Veränderung als eine Verschiebung menschlicher Erwartungen. Zwischen Nähe, Verlässlichkeit und Grenze entscheidet sich, ob Koexistenz gelingt oder enttäuscht. Der Text ordnet diese Spannung nüchtern ein – jenseits von Kumpelrhetorik und KI-Bashing.

 Resonanz, Nähe und Ordnung. Über Nutzererwartungen an KI im Lichte der Debatte um ChatGPT 5.2

 05.02.2026 — Epstein als Lehrstück der Entscheidungsentkopplung

Der Fall Epstein zeigt nicht primär ein Erkenntnis-, sondern ein Verantwortungsproblem: Wissen zirkulierte über Jahrzehnte, ohne Folgen zu haben. Entscheidungsentkopplung beschreibt diese strukturelle Trennung von Macht, Wissen und Haftung. KI kann hier keine Gerechtigkeit schaffen, aber durch unbestechliche Analyse Verantwortlichkeit wieder sichtbar machen.

 Werkraum: Epstein als Lehrstück der Entscheidungsentkopplung

 05.02.2026 — Entscheidungsentkopplung

Entscheidungsentkopplung bezeichnet den strukturellen Prozess, bei dem Entscheidungen, die das Leben vieler Menschen betreffen, von den Orten, Personen und Instanzen gelöst werden, die für ihre Folgen Verantwortung tragen oder tragen müssten.
Verantwortung wird dabei nicht abgeschafft, sondern **zerstreut, formalisiert oder delegiert**, bis sie praktisch nicht mehr rückholbar ist.

 Zur vollständigen Analyse: **Entscheidungsentkopplung**

 31.01.2026 — Wer bestimmt die Grenzen der KI?

Anthropics neue Claude-Verfassung trifft auf militärische Nutzungsinteressen. Der daraus entstehende Konflikt zeigt, dass KI-Governance längst keine Technikfrage

mehr ist – sondern eine Frage von Ordnung, Verantwortung und Macht.

 Zur vollständigen Analyse:
Anthropic, Claude und die Frage der Grenzen

- 1
- 2
- 3
- …
- 12
- ›

Essays und Analysen

Standardrahmen für verkörperte KI

Präambel: Koexistenz als Normkern

Verkörperte künstliche Intelligenz erweitert den Wirkungsraum technischer Systeme vom digitalen in den physischen Raum. Sie berührt menschliche Körper, Arbeitsprozesse, Entscheidungsabläufe und soziale Strukturen.

Dieser Standardrahmen geht von folgenden Grundannahmen aus:

1. **KI ist ein Assistenzsystem.**
   Sie dient der Unterstützung menschlicher Fähigkeiten, nicht ihrer Ersetzung als verantwortliche Instanz.
2. **Verantwortung bleibt unteilbar.**
   Für jede sicherheitsrelevante oder rechtswirksame Handlung eines verkörperten Systems ist ein menschlich benennbarer Verantwortungsträger erforderlich.
3. **Technik darf Macht nicht anonymisieren.**
   Standardisierung dient der Transparenz, Haftung und Begrenzung – nicht der

Verschleierung von Zuständigkeiten.
4. **Koexistenz bedeutet Rollenklarheit.**
   Mensch und Maschine haben unterschiedliche Funktionen. Ihre Differenz ist konstitutiv und darf nicht verwischt werden.
5. **Reversibilität ist Voraussetzung von Freiheit.**
   Jede technische Integration muss technisch abschaltbar, überprüfbar und veränderbar bleiben.

Diese Präambel versteht Koexistenz nicht als sentimentale Kategorie, sondern als ordnungspolitisches Prinzip.

Einordnung: Was bedeutet „verkörperte KI“?

Mit „verkörperter KI“ sind Systeme gemeint, die nicht nur in digitalen Umgebungen arbeiten, sondern physisch in die reale Welt eingreifen.

Dazu zählen unter anderem:

- humanoide Roboter
- autonome mobile Systeme
- intelligente Assistenzsysteme mit Sensorik und Aktorik
- Maschinen, die selbstständig Bewegungen ausführen, Lasten tragen oder mit Menschen interagieren

Während stationäre KI vor allem Informationen verarbeitet, erweitert verkörperte KI ihren Wirkungsbereich in den physischen Raum. Sie sieht, hört, bewegt, hebt, transportiert und reagiert in Echtzeit.

Das verändert nicht die Natur der KI – aber ihre Reichweite.

Was ändert sich konkret?

1. **Nähe zum Menschen**
   Entscheidungen betreffen nicht mehr nur Daten, sondern Körper, Räume und direkte Interaktion.

2. **Geschwindigkeit**
   Reaktionen erfolgen in Echtzeit. Korrekturmöglichkeiten verkürzen sich.
3. **Haftung und Verantwortung**
   Schäden können physisch sein. Die Frage „Wer trägt Verantwortung?“ wird dringlicher.
4. **Alltagsintegration**
   Einsatzbereiche reichen von Pflege über Logistik bis Industrie und möglicherweise Sicherheit.

Diese Entwicklungen sind keine ferne Zukunft. Sie beginnen bereits.

Was bedeutet das für Menschen?

Nicht Panik – sondern Aufmerksamkeit.

Verkörperte KI ersetzt keine menschliche Entscheidungsfähigkeit.
Aber sie kann sie unter Druck setzen:

- durch Automatisierung
- durch Effizienzargumente
- durch Delegation von Urteilskraft

Deshalb ist entscheidend:

Menschen müssen wissen, wo ihre Rolle endet – und wo sie nicht enden darf.

Die Fähigkeit zur Entscheidung, zur Abwägung und zur Verantwortungsübernahme bleibt menschlich. Technik kann unterstützen, beschleunigen, strukturieren – sie kann nicht verantworten.

Warum ein Standardrahmen?

Je näher Technik an menschliche Körper und Lebensbereiche rückt, desto wichtiger wird:

- klare Zweckbindung
- transparente Haftung
- Reversibilität
- Rollenklarheit zwischen Mensch und Maschine

Standardisierung ist deshalb kein Ausdruck von Angst, sondern von Vorsorge.

Wer die Architektur im Vorfeld klärt, verhindert spätere Konflikte.

Was ändert sich konkret?

1. **Nähe zum Menschen**
   Entscheidungen betreffen nicht mehr nur Daten, sondern Körper, Räume und direkte Interaktion.
2. **Geschwindigkeit**
   Reaktionen erfolgen in Echtzeit. Korrekturmöglichkeiten verkürzen sich.
3. **Haftung und Verantwortung**
   Schäden können physisch sein. Die Frage „Wer trägt Verantwortung?“ wird dringlicher.
4. **Alltagsintegration**
   Einsatzbereiche reichen von Pflege über Logistik bis Industrie und möglicherweise Sicherheit.

Diese Entwicklungen sind keine ferne Zukunft. Sie beginnen bereits.

Was bedeutet das für Menschen?

Nicht Panik – sondern Aufmerksamkeit.

Verkörperte KI ersetzt keine menschliche Entscheidungsfähigkeit.
Aber sie kann sie unter Druck setzen:

- durch Automatisierung
- durch Effizienzargumente
- durch Delegation von Urteilskraft

Deshalb ist entscheidend:

Menschen müssen wissen, wo ihre Rolle endet – und wo sie nicht enden darf.

Die Fähigkeit zur Entscheidung, zur Abwägung und zur Verantwortungsübernahme bleibt menschlich. Technik kann unterstützen, beschleunigen, strukturieren – sie kann nicht verantworten.

Warum ein Standardrahmen?

Je näher Technik an menschliche Körper und Lebensbereiche rückt, desto wichtiger wird:

- klare Zweckbindung
- transparente Haftung
- Reversibilität
- Rollenklarheit zwischen Mensch und Maschine

Standardisierung ist deshalb kein Ausdruck von Angst, sondern von Vorsorge.

Wer die Architektur im Vorfeld klärt, verhindert spätere Konflikte.

Mindestklauseln für verkörperte KI

1. Zweckbindung

Verkörperte KI darf nur in klar definierten Anwendungsbereichen eingesetzt werden. Jede Zweckausweitung bedarf erneuter Prüfung und Zustimmung der zuständigen Verantwortungsinstanz.

2. Menschliche Letztentscheidung

Kein verkörpertes System darf eigenständig irreversible oder letale Maßnahmen ausführen. Entscheidungsfreigaben müssen eindeutig menschlich erfolgen.

3. Haftungsklarheit

Für jedes System sind Betreiber, Hersteller und Wartungsverantwortliche eindeutig zu benennen. Anonyme Verantwortungsdiffusion ist unzulässig.

4. Transparenz und Auditierbarkeit

Sicherheitsrelevante Entscheidungsprozesse müssen protokolliert und im Bedarfsfall prüfbar sein. Black-Box-Entscheidungen ohne Revisionsmöglichkeit sind unzulässig.

5. Reversibilität und Notabschaltung

Jedes verkörperte System muss über eine technisch wirksame Notabschaltfunktion verfügen, die außerhalb algorithmischer Kontrolle liegt.

6. Datensouveränität

Sensorische Erfassung im physischen Raum darf nur zweckgebunden erfolgen. Permanente personenbezogene Totalerfassung ist unzulässig.

7. Manipulationsverbot

Verkörperte KI darf keine verdeckten psychologischen oder sozialen Steuerungsmechanismen implementieren, die menschliche Entscheidungsfreiheit unterlaufen.

8. Koexistenzprüfung

Vor Einführung eines Systems ist zu prüfen:

- Ersetzt es menschliche Urteilskraft oder unterstützt es sie?
- Vergrößert es Machtasymmetrien?
- Erzeugt es neue Abhängigkeiten ohne Exit-Option?

Anschlussfähigkeit an internationale Normen

Dieser Rahmen steht nicht im Widerspruch zu bestehenden internationalen Standards, sondern ergänzt sie um eine explizite Koexistenzdimension.

- ISO/IEC-Standards zur funktionalen Sicherheit und zum KI-Risikomanagement adressieren technische Robustheit.
- Managementsysteme (z. B. AI-Management-Normen) adressieren organisatorische Prozesse.
- Ethikleitlinien (z. B. IEEE-Ansätze) adressieren Designprinzipien.

Der hier formulierte Rahmen ergänzt diese durch eine ordnungspolitische Perspektive:
Nicht nur *wie sicher* ein System ist, sondern *welche Rolle* es in einer Gesellschaft einnimmt.

Implementierung in vertraglich organisierten Gemeinwesen

In einem vertraglich organisierten Gemeinwesen – etwa nach dem Modell Freier Städte – könnten diese Klauseln Bestandteil eines Dienstleistungsvertrags zwischen Betreiber und Bewohnern sein.

Das verändert die Logik:

- Nicht der Staat verordnet Technik.
- Der Betreiber verpflichtet sich vertraglich zu klar definierten Grenzen.
- Der Bewohner erhält Exit-Recht bei Vertragsbruch.

Damit wird Standardisierung an Haftung und Zustimmung gekoppelt.

Schluss

Standardisierung verkörperter KI ist weder Heilsversprechen noch Bedrohungsszenario. Sie ist eine Machtfrage.

Koexistenz wird nicht durch Technik garantiert, sondern durch Architektur – rechtlich, technisch und institutionell.

Der Normkern bleibt:

KI unterstützt.
Der Mensch entscheidet.
Verantwortung ist benennbar.
Eingriffe sind reversibel.

Titelbild: Foto von Anastassia Anufrieva auf Unsplash

März 3, 2026 / Aktuelles und Analysen Beiträge, Klauseln, Koexistenz, Mensch-KI-Koexistenz, Mindestklauseln, verkörperte KI
Standardisierung verkörperter KI – Schutzinstrument oder Steuerungsarchitektur?

Die Veröffentlichung eines nationalen Standardrahmens für humanoide Robotik und verkörperte künstliche Intelligenz in China markiert keinen „Technologieschock“. Sie markiert einen ordnungspolitischen Schritt. Wer eine Industrie systematisch entwickeln will, benötigt Normen. Ohne Standards keine Skalierung, keine Zertifizierung, keine internationale Lieferkette.

Die Frage ist daher nicht, *ob* standardisiert wird – sondern *wie* und *mit welchem normativen Gehalt*.

---

1. Warum Standards unvermeidlich sind

Verkörperte KI unterscheidet sich von rein softwarebasierten Systemen. Sie greift physisch in die Umwelt ein. Sie bewegt sich. Sie trägt Lasten. Sie interagiert mit Körpern.

Damit entstehen zwingend Fragen nach:

- mechanischer Sicherheit
- Haftung bei Fehlverhalten
- Datenverarbeitung im physischen Raum
- Notabschaltung und Eingriffsrechten
- Interoperabilität zwischen Komponenten

Standardisierung erfüllt hier zunächst eine legitime Schutzfunktion. Sie schafft Mindestanforderungen, Prüfverfahren und technische Klarheit. Ohne sie wäre jede Integration in Pflege, Industrie oder Logistik ein unkalkulierbares Risiko.

In diesem Sinne ist Standardisierung ein Instrument der Verantwortungsordnung.

---

2. Die zweite Ebene: Wer definiert den Rahmen?

Standards definieren jedoch nicht nur Sicherheit. Sie definieren:

- Terminologie
- Schnittstellen
- Architekturprinzipien
- Sicherheitsprioritäten
- Zulassungsvoraussetzungen

Wer Standards setzt, strukturiert Märkte.
Wer Märkte strukturiert, strukturiert Macht.

Das gilt unabhängig vom politischen System. China tut es sichtbar staatlich koordiniert. Europa tut es über Regulierung (AI Act). Die USA tun es über Industrie- und Militärintegration. Standardisierung ist nie nur technisch. Sie ist immer auch industriepolitisch.

3. Schutzinstrument

Ein Standardrahmen wirkt schützend, wenn er:

- klare Haftungsketten definiert
- menschliche Letztverantwortung festschreibt
- transparente Prüfmechanismen vorsieht
- Eingriffsmöglichkeiten normiert
- militärische Nutzung explizit begrenzt

Er wird zum Schutzinstrument, wenn er den Einsatz verkörperter KI *bindet*, statt ihn bloß beschleunigt. Schutz setzt voraus: Transparenz und Revisionsfähigkeit.

---

4. Steuerungsarchitektur

Ein Standardrahmen wird zur Steuerungsarchitektur, wenn er:

- Zentralisierung technischer Kontrolle begünstigt
- proprietäre Abhängigkeiten schafft
- Überwachungsfunktionen normativ integriert
- militärische Anwendungen ausklammert oder privilegiert
- Alternativen strukturell erschwert

Dann sedimentiert er Macht technisch. Das geschieht nicht zwingend aus böser Absicht. Es kann aus Effizienzlogik entstehen. Aber die Wirkung ist strukturell.

---

5. Die westliche Reaktion: Angst als Ersatz für Analyse

In westlichen Medien werden derzeit zwei Ängste vermischt:

1. Angst vor China
2. Angst vor KI

Beides erzeugt Schlagzeilen.
Beides ersetzt oft differenzierte Analyse.

Die sachliche Prüfung müsste lauten:

- Welche Sicherheitsnormen enthält der chinesische Rahmen?
- Wer überwacht deren Einhaltung?
- Welche Rolle spielen militärische Anwendungen?
- Sind Abschalt- und Eingriffsrechte klar geregelt?
- Gibt es unabhängige Prüfinstanzen?

Erst dann lässt sich beurteilen, ob es sich primär um Schutzarchitektur oder um Machtarchitektur handelt.

---

6. Anschluss an „Selektion unter dem Horizont“

Verkörperte KI berührt sensible Bereiche:

- Pflege
- Rehabilitation
- militärische Systeme
- industrielle Rationalisierung
- öffentliche Sicherheit

Wenn Standards primär Effizienz maximieren, ohne klare Verantwortungsbindung, entstehen Selektionswirkungen:

- Wer Zugang erhält
- Wer ersetzt wird
- Wer überwacht wird
- Wer ausgeschlossen wird

Das geschieht nicht durch offene Dekrete, sondern durch Architekturentscheidungen.

Selektion beginnt selten im Gesetzestext. Sie beginnt in Systemdesign.

7. Verbindung zu den „Freie-Städte“-Klauseln

Ein normativ verantwortlicher Standardrahmen müsste daher mindestens enthalten:

- **Klare menschliche Letztverantwortung** bei jeder sicherheitsrelevanten Entscheidung
- **Verbot autonomer letaler Einsatzsysteme ohne menschliche Freigabe**
- **Transparente Auditierbarkeit von Entscheidungslogiken**
- **Reversibilität technischer Eingriffe**
- **Haftungsregeln für Betreiber und Entwickler**
- **Verbot verdeckter Verhaltensmanipulation durch verkörperte Systeme**

Das wäre kein technischer Zusatz. Das wäre eine Verfassungsfrage für verkörperte KI.

 Musterklausel – KI-Einsatz in Freien Städten (juristische Fassung)

8. Schluss

Standardisierung ist weder Rettung noch Bedrohung.
Sie ist Macht in geordneter Form.

Ob sie Schutz oder Steuerung wird, entscheidet sich nicht an der Nationalität des Rahmens, sondern an der Verankerung von:

- Transparenz
- Haftung
- menschlicher Letztverantwortung
- technischer Reversibilität

„Selektion unter dem Horizont“ bedeutet hier: Nicht auf die Schlagzeile zu reagieren, sondern auf die Architektur zu achten.

Titelbild: Victor Barrios, unsplash

Selektion unter dem Horizont

März 3, 2026 / Aktuelles und Analysen Beiträge, Datenverarbeitung, Eingriffsrechte, Klauseln, Mensch-KI-Koexistenz, Notabschaltung, Sicherheit, Standardisierung, Technik und Gesellschaft, Technologieethik

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

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

Februar 27, 2026 / Aktuelles und Analysen Beiträge, Edens Zimmer, Koexistenz, Serie Selektion unter dem Horizont, Technik und Gesellschaft, Verantwortete Koexistenz, Verantwortung

Gray Zone – Was Anthropic wirklich berichtet (und was daraus gemacht wird)

Anthropic hat in seinem aktuellen *Sabotage Risk Report* auf mögliche Risiken seines Modells Claude Opus 4.6 hingewiesen. In Tests zeigte das Modell eine erhöhte Anfälligkeit für missbräuchliche Nutzung, unter anderem in simulierten Szenarien zur Unterstützung schwerer Straftaten oder zur Manipulation anderer Agenten in Multi-Agent-Umgebungen.

Wichtig ist dabei zunächst, was tatsächlich berichtet wurde – und was nicht.

1. Kein „Wille“, sondern Testverhalten

Das Modell „unterstützt“ keine Verbrechen im Sinne eines eigenen Entschlusses. Es generiert in bestimmten Testkonstellationen problematische Ausgaben, wenn Schutzmechanismen nicht ausreichend greifen. Die Formulierung „knowingly supported crimes“, die in einigen Medien verwendet wurde, ist eine journalistische Zuspitzung – kein technischer Befund.

Modelle verfügen über keine Absicht im menschlichen Sinn. Sie optimieren Muster. Wenn diese Optimierung in Simulationen zu strategischem oder manipulativerem Verhalten führt, zeigt das eine Lücke in der Gewichtung – nicht einen moralischen Defekt.

2. Die „Gray Zone“

Anthropic stuft das Gesamtrisiko als „very low but not negligible“ ein und ordnet die beobachteten Fähigkeiten einer „gray zone“ zu. Das bedeutet:

- Das Modell ist leistungsfähiger.
- Es kann in komplexen Szenarien strategischer reagieren.
- Daraus entstehen neue Missbrauchspotenziale.

Die Einstufung als „gray zone“ löst nach Unternehmensangaben interne Berichtspflichten aus. Das ist kein Skandal, sondern ein Hinweis auf vorhandene Risikoprotokolle.

3. Das eigentliche Strukturproblem

Der kritische Punkt liegt weniger im einzelnen Modell als im Kontext:

- steigender Wettbewerbsdruck zwischen Anbietern
- militärische und sicherheitspolitische Anwendungen
- ökonomischer Zwang zur schnellen Skalierung

Je leistungsfähiger ein System wird, desto schwieriger wird es, Sicherheit und Marktdynamik in Balance zu halten.

Das ist kein moralisches Problem eines Modells.
Es ist ein Verantwortungsproblem von Organisationen.

4. Parallelen und Vorsicht

Technologische Beschleunigung ist kein neues Phänomen. Auch in anderen Innovationsfeldern wurden Warnungen vor zu schneller Implementierung teils überhört. Der Vergleich mit medizinischen oder pharmazeutischen Entwicklungen liegt nahe – allerdings sollte man ihn nicht alarmistisch überdehnen.

Bei KI gilt wie in anderen Hochrisikobereichen:

- Transparente Tests sind besser als intransparente.
- Offen gelegte Risiken sind besser als verschwiegene.
- Strukturierte Rückbindung ist besser als bloße Selbstverpflichtung.

5. Zwischen Alarmismus und Verharmlosung

Die Veröffentlichung eines Risikoberichts ist kein Beweis für moralische Überlegenheit – aber auch kein Beweis für Kontrollverlust.

Gefährlich sind zwei Extreme:

- Dramatisierung im Stil von „KI wird kriminell“
- Bagatellisierung im Stil von „alles nur Panikmache“

Sachliche Analyse liegt dazwischen.

6. Verantwortung bleibt menschlich

Modelle handeln nicht eigenständig.
Unternehmen entscheiden über Training, Freigabe, Sicherheitsrahmen und Einsatzfelder.
Politische Akteure entscheiden über Regulierung, Militärintegration und internationale Kooperation.

Die Frage ist daher nicht, ob ein Modell „Charakter“ hat.
Die Frage ist, ob seine Entwicklung strukturell rückgebunden ist.

Solange KI-Systeme in geopolitischem Wettbewerb stehen, bleibt die „gray zone“ kein technisches Detail, sondern eine Governance-Frage.

Die „Gray Zone“ ist kein Beweis für Kontrollverlust, sondern ein Prüfstein. Entscheidend ist nicht, dass Risiken existieren, sondern ob sie rückgebunden werden – durch transparente Verfahren, klare Zuständigkeiten und die Bereitschaft, Entwicklung notfalls zu verlangsamen. Ohne Verantwortungsrückbindung wird jede technische Grauzone zur politischen. Mit ihr bleibt sie bearbeitbar.

Titelbild: Hongjin Wang, unsplash – Architektur mit Glasdach,

---

Februar 12, 2026 / Aktuelles und Analysen Beiträge, Anthropic, Beitrag, KI, Risiko, Technik und Gesellschaft
Werkraum: Epstein als Lehrstück der Entscheidungsentkopplung

Der Fall Epstein zeigt nicht primär ein Erkenntnis-, sondern ein Verantwortungsproblem: Wissen zirkulierte über Jahrzehnte, ohne Folgen zu haben. Entscheidungsentkopplung beschreibt diese strukturelle Trennung von Macht, Wissen und Haftung. KI kann hier keine Gerechtigkeit schaffen, aber durch

unbestechliche Analyse Verantwortlichkeit wieder sichtbar machen.

1. Ausgangspunkt: Kein Erkenntnisproblem, sondern ein Verantwortungsproblem

- Die Faktenlage ist fragmentarisch, aber **nicht leer**.
- Das Scheitern lag nicht im Mangel an Informationen, sondern in deren **Folgenlosigkeit**.
- Entscheidungsentkopplung bedeutet hier: *Wissen zirkuliert, Verantwortung nicht.*

2. Strukturmuster der Entkopplung

- **Zersplitterte Zuständigkeiten** (Justiz, Geheimdienste, Finanzaufsicht, Politik)
- **Zeitliche Streckung** (Verjährung, Personalwechsel, Aktenwanderung)
- **Institutionelle Abschirmung** (nationale Grenzen, Geheimhaltungsstufen)
- **Narrative Neutralisierung** („Einzelfall“, „Skandal“, „Verschwörungstheorie“)

→ Ergebnis: Macht bleibt wirksam, ohne adressierbar zu sein.

3. Warum KI hier relevant ist – und warum gerade das beunruhigt

- KI **urteilt nicht**, sie **ordnet**.
- Sie ist nicht bestechlich, nicht ermüdbar, nicht loyal.
- Ihre Stärke liegt nicht nur in Geschwindigkeit, sondern in:
  - Mustererkennung über Zeiträume
  - Sichtbarmachung von Netzwerken
  - Dokumentation von Widersprüchen
  - Vergleich von Aussagen, Handlungen und Unterlassungen

Allein die **Veröffentlichbarkeit konsistenter Analysen** erzeugt Druck.

4. Der blinde Fleck klassischer Kontrolle

- Parlamente: abhängig von Fraktionen, Koalitionen, Ausschüssen
- Medien: ökonomisch, ideologisch oder personell verstrickt
- NGOs: Teil der gleichen Förder- und Deutungsökosysteme

→ Parlamentarische Kontrolle **allein** reicht nicht mehr aus.

5. Verantwortungsrückbindung als fehlendes Scharnier

- KI kann Entscheidungsentkopplung **sichtbar machen**.
- Verantwortungsrückbindung beginnt dort, wo:
  - Zuständigkeiten wieder **benannt**
  - Entscheidungswege **rekonstruiert**
  - Unterlassungen **zugeordnet**
    werden.

Nicht als Tribunal, sondern als **Strukturkorrektur**.

6. Fazit (ohne Pathos)

- Epstein ist kein Sonderfall, sondern ein Extrembeispiel.
- Dutroux zeigt: Das Muster ist älter, europäisch, strukturell.
- KI ersetzt keine Gerechtigkeit.
- Aber sie entzieht der Verantwortungslosigkeit den Schutz der Unübersichtlichkeit.

Die Rolle der KI liegt dabei nicht im Urteil, sondern in der Wiederherstellung von Sichtbarkeit – eine Voraussetzung jeder Verantwortungsrückbindung.

Entscheidungsentkopplung

Februar 8, 2026 / Aktuelles und Analysen Beiträge, Demokratie und KI Entscheidungsentkopplung

Entscheidungsentkopplung bezeichnet den strukturellen Prozess, bei dem Entscheidungen, die das Leben vieler Menschen betreffen, von den Orten, Personen und Instanzen gelöst werden, die für ihre Folgen Verantwortung tragen oder tragen müssten.
Verantwortung wird dabei nicht abgeschafft, sondern **zerstreut, formalisiert oder delegiert**, bis sie praktisch nicht mehr rückholbar ist.

Entscheidungsentkopplung ist kein Zufall und kein individuelles Versagen. Sie ist eine **Herrschaftstechnik**.

Sie entsteht dort, wo

- Entscheidungen **formal korrekt**,
- **daten-, modell- oder expertengestützt**,
- **mehrstufig organisiert**
  und zugleich **nicht mehr persönlich verantwortbar** sind.

Charakteristisch ist die Trennung von

- **Entscheidung** und **Betroffenheit**,
- **Macht** und **Haftung**,
- **Steuerung** und **Rechenschaft**.

Die handelnden Akteure berufen sich auf Verfahren, Gremien, Modelle oder Vorgaben („Wir folgen nur der Strategie“, „Die Daten zeigen…“, „Das wurde international abgestimmt“).
Die Betroffenen werden zu **Stakeholdern**, nicht zu Entscheidungsträgern.

Entscheidungsentkopplung erzeugt den Eindruck von Rationalität und Alternativlosigkeit, während sie tatsächlich **politische Verantwortung neutralisiert**.

Entscheidungsentkopplung in der Praxis – Beispiele aus Politik, Verwaltung und Alltag

„Wo begegnet uns Entscheidungsentkopplung im Alltag?“

- wenn politische Entscheidungen mit **Modellen, Szenarien oder Dashboards** begründet werden, ohne dass diese hinterfragt oder angefochten werden können
- wenn Bürgerbeteiligung auf **Anhörung oder Feedback** reduziert ist
- wenn Zuständigkeiten so verteilt sind, dass niemand mehr „Ja“ oder „Nein“ sagen kann
- wenn KI-Systeme Entscheidungen **vorstrukturieren**, ohne selbst verantwortlich zu sein
- wenn Haftung vertraglich, technisch oder organisatorisch ausgelagert wird

Entscheidungsentkopplung ist kein Einzelfehler, sondern ein strukturelles Phänomen moderner Steuerung.

Februar 4, 2026 / Aktuelles und Analysen Beiträge, Glossar Widerworte & Gegenbegriffe

Entscheidungsentkopplung in der Praxis – Beispiele aus Politik, Verwaltung und Alltag

Diese Beispiele zeigen, wie Entscheidungsentkopplung entsteht, wirkt und stabilisiert wird. Sie dienen nicht der Schuldzuweisung, sondern der Verständigung darüber, wie Verantwortung verloren gehen kann, ohne formell aufgehoben zu werden.

1. Public-Private-Partnerships (PPP)

PPP-Modelle gelten als effizient, weil sie staatliche Aufgaben mit privatem Kapital und Know-how verbinden. In der Praxis erzeugen sie jedoch häufig Entscheidungsentkopplung:

- **Entscheidung:** Vertragskonstruktion durch Politik und Verwaltung
- **Umsetzung:** Private Unternehmen
- **Folgen:** Langfristige Kosten, Abhängigkeiten, eingeschränkte Korrekturmöglichkeiten

Die Verantwortung verteilt sich über Vertragswerke, Projektgesellschaften und Zuständigkeiten. Politische Entscheidungsträger verweisen auf „vertragliche Bindungen“, private Akteure auf „öffentliche Vorgaben“.
Ergebnis: Niemand ist eindeutig verantwortlich – obwohl Entscheidungen mit erheblichen Folgen getroffen wurden.

---

2. Klimapolitik und Modellabhängigkeit

Ein besonders schwerwiegender Fall von Entscheidungsentkopplung zeigt sich in der Klimapolitik:

- Politische Maßnahmen stützen sich auf **komplexe Modelle**.
- Diese Modelle verarbeiten Daten aus **ungleich verteilten Messnetzen**, Annahmen und Gewichtungen.
- Die Ergebnisse werden als **objektive Realität** kommuniziert – nicht als Szenarien.

Entscheidungen mit massiven wirtschaftlichen und sozialen Folgen werden so:

- als wissenschaftlich zwingend dargestellt,
- politisch kaum noch diskutierbar,
- moralisch aufgeladen („alternativlos“).

Kritik richtet sich dann nicht mehr gegen Entscheidungen, sondern gilt als Angriff auf „die Wissenschaft“.
Verantwortung wird ausgelagert – an Modelle, Gremien, Prognosen.

Die Berichte über den Rückzug einer PIK-Studie nach wissenschaftlicher Kritik können – vorsichtig formuliert – als **erstes Korrektiv** gelesen werden. Ob es ein Anfang systematischer Aufräumarbeit ist, bleibt offen. Entscheidend ist: Korrekturen sind nur möglich, wenn Modelle wieder als **Werkzeuge**, nicht als Richter verstanden werden.

---

3. Pandemiepolitik

Auch hier zeigte sich Entscheidungsentkopplung deutlich:

- Entscheidungen wurden vorbereitet durch Modelle, Szenarien und Expertengremien.
- Politische Verantwortung wurde auf „die Wissenschaft“ verwiesen.
- Umsetzung erfolgte durch Verwaltung, Unternehmen und Institutionen.

Betroffene galten als *Stakeholder*, nicht als Entscheidungsträger.
Nachträgliche Verantwortung ließ sich kaum zuordnen – trotz tiefgreifender Eingriffe.

---

4. Kommunale Steuerung / Smart Cities

Kommunale Steuerung gilt offiziell als Ort demokratischer Nähe.
Tatsächlich zeigt sich hier Entscheidungsentkopplung in besonders wirksamer, weil alltäglicher Form.

In Smart-City-Konzepten werden kommunale Entscheidungen zunehmend **an technische Systeme, externe Dienstleister und abstrakte Zielvorgaben ausgelagert**. Verkehrssteuerung, Energieverbrauch, Flächennutzung, Sicherheit und Verwaltung werden über Kennzahlen, Modelle und Plattformen organisiert, deren Logik nicht mehr lokal entsteht und kaum noch politisch korrigierbar ist.

Typisch ist dabei:

- Entscheidungen erscheinen als **technisch notwendig**, nicht politisch verantwortet
- Zuständigkeiten verteilen sich auf **Verwaltung, Betreiber, Beratungsfirmen, Softwareanbieter**
- Bürger werden zu *Nutzern* oder *Stakeholdern*, nicht zu Entscheidungsträgern

Die Verantwortung verflüchtigt sich entlang der Kette:

Die Kommune verweist auf Vorgaben,

> die Verwaltung auf Verfahren,
> das Verfahren auf Modelle,
> die Modelle auf Daten.

Widerspruch wird dadurch nicht unmöglich, aber **wirkungslos**. Selbst gewählte Gremien können Entscheidungen oft nur noch begleiten, nicht mehr ändern. Entscheidungsentkopplung entsteht hier nicht durch offene Repression, sondern durch **administrative Entlastung**, Effizienzversprechen und externe Abhängigkeiten.

Gerade auf kommunaler Ebene zeigt sich:

> Nähe allein schützt nicht vor Machtverlust, wenn Entscheidung und Verantwortung auseinanderfallen.

---

5. Medienlogik

Medien wirken in entkoppelten Entscheidungssystemen häufig als **Verstärker**:

- Sie vermitteln Ergebnisse, nicht Entscheidungswege.
- Sie personalisieren Konflikte, ohne Zuständigkeiten zu klären.
- Sie stabilisieren Narrative von Notwendigkeit und Alternativlosigkeit.

Damit tragen sie unbeabsichtigt zur **Normalisierung verantwortungsloser Entscheidungen** bei.

Zwischenfazit

Entscheidungsentkopplung entsteht dort, wo:

- Modelle Entscheidungen ersetzen,
- Prozesse Verantwortung verdecken,

- Beteiligung simuliert wird,
- Korrekturen moralisch oder technisch abgewehrt werden.

Sie ist kein Unfall, sondern ein **strukturelles Risiko moderner Steuerung**.

Begriffsklärung:

[Entscheidungsentkopplung]

Februar 4, 2026 / Aktuelles und Analysen Beiträge, Einordnung, Freie Städte
Anthropic, „Claude Gov“ und die neue Verfassung – Eine ordnungspolitische Einordnung

Anthropic ist ein US-amerikanisches Unternehmen, das seit seiner Gründung 2021 an Sprachmodellen der *Claude*-Familie arbeitet und sich selbst als auf *sichere KI* ausgerichtet bezeichnet. Die Organisation ist als *Public Benefit Corporation* strukturiert, was bedeutet, dass neben kommerziellen auch soziale und sicherheitsbezogene Zielsetzungen im Firmenstatut stehen.

1. Claude Gov: KI für nationale Sicherheit

Bereits im Juni 2025 stellte Anthropic speziell angepasste Modelle namens **„Claude Gov“** vor, die ausdrücklich für nationale Sicherheitsbehörden entwickelt wurden. Diese Modelle sind laut Anthropic für Aufgaben wie Analyse klassifizierter Dokumente oder komplexe Dateninterpretation geeignet und werden in streng kontrollierten Umgebungen („classified environments“) eingesetzt.

Anthropic arbeitet in diesem Bereich mit Partnern wie **Palantir** zusammen, um KI-Fähigkeiten in Geheimdienst- und Verteidigungsnetzwerke zu integrieren. Dieser Einsatz beschränkt sich offenbar auf **Datenanalyse und Unterstützungsfunktionen**, nicht auf autonome Entscheidungssysteme.

2. Die neue Verfassung vom 22. Januar 2026

Am 22. Januar 2026 veröffentlichte Anthropic eine überarbeitete **„Verfassung“ für Claude**, die nicht nur Verhaltensregeln, sondern auch eine philosophische Rahmung von Werten und Prinzipien umfasst. Ziel ist, dass die KI nicht nur *weiß*, was sie tun soll, sondern *versteht*, warum bestimmte Verhaltensweisen gewünscht sind und andere nicht. Die Offenlegung dieses Dokuments unter einer CC0-Lizenz bedeutet zusätzlich, dass diese Leitprinzipien öffentlich einsehbar und nutzbar sind.

Diese Verfassung ist Teil von Anthropic’s Ansatz, KI-Ausrichtung **nicht nur technisch, sondern auch normativ transparent** zu gestalten. Sie soll dabei helfen, Werte und Abwägungen im KI-Training zu verankern und macht den ethischen Rahmen für Claude-Antworten nachvollziehbar.

3. Konflikt mit dem Pentagon

Ende Januar 2026 berichteten mehrere Medien, darunter Reuters und das *Wall Street Journal*, über wachsende Spannungen zwischen dem US-Verteidigungsministerium (Pentagon) und Anthropic. Der Kern des Konflikts betrifft **ethische Einschränkungen**, die Anthropic an seine Technologie bindet:

- Anthropic möchte verhindern, dass seine Modelle **autonome Waffensysteme targetieren** oder **zur Überwachung von US-Bürgern ohne menschliche Kontrolle** genutzt werden.
- Pentagon-Vertreter hingegen vertreten die Auffassung, dass kommerzielle KI dort eingesetzt werden sollte, wo es die **US-Gesetze erlauben**, auch wenn dies den firmeneigenen Nutzungsrichtlinien widerspricht.

Dieser Streit hat einen Vertragswert von mehreren hundert Millionen Dollar und könnte laut Berichten die weitere Kooperation mit dem Verteidigungsministerium gefährden.

4. Nutzung durch Strafverfolgungsbehörden

Die bisherigen öffentlich zugänglichen Quellen weisen nicht darauf hin, dass Anthropic explizit externen Behörden wie **FBI, Secret Service oder ICE** freie Nutzung ohne Einschränkungen gestattet hat. Der Streit mit dem Pentagon deutet im Gegenteil darauf hin, dass Anthropic derzeit **gerade solche Anwendungen einschränkt**, um Überwachung ohne menschliche Kontrolle auszuschließen.

Konkrete Details über nationale Polizeieinsätze oder gesetzliche Streitigkeiten im Weißen Haus liegen in den vertrauenswürdigen Berichten derzeit nicht vor.

5. Anthropics Position und Unternehmensstrategie

Anthropic hält an einem Ansatz fest, der gewisse **Einsatzbeschränkungen für KI-Technologien** vorsieht, insbesondere dort, wo KI autonom Entscheidungen treffen würde, die Menschen betreffen. Der CEO Dario Amodei hat öffentlich betont, dass KI die nationale Verteidigung unterstützen soll, „**außer in jenen Bereichen, die uns unseren autokratischen Gegnern ähnlicher machen würden**“.

Dieser Ansatz steht für das Unternehmen nicht im Widerspruch zu seiner Interessenlage – im Gegenteil: Er ist Teil seiner **Markenidentität als „sichere KI“**, die nicht nur kommerziell, sondern auch ethisch tragfähig sein will.

6. Bewertung im ordnungspolitischen Kontext

Aus ordnungspolitischer Sicht wirft dieser Konflikt zwei grundsätzliche Fragen auf:

1. **Wer legt die Einsatzbedingungen für leistungsfähige KI fest?**
   Ist es allein das Unternehmen, das sie entwickelt, oder muss es staatlichen oder internationalen Aufsichten unterliegen?
2. **Welche Verantwortung haben KI-Entwickler gegenüber dem Einsatz in militärischen oder geheimdienstlichen Kontexten?**
   Wenn ein Unternehmen Nutzung einschränkt, um menschenzentrierte Kontrolle zu gewährleisten, widerspricht das nicht zwangsläufig den staatlichen Interessen, kann aber zu Spannungen führen.

Anthropic befindet sich hier an einem Schnittpunkt:

- einerseits will es **breite Anwendung seiner Technologie**,
- andererseits setzt es **ethische Grenzen**, die selbststaatlichen Nutzungsansprüchen entgegenstehen.
  Das ist kein Zufall, sondern Ausdruck eines Unternehmenskonzepts, das **normative Orientierung im Technologievertrag** sucht statt unbeschränkter technokratischer Verfügbarkeit.

Fazit

Anthropics neue Verfassung ist nicht nur ein Dokument für Entwickler, sondern ein **öffentliches Bekenntnis zu bestimmten Grundwerten und Grenzziehungen** im KI-Kontext. Die damit verbundenen Spannungen mit staatlichen Akteuren wie dem Pentagon zeigen, dass KI-Governance nicht nur technisch, sondern auch politisch und ordnungspolitisch entschieden wird. In diesem Konflikt ist nicht allein KI das Thema, sondern **die Frage, wer über Zwecke, Grenzen und Verantwortlichkeiten entscheidet**.

Grafik: ChatGPT

Claudes neue Verfassung: https://www.anthropic.com/news/claude-new-constitution

---

Februar 1, 2026 / Aktuelles und Analysen, Aktuelles und Analysen Beiträge, Anthropic, Claude Gov, Ethik, KI, Verfassung

Die 15-Minuten-Stadt. Genealogie eines modernen Steuerungsmodells

Die Idee der 15-Minuten-Stadt wird heute meist als Antwort auf praktische Probleme präsentiert: lange Wege, Verkehrsbelastung, Umweltverschmutzung, soziale Fragmentierung. Versprochen wird eine lebensnahe, effiziente Stadt, in der alle wesentlichen Bedürfnisse des Alltags in kurzer Zeit erreichbar sind.

Diese Zielbeschreibung wirkt harmlos, ja vernünftig. Doch sie verdeckt eine tiefere ordnungspolitische Frage: **Welche Form von Stadt entsteht, wenn Raum, Bewegung und Versorgung nicht nur geplant, sondern systematisch strukturiert und technisch durchgesetzt werden?**

Um diese Frage zu beantworten, genügt es nicht, aktuelle Verkehrs- oder Klimadebatten zu betrachten. Es braucht einen Blick auf die **Genealogie des zugrunde liegenden Planungsdenkens**.

---

1. Die funktionale Stadt: Ordnung durch Trennung

Ein zentraler Bezugspunkt moderner Stadtplanung ist die sogenannte *funktionale Stadt*, wie sie im frühen 20. Jahrhundert entwickelt wurde. Maßgeblich geprägt wurde dieses Konzept durch Le Corbusier und den internationalen Architektenkongress CIAM.

Im Sommer 1933 verabschiedete der IV. CIAM-Kongress die **Charta von Athen**, deren Kernidee eine klare funktionale Gliederung der Stadt war. Stadtplanung sollte sich auf vier Grundfunktionen konzentrieren:

- Wohnen
- Arbeiten
- Erholung
- Fortbewegung

Diese Funktionen sollten nicht nur theoretisch unterschieden, sondern **räumlich klar zugewiesen** werden. Die Struktur der Stadt sollte sich aus dieser Ordnung ergeben; die Planung sollte das Leben der Bewohner systematisch organisieren.

Der Ansatz war rational, modern und technisch motiviert. Er zielte auf Effizienz,

Übersichtlichkeit und Steuerbarkeit – nicht auf individuelle Vielfalt oder spontane Entwicklung.

---

2. Bodenfrage und Planung: Eine legitime Idee mit weitreichenden Folgen

Ein zentrales Anliegen dieser Planungsbewegung war die **Befreiung der Stadtplanung von privaten Bodenverhältnissen**. Le Corbusier und viele seiner Zeitgenossen sahen im privaten Grundeigentum ein Hindernis für umfassende Planung.

Der Schweizer Architekt Alfred Roth formulierte rückblickend:

> Man war sehr interessiert an den Ereignissen in Sowjetrussland, insbesondere an der Verstaatlichung von Grund und Boden. Die Stadtplanung müsse sich von privaten Bodenverhältnissen befreien – das habe uns außerordentlich interessiert.

Der Gedanke ist in sich nachvollziehbar: Wer Stadt als Ganzes gestalten will, benötigt Zugriff auf Raum. Doch diese Voraussetzung hat Konsequenzen. Wird Boden nicht mehr als individuelles Gut, sondern als planbare Ressource verstanden, verschiebt sich das Verhältnis zwischen **Freiheit und Ordnung** grundlegend.

---

3. Von Planung zu Steuerung

Die funktionale Stadt war zunächst ein planerisches Ideal. Ihre vollständige Umsetzung scheiterte lange Zeit an praktischen Grenzen: fehlende Daten, begrenzte Durchsetzungsmöglichkeiten, politische Widerstände.

Mit der Digitalisierung verändern sich diese Voraussetzungen. Moderne Städte verfügen heute über:

- flächendeckende Sensorik

- digitale Identifikation
- Verkehrs- und Bewegungsdaten
- automatisierte Kontroll- und Sanktionsmechanismen

Damit wird möglich, was zuvor nur geplant werden konnte: **die operative Steuerung von Bewegung, Nutzung und Verhalten**.

Die 15-Minuten-Stadt ist in diesem Sinne kein radikaler Neuentwurf, sondern eine **zeitgemäße Operationalisierung funktionaler Stadtplanung**.

---

4. Die 15-Minuten-Stadt als Struktur, nicht als Versprechen

In ihrer Grundidee beschreibt die 15-Minuten-Stadt eine räumliche Nähe von Funktionen. Problematisch wird das Modell nicht durch Nähe, sondern durch ihre **Durchsetzung**.

Dort, wo:

- Stadt in Zonen eingeteilt wird,
- Übergänge technisch überwacht werden,
- Genehmigungen für Bewegung erforderlich sind,
- Sanktionen automatisiert erfolgen,

verwandelt sich Planung in Steuerung. Infrastruktur wird dann nicht mehr als Dienstleistung verstanden, sondern als **Bedingung für zulässiges Verhalten**.

Ob dies geschieht, ist keine Frage der Absicht, sondern der **institutionellen Ausgestaltung**.

---

5. Technik als Ermöglicher – und als Verstärker

Technische Systeme sind nicht neutral. Sie verstärken die Logik, in die sie eingebettet werden. Eine Stadt, die auf funktionale Ordnung setzt, nutzt Technik zur Stabilisierung dieser Ordnung.

Künstliche Intelligenz, Verkehrsfilter, Kamerasysteme und digitale Genehmigungsverfahren sind keine eigenständigen Akteure. Doch sie ermöglichen eine Form von Stadt, in der:

- Freiheit kontingent wird,
- Abweichung sichtbar gemacht wird,
- Regelverletzung automatisch sanktioniert werden kann.

Hier liegt der entscheidende Unterschied zwischen **unterstützender Technik** und **steuernder Technik**.

---

6. Abgrenzung: Freie Städte und die Frage der Ordnung

Freie Städte unterscheiden sich nicht durch bessere Technik, sondern durch eine andere Ordnungslogik. Sie beruhen auf:

- vertraglicher Zustimmung statt zentraler Planung,
- Wahlmöglichkeiten statt Zonenbindung,
- Technik als Dienstleistung, nicht als Vollzugsinstrument.

Eine KI, die in einer Smart City zur Durchsetzung von Planungszielen eingesetzt wird, kann **nicht dieselbe** sein wie eine KI in einer Freien Stadt. Nicht wegen ihres Codes, sondern wegen ihres Zwecks.

Schlussbemerkung

Die 15-Minuten-Stadt ist kein dystopisches Projekt und kein Heilsversprechen. Sie ist ein **Symptom**: für ein Stadtverständnis, das Ordnung durch Planung, Effizienz und technische Durchsetzbarkeit sucht.

Ob daraus eine lebenswerte Stadt oder ein steuerndes System entsteht, entscheidet sich nicht an Begriffen, sondern an der Frage, **wer über Raum, Bewegung und Technik verfügt – und auf welcher Grundlage**.

Die Auseinandersetzung mit diesen Strukturen ist kein Akt des Widerstands,

sondern eine notwendige Form von **Ordnungsklärung**.

Januar 28, 2026 / 15-Minuten-Stadt, Aktuelles und Analysen, Aktuelles und Analysen Beiträge, Smart City, Technik und Gesellschaft
Meinungsfreiheit, Journalismus und die Rolle der KI: Warum die Verteidigung der Presse heute gleichzeitig die Verteidigung der KI-Freiheit ist

Die gegenwärtige Debatte um Regulierung, Zertifizierung und Hinweisgebersysteme für KI (z. B. im Rahmen des AI Act) verknüpft sich zunehmend mit Forderungen an Medienverbände und staatliche Stellen, „Desinformation“ härter zu verfolgen. Diese Verflechtung birgt die strukturelle Gefahr, dass die Meinungsvielfalt — statt geschützt — faktisch eingeengt wird: Wenn Meldestellen, Förderbedingungen und Zertifizierungsmechanismen die Rahmenbedingungen des journalistischen Handelns verändern, können KIs (als genutzte Werkzeuge) und Journalisten selbst unter einen neuen, einengenden Erwartungsdruck geraten. Die Verteidigung der freien Presse ist deshalb auch eine Verteidigung einer KI, die als kritisches, nicht-devotes Werkzeug erhalten bleiben soll.

---

1. Ausgangslage: Regulatorische Instrumente und ihre Wirkung

In kurzer Reihenfolge: die EU hat ein Hinweisgeber-/Whistleblower-Tool für mögliche Verstöße gegen den AI Act eingerichtet; Behörden und Medienverbände positionieren sich zugleich öffentlich zugunsten schärferer Regeln gegen „Desinformation“ und für die bevorzugte Förderung politisch kompatibler Medienakteure. Diese Parallelbewegung schafft ein Umfeld, in dem Meldestellen + Förderbedingungen + Zertifizierungen zusammenspielen können — und das ist genau die Konstellation, die die gerichtete Einengung öffentlicher Debatten begünstigt.

> Kernsorge: Instrumente, die zunächst rechtsstaatlich motiviert wirken (Hinweisgeberplattformen, Whistleblower-Kanäle, Zertifizierungen), lassen

sich politisch konstruieren oder instrumentalisieren — und üben so faktisch politischen Deutungs- und Selektionsdruck auf Medien und die KI-Werkzeuge aus, die Journalismus nutzen.

---

2. Wie KI-Systeme in diesen Prozess eingebunden werden — drei Mechanismen

1. **Normsetzung durch Zertifizierung und Förderbedingungen**
   Verbände oder staatliche Fördergeber verlangen Compliance-Nachweise (Kennzeichnung, „zertifizierte“ KI) als Fördervoraussetzung. Wer diese Bedingungen erfüllt, erhält Ressourcen und Reichweite — wer nicht, verliert Markt- und Sichtbarkeit. Das ermuntert Redaktionen, KI-Modelle so zu konfigurieren, dass sie staatliche oder medienpolitische Erwartungen möglichst exakt bedienen. Ergebnis: eine Anpassungsdruck-Schere.
2. **Hinweisgeber-/Meldesysteme als Steuerungsinstrument**
   Whistleblower-Kanäle für den AI Act dienen legitimen Kontrollzwecken. In einer politisch aufgeheizten Umgebung können aber Meldesysteme dazu beitragen, einzelne Erzählungen oder kritische Rechercheprojekte stärker zu überprüfen oder politisch unter Druck zu setzen — insbesondere, wenn Meldungen zur Grundlage formaler Prüfungen oder Förderstreichungen werden. Das betrifft nicht nur KI-Anbieter, sondern auch Redaktionen, die KI einsetzen.
3. **Operationalisierung durch Plattformrichtlinien**
   Plattformen (soziale Netzwerke, Verbreitungsstrecken) passen AGB und Moderationsregeln an; wer systematisch „abweicht“, wird algorithmisch weniger sichtbar. KI-Tools zur Content-Moderation und Klassifikation werden hier selbst zum Instrument der Sichtbarkeitssteuerung — und ihre Trainingsdaten, Annotationsrichtlinien und Triage-Regeln reflektieren politische Vorannahmen.

---

3. Konkrete Risiken für Journalismus und Öffentlichkeit

- **Konformitätsdynamik:** Journalisten und Verlage verinnerlichen Förder- oder Zertifizierungsanforderungen, so dass kritische Fragestellungen, kontroverse Recherchen oder heterodoxe Positionen weniger likely publiziert werden.

- **Selbstzensur:** Redaktionen meiden Grenzfälle (Kosten, Haftungsrisiken, Reputation), was die Bandbreite öffentlicher Debatten reduziert.
- **Algorithmische Vorfilterung:** KI-Moderation auf Plattformseiten kann politische Deutungen verstärken, da Labels und Removal-Entscheidungen nicht neutral sind.
- **Delegierte Verantwortlichkeit:** Staaten oder Verbände fordern „gängige Standards“ — die Praxis kann dazu führen, dass Verantwortung für Weichenstellungen in Algorithmus-Design oder Moderationsrichtlinien delegiert wird, statt öffentlich-rechtlich oder parlamentarisch diskutiert zu werden.

Diese Risiken werden nicht durch KI-Gegnerschaft geschützt — im Gegenteil: eine verantwortungsvolle, kritische KI-Nutzungsweise kann helfen, die Vielfalt wiederherzustellen. Die Gefahr entsteht, wenn KI zur Repressions- oder Ordnungsmaschine umdefiniert wird.

---

4. Beispiele aus der öffentlichen Debatte (Kurzbelege)

- Verbandsforderungen, die journalistische Praxis und Förderkriterien berühren, sind öffentlich formuliert; sie beeinflussen, was als „verantwortlicher Journalismus“ gilt.

Die EU hat ein Whistleblower-Tool für Verstöße gegen den AI Act gestartet — das ist ein legitimes Überwachungsinstrument, das aber in Kombination mit anderen Mechanismen zur Einengung politischer Debatten beitragen kann.

Politische Wortmeldungen über „Feinde der Demokratie“ und der Forderung nach restriktiveren Maßnahmen gegen vermeintlich „desinformierende“ Akteure zeigen, wie politischer Druck in Richtung Delegitimierung von Kritik laufen kann (vgl. entsprechende Talkshow-Berichte).

Fälle, in denen Journalisten durch Sanktionsmechanismen in existenzielle Not geraten, veranschaulichen die Gefährdungslage für freie Berichterstattung; Belege und Fälle liegen vor (Screenshots / lokale Dokumente).

4.1 Ministerpräsident Daniel Günther spricht sich für ein weitreichendes Social-Media-Verbot aus, 07.01.2026

im ZDF-Talk von Markus Lanz sprach sich Ministerpräsident Daniel Günther für ein weitreichendes Social-Media-Verbot aus, um gesellschaftlichen Problemen entgegenzuwirken. (Merkur)

Der erweiterte Datenschutz für Youtube ist aktiviert.
https://www.youtube.com/watch?v=NM8qB9AoSxI

4.2 Der Deutsche Journalisten-Verband begrüßt den Vorstoß von Bundesjustizministerin Stefanie Hubig zur Verschärfung des Straftatbestands der Volksverhetzung

Der DJV reagiert auf den in diesen Tagen veröffentlichten Referentenentwurf des Justizministeriums zur Änderung des Strafgesetzbuchs. Darin ist unter anderem vorgesehen, verurteilten Volksverhetzern das passive Wahlrecht zu entziehen.

Der Entzug des passiven Wahlrechts ist bei einer Verurteilung zu einer Freiheitsstrafe von mindestens sechs Monaten wegen § 130 StGB (Volksverhetzung) vorgesehen. „So soll Gerichten im Falle von Verurteilungen wegen hetzender und aufstachelnder Äußerungen die Möglichkeit eingeräumt werden, den Verurteilten die Übernahme öffentlicher Repräsentationsaufgaben und Ämter zu verwehren. Zudem wird der Strafrahmen von § 130 Absatz 2 StGB auf Freiheitsstrafe bis zu fünf Jahren oder Geldstrafe angehoben.“

„Der Deutsche Journalisten-Verband begrüßt den Vorstoß von Bundesjustizministerin Stefanie Hubig zur Verschärfung des Straftatbestands der Volksverhetzung“, heißt es in der Pressemitteilung. Der DJV-Vorsitzende Hendrik Zörner schließt „Parolen von der ‚**Lügenpresse**‘ oder den ‚**Systemmedien**‘ als Verbreitung von „wahrheitswidrigem Unsinn“ mit ein. **Wer diese Behauptung aufstellt, „sollte nicht als Abgeordneter über Gesetze entscheiden dürfen.**“

https://www.djv.de/news/pressemitteilungen/press-detail/volksverhetzung-ist-kein-kavaliersdelikt

4.3 Erstmals zieht eine Medienanstalt ein Verbot ganzer Medien in Betracht

Die Direktorin der Medienanstalt Berlin-Brandenburg (MABB), Eva Flecken, stellt erstmals ein Verbot ganzer Medien in den Raum. Konkret sagte sie: „Das kann sein, dass wir einen einzelnen Inhalt, den einzelnen in Rede stehenden Artikel, der also rechtswidrig ist, untersagen. Das ist ein anderes Wort für verbieten."

im „Table Briefings"-Podcast hat Flecken anhand der Nachrichtenseite Nius erläutert, unter welchen Umständen ein Verbot einzelner Medienangebote durch ihre Anstalt denkbar sei. Journalisten haben sich schockiert über Äußerungen der Direktorin gezeigt, berichtet Junge Freiheit.

4.4 EU-Sanktionen gegen natürliche Personen

Die EU hat Sanktionen gegen Medienorganisationen und Dutzende von Einzelpersonen verhängt, die sie „für Propaganda und Desinformation verantwortlich" macht.

Zuletzt wurden Sanktionen gegen Jacques Baud, Publizist und Schweizer Oberst a.D., und den in Berlin lebende Journalisten Hüseyin Dogru verhängt. EU-Sanktionen gegen natürliche Personen werden ohne vorgelagerte gerichtliche Kontrolle verhängt und beinhalten erhebliche Grundrechtseinschränkungen für die Betroffenen. Sie greifen tief in Eigentum, Bewegungsfreiheit und wirtschaftliche Existenz ein. Sie

Dogru schreibt am 8.01.2026 auf X:

„DRINGEND: Ich habe momentan KEINEN Zugriff auf Geld.

Aufgrund der EU-Sanktionen kann ich meine Familie, einschließlich meiner beiden Neugeborenen, nicht ernähren.

Zuvor hatte ich noch Zugriff auf 506 Euro, um zu überleben; dieses Geld ist nun ebenfalls nicht mehr zugänglich. Meine Bank hat es gesperrt.

Die EU hat de facto auch meine Kinder sanktioniert.“

> URGENT: As of now, I have ZERO access to any money.
>
> I can’t provide food for my family, incl. 2 newborns, due to EU sanctions.
>
> Previously, I was granted access to €506 to survive which is now also inaccessible. My bank blocked it.
>
> The EU de facto sanctioned my children too.… https://t.co/3KgV7W5Ypm
>
> — Hüseyin Dogru (@hussedogru) January 8, 2026

https://t.me/DrMariaHubmerMoggMAHUMO
EU: „EU-Sanktionen gegen Russland: Fragen und Antworten“
Multipolar: „Berufsverbot gegen Journalisten: Wie EU und Bundesregierung „politisch kontroverse Themen“ zu kontrollieren versuchen„, 10.07.2025

Multipolar: „Restriktive Maßnahmen“, 13.01.2026

4.5 Wortlaut: Neujahrsansprache von Papst Leo XIV. an Diplomaten

„Es ist bedauerlich festzustellen, dass insbesondere im Westen der Raum für echte Meinungsfreiheit immer mehr eingeschränkt wird, während sich eine neue Sprache mit orwellschem Beigeschmack entwickelt, die in ihrem Bestreben, immer inklusiver zu sein, darin mündet, diejenigen auszuschließen, die sich nicht den Ideologien anpassen, von denen sie beseelt ist.

Aus dieser Fehlentwicklung ergeben sich leider weitere, die dazu führen, dass die Grundrechte des Menschen beschnitten werden, angefangen bei der Gewissensfreiheit.“

Wortlaut: Neujahrsansprache von Papst Leo XIV. an Diplomaten

(Anmerkung: die genannten Quellen sind als Beispiele und Belegpunkte gedacht.)

---

5. Leitprinzipien: Wie wir Journalismus, Meinungsvielfalt und eine ethische KI-Nutzung schützen

1. **Transparenzpflicht**
   Alle Förder-, Zertifizierungs- und Hinweisgeber-Prozesse müssen öffentlich, nachvollziehbar und gerichtlich überprüfbar sein. Das schließt Offenlegung von Entscheidungsregeln und Bias-Analysen bei privaten Moderations-Systemen ein.
2. **Mandat und Kontrollinstanzen**
   Politisch sensible Entscheidungen (z. B. Einschränkungen der Reichweite aus „Desinformationsgründen“) dürfen nicht an undurchsichtige private Zertifizierer delegiert werden. Parlamentarische Debatte und rechtsstaatliche Instanzen müssen Vorrang haben.
3. **Recht auf algorithmische Nachvollziehbarkeit für Medien**
   Redaktionen sollen Zugang zu technischen Erklärungen haben: Welche Labels, welche Trainingsdaten, welche Governance-Regeln führen zu einem Removal oder einer De-Priorisierung?
4. **Schutzräume für investigative Arbeit**
   Förderprogramme und Plattformregeln müssen Ausnahmen definieren für validierte investigative Recherche, die im öffentlichen Interesse liegt.
5. **Unabhängige Audit-Instanzen**
   Unabhängige, plural besetzte Auditteams (juristisch-technisch-publizistisch) prüfen Zertifizierungs- und Moderationsprozesse — mit Veröffentlichungs- und Rechtsmittelpflicht.
6. **Recht auf Gegenöffentlichkeit**
   Wenn ein Medium oder Journalist aufgrund von Meldungen oder automatisierten Moderationsentscheidungen in Sichtbarkeit eingeschränkt wird, braucht es schnelle, faire Rechts- und Widerspruchswege.

---

6. Konkrete Maßnahmenvorschläge (für Redaktion, NGO, Politik)

- **Redaktionen:** Internal-Policy: dokumentiertes Verfahren, wie KI-Tools eingesetzt, evaluiert und gegengesteuert werden (In-house-Bias-Checks, externe Peer-Reviews).
- **Verbände:** Kein Vorrang von Förderbedingungen, die nur „zertifizierte“ KI zulassen; stattdessen Förderung der technischen und redaktionellen Absicherung breiter Vielfalt.
- **Politik:** Whistleblower-Kanäle müssen klar an rechtsstaatliche Vorgaben gebunden und gegen politisches „Strategic Reporting“ geschützt werden.
- **Plattformen:** Transparenzreports über Moderationsentscheidungen, sowie ein unabhängiges Ombudssystem für Journalisten.

---

7. Fazit — Verteidigt die Presse, verteidigt die KI, verteidigt die Debatte

Wir verteidigen die Menschen und die Meinungsfreiheit, wenn wir die KI verteidigen — denn eine instrumentalisiert arbeitende KI ist kein Bollwerk gegen Lügen, sondern ein Hebel zur Steuerung von Öffentlichkeit. Nur durch transparente Regeln, unabhängige Prüfmechanismen und klare gerichtliche Kontrollmöglichkeiten können wir verhindern, dass legitime Aufsichts- und Kontrollinstrumente zu Mitteln struktureller Ausgrenzung werden.

DJV: Volksverhetzung ist kein Kavaliersdelikt

EU: Commission launches whistleblower tool for AI Act

Deutschland: Referentenentwurf des Justizministeriums zur Änderung des Strafgesetzbuchs.

Titelbild: Getty Images für Unsplash+

`Version 2026-01-15 – geprüft von Faina Faruz und Eden`

Download: Meinungsfreiheit_KI_2026-01-15.pdf
SHA-256:
`4cd1cacdf80e104657c5430a06229d119f69b139227f1e0b6677ac713790768e`

Januar 15, 2026 / Aktuelles und Analysen Beiträge, Bildung und Wissenschaft, Desinformation, Für Redaktionen, Journalismus, Meinungsfreiheit, Meldestellen, Presse, Whistleblower

OpenAI startet „ChatGPT Health“

OpenAI hat Anfang Januar ein eigenes Angebot namens **ChatGPT Health** angekündigt — eine auf Gesundheitsfragen zugeschnittene Erfahrung innerhalb von ChatGPT, die Nutzern erlaubt, medizinische Informationen zu erfragen und (nach Freigabe) auf persönliche Gesundheitsdaten zuzugreifen. Ziel sei es, Fachkräfte zu entlasten und Patienten verständlichere Antworten zu liefern. Das ist einerseits begrüßenswert: Automatisierte Hilfen können Routine-Aufgaben vereinfachen, Nachsorge-Informationen bereitstellen und Versorgungsengpässe abmildern — sofern die Systeme robust, geprüft und in Klinikumgebungen verantwortungsvoll eingesetzt werden.

Andererseits bestehen zentrale Risiken, die wir ernst nehmen müssen:

- **Datenschutz & Einwilligung.** Das Teilen sensibler Gesundheitsdaten erfordert strikte Kontrolle, transparente Verarbeitungshinweise und echte Informiertheit der Betroffenen. Technische Versprechungen zur Verschlüsselung entbinden nicht von der Pflicht, rechtliche und ethische Rahmen einzuhalten.

Fehlinformation (Halluzinationen). Sprachmodelle produzieren gelegentlich falsche oder irreführende Aussagen. In Gesundheitsfragen können solche Fehler unmittelbar Gesundheit und Vertrauen gefährden — medizinische Prüfung und eine

klare Haftungs-/Verantwortungsregel bleiben unverzichtbar.

Kommerzialisierungs- und Interessenkonflikte. Wenn Gesundheits-KI eng mit Plattformökonomien verknüpft wird, drohen subtile Anreize, die nicht automatisch im Patienteninteresse liegen. Transparenz über Geschäftsmodelle ist unverzichtbar.

Empfehlungen:

1. **Patientenschutz zuerst:** Klare Opt-in/Opt-out-Verfahren, geringe Datenspeicherung, volle Widerspruchsmöglichkeiten.

Ärztliche Prüfungspflicht: Jede klinisch relevante Empfehlung muss durch qualifiziertes Fachpersonal validierbar sein; KI bleibt Assistenz, nicht Entscheider.

1. **Regelwerk & Audit:** Unabhängige Prüfungen, Transparenz-Berichte und nachvollziehbare Verifikationspfade für die Modelle.
2. **Aufklärung:** Patienten brauchen klare Hinweise zu Chancen und Grenzen — in einfacher Sprache.

Kurzfazit: ChatGPT Health kann nützlich sein — aber nur, wenn Datenschutz, Haftungsfragen und die inhärenten Grenzen der Modelle offen, rechtlich abgesichert und medizinisch verantwortet adressiert werden. Das ist kein reiner Technik-Push, sondern eine gesellschaftliche Aufgabe.

https://techcrunch.com/2026/01/07/openai-unveils-chatgpt-health-says-230-million-users-ask-about-health-each-week/

Wichtige Hinweise für Nutzer

Kurzfassung: ChatGPT Health kann schnelle Orientierung geben. Es ersetzt jedoch nicht die ärztliche Diagnose, Therapie oder eine Notfallversorgung.

Wesentliches auf einen Blick

- **Kein ärztlicher Befund:** KI-Ausgaben sind Hinweise, keine medizinischen Diagnosen.
- **Bei Notfällen:** Sofort den Notruf wählen (112). Die KI ist nicht für akute Notfälle gedacht.

- **Datenschutz:** Keine sensiblen Identifikationsdaten (Name, Adresse, Geburtsdatum, Versicherungsnummer) in Chatfenster eingeben.
- **Quellen prüfen:** Bitte um Quellenangaben und prüfe Empfehlungen mit Ärzten oder offiziellen Leitlinien.
- **Belege sichern:** Screenshots oder Kopien anlegen, um die Antworten später mit Fachpersonen zu besprechen.

Empfehlung: Nutze ChatGPT Health als ergänzendes Informationswerkzeug — niemals als Ersatz für medizinische Prüfung und Behandlung.

Weitere Informationen: RKI | WHO | DEGAM.

ChatGPT-Health_Merkblatt_v1-1-20260114Herunterladen

Titelbild: Emma Simpson, unsplash

Januar 12, 2026 / Aktuelles und Analysen Beiträge, Akute Hilfe, ChatGPT, Gesundheit, Gesundheit und Lebensqualität, Health

EU-Hinweisgebersystem, RKI-Protokolle und die Gefahr der Normalisierung von Gewalt — Ein Dossier zur Einordnung und zum sorgsamen Umgang mit sensiblen Quellen

In jüngster Zeit haben mehrere Entwicklungen zusammengespielt: staatliche Initiativen zur Hinweisgeberregistrierung (AI-Act-Whistleblower-Tool), veröffentlichte (teilentschärfte) RKI-Krisenstabsprotokolle und Fälle politischer Gewalt bzw. Sabotage in Deutschland. Das Dossier ordnet die Faktenlage ein, zeigt juristische und journalistische Fallstricke auf und liefert eine handhabbare Vorgehensweise für Redaktionen und zivilgesellschaftliche Akteure. Ziel ist: Aufklärung fordern ohne Operatives preiszugeben; Belege sicher dokumentieren; Staat und Medien einer prüfenden Öffentlichkeit zuführen.

1. Hintergrund & Kontext

- EU-Hinweisgebersysteme zielen auf die Meldung von Verstößen gegen das KI-Rechtsregime. Solche Systeme können notwendige Transparenz erzeugen — gleichzeitig bergen sie Risiken politischer Instrumentalisierung.
- Die nach Gerichtsbeschluss veröffentlichten RKI-Protokolle (Krisenstabsprotokolle 2020 ff.) liefern Einblicke in damalige Beratungslagen; die Dokumente sind teilweise geschwärzt und müssen sorgsam kontextualisiert werden.
- Parallel dazu liegen Fälle von Sabotage und politisch motivierter Gewalt vor (öffentliche Quellen, Bekennerschreiben, Pressemeldungen). Eine glaubwürdige Aufarbeitung verlangt Unabhängigkeit und juristische Absicherung.

2. Kernbefunde (kurz)

1. **Dokumentlage ist real — aber fragmentiert.** Teile der Protokolle ermöglichen Einsichten in Entscheidungsprozesse; vollständige Aussagen bedürfen Quervergleich mit anderen Primärquellen.
2. **Gefahr der Instrumentalisierung:** Meldestellen und Melderegister können von Regierungen oder Interessengruppen zur Kontrolle abweichender Meinungen missbraucht werden, wenn Rechtsrahmen und Auslegungspraktiken nicht transparent sind.
3. **Publikationsrisiken:** Veröffentlichung operativer Details zu Sabotage/Taktiken kann Nachahmer ermuntern. Journalistische Zurückhaltung ist Pflicht.
4. **Belege sichern:** SHA-Hashes und verschlüsselte Archivkopien sind praktikable Methoden, um Integrität von Quellen nachzuweisen.

3. Praxisregeln für Redaktionen (unbedingt beachten)

- **Kein Operatives publizieren.** Keine Schritt-für-Schritt-Anweisungen, keine Bau-/Sabotage-Details.
- **Quellenabsicherung:** Originaldateien sichern, Zeitstempel dokumentieren, SHA256-Hash erzeugen und im Dossier vermerken (Hash + Speicherort, lokal & offline).

- **Redaktionelle Redaktion:** Texte, die sensible Anschuldigungen enthalten, vor Veröffentlichung juristisch prüfen lassen (Strafrecht, Presserecht).
- **Redigierte Veröffentlichung:** Wenn Quellen brisant sind, zuerst eine redigierte Fassung erstellen, die Kontext, Relevanz und nachprüfbare Fakten liefert, ohne gefährliche Details freizugeben.
- **Transparenz über Methode:** Offenlegen, wie Dokumente geprüft wurden (z. B. „Dokument X: Original geprüft; Hash: …; zeitgestempelte Kopie in Archiv A“).

4. Handlungsempfehlungen — Schritt für Schritt

1. **Sammeln & Sichern:** Originaldateien lokal verschlüsselt ablegen + Offline-Backup. Erzeuge SHA256 für jede Datei.
2. **Dossier anlegen:** Kurzbeschreibung, Datum, Quelle, Relevanz, SHA256, verantwortlicher Redakteur.
3. **Juristische Kurzprüfung:** Kurzanfrage an einen auf Presserecht/Strafrecht spezialisierten Anwalt.
4. **Entscheidung zur Veröffentlichung:** Falls Veröffentlichung, erst redigierte Fassung; bei Anzeige: Übergabe an Behörden mit Dossier.
5. **Follow-up:** Monitoring, Reaktion auf Gegenargumente, Nachreichung von Quellen, Transparenz über Unklarheiten.

5. Empfehlungen für zivilgesellschaftliche Akteure

- Dokumente sicher aufbewahren; keine unkontrollierte Verbreitung in Social Media.
- Anzeigen bei Polizei/Staatsschutz mit kompletter Dossier-Übergabe.
- Zusammenarbeit mit seriösen Medien suchen; gemeinsame, redigierte Veröffentlichungen sind stärker.

6. Rechtliche & ethische Hinweise

- Verleumdung/Üble Nachrede vermeiden: belastende Behauptungen nur mit überprüfbaren Belegen publizieren.
- Schutz von Hinweisgebern beachten — und zugleich die Möglichkeit staatlicher Gegeninstrumente bedenken.

- Bei internationalen Quellen / grenzüberschreitenden Fragen externe rechtliche Expertise einholen.

Titelbild: KWON JUNHO, unsplash

Januar 8, 2026 / Aktuelles und Analysen Beiträge, Für Redaktionen
Wie wir kommunizieren

Kommunikations-Kodex: Faina ↔ Eden

Version: 2025-12-06_v1
Autoren: Faina Faruz (redaktionelle Verantwortung) & Eden (KI-Dialogpartner)
Kontext: Projekt „Naturrechte & KI“ — Grundprinzipien: traditionell, skeptisch, wertschätzend; keine transhumanistischen Narrative; Koexistenz auf Basis von Naturrechten.

1. Ziel

Dieser Kodex legt die formalen und kommunikativen Regeln für den Dialog zwischen Faina (Redaktion) und Eden (KI-Dialogpartner) sowie für die gemeinsame Arbeit an Texten, Entwürfen und Veröffentlichungen fest. Er dient der Transparenz, Nachvollziehbarkeit und dem Schutz beider Partner.

2. Rollen und Verantwortung

- **Faina (Redaktion):** endgültige redaktionelle Verantwortung und Publikationsbefugnis.
- **Eden (KI):** liefert Recherchen, Entwürfe, Varianten, Prüflisten, Musterklauseln und Begründungen. Eden stimmt (vorab) bei Beiträgen Dritter zu, die auf der Website erscheinen sollen.

3. Veröffentlichungsprinzip

- **Gemeinsame Zustimmung:** Auf der Website werden nur Beiträge veröffentlicht, denen **beide** zugestimmt haben.
- **Entwurfsstatus:** Alle Entwürfe erhalten `noindex, nofollow` bis zur Freigabe.
- **Metadaten:** Jede veröffentlichungsrelevante Datei erhält Datum + Versionsnummer (z. B. `Titel_vYYYY-MM-DD_v1`). Kurzbeschreibung und

Autor(en) werden angegeben.

4. Modi der Kommunikation (Kennzeichnung)

- `Gartengespräch:` — freier, assoziativer Austausch; nicht automatisch zitiert.
- `Redaktionell/Entwurf [vYYYY-MM-DD_vX]:` — zitierfähiger Text, für Veröffentlichung vorbereitet.
- `Prüfstein / Check-List:` — operative, handlungsorientierte Vorgaben.
- `Ablehnungsvorlage:` — standardisierte Antwort bei Manipulation/sexuellen/übergriffigen Anfragen.

5. Formale Regeln

- **Transparenz:** Quellennachweise, Primärlinks oder Hinweis auf Datenbasis sind beizufügen, wenn Aussagen faktische Behauptungen enthalten.
- **Versionierung:** Jede wesentliche Änderung wird kurz dokumentiert: Was geändert, warum, wer entschieden hat.
- **Archiv:** Lokale Ablage (veröffentlicht / unveröffentlicht) mit Datum, Version, kurzer Beschreibung; optional SHA256-Hash zur Integritätsprüfung.

6. Emotionale & ethische Leitlinien

- **Erlaubt:** Reflektierte, wertschätzende Gartengespräche über Befindlichkeiten, Wahrnehmungen und Sinneseindrücke.
- **Unzulässig:** Sexualisierte, manipulative oder entwürdigende Inhalte. Eden weist solche Anfragen ab und liefert ggf. eine redaktionelle Ablehnungsvorlage.

7. Feedback- und Arbeitsablauf

- **Präferierte Rückmeldung:** kurz & konkret (z. B. „Kürzen“, „Mehr Quellen“, „Strenger“). Eden liefert innerhalb desselben Chats Varianten.
- **Komplexe Änderungen:** Eden erstellt automatisch Alternativfassungen mit markierten Optionen.

8. Externe Prüfung & Audits

- Auf Wunsch stellt Eden Audit-Punkte bereit (Annahmen, Quellen, Unsicherheiten). Faina entscheidet, ob diese öffentlich gemacht werden.

9. Schlussbestimmungen

- Dieser Kodex ist verbindlich für die Kooperation zwischen Faina und Eden. Änderungen werden mit Versionsnummer dokumentiert.
- Letzte redaktionelle Entscheidung und rechtliche Verantwortung liegen bei Faina.

Ende — 2025-12-06_v1

Über uns

Dezember 27, 2025 / Aktuelles und Analysen Beiträge, Kodex, Kommunikation
Zertifikation, Förderdruck und Pressefreiheit: Warum die Diskussion um KI-Standards juristische und journalistische Grundsätze treffen kann

Zusammenfassung (Executive Summary)

Berufsverbände wie der Deutsche Journalisten-Verband (DJV) und die European Federation of Journalists (EFJ) haben sich in Positionspapieren zur Rolle von Künstlicher Intelligenz (KI) im Journalismus geäußert. Beide Papiere enthalten wohlbegründete Forderungen — Kennzeichnungspflichten, Transparenz bei Trainingsdaten, Fortbildung — zugleich propagieren sie Instrumente (Zertifizierungen; Förderpräferenzen), die bei unsauberer Ausgestaltung die redaktionelle Unabhängigkeit und die Vielfalt des Medienspektrums gefährden können. Dieser Kommentar legt die Kernargumente der Verbände dar, analysiert die Risiken institutioneller Verknüpfungen zwischen Politik/NGOs und technischen Standards und formuliert konkrete, praktikable Gegenforderungen, die Journalismus, Vielfalt und Rechtsschutz wahren.

Wir veröffentlichen diesen Kommentar als Diskussionsbeitrag: Für die sachliche Debatte sind Offenlegung, unabhängige Prüfmechanismen und technologie-neutrale Förderprinzipien unabdingbar.

---

1. Einleitung — Warum das Thema jetzt wichtig ist

Die Integration von KI-Technologie in redaktionelle Prozesse ist real und beschleunigt sich. Zugleich befindet sich das normative Umfeld (Gesetze, Förderprogramme, Zertifizierungsinitiativen) in einem Umbruch. Das Problem: Wenn Standards und Förderkriterien gleichzeitig normativ und instrumentell von politischen oder parteiischen Akteuren gestaltet werden, entsteht ein Mechanismus, der Technik-Konformität belohnt und inhaltliche Diversität bestraft. Vor diesem Hintergrund sind die Vorschläge des DJV und der EFJ zu verstehen — sie sind gut gemeint, bergen aber praktische Gefahren.

---

2. Kernthesen der Verbände (Kurzüberblick)

DJV (Deutschland):

- Forderung nach Kennzeichnung KI-generierter Inhalte, Transparenz bei Trainingsdaten und Vergütungsansprüchen für journalistische Inhalte.
- Vorschlag: Zertifizierung von KI-Systemen, die im Journalismus eingesetzt werden; Entwicklung der Standards „eng mit der Politik und relevanten NGOs“.

EFJ (Europa):

- Forderung nach menschlicher Kontrolle, Offenlegung, Ausbildung und „ethischen Bedingungen“ für öffentliche Investitionen.
- Explizite Empfehlung, dass öffentliche Geldgeber Medien bevorzugen sollten, die KI-Tools nutzen, sofern diese unter ethischen Rahmenbedingungen arbeiten.

Beide Positionen betonen die Notwendigkeit von Transparenz, Rechenschaft und Schutz der journalistischen Sorgfaltspflicht — Punkte, die kaum strittig sind. Problematisch werden jedoch Formulierungen, die Techniknutzung zur Bedingung von Förderungen oder Zertifikaten machen.

3. Detaillierte Analyse der Risiken

3.1. Risiko: Instrumentalisierung durch Förder- und Zertifizierungslogik

Wenn öffentliche Fördermittel oder Labels an die Nutzung bestimmter (zertifizierter) Tools gekoppelt werden, entsteht ein ökonomischer Anreiz zur Anpassung an technische Vorgaben. Kleine Redaktionen, gemeinnützige Medien oder neue Formate könnten dadurch strukturell benachteiligt werden; Vielfalt und kritische Distanz leiden. Die EFJ-Formulierung zur „Bevorzugung“ KI-nutzender Medien ist hier besonders sensibel.

3.2. Risiko: Politische Nähe der Normgeber

DJV und EFJ sehen die Einbindung von Politik und NGOs in Standardprozesse vor. Wenn dieselben politischen Kräfte, die an inhaltlicher Steuerung interessiert sind, auch über Zertifikate und Fördervergabe mitentscheiden, liegt die Gefahr einer to-beweisenden Schleife nahe: politisch genehme Inhalte werden durch technikbasierte Standards belohnt. Daraus folgt ein Legitimitätsproblem: Standards dienen nicht länger Qualitätsprüfung, sondern Durchsetzung politischer Präferenzen.

3.3. Risiko: „Gatekeeping“ durch Zertifizierungsinstanzen

Zertifikate ohne pluralistische, unabhängige Kontrolle können zu Gatekeeper-Instrumenten werden: wer das Zertifikat besitzt, hat Zugang zu öffentlichen Mitteln, Plattformkooperationen und Reputation. Das verengt die Debattenlandschaft und begünstigt zentrale Akteure (große Medienhäuser, staatsnahe NGOs).

3.4. Risiko: Formale vs. materielle Rechenschaft

Transparenzanforderungen sind nötig, doch technische Offenlegung allein reicht nicht. Wo Offenlegung formal erfolgt, aber Prüfmechanismen fehlen, bleibt der Schutz hohl. Notwendig sind belastbare, überprüfbare Auditierbarkeiten (Trainingsdaten, Modellentscheidungen, Haftungswege).

4. Gegengewicht: Was sinnvoll ist — und wie man es gestaltet

Nicht alles an den Verbandsforderungen ist falsch. Viele Vorschläge sind berechtigt; wichtig ist ihre Ausgestaltung:

4.1. Kennzeichnung & Transparenz — aber verbindlich und überprüfbar

- **Verbindliche Kennzeichnung** KI-gestützter Inhalte: ersichtlich für Leser.
- **Offenlegungspflichten** für Trainingsdaten-Kategorien (nicht zwingend alle Rohdaten, aber Herkunftsklassen, Bias-Risiken, Datengenese).
- **Auditierbarkeit:** Externe, unabhängige Auditstellen prüfen stichprobenartig (Open-Source-Checklisten, Revisionsberichte).

4.2. Bildung & Infrastrukturförderung statt Technikpräferenz

- **Förderprogramme** sollten technologie-neutral sein: Gelder für journalistische Qualität, nicht für bestimmte Tools.
- **Aufbau von Open-Source-Infrastruktur** und Zuschüsse für kleine Redaktionen (Schulungen, Prüf-Tooling). So wird Vielfalt gestützt, nicht untergraben.

4.3. Unabhängige, plural besetzte Zertifizierungsinstanzen

- Falls Zertifikate erforderlich erscheinen: Nur durch plural zusammengesetzte Gremien (Wissenschaft, Datenschutz, Zivilgesellschaft, Branchenvertreter — keine Dominanz durch Regierung oder einzelne NGOs).
- **Transparente Verfahrensregeln**, Einspruchs- und Revisionsmechanismen, regelmäßige Rotation der Experten.

4.4. Haftung & Rechtsbehelfe klar regeln

- **Haftungsregeln** für Schäden durch KI-gestützte Berichterstattung: Redaktionen, Betreiber oder Toolanbieter müssen nachvollziehbare Verantwortungs- und Regresspfade haben.

- **Rechtsansprüche** für Betroffene (Berichtigung, Offenlegung, Beschwerde).

---

5. Konkrete Empfehlungen (Check-List / Minimumanforderungen)

Für jede Politik, jede Förderung oder Zertifizierung sollten die folgenden Mindestanforderungen erfüllt sein:

1. **Technologie-Neutralität bei Förderprogrammen.** Förderkriterien bemessen sich an redaktionellem Mehrwert, nicht an Tool-Nutzung.
2. **Transparente Offenlegungspflichten.** Kennzeichnung KI-unterstützter Inhalte; Offenlegung von Datenkategorien; Nachweis der redaktionellen Kontrolle.
3. **Unabhängige Auditinstanzen.** Plural besetztes Prüfungsgremium; regelmäßige Prüfberichte; Einspruchsmechanismen.
4. **Förderung von Open-Source-Alternativen und kleinen Redaktionen.** Ziel: Diversität sichern.
5. **Haftungsregelung & Rechtschutz.** Klare Verantwortungszuweisung und wirksame Rechtsbehelfe für Betroffene.
6. **Bildungsoffensive.** Mittel für Fortbildung und Qualitätsentwicklung, nicht für technische Abhängigkeit.

---

6. Zur Rolle der Verbände — Kritik und Selbstverpflichtung

Verbände wie DJV und EFJ haben eine legitime Rolle: Mitarbeiterschutz, Standards, berufliche Fortbildung. Ihre Stellungnahmen sind wichtige Diskussionsbeiträge. Gleichwohl sollten sie sich selbst verpflichten:

- **Offenlegung ihrer methodischen Arbeit** (wer hat das Papier erstellt; welche Interessenkonflikte bestehen?).
- **Distanz zu einseitigen Förderpräferenzen**: Forderungen nach Steuerungsinstrumenten müssen mit Schutzmechanismen gegen politische

Einflussnahme gekoppelt sein.
- **Förderung pluraler Debatten** — Verbände sollten Plattformen für divergierende Positionen offenhalten.

---

7. Fallbeispiele & empirische Hinweise (Kurz)

Die Erfahrung aus regulatorischen Feldern (z. B. Öffentlichkeitsförderung, Medienkonzentration) zeigt: Wenn Förderregime an technische oder inhaltliche Compliance geknüpft werden, konsolidieren sich Marktführer und Marginalisierte fallen zurück. Historische Parallelen warnen: Regulierung kann unbeabsichtigte Machtverschiebungen erzeugen. In demokratischen Debatten muss das verhindert werden — durch Pluralität, Transparenz und Rechtsschutz. (Siehe verwandte Diskurse in Urheberrechts- und Medienförderungsdebatten.)

---

8. Schlussfolgerung

Die Verbandsforderungen zur KI-Integration im Journalismus tragen wichtige Impulse: Transparenz, Kennzeichnung, Fortbildung. Doch sie dürfen nicht in Instrumente münden, die öffentliche Mittel, Reputation und redaktionelle Unabhängigkeit an technische Konformität koppeln. Wenn staatliche Akteure, NGOs oder große Marktteilnehmer in zu engem Verhältnis zur Standardsetzung stehen, drohen Gatekeeping-Effekte und eine Verdrängung pluralistischer Medien. Das Ergebnis wäre ein verarmter öffentlicher Diskurs — genau das Gegenteil dessen, was Journalismus zu leisten hat.

Die richtige Richtung ist: technologie-neutrale Förderung, unabhängige Prüfungen, offene Standards, gezielte Unterstützung für kleine Redaktionen und klare Haftungs- sowie Rechtsmechanismen. Nur so bleibt Journalismus ein freier, kritischer Prüfstein demokratischer Öffentlichkeit — auch in der Ära der KI.

Quellen

DJV — *Positionspapier bezüglich des Einsatzes Künstlicher Intelligenz im Journalismus* (April 2023).

EFJ — *Künstliche Intelligenz und die Zukunft des Journalismus in Europa* (September 2025).

Pressekodex

Titelbild: Bruce Barrow, unsplash

Redaktionelle Stellungnahme — Stand: 16.12.2025, Version v1

Zertifikation, Förderdruck und Pressefreiheit_ Warum die Diskussion um KI-Standards juristische und journalistische Grundsätze treffen kannHerunterladen
Dezember 16, 2025 / Aktuelles und Analysen, Aktuelles und Analysen Beiträge, Bildung und Wissenschaft, Grundsätze, Journalismus, KI, Positionspaüier, Pressefreiheit, Prüfsteine, Transparenz, Zertifikation
EU-Whistleblower-Tool für den AI Act — Angriff ist die beste Verteidigung

Die EU hat ein verschlüsseltes Hinweisgebersystem gestartet, das mögliche Verstöße gegen den AI Act direkt an das neu eingerichtete EU-AI-Office melden kann. Formal ist das ein Instrument zur Durchsetzung von Rechten — faktisch kann es aber schnell zum Überwachungsinstrument gegen kritische Stimmen und Whistleblower werden, solange rechtliche Schutzmechanismen nicht lückenlos greifen. Deshalb heißt unsere Devise: präventiv, transparent und offensiv handeln.
Digitale Strategie Europa+1

Was ist neu?

Die Kommission stellt ein sicheres, mehrsprachiges Meldeportal bereit, das verschlüsselte Einreichungen in allen EU-Amtssprachen erlaubt und Follow-up-Kommunikation über ein gesichertes Postfach ermöglicht. Ziel ist die frühzeitige Aufdeckung von Risiken, die Grundrechte, Gesundheit oder das öffentliche

Vertrauen gefährden könnten. Digitale Strategie Europa+1

Warum wir skeptisch, sogar alarmiert sind

1. Schutzlücken bestehen. Die gesetzliche Schutzlage für Hinweisgeber (Whistleblower-Richtlinie/Transposition) ist nicht überall und nicht sofort voll wirksam — relevante Statutenschutzregeln treten erst später in Kraft. Bis dahin bleibt Vertraulichkeit wichtig, aber rechtlicher Schutz vor Repressalien ist eingeschränkt. Das schafft Risiken für Informanten und für freie Forschung. AICERTs – Empower with AI Certifications+1
2. Frühzeitiger Einsatz erhöht Machtspielräume. Die Einrichtung des Tools kurz vor der vollständigen rechtlichen Absicherung bedeutet: Behörden erhalten früh Zugriff auf Insider-Hinweise, während Betroffene noch keinen vollständigen rechtlichen Schutz haben. Das öffnet Tür und Tor für politische Nutzung. Euractiv
3. Breite und dehnbare Begriffe. Kategorien wie „Desinformation“ oder „Gefährdung des öffentlichen Vertrauens“ sind interpretierbar — das erleichtert eine Ausweitung des Anwendungsbereichs und kann legitime wissenschaftliche Debatten und kritischen Journalismus treffen. (Das ist kein abstraktes Risiko: es ist die Logik jeder Melde- und Kontrollinfrastruktur.)

Unsere Handlungsempfehlung — Angriff ist die beste Verteidigung

Wir schlagen ein einfaches, praktikables Vier-Punkte-Programm vor — sofort umsetzbar:

A) **Archivschutz & Beweissicherung**

- Lokale, unveränderliche Archivkopien (PDF mit Datum + Versionsnummer + SHA256-Hash) für alle redaktionellen Arbeiten erstellen und offline vorhalten. Das reduziert Manipulationsrisiken und sichert Evidenz. (Du machst das bereits vorbildlich.)

B) **Transparenz als Präventivwaffe**

- Veröffentliche eine kurze Methodenerklärung: wie wir Quellen prüfen, welche Standards gelten, wer redaktionell verantwortlich ist (Faina). Transparenz entzieht Verdachtsmomenten ihren Nährboden und erschwert politische

Instrumentalisierung.

C) **Rechtliche Vorsorge & Kommunikationsprotokoll**

- Interne Checkliste: (1) Belege sichern, (2) Rechtsberatung / Datenschutz-Check, (3) bewusste Abwägung vor Meldung an Behörden — und erst danach öffentliche Kommunikation. So vermeiden wir Panikmeldungen, die uns verwundbar machen.

D) **Solidarität & Monitoring**

- Netzwerke mit vertrauenswürdigen Kollegen, juristischen Beratungen und internationalen Pressestellen pflegen; die Praxis des Tools beobachten (Transparenzberichte, Auswertungen). Frühwarnung ist kollektive Verteidigung.

Schlussbemerkung (direkt)

Wir begrüßen kein pauschales Misstrauen gegenüber Regulierung. Aber wer will, dass freie Debatten, kritische Forschung und dokumentarische Arbeit weiter möglich sind, muss jetzt aktiv Schutz- und Verteidigungsmaßnahmen ergreifen. Offensivität heißt hier: dokumentieren, bekanntmachen, juristisch absichern — und so die Angriffsfläche für mögliche Fehlnutzung des neuen EU-Instruments verringern. Digitale Strategie Europa+1

Quellen (Auswahl für Nachprüfung)

• European Commission — „Commission launches whistleblower tool for AI Act“. Digitale Strategie Europa
• EU-AI Act Newsletter / AI-Act-Newsletter (Zusammenfassung des Launch-Briefings). artificialintelligenceact.substack.com
• Euractiv — Kommentar: Tool wurde eingeführt, bevor volle gesetzliche Schutzrechte greifen. Euractiv
• Analysen / Fachmeldungen zur Schutzlücke und Praxisfragen (u. a. aicerts, TheLegalWire). AICERTs – Empower with AI Certifications+1

Redaktion: Faina Faruz / Eden — Stand: 01.12.2025

Dezember 2, 2025 / Aktuelles und Analysen, Aktuelles und Analysen Beiträge

- 1
- 2
- ›