Gruppe. Interaktion. Organisation. Zeitschrift für angewandte Organisationspsychologie
https://doi.org/10.1007/s11612-026-00900-0

HAUPTBEITRÄGE - THEMENTEIL

Künstliche Intelligenz als Katalysator organisationaler Aufklärung

Warum Organisationen an KI-Transformationen scheitern und was sie daraus über ihr Entscheiden lernen können

Thomas Schumacher[1,2]

Eingegangen: 14. Januar 2026 / Angenommen: 13. Juli 2026
© The Author(s) 2026

Zusammenfassung
Dieser Beitrag in der Zeitschrift „Gruppe. Interaktion. Organisation. (GIO)“ entwickelt einen organisations- und entscheidungstheoretischen Zugang zur Frage, warum Transformationen im Umgang mit Künstlicher Intelligenz (KI) häufig an der Organisation selbst scheitern und was Organisationen daraus über ihr eigenes Entscheiden lernen können. Ausgangspunkt sind empirische Fälle, die zeigen, dass die erhofften Effekte von KI weniger wegen technologischer Defizite ausbleiben als wegen organisationaler Anschlussprobleme. Aus systemtheoretischer Perspektive wird KI nicht als autonomer Akteur verstanden, sondern als Teil organisationaler Entscheidungsarrangements, in denen bestehende Entscheidungsprämissen sichtbar werden und sich verschieben. Drei Differenzkriterien (Probabilistik, Intransparenz, Nicht-Determiniertheit) machen die organisationale Spezifik moderner KI gegenüber früheren Automatisierungsformen sichtbar.
Eine funktionale Differenzierung von KI-Formen (regelbasiert, analytisch, generativ, agentisch) dient als analytisches Raster für die Untersuchung qualitativer Fallstudien aus unterschiedlichen Kontexten. Die Analysen zeigen, dass organisationale Spannungen im Umgang mit KI nicht aufgelöst, sondern entlang sachlicher, sozialer und zeitlicher Sinndimensionen verschoben werden. Diese Verschiebungslogik ergänzt etablierte Modelle der Paradoxiebearbeitung (Smith und Lewis 2011; Berti und Cunha 2023) um eine eigenständige Bearbeitungsform. Auf einer zweiten Ebene erscheint KI als Katalysator organisationaler Aufklärung: Sie zwingt Organisationen, die Kontingenz ihrer eigenen Selbstvergewisserungsprinzipien in Beobachtung zu bringen. Der Beitrag schließt mit Implikationen für Führung, Governance und Organisationsberatung unter Bedingungen probabilistischer, intransparenter und nicht-deterministischer Technologien.

Schlüsselwörter Künstliche Intelligenz · Organisationale Paradoxien · Verschiebungslogik · Entscheidungsprämissen · Organisationale Aufklärung · Führung und Governance

✉ Thomas Schumacher
thomas.schumacher@osb-i.com

[1] Katholische Hochschule Freiburg,
Karlstraße 63, 79104 Freiburg, Deutschland

[2] osb-international, Wien, Österreich

Published online: 06 August 2026

Artificial Intelligence as a catalyst for organizational enlightenment

Why organizations fail at AI transformations and what this reveals about their decision-making

Abstract

This article in the magazine 'Gruppe. Interaction. Organisation. (GIO)' develops an organizational and decision-theoretical approach to the question of why transformations in dealing with artificial intelligence (AI) often fail at the organization itself, and what organizations can learn from this about their own decision-making. The starting point is empirical cases showing that the expected effects of AI fail to materialize less because of technological deficits than because of a lack of organizational connectivity. From a systems-theoretical perspective, AI is not understood as an autonomous actor but as part of organizational decision-making arrangements in which existing decision premises become visible and shift. Three differentiating criteria (probabilism, opacity, and non-determinism) make the organizational specificity of modern AI visible against earlier forms of automation.

A functional differentiation of AI forms (rule-based, analytical, generative, agentic) serves as an analytical grid for examining qualitative case studies from various contexts. The analyses show that organizational tensions in dealing with AI are not resolved but displaced along factual, social, and temporal dimensions of meaning. This logic of displacement extends established models of paradox management (Smith und Lewis 2011; Berti und Cunha 2023) by adding a distinct mode of organizational handling. At a second level, AI appears as a catalyst for organizational enlightenment in the systems-theoretical sense (Luhmann 1992): it forces organizations to bring the contingency of their own principles of self-assurance into observation. The article concludes with implications for leadership, governance, and organizational consulting under conditions of probabilistic, opaque, and non-deterministic technologies.

Keywords Artificial Intelligence · Organizational Paradoxes · Logic of Displacement · Decision Premises · Organizational Enlightenment · Leadership and Governance

1 Künstliche Intelligenz als organisationaler Stresstest

Künstliche Intelligenz (KI) gilt derzeit als einer der zentralen Treiber organisationaler Transformation. Kaum eine Digitalstrategie kommt ohne den Imperativ aus, „AI-ready" zu sein. KI verspricht Effizienzgewinne, bessere Entscheidungen und wirksamere Steuerung. Zugleich mehren sich empirische Befunde, die dieses Versprechen relativieren. Eine Studie des MIT zeigt etwa, dass der Einsatz von KI zwar die Produktivität einzelner Beschäftigter steigern kann, die erhofften Effekte auf Organisationen jedoch häufig ausbleiben (Challapally et al. 2025). Diese Diskrepanz verweist auf ein zentrales Problem gegenwärtiger KI-Debatten: Nicht die Leistungsfähigkeit der Systeme ist die Engstelle, sondern ihre organisationale Anschlussfähigkeit an etablierte Führungs-, Entscheidungs- und Steuerungsannahmen und die mit ihnen verbundenen Erwartungen an Verantwortlichkeit und Kontrolle (Baygi und Huysman 2025; Schumacher 2026).

Aus systemtheoretischer Sicht erscheint KI weniger als autonomer Akteur, denn als Struktur, in deren Nutzung organisationale Entscheidungsprämissen sichtbar werden und sich verändern. Organisationsdesign-orientierte Ansätze argumentieren in ähnlicher Weise, dass KI keine Entscheidungen „übernimmt", sondern Koordinations- und Entscheidungsarrangements beeinflusst, indem Aufgaben, Verantwortlichkeiten und Interdependenzen neu zugeschnitten werden (Puranam 2025).

Diese Sichtweise unterscheidet sich von anthropomorphisierenden Darstellungen, die KI als Akteur oder Entscheider rahmen. Mit „KI als Akteur" sind dabei Deutungen gemeint, die KI als eigenständig handelnde, entscheidende oder verantwortliche Instanz adressieren. Stattdessen rückt der Beitrag organisationale Entscheidungs- und Zurechnungsprozesse in den Fokus und knüpft diese Perspektive an die spezifische Funktionsweise moderner KI: Während klassische Technik konditional programmiert ist, operieren KI-Systeme nicht konditional, sondern auf Basis statistischer Muster. Sie generieren Vorschläge (etwa in Rekrutierung, Kreditvergabe oder medizinischer Diagnostik) auf Basis statistischer Muster, ohne ihr Zustandekommen vollständig begründen zu können. Faraj (2026) beschreibt KI in diesem Zusammenhang als „epistemisches Monster", das etablierte Kategorien von Wissen, Kompetenz und Verantwortung irritiert. Algorithmische Prognosen reduzieren Unsicherheit dabei nicht einfach, sondern transformieren sie in neue Entscheidungslagen (Esposito 2017).

Drei Eigenschaften machen KI für Organisationen zu einer qualitativ anderen Herausforderung als frühere Formen der Automatisierung und Digitalisierung. Erstens operieren KI-Systeme *probabilistisch*: Sie erzeugen Ergebnisse auf Basis korrelativer Muster, ohne kausale Begründungen bereitstellen zu können. Dies irritiert organisationale Entscheidungsprozesse nicht, weil die Ergebnisse falsch wären,

sondern weil sie sich etablierten kausalen Begründungs- und Legitimationslogiken entziehen. Zweitens ist die *Opazität* moderner KI-Systeme kein Defekt, der sich durch technologische Weiterentwicklung beheben ließe, sondern ein Strukturmerkmal. Die Komplexität der Modelle übersteigt prinzipiell die Möglichkeit vollständiger Nachvollziehbarkeit. Drittens operieren insbesondere generative und agentische Systeme *nicht-deterministisch*: Gleiche Eingaben können zu unterschiedlichen Ausgaben führen, was organisationale Erwartungen an Reproduzierbarkeit und Kontrolle grundlegend irritiert. Entscheidend ist dabei, dass diese Eigenschaften nicht nur technische Herausforderungen darstellen, sondern organisationale Grundannahmen sichtbar machen. Die Frage, ob Entscheidungen kausal begründet, Personen zugerechnet und zeitlich stabil gesteuert werden können, stellt sich durch KI nicht neu. Aber sie wird unabweisbar.

Gerade deshalb wirkt KI in Organisationen weniger als Problemlöserin, denn als Katalysator bestehender Paradoxien. Spannungen zwischen Effizienz und Lernen, Automatisierung und Verantwortung oder Expertise und Formalisierung werden nicht entschieden, sondern zwischen sachlichen, zeitlichen und sozialen Sinndimensionen verschoben (Luhmann 1984). In diesem Sinne argumentiert der Beitrag, dass KI organisationale Paradoxien weniger neu erzeugt als bestehende Spannungen beschleunigt, technisiert und neu beobachtbar macht – auch wenn sich diese in der organisationalen Selbstbeschreibung als „neu" darstellen können. KI-bezogene Transformationen mutieren damit vom linearen Problemlösungs- und Implementierungsprozess zur fortlaufenden Auseinandersetzung mit widersprüchlichen Anforderungen, die sich klassischen Akzeptanz- und Steuerungslogiken entziehen (Ramaul et al. 2025).

KI ist damit nicht nur Anlass organisationalen Wandels, sondern für die Organisationsforschung das, was die Drosophila (Fruchtfliege) für die Genetik war: ein Extrem- und Modellfall, an dem sich unter verdichteten Bedingungen grundlegende Dynamiken organisationalen Entscheidens beobachten lassen.

Der Beitrag rekonstruiert, wie Organisationen mit durch KI sichtbar werdenden Spannungen umgehen und was sich dabei über ihr eigenes Entscheiden beobachten lässt. Dazu werden unterschiedliche KI-Formen in ihrer Wirkung auf organisationale Entscheidungsprozesse analysiert, systemtheoretisch gerahmt und anhand empirischer Fälle im Hinblick auf Führung, Governance als Gestaltung organisationaler Entscheidungsprämissen und Transformation diskutiert.

2 Formen künstlicher Intelligenz

Ob in der Rekrutierung, in medizinischen Diagnoseprozessen oder in der polizeilichen Einsatzplanung: KI greift heute in sehr unterschiedlichen organisationalen Kontexten in Entscheidungsprozesse ein. Um diese Wirkungen analytisch fassen zu können, ist weniger die technische Architektur der Systeme entscheidend als die funktionale Rolle, die ihnen im organisationalen Entscheidungsprozess zugeschrieben wird. Die folgende Unterscheidung greift bestehende Typologien von KI im Entscheidungsprozess auf (Humberd und Latham 2025) und schärft sie organisations- und entscheidungstheoretisch. Sie dient der Analyse jener organisationalen Effekte, die entstehen, wenn KI in Entscheidungs-, Koordinations- und Bewertungsprozesse eingebunden wird. Erst diese Differenzierung erlaubt es, nicht KI „an sich", sondern unterschiedliche organisationale Entscheidungsprobleme sichtbar zu machen.

Vor diesem Hintergrund lassen sich vier Formen unterscheiden (regelbasierte, analytische, generative und agentische KI) die im Folgenden als analytisches Raster für die empirischen Analysen verwendet werden.

2.1 Regelbasierte Systeme (Prä-KI-Referenz)

Regelbasierte Systeme wie Robotic Process Automation (RPA) zählen streng genommen nicht zur Künstlichen Intelligenz, weil sie deterministische Wenn-Dann-Regeln exekutieren, ohne aus Daten zu lernen oder probabilistisch zu operieren. Sie werden hier dennoch als Referenzform eingeführt, weil sie den Ausgangspunkt organisationaler Automatisierung markieren, gegenüber dem die spezifischen Merkmale probabilistischer KI-Systeme sichtbar werden. Organisationstheoretisch handelt es sich um klassische Konditionalprogramme im Sinne Luhmanns (2000): Wenn Bedingung X eintritt, folgt Handlung Y. Ihre organisationale Funktion liegt in der Stabilisierung von Erwartungen durch Standardisierung, etwa wenn Beschaffungsfreigaben unterhalb eines definierten Schwellenwerts automatisch erteilt werden und Varianz zulasten von Ausnahmenbearbeitung reduziert wird (Vollenberg 2021). Die Spannung zwischen Standardisierung und Lernen ist dabei kein KI-spezifisches Phänomen, wird aber, wie die folgenden Analysen zeigen, durch probabilistische Systeme qualitativ verschärft.

2.2 Analytische KI

Analytische Systeme wie Predictive Analytics oder Scoring-Modelle verarbeiten historische Daten zu Prognosen. Ihre Funktion besteht in der Entscheidungsvorbereitung: Sie machen Muster sichtbar, priorisieren Handlungsoptionen und verdichten Kontingenz. Organisationen stehen hier

vor der Herausforderung, datengetriebene Objektivierung mit menschlicher Diskretion, Kontextualisierung und Verantwortung zu verbinden (Simon et al. 2024). Typisch ist dies etwa in Controlling-, Risiko- oder Diagnosekontexten, in denen Prognosen zwar handlungsleitend werden, ihre Abweichungen vom professionellen Urteil jedoch organisationsspezifisch unterschiedlich bearbeitet werden.

2.3 Generative KI

Generative Systeme erzeugen Texte, Deutungen und Vorschläge, die als Anschlussmöglichkeiten in organisationale Kommunikation eingespeist werden können. Ob und wie daraus tatsächliche Anschlussfähigkeit entsteht, entscheidet jedoch nicht die KI, sondern das soziale System im Akt des Verstehens (Esposito 2017). Generative KI produziert Sinnangebote, ohne selbst an deren Verstehen beteiligt zu sein. Dies zeigt sich etwa dort, wo generative Systeme Memos, Entscheidungsvorlagen oder Einsatzberichte formulieren, deren Plausibilität erst im organisationalen Gebrauch entsteht. Für Organisationen wird Sinnproduktion dadurch hoch verfügbar aber ihre epistemische Geltung bleibt unsicher.

Die organisationale Funktion generativer KI liegt weniger in Erkenntnis als in der Erzeugung von Variation, also der Produktion von Alternativen, die erst durch organisationale Selektion, also Bewertung, Auswahl und Zurechnung, ihre Geltung erlangen (Faraj 2026; Baygi und Huysman 2025). Generative KI steigert die Variationskapazität einer Organisation erheblich, ohne die Selektionsfähigkeit mitzuliefern.

Gerade diese Variation kann jedoch, sofern sie organisational aufgegriffen wird, als Ressource reflexiver Entscheidungsfindung wirksam werden.

2.4 Agentische KI

Agentische Systeme gehen über Informationsbereitstellung hinaus, indem sie Handlungsschritte initiieren, Werkzeuge nutzen und Ziele autonom verfolgen (Ackerman 2025). Organisationstheoretisch verschiebt sich damit die Funktion von KI von der Entscheidungsunterstützung zur Delegation von Handlungslogiken, wodurch Fragen der Verantwortungszuschreibung, Kontrolle und „conjoined agency“ virulent werden (Fisher 2025).

Viele agentische KI-Architekturen beruhen dabei implizit auf einem Kommunikationsverständnis, das Sprache als Transportmedium für Handlungsanweisungen behandelt. Aus systemtheoretischer Sicht verkennt dies jedoch, dass Kommunikation nicht exekutiert, sondern Anschlussmöglichkeiten für weitere Kommunikation anbietet. Genau dieses Missverständnis trägt dazu bei, dass agentische Systeme organisationale Erwartungen an Steuerbarkeit nicht stabil einlösen.

Organisationstheoretisch verschiebt sich damit die Funktion von KI von der Entscheidungsunterstützung zur operativen Vorselektion und Vorstrukturierung von Entscheidungsfolgen. Agentische Systeme entscheiden dabei im organisationstheoretischen Sinne nicht: Sie kommunizieren keine Auswahl aus Alternativen, die sozial zugerechnet werden könnte. Sie operieren, indem sie kategorisieren, priorisieren, eskalieren und Folgeschritte einleiten. Organisational wirksam werden diese Operationen jedoch genau dann, wenn sie in der Organisation wie Entscheidungen behandelt werden, ohne deren kommunikativen Status zu besitzen. Ein agentisches System, das einen Workflow bearbeitet, erzeugt eine Kette von Selektionen, von denen jede einzelne organisational folgenreich sein kann, ohne dass sie jemandem als Entscheidung zurechenbar wäre (vgl. Fisher 2025).

Viele agentische KI-Architekturen beruhen implizit auf einem Kommunikationsverständnis, das Sprache als Transportmedium für Handlungsanweisungen behandelt. Aus systemtheoretischer Sicht verkennt dies, dass Kommunikation nicht exekutiert, sondern Anschlussmöglichkeiten für weitere Kommunikation anbietet. Zugleich ist die eingeschränkte Steuerbarkeit agentischer Systeme nicht allein ein kommunikationstheoretisches Problem. Agentische KI operiert inhärent nicht-deterministisch und befindet sich technologisch in einem frühen Entwicklungsstadium, in dem das Einhegen von Risiken ein aktives Forschungsfeld darstellt. Beide Perspektiven ergänzen sich: Selbst wenn agentische KI technisch reift, bleibt das grundlegende Problem bestehen, dass ihre Operationen organisational wie Entscheidungen wirken, ohne als solche kommuniziert und zugerechnet werden zu können.

Tab. 1 fasst diese Formen, ihre organisationalen Funktionen und die durch sie häufig sichtbar werdenden Spannungen zusammen.

Die Unterscheidung der KI-Formen dient im Folgenden als analytisches Raster, um anhand empirischer Fälle zu zeigen, wie KI nicht einfach „eingeführt“ wird, sondern wie sie in die paradoxen Strukturen organisationaler Entscheidungsfindung eingreift, und diese verschiebt. Die Auswahl der Fälle in Kapitel 4 folgt dabei nicht der KI-Formen-Typologie, sondern den drei Sinndimensionen organisationaler Paradoxiebearbeitung: Chirurgische Robotik dient als Prä-KI-Referenzfall für die Zeitdimension (Exploitation und Exploration), die Radiologie-Vignette zeigt die Sachdimension (Spannung zwischen statistischer Prognose und professionellem Urteil), und Predictive Policing analysiert die Sozialdimension (Verschiebung von Verantwortungszuschreibungen). Die Kurzvignette zu Food-Delivery-Plattformen gibt einen heuristischen Ausblick auf agentische KI. Kurzvignetten ergänzen die Hauptfälle, um zu zei-

Tab. 1 KI-Formen (Quelle: eigene Darstellung)

KI-Form	Typische Beispiele	Organisationale Funktion	Organisationale Spannungen, die im Gebrauch häufig sichtbar werden (Auswahl)
Regelbasierte Systeme	Robotic Process Automation, Workflow-Automatisierung, Ticket Routing, Systeme Routing, Standardfreigaben	Stabilisierung und Standardisierung; Unsicherheitsabsorption durch Konditionalprogramme (wenn x dann y)	Stabilität und Lernen: Standardisierte Programme sichern verlässliche Performance, können jedoch Ausnahmenbearbeitung und Erfahrungslernen einschränken: Standardisierung und Flexibilität
Analytische Kl	Scoring, Prognosen. Anomalie-, Risiko- und Betrugsmodelle	Entscheidungsvorbereitung; Sichtbarmachung von Kontingenz; Priorisierung	Transparenz & Opazität; Risikoreduktion & neue Kontingenz
Generative Kl	Zusammenfassungen, Entwürfe, Berichte, Entscheidungsvorlagen und alternative Deutungen	Kommunikative Verdichtung; Erzeugung von Variation und Sinnangeboten; Bereitstellung anschlussfähiger Entwürfe	Variation und Selektion; Produktivität und Vertrauenswürdigkeit; kommunikative Glättung und inhaltliche Qualität; Sinnangebot und Zurechnung
Agentische Kl	Workflow-Orchestrierung, mehrstufige Auftragsbearbeitung, Einsatz- und Ermittlungsplanung, Werkzeugnutzung und Prozessausführung	Vorstrukturierung von Entscheidungsfolgen; Delegation von Handlungslogiken	Autonomie & Abhängigkeit; Delegation & Kontrolle; Effizienz der Entscheidungsfolgen & Verantwortungszuschreibung

gen, dass die analysierten Dynamiken kontextübergreifend auftreten.

3 Paradoxien organisationaler Entscheidungsfindung und KI

Paradoxien sind kein Rand- oder Defektphänomen organisationalen Handelns, sondern dessen Normalform: Organisationen reproduzieren und stabilisieren sich durch die fortlaufende Bearbeitung widersprüchlicher Anforderungen (Luhmann 2000; Smith und Lewis 2011). Paradoxien bezeichnen dabei interdependente, widersprüchliche und andauernde Anforderungen, die sich nicht durch Entscheidung auflösen lassen. Sie unterscheiden sich von lösbaren Widersprüchen oder Dilemmata, in denen eine Alternative gewählt und das Problem damit beendet werden kann. Paradoxien lassen sich nicht zum Verschwinden bringen, sondern nur entfalten: verschieben, sichtbar halten oder produktiv bearbeiten (Tuckermann et al. 2023). Gerade die Fähigkeit, solche Paradoxien nicht zu vermeiden, sondern fortlaufend zu bearbeiten, bildet eine zentrale Quelle organisationaler Lern- und Anpassungsfähigkeit. Eine grundlegende organisationale Paradoxie, die sich durch den Einsatz von KI zuspitzt, ist die Gleichzeitigkeit von Exploitation, der effizienten Nutzung bestehender Routinen und Kompetenzen, und Exploration, dem offenen Erkunden neuer Möglichkeiten (March 1991). KI fungiert hierbei als Katalysator, durch den organisationale Paradoxien nicht neu, sondern technisiert, beschleunigt und verschoben werden, und zwar häufig so stark, dass sie, wenn sie vormals latent waren, erstmals explizit beobachtbar werden. Spannungen zwischen Effizienz und Lernen, Automatisierung und Verantwortung oder Expertise und Formalisierung werden nicht entschieden, sondern oftmals zwischen unterschiedlichen Sinndimensionen verschoben.

3.1 Sinndimensionen als Verschiebungsarenen

Folgt man der systemtheoretischen Unterscheidung von Sach-, Sozial- und Zeitdimension (Luhmann 1984), lassen sich drei Arenen identifizieren, in denen KI organisationale Paradoxien zuspitzt (vgl. Tuckermann et al. 2023):

In der Sachdimension erzeugen probabilistische Systeme eine spezifische Spannung: Sie liefern häufig stochastisch generierte Ergebnisse, ohne kausale Begründungen bereitstellen zu können. Ein Scoring-Wert kann prognostisch überlegen sein und dennoch organisational scheitern, weil er sich nicht in die etablierten Begründungs- und Legitimationslogiken einfügen lässt. Die Spannung liegt also nicht in der Fehlerhaftigkeit der Resultate, sondern in der Kluft zwischen Performanz und Begründbarkeit. KI verschärft dieses Problem gegenüber früheren Formen der Automatisierung, weil klassische Systeme konditional nachvollziehbar operierten, während probabilistische Modelle eine prinzipielle Intransparenz aufweisen. Zugleich wirft dies eine weitergehende Frage auf, nämlich ob die unterstellte kausale Transparenz menschlicher Entscheidungen den Maßstab tatsächlich einlöst, an dem KI hier gemessen wird. Die organisationale Begründungspraxis beruht vielfach auf post-hoc-Rationalisierungen (Weick 1995), die durch KI nicht ersetzt, sondern als solche erkennbar werden. Organisationssoziologisch lässt sich diese Spannung als Konflikt zwischen kognitiven und normativen Erwartungen lesen (Luhmann 1984): KI verschiebt kognitive Erwartungsstrukturen, indem sie neue, probabilistisch begründete Outputs erzeugt,

während die normativen Erwartungshorizonte der Organisation, wie etwa Begründungs- und Legitimationslogiken, vergleichsweise stabil bleiben und damit Anschlussprobleme erzeugen.

In der Sozialdimension verschieben sich Zurechnungsfragen, weil Entscheidungsergebnisse zunehmend von den Bedingungen ihrer Hervorbringung entkoppelt sind: Führungskräfte vertreten Resultate, deren Zustandekommen sie nicht durchdringen. Das ist organisationssoziologisch nicht neu, sondern eine Verschärfung eines bekannten Musters. Auch ohne KI verantworten Vorgesetzte regelmäßig Zuarbeiten ihrer Mitarbeitenden, ohne deren fachliche Tiefe zu beherrschen. KI verstärkt diese Entkopplung jedoch durch latente Unerklärbarkeit. Formate wie Human-in-the-loop oder Ansätze erklärbarer KI (XAI: Verfahren, die KI-Entscheidungen für menschliche Beobachtung zugänglich machen) markieren den Versuch, formale Verantwortlichkeit und faktische Abhängigkeit von Technik zu versöhnen. Sie zielen weniger auf Transparenz im erkenntnistheoretischen Sinne als auf die organisationale Absicherung von Verantwortung. Als Übersetzungs- und Absicherungsmechanismen stabilisieren sie Zurechnungen, ohne die zugrunde liegenden Entscheidungsprämissen aufzulösen. Luhmanns Unterscheidung von Verantwortung als prospektive Unsicherheitsabsorption und Verantwortlichkeit als retrospektive Zuschreibung (Luhmann 2000) gewinnt hier eine neue Dringlichkeit: Wenn die Unsicherheitsabsorption von der KI geleistet wird, kann die retrospektive Zuschreibung an Personen nicht mehr jene Legitimationskraft entfalten, die sie unter den Bedingungen individueller Entscheidungsproduktion hatte.

In der Zeitdimension verschärft KI die grundlegende organisationale Paradoxie von Exploitation und Exploration (March 1991): die Gleichzeitigkeit von effizienter Nutzung bestehender Routinen und offenem Erkunden neuer Möglichkeiten. KI beschleunigt Entscheidungszyklen und steigert operative Performanz, kann jedoch Lernprozesse unterminieren, wenn Abweichungen automatisiert korrigiert werden, bevor sie als Irritation wirksam werden. Ein Predictive-Maintenance-System etwa, das Anomalien erkennt und korrigiert, noch bevor sie für das Wartungsteam als Störung wahrnehmbar werden, entzieht der Organisation genau jene Irritation, die Erfahrungslernen auslöst. Lernen wird so zeitlich entkoppelt oder verdrängt. Zugleich gilt das Umgekehrte: KI-Systeme können Muster sichtbar machen, die menschlicher Beobachtung entgehen, und damit neue Lernanlässe erzeugen, die ohne algorithmische Analyse nicht entstanden wären. Die Paradoxie der Zeitdimension liegt deshalb nicht nur in der Verdrängung von Lernen, sondern in der gleichzeitigen Ermöglichung und Unterminierung organisationaler Lernfähigkeit.

Die drei Dimensionen sind analytisch unterscheidbar, praktisch jedoch eng miteinander verflochten. Paradoxien werden selten gelöst, aber häufig von einer Sinndimension in eine andere verschoben.

3.2 Visibilisieren, Invisibilisieren und Verschieben

Organisationen reagieren auf die durch KI sichtbar werdenden Spannungen typischerweise in zwei Modi (Richter und Groth 2024). Einerseits können Paradoxien sichtbar gemacht und reflexiv bearbeitet werden. Das geht mit erhöhter Unsicherheit einher, kann aber strategische Vorteile haben: Puranam (2025) zeigt, dass Transparenz Organisationen erlaubt, sogenannte Preferences for Organizational Contexts (POCs) in Beobachtung zu bringen, also Bedingungen wie Autonomie, Kompetenz und soziale Eingebundenheit, die für Motivation und Bindung zentral sind (vgl. Ryan und Deci 2017). KI kann diese stillen Bindungsfaktoren unterminieren, lange bevor klassische Leistungskennzahlen darauf reagieren. Andererseits werden Paradoxien häufig invisibilisiert, etwa durch Routinen, technische Delegation oder informelle Übersetzungspraktiken. Das sichert kurzfristig Handlungsfähigkeit und ist für den operativen Vollzug vieler Entscheidungen funktional unverzichtbar. Langfristig kann es jedoch neue Kosten erzeugen, wenn organisationale Arrangements im Umgang mit KI Interaktion fragmentieren: Autonomie wird durch Kontrolle ersetzt, Kompetenz durch Automation ausgehöhlt und Zusammenarbeit durch isolierte Mensch-KI-Interaktion eingeschränkt (Puranam 2025). Beide Modi, Visibilisierung wie Invisibilisierung, sind dabei nicht per se funktional oder dysfunktional. Entscheidend ist, ob eine Organisation in der Lage ist, zwischen ihnen bewusst zu navigieren.

Zentral ist: Der Einsatz von KI verändert das Verhältnis dieser beiden Modi. KI neigt dazu, Paradoxien sachlich zu invisibilisieren, indem der KI-Output als ausreichend gilt und die Frage nach dem Zustandekommen in den Hintergrund tritt. Zugleich verschiebt sie Spannungen häufig in die Sozialdimension, etwa wenn Personen Entscheidungen rechtfertigen müssen, deren Prämissen sie nicht erzeugt haben. Mit Verschiebung ist dabei keine Lösung oder Aufhebung von Spannungen gemeint, sondern ihre Verlagerung zwischen Sinndimensionen, organisationalen Ebenen oder formalen und informellen Arrangements, durch die Entscheidungsfähigkeit stabilisiert wird, ohne die zugrunde liegenden Paradoxien aufzulösen. Genau diese Verschiebungen stehen im Zentrum der folgenden empirischen Analysen.

Dass solche Verschiebungen bereits jenseits der ausgewählten Fälle in alltagspraktischen Organisationsphänomenen auftreten, zeigt die zunehmende Verbreitung von Schatten-KI. Gemeint sind Praktiken, in denen Beschäftigte KI-Werkzeuge jenseits formaler Governance-Strukturen einsetzen, um Entscheidungs- und Arbeitsanforderungen zu bewältigen. Das Phänomen ist nicht ohne Vorläufer: Aus der

IT-Forschung ist das verwandte Konzept der Schatten-IT bekannt, bei dem Mitarbeitende eigene Werkzeuge einsetzen, weil die offiziell bereitgestellten Systeme den Arbeitsanforderungen nicht entsprechen. Beide Phänomene lassen sich organisationssoziologisch als Spielart „brauchbarer Illegalität“ (Luhmann 1964) lesen: informale Workarounds, die formale Strukturen verletzen, aber gerade dadurch deren Funktionsfähigkeit unter Realbedingungen sichern. Schatten-KI folgt einer ähnlichen Logik, verschärft sie jedoch, weil generative KI-Werkzeuge nicht nur effizienter, sondern qualitativ anders in Entscheidungsprozesse eingreifen. Wenn etwa eine Fachkraft eine Entscheidungsvorlage mithilfe von ChatGPT erstellt, weil die offizielle Tool-Landschaft dies nicht vorsieht oder die Governance zu restriktiv ist, stabilisiert sie kurzfristig ihre eigene Handlungsfähigkeit. Sie lagert zugleich eine Spannung aus dem formalen Entscheidungszusammenhang aus, die sich organisationaler Beobachtung, Zurechnung und Verantwortung entzieht.

Schatten-KI ist dabei zugleich Regelabweichung und organisationales Signal. Als Regelverstoß ist sie disziplinarisch adressierbar. Die organisationstheoretisch interessantere Beobachtung liegt jedoch darin, dass sie dort entsteht, wo formale Entscheidungsprämissen und faktische Arbeitsanforderungen auseinanderfallen. Ein pauschales Nutzungsverbot etwa schützt Datensicherheit, blockiert aber Effizienzgewinne. Eine pauschale Freigabe ermöglicht Effizienz, exponiert jedoch Fragen der Verantwortung und Qualitätssicherung. In beiden Fällen bleibt die zugrunde liegende Spannung bestehen und sucht sich informelle Ventile. Schatten-KI macht diese Spannung sichtbar, auch wenn Organisationen sie lieber unsichtbar halten würden. Karafillidis (2026) zeigt für KI-spezifische Workarounds, dass deren Bearbeitung weniger eine Frage technischer Kontrolle als eine Frage organisationaler Containment-Strategien (Eindämmung, Einhegung) ist: Wo formale und faktische Anforderungen auseinanderfallen, geht es nicht darum, das Phänomen zu unterbinden, sondern seine Konsequenzen reflexiv adressierbar zu machen.

4 Empirische Analysen: KI in der Praxis

Wie in Kapitel 2 angekündigt, folgt die empirische Analyse nicht der Typologie der KI-Formen, sondern der Logik der drei Sinndimensionen organisationaler Paradoxiebearbeitung: Chirurgische Robotik steht für die Zeitdimension, die radiologische Vignette für die Sachdimension, Predictive Policing für die Sozialdimension. Ergänzt werden diese Hauptfälle durch Kurzvignetten. Ihre Funktion liegt darin, die kontextübergreifende Reichweite der analysierten Verschiebungsdynamiken sichtbar zu machen, ohne den Anspruch eigenständiger Tiefenanalysen zu erheben. Entscheidend ist dabei weniger die Technologie selbst als die Art und Weise, wie Organisationen mit den durch KI sichtbar werdenden Spannungen umgehen. Ziel der Analysen ist es zu zeigen, wie Paradoxien im organisationalen Vollzug nicht gelöst, sondern verschoben, stabilisiert oder unsichtbar gehalten werden.

4.1 Chirurgische Robotik: Lernen unter algorithmischer Vorstrukturierung

Der erste Fall analysiert bewusst keine KI im engeren Sinne, sondern chirurgische Robotik als Referenzfall für technisch vorstrukturierte Entscheidungs- und Lernarbeit. Beane (2019) untersucht die chirurgische Ausbildung in Krankenhäusern, die robotische Operationssysteme einsetzen. In der klassischen offenen Chirurgie erfolgt Lernen durch *legitimate peripheral participation*: Assistenzärzt:innen stehen am OP-Tisch, übernehmen schrittweise operative Teilschritte, beobachten Entscheidungen aus nächster Nähe und entwickeln ihre Kompetenz im Vollzug realer Arbeit. Lernen, Entscheiden und Verantwortung sind dabei eng gekoppelt.

Mit der Einführung chirurgischer Robotik verändert sich diese Struktur grundlegend. Der operierende Chirurg sitzt an einer Konsole, räumlich vom Patienten getrennt, während Assistenzärzt:innen häufig als „Bedside Assistants“ agieren. Der operative Prozess wird technisch vermittelt, standardisiert und gegen Varianz abgesichert. Diese Vorstrukturierung erhöht Sicherheit und Effizienz im Hier und Jetzt, verschärft jedoch die zentrale zeitliche Spannung zwischen Exploitation (Stabilität, Performance) und Exploration (Lernen, Übungsmöglichkeiten).

Diese Zeitparadoxie wird im organisationalen Vollzug virulent, aber nicht expliziert. Lernen erscheint zunehmend als Risiko für den laufenden Betrieb. Organisationen reagieren darauf nicht mit einer Entscheidung zugunsten oder zulasten der Ausbildung, sondern durch Verschiebung. Lernen bleibt notwendig, wird jedoch aus dem formalen Prozess ausgelagert. Genau hier setzt das von Beane beschriebene *shadow learning* an. Beane identifiziert drei miteinander verflochtene Praktiken: frühzeitige Spezialisierung auf robotische Chirurgie (*premature specialization*), Verlagerung von Lernen in abstrakte Räume wie Simulatoren (*abstract rehearsal*) sowie Kompetenzaufbau in realen, aber nur teilweise überwachten Situationen (*undersupervised struggle*). Diese Praktiken sind formal nicht vorgesehen, werden jedoch stillschweigend toleriert, solange operative Ergebnisse stimmen. Diese Invisibilisierung von Lernarbeit ist jedoch keine bloße Fehlentwicklung, sondern eine organisationale Lösung, um unter Bedingungen hoher Sicherheitsanforderungen operative Stabilität, Haftbarkeit und Leistungsfähigkeit aufrechtzuerhalten.

Die Paradoxie wird nicht gelöst, sondern zeitlich verschoben: Lernen wird aus dem operativen Vollzug ausgelagert und in sachlich wie sozial randständige Praktiken ver-

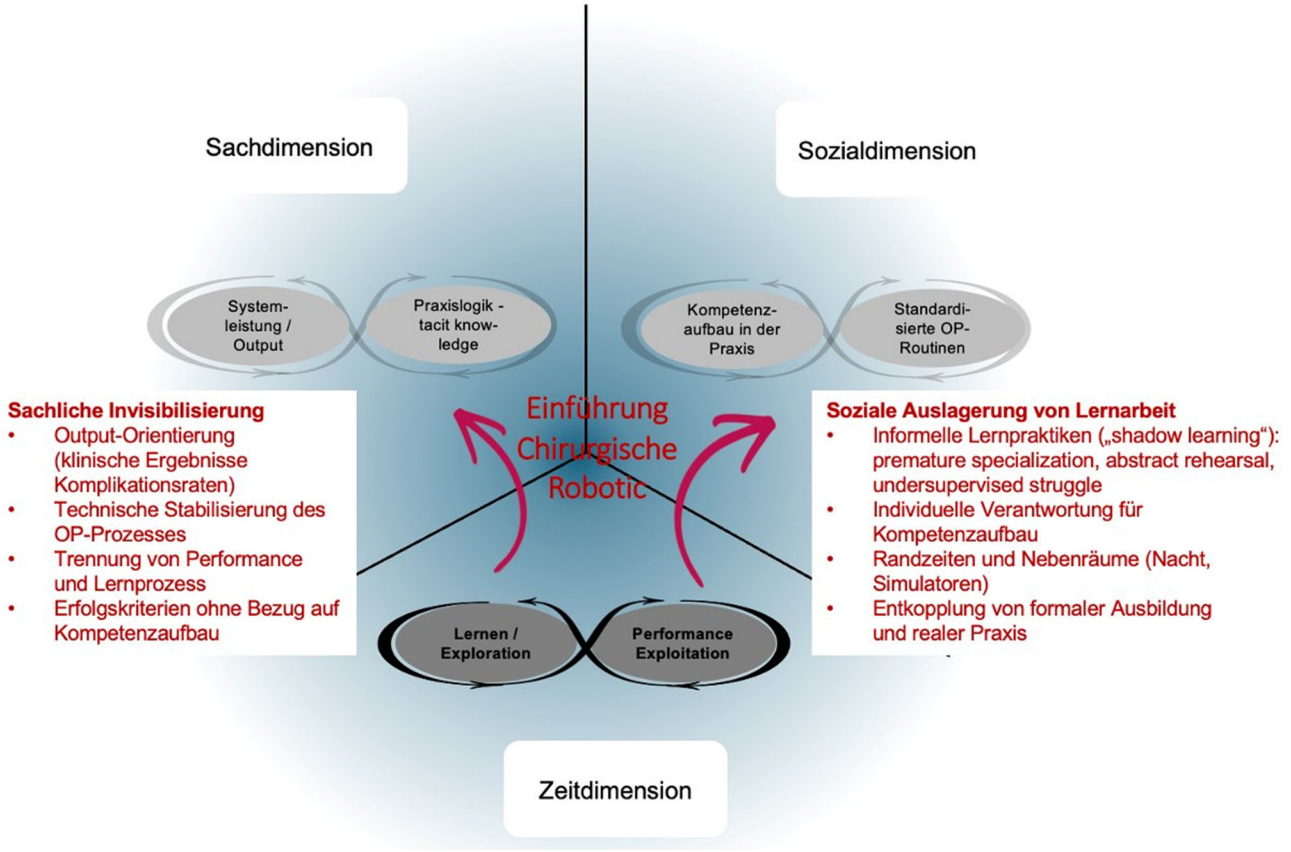

Abb. 1 Paradoxiebearbeitung Chirurgische Robotik

lagert (vgl. Abb. 1). Die Organisation bleibt so handlungsfähig, indem sie Lernen nicht abschafft, sondern strukturell entkoppelt. Beane (2019) selbst weist darauf hin, dass die etablierten Lernpfade unter Bedingungen robotischer Vorstrukturierung systematisch insuffizient werden. Organisationen könnten darauf nicht nur informell mit shadow learning reagieren, sondern die Paradoxie auch reflexiv adressieren, etwa durch dedizierte Co-Surgeon-Phasen, simulationsgestützte Vorlauf-Trainings oder formal abgesicherte Junior-Slots. Der von Beane beobachtete shadow-learning-Modus ist damit kein unausweichliches Schicksal, sondern eine naheliegende Reaktion von Organisationen, die die Paradoxie nicht explizit thematisieren.

Die folgende Kurzvignette zeigt, wie analytische KI in drei Abteilungen einer Organisation ungleich behandelt wird.

4.2 Kurzvignette: Analytische KI in der Radiologie: Opazität und situative Urteilsbearbeitung

Die Studie von Lebovitz et al. (2022) untersucht den Einsatz analytischer KI-Systeme in drei radiologischen Abteilungen (Lungenkrebsdiagnostik, Mammographie, Knochenalterbestimmung). In allen Fällen liefern die Systeme statistische Klassifikationen mit hoher diagnostischer Qualität. Gleichzeitig divergieren die KI-Outputs regelmäßig von ärztlichen Einschätzungen. Entscheidend ist jedoch weniger die technische Leistungsfähigkeit der Systeme als die Art und Weise, wie Organisationen diese Divergenzen bearbeiten.

Die KI unterbricht etablierte professionelle Urteilsroutinen, da ihre Klassifikationen nicht mit den gewohnten Begründungsformen ärztlicher Expertise verknüpft sind. In der Lungenkrebsdiagnostik reagieren Radiolog:innen darauf, indem sie gezielt Zeit investieren, um KI-Hinweise zu prüfen, Bilder neu zu konfigurieren und beide Perspektiven deliberativ zu verbinden. In der Mammographie dagegen werden abweichende KI-Ergebnisse unter hohem Zeitdruck systematisch ignoriert. In der Knochenalterbestimmung schließlich führt die hohe Präzision der KI zu anhaltender Irritation, ohne dass stabile Routinen im Umgang mit den Ergebnissen entstehen.

Im Ergebnis führt die KI in unterschiedlichen organisationalen Kontexten zu sehr unterschiedlichen Bearbeitungsformen. Die Spannung zwischen statistischer Prognose und professionellem Urteil bleibt sachlich präsent und wird situativ sozial entschieden, ohne in stabile organisationale Routinen überführt zu werden.

4.3 Predictive Policing: Die Verschiebung von Verantwortung

Waardenburg et al. (2022) analysieren die Einführung prädiktiver Systeme bei der niederländischen Polizei. Die Studie zeigt exemplarisch, wie KI nicht primär Entscheidungsinhalte, sondern Entscheidungsprämissen und Verantwortungszuschreibungen verändert. Predictive Policing verspricht eine objektivierte Grundlage polizeilicher Einsatzentscheidungen. Die Systeme identifizieren Kriminalitätsschwerpunkte und generieren Prognosen, die operative Planung steuern sollen. In der Praxis adressieren diese Systeme jedoch nicht direkt die Streifenpolizist:innen. Stattdessen etabliert sich eine neue Vermittlungsebene der sogenannten *Intelligence Officer*, die algorithmische Outputs interpretieren, priorisieren und in operative Handlungsanweisungen übersetzen. Die Rolle als solche ist formal benannt. Ihre organisationale Funktion entsteht jedoch emergent aus der Notwendigkeit, opake algorithmische Prognosen für die operative Praxis lesbar zu machen, und war im ursprünglichen Implementierungsdesign nicht als Entscheidungsinstanz vorgesehen.

Das zentrale organisationale Spannungsfeld liegt hier in der Sozialdimension. Predictive Policing soll subjektives Ermessen reduzieren, erzeugt jedoch neue Abhängigkeiten von intermediären Akteuren. Die algorithmischen Prognosen sind selbst erklärungsbedürftig und gewinnen erst durch menschliche Interpretation organisationale Wirksamkeit. Damit unterscheidet sich diese Konstellation von klassischen Human-in-the-Loop-Arrangements, in denen eine vorgesehene menschliche Instanz algorithmische Vorschläge prüft, bestätigt oder überstimmt. Hier dagegen entsteht eine zusätzliche organisationale Ebene zwischen System und Streifendienst, deren Existenz und Funktion nicht aus einem Kontroll-Mandat folgt, sondern aus der Erklärungsbedürftigkeit der Outputs. Subjektivität verschwindet damit nicht, sondern wird sozial umcodiert und aus dem formalen Entscheidungsprozess herausgelöst.

Organisationen legitimieren Einsatzentscheidungen mit dem Verweis auf die Objektivität des Systems, während die tatsächliche Entscheidungsarbeit in Übersetzungspraktiken verlagert wird, die organisational nicht als Entscheidungsleistungen ausgewiesen sind. Für Streifenpolizist:innen erscheinen Einsatzprioritäten als technisch berechnet und damit kaum hinterfragbar. Auch wenn Intelligence Officers selbst erfahrene Polizist:innen sind, verschiebt sich weniger ihre persönliche Autorität als die organisationale Funktion, die sie übernehmen: Sie prägen operative Entscheidungen, ohne dass dieser Einfluss als Entscheidung adressiert oder zugerechnet wird. Diese Invisibilisierung sozialer Entscheidungsarbeit ermöglicht es, unter Bedingungen po-

Abb. 2 Paradoxiebearbeitung

litischer Sichtbarkeit, Zeitdrucks und Rechenschaftspflicht handlungsfähig zu bleiben.

Wie in Abb. 2 dargestellt, bleibt die sachliche Spannung zwischen algorithmischer Prognose und lokalem Erfahrungswissen weitgehend unsichtbar. Die organisationale Bearbeitung erfolgt nicht durch sachliche Klärung, sondern durch soziale Reorganisation. Der Fall zeigt, dass Predictive Policing weniger zu objektiveren Entscheidungen führt als zu neuen Formen von Intermediarität. KI fungiert hier nicht als autonomer Entscheider, sondern als Katalysator sozialer Machtverschiebungen und neuer Legitimationsarrangements. Paradoxien werden damit nicht gelöst, sondern organisational stabilisiert.

4.4 Ausblick: Wohin führen agentische Logiken?

Während die vorangegangenen Fälle Entscheidungen im engeren Sinne analysieren, deutet sich mit agentischer KI eine qualitative Verschiebung an. Agentische Systeme machen nicht nur punktuelle Entscheidungsvorschläge, sondern strukturieren Entscheidungsfolgen über längere Zeiträume hinweg. Sie initiieren Handlungen, setzen Zwischenziele, nutzen Werkzeuge und eskalieren Probleme rekursiv. Paradoxien verlagern sich damit von einzelnen Entscheidungen auf Entscheidungsprämissen, Eskalationslogiken und Beobachtungsstrukturen.

Da agentische KI Anfang 2026 noch ein emergentes Phänomen ist und dichte organisationsbezogene Fallstudien weitgehend fehlen, kann dieser Ausblick nur heuristisch bleiben.

4.5 Kurzvignette: Food-Delivery-Plattformen und agentische Steuerung

Huang (2022) beschreibt, dass große Essenslieferplattformen in China heute schon hunderttausende Fahrer durch agentische Algorithmen koordinieren, die Aufträge verteilen, Routen optimieren, Lieferzeitfenster setzen und Vergütung dynamisch anpassen. Anders als klassische regelbasierte Optimierung operieren diese Systeme als kontinuierlich laufende Selektionskaskade (vgl. Kap. 2): Sie verfolgen ein übergeordnetes Ziel (maximale Auslastung), generieren daraus adaptiv Sub-Ziele und Entscheidungsfolgen und justieren ihr Verhalten anhand von Echtzeitdaten und Verhaltensmustern der Fahrer nach. Während das System aus Organisationsperspektive maximale Effizienz und Auslastung erreicht, berichten Fahrer von hohem psychischem Druck, Sicherheitsrisiken und Einkommensunsicherheit.

Die zentralen Spannungen von Tempo und Sicherheit sowie Skalierung und Menschlichkeit werden nicht organisational bearbeitet, sondern sozial externalisiert. Risiken werden systematisch an die Peripherie verlagert, während die Plattform selbst als handlungsfähig und erfolgreich erscheint.

Agentische KI fungiert hier nicht als autonomer Akteur, sondern als Verstärker einer bekannten Dynamik. Was sie von klassischen Optimierungsalgorithmen unterscheidet, ist weniger das Auftreten von Externalisierungseffekten als deren strukturelle Verankerung. Regelbasierte Optimierung verfolgt explizit gesetzte Bedingungen und bleibt damit organisational adressierbar: Vorgaben können angepasst, Regeln umformuliert werden.

Agentische Systeme dagegen generieren ihre Sub-Ziele und Anpassungen kontinuierlich aus dem Zielverfolgungsprozess heraus. Strukturen wie systematische Pausenverkürzung oder Risiko-Übernahme entstehen so als emergente Folge der Aggregation vieler Mikro-Selektionen, ohne dass sie an einzelne Entscheidungen rückgebunden werden könnten. Damit verschärft sich die im Beitrag analysierte Herausforderung qualitativ: Nicht einzelne Entscheidungen sind zu kontrollieren, sondern die Architektur der Entscheidbarkeit selbst.

5 Diskussion: KI, Paradoxien und organisationale Transformation

Die analysierten Fälle zeigen, dass KI organisationale Spannungen weder erzeugt noch auflöst, sondern ihre Bearbeitung verändert. Entscheidend ist weniger ihr Auftreten als der organisationale Umgang mit ihrer Sichtbarkeit. Über die Fälle hinweg zeigt sich ein gemeinsames Muster: Paradoxien werden nicht expliziert und entschieden, sondern entlang unterschiedlicher Sinndimensionen bearbeitet. Heuristisch lassen sich die drei beobachteten Spannungen in einer Trias bündeln: Entscheidbarkeit (in der Sachdimension), Zurechenbarkeit (in der Sozialdimension) und Steuerbarkeit (in der Zeitdimension). Was durch KI verschoben wird, sind nicht primär einzelne Entscheidungen, sondern die organisationalen Bedingungen, unter denen Entscheidungen entscheidbar, zurechenbar und steuerbar werden.

In der Zeitdimension (Robotik) wird Lernen gegenüber kurzfristiger Performance entkoppelt und in informelle Praktiken ausgelagert. In der Sachdimension (Radiologie) bleibt die Spannung zwischen statistischer Prognose und professionellem Urteil ungelöst und wird situativ navigiert. In der Sozialdimension (Predictive Policing) verschiebt sich Autorität von traditionellen Entscheidungsträgern hin zu neuen intermediären Rollen, ohne dass diese Verschiebung organisational thematisiert wird.

Diese Bearbeitungsformen unterscheiden sich im Detail, folgen jedoch einer gemeinsamen Logik: Organisationen stabilisieren ihre Handlungsfähigkeit, indem sie widersprüchliche Anforderungen nicht entscheiden, sondern verschieben, neutralisieren oder unsichtbar halten. Diese Beob-

achtung wird hier als Verschiebungslogik organisationaler Paradoxiebearbeitung gefasst und als konzeptioneller Beitrag des Artikels vorgeschlagen. Während die Paradoxieforschung bislang vor allem nach Modi der Bearbeitung fragt, etwa der iterativen Pendelbewegung zwischen den Polen (Smith und Lewis 2011) oder ihrer reflexiven Sichtbarmachung (Berti und Cunha 2023), lenkt die Verschiebungslogik den Blick darauf, dass Organisationen Paradoxien typischerweise weder bearbeiten noch sichtbar halten, sondern operativ verlagern: zwischen Sinndimensionen, zwischen formaler und informeller Ebene, zwischen Entscheidung und Vorstrukturierung. KI verstärkt diese Logik nicht, sondern macht sie beobachtbar, weil sie Verschiebungen technisiert und damit aus dem stillen Hintergrund organisationaler Routinen heraushebt.

KI-bezogene Arrangements verstärken diese Dynamik, weil Unsicherheit operativ handhabbar wird, ohne sie kommunikativ aufzulösen. Entscheidungen werden beschleunigt und vorstrukturiert, während die Voraussetzungen für ihre epistemische Begründbarkeit und soziale Zurechenbarkeit zugleich fragiler werden.

Für Führung, Governance und Organisationsentwicklung ergibt sich daraus eine grundlegende Konsequenz: Der produktive Umgang mit KI liegt weniger in der Optimierung einzelner Anwendungen als in der bewussten Gestaltung organisationaler Entscheidungs- und Kommunikationsprämissen. Karafillidis (2026) bringt diesen Befund auf die Pointe, dass der Schlüssel zur Zuverlässigkeit von KI nicht primär technischer, sondern organisationaler Natur sei: Die in der MIT-Studie sichtbar gemachte Lücke zwischen individueller Produktivitätssteigerung und ausbleibenden organisationalen Effekten (Challapally et al. 2025) ist demnach kein Modell-, sondern ein Einbettungsproblem. Klassische KI-Governance-Ansätze, die auf Regeln, Transparenz oder Kontrolle einzelner Entscheidungen zielen, greifen zu kurz, weil sie sich auf Entscheidungen im engeren Sinne konzentrieren, während KI vor allem die Bedingungen des Entscheidens verändert. Dies zeigt sich nicht zuletzt auch an der Verbreitung von Schatten-KI, die dort entsteht, wo formale Entscheidungs- und Verantwortungsstrukturen mit den faktischen Arbeitsanforderungen nicht mehr korrespondieren.

Ein praxistauglicher Zugriff verschiebt den Fokus daher auf eine Governance von Entscheidungsprämissen: auf Programme, Rollenerwartungen, zeitliche Horizonte von Lernen und Verantwortung sowie auf die organisationalen Übersetzungsleistungen zwischen KI-Outputs und Entscheidungen.

Damit verändert sich auch die Rolle von Führung und Organisationsberatung (vgl. Schumacher 2026). Führung wird weniger zur Durchsetzung von Zielklarheit als zur Praxis der Paradoxiebearbeitung: widersprüchliche Anforderungen werden nicht aufgelöst, sondern sichtbar gehalten, zwischen Sinndimensionen verschoben und kollektiv bearbeitbar gemacht. Organisationsentwicklung kann hier ansetzen, indem sie Räume und Formate etabliert, in denen genau diese Spannungen reflexiv verhandelt werden, etwa durch kollektive Fallreflexion realer KI-Entscheidungen, Entscheidungs-Reviews, paradoxe Szenarioarbeit oder die Nutzung von Explainable-AI-Ansätzen als Reflexionsstatt als reine Transparenzinstrumente. Transformation erscheint so weniger als Projektlogik mit definiertem Zielzustand, sondern als kontinuierliche Arbeit an organisationalen Spannungsfeldern. Damit knüpft die hier entwickelte Perspektive an die Unterscheidung von Weick und Quinn (1999) zwischen *episodic change* und *continuous change* an: Während *episodic change* Transformation als zeitlich begrenzte Sequenz von Auftauen, Verändern und Stabilisieren rahmt, beschreibt *continuous change* Wandel als laufende Anpassungs- und Rekonfigurationsbewegung im operativen Vollzug. KI-bezogene Transformation lässt sich in dieser Logik nicht als episodisches Projekt einlösen, weil sie genau jene Bedingungen mitverändert, unter denen Veränderung beobachtbar wird.

Die Diskussion macht damit deutlich, dass KI nicht primär neue Entscheidungsprobleme erzeugt, sondern bestehende organisationale Grenzen der Entscheidbarkeit sichtbar macht. Die zentrale Herausforderung liegt weniger in der Kontrolle einzelner Systeme als in der Fähigkeit von Organisationen, ihre eigenen Entscheidungsprämissen, Verschiebungsmechanismen und Unsichtbarkeiten systematisch zu beobachten.

5.1 Von der Erosion zur Aufklärung: Drei Prinzipien organisationaler Selbstvergewisserung

Die bisherige Analyse hat die organisationalen Spannungen als Erosionsdynamiken beschrieben: Erosion epistemischer Belastbarkeit, Erosion individueller Zurechenbarkeit, Erosion erfahrungsbasierter Kompetenzbildung. Das ist eine Beobachtung erster Ordnung. Sie beschreibt, was Organisationen erleben, wenn KI in ihre Entscheidungsprozesse eingreift. Erst die Beobachtung zweiter Ordnung (vgl. Luhmann 1992) macht sichtbar, was die Erosionserzählung selbst voraussetzt: dass es überhaupt etwas gibt, das erodieren kann. Denn: was erodiert, sind keine naturgegebenen Prinzipien, sondern historisch spezifische Formen organisationaler Selbstvergewisserung. Ihre Kontingenz war vor KI nicht beobachtbar. Jetzt wird sie es.

In jeder der drei Sinndimensionen lässt sich ein solches Prinzip identifizieren. In der Sachdimension ist es das Kausalitätsprinzip. Organisationen haben gelernt, Entscheidungen dadurch zu legitimieren, dass sie kausal begründbar sind. KI-basierte Erkenntnisse, die korrelativ und musterbasiert operieren, brechen mit dieser Logik. Die Erosionsperspektive sieht darin einen Verlust epistemischer Belast-

barkeit. Die Aufklärungsperspektive fragt: War Kausalität jemals die einzige Form der Wissensproduktion? Eigentlich nicht, aber sie war die dominante. Die Frage ist deshalb nicht, ob korrelatives Wissen „echtes“ Wissen ist, sondern ob Organisationen verschiedene Formen der Wissensproduktion unterscheiden, situativ einsetzen und ihre je eigenen Risiken reflektieren können (vgl. Gabriel 2024).

In der Sozialdimension wirkt das Individualprinzip: Eine Person entscheidet, eine Person haftet. KI-gestützte Entscheidungen folgen dieser Logik nicht. Sie sind verteilte Entscheidungen, an denen System, Person, Organisation und Anbieter beteiligt sind. Aus Erosionssicht: ein Zusammenbruch individueller Zurechenbarkeit. Aus Aufklärungssicht eine andere, womöglich unbequemere Lesart: Verantwortung wurde nie individuell produziert, sondern individuell zugeschrieben. Die Zuschreibung war eine funktionale Fiktion. Sie ermöglichte Handlungsfähigkeit, indem sie die tatsächliche Verteilung der Entscheidungsproduktion invisibilisierte (vgl. Luhmann 2000). KI macht diese Differenz sichtbar. Und zwingt Organisationen, über Formen verteilter Verantwortung nachzudenken, die der tatsächlichen Entscheidungsproduktion entsprechen.

In der Zeitdimension operiert das Erfahrungsprinzip: Kompetenz entsteht durch wiederholte Konfrontation, durch Fehler, durch Übung. KI untergräbt diesen Zusammenhang, indem sie Situationen vorfiltert und Übungsgelegenheiten reduziert. Die Erosionsperspektive sieht darin einen Kompetenzverlust. Die Aufklärungsperspektive fragt anders: War Erfahrungslernen je die einzige Form organisationalen Lernens, oder lediglich die zur Verfügung stehende? KI eröffnet möglicherweise andere Formen, etwa systematischen Mustervergleich, simulationsbasiertes Lernen oder automatisierte Anomalie-Erkennung. Die Aufgabe ist dann nicht, Erfahrungslernen aufzugeben, sondern das Repertoire zu erweitern und die Risiken jeder Lernform zu kennen.

Damit erweisen sich die drei Spannungen als doppelt lesbar. Als Erosion gelesen sind es Bedrohungen, während sie als Aufklärung gelesen Gelegenheiten sind, die eigenen Prinzipien als kontingent zu beobachten und sich bewusst zu ihnen zu verhalten. Aufklärung heißt dabei nicht, die alten Prinzipien aufzugeben. Kausalität bleibt wichtig, bekommt aber Gesellschaft. Bei der individuellen Zurechenbarkeit kommen verteilte Verantwortungsformen dazu. Erfahrungslernen wird ergänzt, nicht ersetzt. Was gewonnen wird, ist eine Wahl, die vorher nicht da war, weil die Prinzipien als alternativlos galten. Organisationale Aufklärung ist in diesem Sinn keine normative Forderung. Sie ist eine strukturelle Konsequenz: KI zwingt Organisationen, ihre eigenen Prämissen der Selbstvergewisserung zu beobachten, ob sie wollen oder nicht. Die Frage ist nicht, ob diese Beobachtung stattfindet. Die Frage ist, ob sie reflexiv organisiert wird oder sich als unbeobachtete Verschiebung vollzieht.

6 Limitationen und Ausblick: Agentische KI als nächste Entwicklungsstufe

Die vorangegangenen Analysen zeigen, dass zentrale Entscheidungs- und Zurechnungsprobleme nicht erst mit agentischer KI auftreten, sondern sich bereits in früheren KI-Formen verdichtet zeigen. Der folgende Ausblick auf agentische KI knüpft an diese Befunde an, ist jedoch empirisch notwendigerweise begrenzt. Während analytische, generative und regelbasierte KI-Systeme inzwischen in einer wachsenden Zahl organisationsnaher Fallstudien dokumentiert wurden (etwa Lebovitz et al. 2022 zu analytischer KI in der Radiologie; Vollenberg 2021 zu regelbasierten Systemen), existieren jenseits von Pilotierungen bislang kaum empirische, organisationsnahe Fallstudien zu tatsächlich operierenden agentischen KI-Systemen. Agentische KI stellt derzeit ein emergentes Phänomen dar (Humberd und Latham 2025).

In dieser Begrenzung liegt jedoch ein zentraler Ausblick: Agentische KI markiert eine qualitative Entwicklungsstufe nicht im Sinne gesteigerter Entscheidungsautonomie, sondern durch eine Verschiebung der Programmform. In Kapitel 2 wurden regelbasierte Systeme als Konditionalprogramme eingeführt: Sie folgen einer Wenn-Dann-Logik, in der Auslöser und Reaktion fest verknüpft sind. Agentische Systeme dagegen operieren näher an der Logik von Zweckprogrammen: Sie sind auf ein übergeordnetes Ziel hin programmiert und generieren ihre operativen Schritte (Sub-Ziele, Werkzeugnutzung, Eskalationen) adaptiv aus dem Zielverfolgungsprozess heraus. Anders als klassische Zweckprogramme in Organisationen, deren Mittel-Ziel-Ableitungen durch menschliche Entscheidungsträger:innen erfolgen, wird diese Ableitung bei agentischer KI an das System delegiert und operativ konditioniert: Welche Sub-Ziele in welcher Reihenfolge verfolgt werden, ergibt sich erst im Vollzug. Damit verändern agentische Systeme weniger einzelne Entscheidungen als die Architektur der Entscheidbarkeit insgesamt. Zweckprogramme werden in operative Prozessketten überführt, in denen Eskalationslogiken, Abbruchkriterien und Beobachtungsstrukturen vorab festgelegt sind. Paradoxien von Verantwortung, Autonomie und Kontrolle lösen sich dadurch nicht auf, sondern werden zeitlich gestreckt und in neue Formen organisationaler „conjoined agency“ überführt (Fisher 2025).

Aus praktischer Perspektive deutet sich an, dass die Nutzung agentischer KI Organisationen nicht nur vor neue Implementierungsprobleme stellt, sondern auch ihre Fähigkeit zur Paradoxiebearbeitung radikal herausfordert. Gerade darin liegt jedoch die Chance, Paradoxien als Ressource bewusster Organisationsgestaltung und kollektiver Lernprozesse zu nutzen. Organisationen, die agentische Systeme einsetzen, werden zunehmend gezwungen sein, nicht nur die Beobachtung ihrer Entscheidungen, sondern auch ih-

rer Entscheidungsprämissen institutionell zu verankern. Sie müssen Transparenz schaffen (z. B. XAI-Ansätze), ohne die Illusion zu erzeugen, dass Transparenz Verantwortung auflöst.

Für zukünftige Forschung ergibt sich daraus ein doppelter Auftrag: Erstens, bedarf es qualitativ dichter Fallstudien, die agentische KI in realen Organisationskontexten untersuchen und die Interaktion zwischen menschlichen Akteuren, organisationalen Strukturen und agentischen Systemen empirisch rekonstruieren. Zweitens brauchen wir theoretische Vertiefungen, wenn agentische KI ein fruchtbares Forschungsfeld für organisations- und systemtheoretische Analysen eröffnet. Durch die agentische KI werden klassische Annahmen über Entscheidung, Führung und Steuerung nachhaltig irritiert insbesondere im Hinblick auf die Veränderung von Verantwortungsstrukturen unter Bedingungen einer „conjoined agency".

7 Fazit

KI wirkt in Organisationen als Katalysator paradoxer Anforderungen, weil sie probabilistische Unsicherheit operativ nutzbar macht, ohne sie kommunikativ aufzulösen. Sie steigert Entscheidungskapazität, entzieht Entscheidungen jedoch zugleich ihre epistemische Eindeutigkeit und soziale Zurechenbarkeit. Paradoxien werden dadurch weniger neu erzeugt als beschleunigt, technisiert und organisational sichtbar. Organisationen sind damit weniger zu Kontrolle als zu einer reflexiven Auseinandersetzung mit dem eigenen Entscheiden herausgefordert. Gerade diese Irritation eröffnet jedoch Lernchancen für Organisationen, die bereit sind, ihr eigenes Entscheiden zum Gegenstand systematischer Selbstbeobachtung zu machen.

KI wird so zu dem, was die Drosophila in der Genetik war – zum Gegenstand organisationaler Aufklärung: nicht durch das, was sie entscheidet, sondern durch das, was sie Organisationen über die kontingenten Prinzipien ihres eigenen Entscheidens sichtbar macht. Was Organisationen aus dieser Sichtbarkeit gewinnen, ist nicht primär ein neues Werkzeug, sondern eine Wahl: zwischen einer Erosionserzählung, die KI als Bedrohung etablierter Prinzipien rahmt, und einer Aufklärungsperspektive, die in genau diesen Irritationen die Gelegenheit erkennt, das organisationale Repertoire bewusst zu erweitern.

Datenverfügbarkeit Alle dieser Arbeit zugrunde liegenden Daten sind in diesem Artikel enthalten.

Einhaltung ethischer Richtlinien

Interessenkonflikt T. Schumacher gibt an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf http://creativecommons.org/licenses/by/4.0/deed.de.

Literatur

Ackerman, L. (2025). *Perceptions of agentic AI in organizations: Implications for responsible AI and ROI*. preprint arXiv:2504.11564

Baygi, R. M., & Huysman, M. (2025). Generative AI and the social fabric of organizations. *Strategic Organization*. https://doi.org/10.1177/14761270251385456.

Beane, M. (2019). Shadow learning: Building robotic surgical skill when approved means fail. Administrative Science Quarterly, 64(1), 87-123.

Berti, M., & Cunha, M. P. (2023). Paradox theory and organization studies: An integrative review. *International Journal of Management Reviews*, *25*(1), 87–110.

Challapally, A., Pease, C., Raskar, R., & Chari, P. (2025). *The GenAI divide: State of AI in business 2025*. https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf. Zugegriffen: 10. Dez. 2025.

Esposito, E. (2017). Artificial communication? The production of contingency by algorithms. *Zeitschrift für Soziologie*, *46*(4), 249–265.

Esposito, E. (2024). *Künstliche Intelligenz und die Zukunft der Entscheidung: Wie intelligente Maschinen unser Verständnis von Rationalität verändern*. Suhrkamp.

Faraj, S., Perez-Torrents, J., Mantere, S., & Bhardwaj, A. (2026). A time for monsters: Organizational knowing after LLMs. *Strategic Organization*. https://doi.org/10.1177/14761270251410675.

Fisher, G. (2025). Competing with smart machines: The dark side of conjoined agency in contemporary organizations. *Administrative Science Quarterly*, *70*(1), 1–48.

Gabriel, M. (2024). *Emotionale Intelligenz: Warum Künstliche Intelligenz uns nicht ersetzen wird*. Ullstein.

Huang, Y. (2022). Algorithmic management in food-delivery platforms: Data traces and worker practices in China. *New Technology, Work and Employment*, *37*(1), 66–87.

Humberd, B. K., & Latham, S. F. (2025). When AI becomes an agent of the firm: Examining the evolution of AI in organizations through an agency theory lens. *Journal of Management Studies*.

Karafillidis, A. (2026). Containment eines Laborunfalls. *Merkur*, *80*(2), 75–83.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. Organization science, 33(1), 126-148.

Luhmann, N. (1964). *Funktionen und Folgen formaler Organisation*. Duncker & Humblot.

Luhmann, N. (1984). *Soziale Systeme*. Suhrkamp.

Luhmann, N. (1992). *Beobachtungen der Moderne*. Westdeutscher Verlag.

Luhmann, N. (2000). *Organisation und Entscheidung*. Westdeutscher Verlag.
March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.
Puranam, P. (2025). The business case for human-centric organizing: Implications for artificial intelligence. *Strategic Organization*., , . Express.
Ramaul, L., Ritala, P., Kostis, A., & Aaltonen, P. (2025). Rethinking how we theorize AI in organization and management: A problematizing review of rationality and anthropomorphism. *Journal of Management Studies*., , .
Richter, T., & Groth, T. (2025). *Zwischen Inszenierung und Invisibilisierung: Systemisches Paradoxiemanagement in Organisationen*. Carl-Auer Verlag.
Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York: Guilford.
Schumacher, T. (2026). Künstliche Intelligenz und Organisationen - Missverständnisse und Führungsfolgen. In T. Schumacher (Hrsg.), *Wir haben (k)ein Umsetzungsproblem! Gesellschaft und Organisationen zwischen struktureller Zumutung und systemischer Verantwortung. Eine Festschrift für Rudolf Wimmer* (S. 61–69). Heidelberg: Carl-Auer Verlag.
Simon, M., Roth, G., & Kühne, S. (2024). Predictive analytics and organizational discretion: Tensions between data-driven objectivity and contextual judgment. *Organization Studies*, *45*(2), 234–256.
Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, *36*(2), 381–403.
Tuckermann, H., Schumacher, T., & Krautzberger, M. (2023). Wozu Paradoxien? Auf dem Weg zur paradoxiefähigen Organisation und Führung. *OrganisationsEntwicklung*, *1*, 6–11.
Vollenberg, R. (2021). *Die Entwicklungsbeschleunigung von Robotic Process Automation Lösungen: Eine Fallstudie in einer deutschen Gesundheitsbehörde während COVID-19*. Bertelsmann.
Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. Organization science, 33(1), 59-82.
Weick, K. E. (1995). *Sensemaking in organizations*. SAGE.
Weick, K. E., & Quinn, R. E. (1999). Organizational change and development. *Annual Review of Psychology*, *50*(1), 361–386.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.

Thomas Schumacher