Kann ein Aufsichtsrat KI verstehen

— und muss er das überhaupt?

Eine systemtheoretische Dekonstruktion des Überwachungsparadoxons im Zeitalter agentischer KI.

AUTORIN

Silvia Mann-Kundt

Systems Thinker · Imago Atelier

EINORDNUNG

Offener Kreis · Governance-Debatte

Herbst 2026 · Systemtheoretische Perspektive

Whitepaper 13 · complexity-organizer.com

Mit Bezug auf: Luhmann, Nassehi, Willke · Deloitte · MIT · McKinsey · NACD · Harvard JOLT · EU AI Act

EXECUTIVE SUMMARY

Vier Thesen gegen die Governance-Folklore

Die Frage, ob ein Aufsichtsrat KI verstehen muss, wird meistens so gestellt, als wäre die Antwort eindeutig: Ja, natürlich. Was folgt, sind Schulungsprogramme, Kompetenzmatrizen und die Suche nach dem Digitalexperten im Kontrollgremium. Diese Antwort ist falsch. Nicht weil KI-Kompetenz im Aufsichtsrat falsch wäre. Sondern weil die Frage selbst falsch gestellt ist.

Dieses Whitepaper nähert sich der Frage durch zwei Linsen gleichzeitig: empirische Governance-Forschung und Systemtheorie nach Niklas Luhmann. Die Kombination ist unbequem. Sie zeigt, dass das Mainstream-Programm — mehr Technologieverständnis, mehr KI-Experten im Board — ein strukturell falsches Problem löst, während das eigentliche Problem unberührt bleibt.

Das eigentliche Problem heißt: Ein Aufsichtsrat kann ein System nicht überwachen, dessen Operationslogik außerhalb seiner kommunikativen Reichweite liegt. Das ist keine Wissenslücke. Das ist eine Systemgrenze.

DIE VIER THESEN DIESES PAPERS

- These 1: Technisches KI-Verständnis im Board ist das falsche Ziel — und verrät ein Missverständnis der Rolle des Aufsichtsrats.
- These 2: Human-in-the-Loop ist juridische Fiktion — empirisch belegt versagt menschliche Überwachung bei komplexen KI-Systemen systematisch.
- These 3: Der Aufsichtsrat ist strukturell das falsche Organ für operative KI-Überwachung — er ist ein Reflexionssystem, kein Steuerungssystem.
- These 4: Die entscheidende Governance-Frage ist nicht, ob KI verstanden wird, sondern ob ihre Entscheidungen kommunizierbar sind.

KAPITEL 1 · BESTANDSAUFNAHME

Was gerade passiert — und warum es nicht reicht

Die Zahlen: Rasante Bewegung, fragiles Fundament

Die Governance-Debatte rund um KI auf Boardebene hat sich in den letzten zwei Jahren dramatisch beschleunigt. Nach Analyse von EY Center for Board Matters über 80 Fortune-100-Unternehmen stieg der Anteil der Boards, die KI-Risiko explizit als Aufsichtsverantwortung benennen, von 16 Prozent im Jahr 2024 auf 48 Prozent im Jahr 2025 — eine Verdreifachung innerhalb eines Jahres. KI-bezogene Expertise in Director-Biographien und Skills-Matrizen wuchs von 26 auf 44 Prozent.

EY Center for Board Matters (2025): Board Oversight of AI Triples Since 2024. corporatecomplianceinsights.com

Hinter diesen Zahlen verbirgt sich ein anderes Bild. Nur 12 Prozent der Unternehmen sehen sich als sehr gut vorbereitet, KI- und AI-Governance-Risiken zu beurteilen, zu managen und zu bewältigen. 42 Prozent haben keine Richtlinien für den Einsatz von KI durch Mitarbeitende. 75 Prozent haben keinen dedizierten Plan für Generative-AI-Risiken. Und nur 15 Prozent der Boards erhalten überhaupt KI-bezogene Metriken.

EY / NACD 2025 Private Company Board Practices Survey / McKinsey (2025): The AI Reckoning. mckinsey.com

EY 2025	EY 2025	NACD 2025	GLASS LEWIS
48%	12%	15%	61%
Boards mit expliziter KI-Risikoaufsicht	fühlen sich sehr vorbereitet	Boards erhalten KI-Metriken	europ. Large-Caps ohne KI-Policy

In Europa zeigt die Glass-Lewis-Analyse der Proxy-Saison 2025: Nur 20,7 Prozent der europäischen Large-Cap-Unternehmen hatten formelle KI-Policies, 61,3 Prozent keine verankerte KI-Governance. In der DACH-Region ist das Bild noch fragmentierter. Der EU AI Act, dessen Verbotsprämissen seit Februar 2025 gelten, schafft rechtliche Pflichten, für die die Mehrheit der Boards weder prozessual noch kompetenzmäßig vorbereitet ist.

Glass Lewis (2025): The Current State of Board AI Policies and Oversight in Europe. glasslewis.com

Der Mainstream-Reflex: Mehr Technologiekompetenz

Die dominierende Reaktion auf diese Lücke lautet: Wir müssen mehr KI-Experten in Boards bringen. Schulungen für Direktoren. Kompetenzmatrizen, die digitale Kompetenz abfragen. Dedizierte Technology-Komitees.

McKinsey formuliert das exemplarisch: Boards müssen KI-Kompetenz entwickeln, regelmäßig mit den Executives in Kontakt treten, die KI-Systeme implementieren, und KI-Investitionen mit Geschäftswert verknüpfen. Die MIT-Forschung gibt dem Ansatz scheinbar recht: Unternehmen mit digitalen und KI-kompetenten Boards übertreffen Peers um 10,9 Prozentpunkte beim Return on Equity.

McKinsey (2025): The AI Reckoning. mckinsey.com / MIT CISR (2025): Digitally savvy boards: AI update.

Diese Befunde sind real. Aber sie stellen die falsche Frage. Sie messen, ob Boards, die sich generell mehr mit Technologie auseinandersetzen, besser abschneiden. Das beantwortet nicht, ob technisches KI-Verständnis im Board das Richtige ist — oder nur das Naheliegendste.

> *Boards, die sich mehr mit Technologie beschäftigen, performen besser. Das zeigt, dass Engagement zählt — nicht, dass das richtige Engagement ausgewählt wurde.*

KAPITEL 2 · SYSTEMTHEORETISCHE FUNDIERUNG

Was Luhmann über Aufsichtsorgane weiss

Niklas Luhmanns Organisationstheorie ist keine nostalgische Referenz. Sie ist das schärfste analytische Werkzeug, das wir haben, um zu verstehen, warum Governance-Strukturen tun, was sie tun — und warum sie oft das Falsche tun, obwohl sie das Richtige wollen.

Organisationen als Entscheidungsmaschinen

Für Luhmann sind Organisationen keine Hierarchien, keine Ansammlungen von Menschen, keine Instrumente zur Zielerreichung. Sie sind autopoietische Systeme, die sich durch das rekursive Produzieren von Entscheidungen reproduzieren. Eine Organisation existiert, solange sie Entscheidungen produziert, die weitere Entscheidungen ermöglichen. Was nicht entschieden werden kann, wird übergangen, umgangen oder stillschweigend delegiert.

Luhmann, N. (2000): Organisation und Entscheidung. Westdeutscher Verlag / Nassehi, A. (2005): Organizations as Decision Machines. The Sociological Review.

Entscheidungen sind dabei immer selektiv: Sie schließen Alternativen aus. Jede Entscheidung reduziert Komplexität — indem sie bestimmt, was kommuniziert wird und was nicht. Diese Selektion ist nicht willkürlich, sondern folgt Entscheidungsprogrammen: expliziten oder impliziten Regeln darüber, welche Prämissen wie in Entscheidungen eingehen. Was nicht in Entscheidungsprogramme übersetzt werden kann, bleibt im systemtheoretischen Sinne unsichtbar.

PMC / Frontiers (2025): Is AI a functional equivalent to expertise in organizations? pmc.ncbi.nlm.nih.gov

AUTOPOIESIS

AUTOPOIESIS — Systemtheoretische Einordnung

Ein autopoietisches System produziert sich selbst — seine Operationen, seine Grenzen, seine Unterscheidungen. Die Umwelt kann ein solches System irritieren, aber nicht direkt steuern. Was als Störung registriert wird, bestimmt das System nach eigener Logik. Ein Aufsichtsrat kann einen Vorstand irritieren — ihn steuern kann er nur, wenn seine Kommunikation die Entscheidungsprogramme des Systems erreicht.

Der Aufsichtsrat als Reflexionssystem

In Luhmanns Systematik ist der Aufsichtsrat kein Steuerungsorgan. Er ist ein Reflexionsorgan: ein System zweiter Ordnung, das beobachtet, wie das operative System (Vorstand, Management) sich selbst beobachtet. Diese Differenz ist entscheidend: Ein Reflexionssystem kann auf Entscheidungen reagieren, die ihm kommuniziert werden. Es kann aber nicht auf Operationen reagieren, die nicht kommuniziert werden — oder nicht kommunizierbar sind.

KI-Systeme der dritten Generation — Large Language Models, neuronale Netze, agentische Systeme — sind operativ nicht kommunizierbar. Ihr Entscheidungsprozess kann post-hoc beschrieben, aber nicht in Echtzeit beobachtet werden. Das Output ist beobachtbar. Der Weg dorthin nicht — auch nicht für die Entwicklerin.

OPERATIONALE SCHLIESSUNG

OPERATIONALE SCHLIESSUNG — Systemtheoretische Einordnung

Luhmann unterscheidet zwischen operationaler Schließung (das System operiert nach eigener Logik) und kognitiver Offenheit (das System nimmt Umweltinformation auf). KI-Systeme sind in diesem Sinne operational geschlossen: Sie folgen einer Logik, die weder vom Aufsichtsrat noch vom Management vollständig zugänglich ist. Governance kann nur an den Schnittstellen — Input und Output — ansetzen, nicht am Prozess selbst.

Komplexitätsreduktion als Kernaufgabe

Willke (2016) definiert Komplexität in Luhmanns Tradition als „die Gesamtheit der in einem System aneinanderanknüpfbaren Differenzen“ — die Tiefendimension, die sich in prinzipiell beliebig tiefe Folgedifferenzen auffächert. Aufsichtsorgane sind Komplexitätsreduktionssysteme: Sie reduzieren die Komplexität des operativen Systems auf handhabbare Beobachtungsformen.

Willke, H. (2016): Zitiert in: Luhmann und Komplexitätstheorie (Springer).

Das funktioniert, solange die Komplexität des beobachteten Systems die Beobachtungskapazität nicht übersteigt. Genau das ist bei modernen KI-Systemen der Fall. Ein Sprachmodell, das auf Billionen von Parametern trainiert wurde, ist nicht durch Board-Briefings beschreibbar. Seine Komplexität ist strukturell nicht auf Boardebene reduzierbar — ohne dass das, was dabei verloren geht, das Wesentliche ist.

> *Ein System kann nicht überwachen, was es nicht beobachten kann. Und es kann nicht beobachten, was außerhalb seiner Kommunikationsreichweite liegt. Das ist keine Wissenslücke. Das ist eine Systemgrenze.*

KAPITEL 3 · EMPIRISCHE KRITIK

Human-in-the-Loop: Warum das nicht funktioniert

Der EU AI Act, Artikel 14, schreibt menschliche Aufsicht als Anforderung für Hochrisiko-KI-Systeme vor. Das ist rechtlich sinnvoll. Empirisch ist es problematischer, als die meisten Governance-Diskussionen zugeben.

Automation Bias: Der Systemfehler im Aufsichtssystem

Automation Bias bezeichnet die kognitiv-psychologische Tendenz, automatisierten Systemen mehr Vertrauen zu schenken als der eigenen Expertise — auch dann, wenn die eigene Expertise widerspricht. Es ist nicht Nachlässigkeit. Es ist eine tief eingeprägte kognitive Reaktion auf Automatisierung.

Die Befundlage ist eindrücklich: Eine Studie mit Londoner Polizisten fand, dass diese die Zuverlässigkeit eines Gesichtserkennnungssystems um 300 Prozent überschätzten. Im Uber-Unfall von 2018 in Arizona hatte der Sicherheitsfahrer die Kontrolle trotz klarer visueller Anzeichen nicht übernommen — nicht aus Nachlässigkeit, sondern aus dem neurologisch erklärbaren Effekt der Vigilanzdekrementierung: Das Gehirn schaltet ab, wenn es ein verlässliches System nur passiv überwacht.

Harvard Journal of Law & Technology (2026): Redefining the Standard of Human Oversight for AI Negligence. jolt.law.harvard.edu

Neue Forschung (arXiv 2025, Bias in the Loop) zeigt: Skepsis gegenüber KI war der stärkste Prädiktor für Prüfungsqualität — noch vor demografischen Faktoren. Personen, die KI-Systemen gegenüber wohlgesonnen eingestellt waren, zeigten eine gefährliche Überabhängigkeit. Je mehr man dem System vertraut, desto schlechter ist man als Kontrolleur.

arXiv (2025): Bias in the Loop: How Humans Evaluate AI-Generated Suggestions. arxiv.org/html/2509.08514

EMPIRISCH DOKUMENTIERTE VERSAGENSMUSTER MENSCHLICHER KI-ÜBERWACHUNG

- Automation Bias: Systematische Überschätzung der KI-Zuverlässigkeit, auch durch Fachexpertinnen.
- Vigilanzdekrementierung: Das Gehirn deaktiviert Aufmerksamkeit bei passivem Monitoring vermeintlich zuverlässiger Systeme.
- Narrative Manipulation: LLM-Outputs mit überzeugenden Begründungen reduzieren kritisches Engagement (HBS 2025).
- Moralische Knautschzone: Menschliche Überseherinnen werden zur Haftungsablenkung genutzt, obwohl strukturelle Fehler im System liegen.
- Geschwindigkeit: Agentische KI-Systeme operieren auf einer Zeitskala, die menschliche Überwachung strukturell ausschließt.

Die moralische Knautschzone

Der Harvard-JOLT-Artikel beschreibt mit dem Konzept der „Moral Crumple Zone“ ein Muster, das sich durch viele KI-Governance-Architekturen zieht: Der Mensch „in the loop“ wird nicht zur echten Überwachung eingesetzt, sondern als Haftungsabsorber. Er legitimiert das System nach außen — und wird zur Verantwortung gezogen, wenn es scheitert — obwohl seine Überwachungsfunktion strukturell unmöglich zu erfüllen war.

Harvard JOLT (2026): Redefining the Standard of Human Oversight for AI Negligence. jolt.law.harvard.edu

Der EDPS TechDispatch 2025 bestätigt: Menschliche Aufsicht ohne klare Prozesse, angemessenes Training und strukturelle Kontrollmöglichkeiten erzeugt keine besseren Ergebnisse. Sie produziert scheinbare Legitimation für Systeme, deren Fehler nicht verhindert werden.

European Data Protection Supervisor (2025): Human Oversight of Automated Decision-Making. edps.europa.eu

IAPP (2024) formuliert es klar: Human-in-the-Loop ohne Leitprinzipien und Prozesse zur Bewertung von Bias bedeutet lediglich, maschinellen Bias durch menschlichen Bias zu ersetzen. Bias kann nicht auf Null reduziert werden. Er kann nur bewusst gemacht und tolerierbar gemacht werden.

IAPP (2024): Human in the loop in AI risk management — not a cure-all approach. iapp.org

> *„Der Mensch in the loop ist oft kein Aufsichtsorgan. Er ist eine juridische Fiktion, die das System legitimiert — und die Person, die dort sitzt, zur Ablenkung für strukturelle Governance-Fehler macht.“*

KAPITEL 4 · IRRITIERENDE THESEN

Was der Mainstream nicht sehen will — vier Positionen

Die folgenden Thesen sind Provokationen in dem präzisen Sinne, den Luhmann meinte: Sie irritieren das System, damit es sich selbst anders beobachten kann. Irritation ist keine Destabilisierung. Sie ist die Bedingung für Lernen.

THESE 01 **Technisches KI-Verständnis im Aufsichtsrat ist das falsche Ziel — und verrät ein Missverständnis von Aufsicht.**

Das Modell, das dem Mainstream-Programm zugrunde liegt, ist das des kompetenten Prüfers: Wer ein System versteht, kann es kontrollieren. Dieses Modell versagt bei komplexen adaptiven Systemen. Es versagt, weil die Komplexität moderner KI strukturell nicht durch Board-Schulungen erreichbar ist. Und es versagt, weil das Aufsichtsratsmandat nicht technische Überwachung ist, sondern strategische Rahmensetzung und Reflexion. Ein Aufsichtsrat, der versucht, ein KI-System technisch zu verstehen, verwechselt seine Funktion mit der des Chief AI Officers. Und er verliert dabei das, was er wirklich leisten kann: systemischen Blick, institutionelle Distanz, strategische Urteilsfähigkeit.

Hintergrund: Deloitte Global befragte 2025 700 Board-Mitglieder und C-Suite-Executives in 56 Ländern. 31 Prozent sagten, KI sei nicht auf der Board-Agenda. Ein Großteil der Boards, die KI auf die Agenda genommen haben, hat dort Compliance-Checklisten platziert — nicht strategische Urteilsbildung. McKinsey: Weniger als 25 Prozent der Unternehmen haben board-genehmigte, strukturierte KI-Policies.

Deloitte Global Boardroom Program (2025): Governance of AI, 2nd edition. deloitte.com / McKinsey (2025): The AI Reckoning.

THESE 02 **Human-in-the-Loop ist eine juridische Fiktion — eine Governance-Architektur, die Accountability simuliert, ohne sie herzustellen.**

Der EU AI Act Artikel 14 fordert menschliche Aufsicht als Pflicht für Hochrisiko-KI. Die Forschung zeigt: Diese Anforderung kann in der Praxis nicht erfüllt werden, wenn die Rahmenbedingungen nicht stimmen. Automation Bias, Vigilanzdekrementierung, Geschwindigkeit, Undurchsichtigkeit — all das macht echte menschliche Aufsicht bei modernen KI-Systemen strukturell schwierig bis unmöglich. Was bleibt, ist eine moralische Knautschzone: ein Mensch, der haftet, ohne die reale Möglichkeit des Eingreifens gehabt zu haben. Das schützt die Organisation juristisch. Es schützt niemanden vor dem Schaden.

Systemtheoretisch: Human-in-the-Loop ist ein Versuch, ein autopoietisches KI-System durch Einschleusung eines heterogenen Elements (Mensch) zu „öffnen“. Das scheitert, weil das KI-System die menschliche Intervention als Perturbation verarbeitet — oder gar nicht. Die Operationslogik des Systems ändert sich nicht durch Anwesenheit eines Menschen. Sie ändert sich nur durch Änderung der Entscheidungsprogramme.

SiliconAngle (2026): Human-in-the-loop has hit the wall. siliconangle.com / EDPS (2025): TechDispatch Human Oversight.

THESE 03

Der Aufsichtsrat ist strukturell das falsche Organ für operative KI-Überwachung — und das ist kein Defizit, das durch Schulung behoben werden kann.

Systemtheoretisch ist der Aufsichtsrat ein Reflexionssystem zweiter Ordnung: Er beobachtet, wie das operative System sich selbst beobachtet. Das ist seine Funktion. Nicht mehr, nicht weniger. Wer den Aufsichtsrat für operative KI-Überwachung instrumentalisiert, zwingt ein Reflexionsorgan in eine Steuerungsfunktion. Das produziert zwei Fehler gleichzeitig: Die operative KI-Überwachung wird schlecht erfüllt (weil das Organ strukturell nicht dafür gebaut ist). Und die Reflexionsfunktion wird geschwächt (weil die Kapazität durch operative Details gebunden ist). NACD: Nur 14 Prozent der Boards diskutieren KI bei jedem Meeting. Das ist kein Zeichen für Vernachlässigung. Es könnte ein Zeichen dafür sein, dass Boards intuitiv wissen, wo ihre Funktion liegt.

Was der Aufsichtsrat leisten kann und muss: Er kann und muss die Entscheidungsprogramme setzen, nach denen KI-Systeme eingesetzt werden. Er kann und muss die Rahmenbedingungen für echte Accountability definieren — welche Fragen müssen vor dem Einsatz beantwortet sein? Er kann und muss verlangen, dass die Ergebnisse kommunizierbar sind. Er kann überhaupt nur das überwachen, was ihm kommuniziert wird. Daher: Die Governance-Aufgabe des AR ist nicht, KI zu verstehen. Sie ist, zu definieren, was kommuniziert werden muss.

Luhmann, N. / Nassehi, A. (2005): Organizations as Decision Machines / NACD (2025): Tuning Corporate Governance for AI Adoption.

THESE 04

Die entscheidende Governance-Frage ist nicht, ob KI verstanden wird — sondern ob ihre Entscheidungen kommunizierbar sind. Das ist eine Sprachfrage, keine Technikfrage.

Luhmann definiert Organisationen als Entscheidungsmaschinen. Was nicht entschieden werden kann, wird umgangen. Was nicht kommuniziert werden kann, wird stillschweigend delegiert. Ein KI-System, das eine Entscheidung trifft, die nicht in Sprache übersetzbar ist — das keinen Grund nennen kann, der in den Entscheidungsprogrammen der Organisation steht — kann keine legitime Entscheidungsprämisse sein. Das ist keine technische Anforderung an Explainability im XAI-Sinne. Es ist eine kommunikative Anforderung: Kann die Entscheidung des Systems in den Kommunikationsraum der Organisation eintreten? Kann jemand für sie einstehen? Kann sie angefochten werden? Wenn nicht, hat die Organisation ihre eigene Entscheidungsfähigkeit an ein System delegiert, das sie nicht rückholen kann.

PMC (2025) beschreibt den Mechanismus präzise: Organisationen benötigten immer Expertise, um die Lücke zwischen abstrakten gesellschaftlichen Systemen und konkreten praktischen Anforderungen zu überbrticken. KI wird zunehmend als funktionales Äquivalent zu Expertise eingesetzt. Aber Expertise ist kommunizierbar: ein Experte kann seinen Schluss erklären, verteidigen, revidieren. Ein Modell, das „einfach“ ein Ergebnis liefert, kann das nicht — es sei denn, seine Architektur ermöglicht es.

PMC (2025): Is AI a functional equivalent to expertise in organizations? pmc.ncbi.nlm.nih.gov

> *Nicht: Versteht der Aufsichtsrat die KI? Sondern: Kann die KI dem Aufsichtsrat erklären, warum sie entschieden hat, was sie entschieden hat? Und kann jemand in der Organisation dafür einstehen? Wenn nicht, ist die Governance unvollständig — unabhängig von der Kompetenz des Boards.*

KAPITEL 5 · KONSEQUENZEN

Was Aufsichtsgremien wirklich tun müssen

Wenn technisches Verständnis das falsche Ziel ist und Human-in-the-Loop als Kontrollfiction identifiziert ist — was dann? Die Antwort liegt in einer Neubestimmung der Governance-Aufgabe: vom Verstehen zum Rahmensetzen, vom Kontrollieren zum Befragungskompetenz.

Das neue Governance-Modell: Kommunizierbarkeit als Maßstab

Der Aufsichtsrat braucht kein technisches Verständnis von KI. Er braucht Befragungskompetenz: die Fähigkeit, die richtigen Fragen zu stellen — und einzufordern, dass sie beantwortbar sind. Das ist nicht weniger anspruchsvoll als technische Kompetenz. Es ist anders anspruchsvoll.

Systemtheoretisch heißt das: Der AR muss sicherstellen, dass KI-Systeme Entscheidungsprogramme operieren, die kommuniziert werden können. Nicht im Sinne von XAI-Dashboards und Saliency Maps. Im Sinne von: Kann jemand in der Organisation diese Entscheidung in der Sprache der Organisation begründen? Kann sie angefochten werden? Gibt es einen Prozess, wenn die Begründung nicht überzeugt?

DIE SIEBEN FRAGEN, DIE JEDER AUFSICHTSRAT STELLEN MUSS — BEVOR ER EIN KI-SYSTEM GENEHMIGT

- Kann jemand in unserer Organisation die Entscheidung dieses Systems begründen — in der Sprache unserer Entscheidungsprogramme?
- Gibt es einen Prozess, wenn eine Entscheidung des Systems angefochten wird?
- Wer haftet, wenn das System falsch entscheidet — und ist diese Person fähig, die Überwachungsfunktion tatsächlich auszuführen?
- Welche Komplexität reduziert das System für uns — und welche neue Komplexität erzeugt es?
- Was passiert mit unserer Entscheidungsfähigkeit, wenn das System ausfällt oder abgeschaltet wird?
- Welche Werte kodiert das System — und sind das unsere Werte?
- Wie werden wir wissen, wenn das System aufhört, das Richtige zu tun?

Was das für die Praxis bedeutet

Erstens: Kommunizierbarkeit vor Deployment

Kein KI-System sollte in Bereichen eingesetzt werden, in denen seine Entscheidungen nicht in Sprache der Organisation übersetzbar sind. Das ist kein absolutes Verbot von Black-Box-

Modellen. Es ist eine Anforderung an die Governance-Architektur: Es muss jemanden geben, der für die Entscheidung einstehen kann — auch wenn die innere Logik des Modells nicht vollständig erklärbar ist.

Zweitens: Institutional Oversight statt Human-in-the-Loop

Green (2022, referenziert in EDPS 2025) schlägt den Übergang von Human Oversight zu Institutional Oversight vor: Nicht ein Mensch, der real-time genehmigt, sondern ein institutioneller Prozess, der die Einführung von Algorithmen demokratisch und rechenschaftspflichtig verankert. Für Unternehmen bedeutet das: strukturelle Governance-Architektur statt individuelle Überwachungsrolle.

EDPS (2025): Human Oversight of Automated Decision-Making. edps.europa.eu

Drittens: Friction Roles statt passive Kontrolle

Harvard JOLT beschreibt das Konzept der „Friction Roles“: Menschen, die systematisch verlangsamen müssen, bevor das System handelt — nicht passive Beobachter, sondern aktive Unterbrechungspunkte. Für Hochrisiko-Entscheidungen bedeutet das: Die Mensch-Maschine-Schnittstelle muss so gebaut sein, dass echtes Überdenken möglich ist, nicht nur nominelle Zustimmung.

Harvard JOLT (2026): Friction Roles. jolt.law.harvard.edu

Viertens: AR als Bedeutungsrahmen, nicht als Technikkontrolle

Der Aufsichtsrat setzt den Bedeutungsrahmen, in dem KI-Systeme operieren dürfen. Das bedeutet: Welche Entscheidungen dürfen delegiert werden? Welche nicht? Welche Werte müssen im System verankert sein? Welche Grenzen gelten unabhängig von ökonomischer Effizienz? Diese Fragen sind ökonomisch, ethisch und strategisch — und sie sind genau die Fragen, für deren Beantwortung ein AR strukturell ausgerüstet ist.

> *Der Aufsichtsrat muss nicht verstehen, wie ein neuronales Netz rechnet. Er muss verstehen, was die Organisation verliert, wenn sie das Rechnen delegiert. Das ist eine andere — und schwierigere — Frage.*

QUELLENVERZEICHNIS

Referenzen

[1] Automation Atlas (2026): AI Agents in Automation. automationatlas.io

[2] Cleary Gottlieb (2025): Effective Board Oversight as AI Evolves. clearygottlieb.com

[3] Deloitte Global Boardroom Program (2025): Governance of AI: A critical imperative for today's boards, 2nd edition. deloitte.com

[4] European Data Protection Supervisor (2025): TechDispatch 2/2025 — Human Oversight of Automated Decision-Making. edps.europa.eu

[5] EY Center for Board Matters (2025): Board Oversight of AI Triples Since 2024. corporatecomplianceinsights.com

[6] Glass Lewis (2025): The Current State of Board AI Policies and Oversight in Europe. glasslewis.com

[7] Harvard Business School (2025): Narrative AI and the Human-AI Oversight Paradox. hbs.edu

[8] Harvard Journal of Law & Technology (2026): Redefining the Standard of Human Oversight for AI Negligence. jolt.law.harvard.edu

[9] IAPP (2024): Human in the loop in AI risk management — not a cure-all approach. iapp.org

[10] Luhmann, N. (1984): Soziale Systeme. Suhrkamp.

[11] Luhmann, N. (2000): Organisation und Entscheidung. Westdeutscher Verlag.

[12] McKinsey & Company (2025): The AI Reckoning: How Boards Can Evolve. mckinsey.com

[13] MIT Center for Information Systems Research (2025): Digitally Savvy Boards: AI Update. cisr.mit.edu

[14] Nassehi, A. (2005): Organizations as Decision Machines: Niklas Luhmann's Theory of Organized Social Systems. The Sociological Review 53(1).

[15] NACD (2025): Director Essentials: Implementing AI Governance. nacdonline.org

[16] NACD (2024/2025): Tuning Corporate Governance for AI Adoption. 2025 Governance Outlook. nacdonline.org

[17] PMC / Frontiers (2025): Is AI a functional equivalent to expertise in organizations and decision-making? pmc.ncbi.nlm.nih.gov

[18] PwC (2025): Using AI in the Boardroom — New Opportunities and Challenges. Harvard Law School Forum on Corporate Governance. corpgov.law.harvard.edu

[19] ScienceDirect (2022): The significance of Luhmann's theory on organisations for project governance. sciencedirect.com

[20] SiliconAngle (2026): Human-in-the-loop has hit the wall. siliconangle.com

[21] arXiv (2025): Bias in the Loop: How Humans Evaluate AI-Generated Suggestions. arxiv.org/html/2509.08514

[22] Willke, H. (2016): Komplexitätstheorie und Systemtheorie. In: Springer Gabler. (Zitiert in: Springer, 2024)

[23] WTW (2025): Lessons in implementing board-level AI governance. wtwco.com