// mib

MANAGEMENT INSTITUT BOCHUM

Managementsysteme im KI-Zeitalter

Verantwortung ohne Ausführung: wie Organisationen eine hybride Wertschöpfung führbar und nachweisbar halten

Mike Emenako · mib Management Institut Bochum

Arbeitspapier · Stand: 29. August 2026, korrigiert am 30. September 2026

Kurzfassung

Die Ausgangsfrage lautete: *Brauchen wir Managementsysteme noch, wenn Maschinen die Prozesse ausführen?* Die Antwort dieses Papiers ist ein Ja mit vier Präzisierungen, ergänzt um eine ausführlich ausgearbeitete Gegenposition, an der sie sich messen lassen muss.

Die These. Managementsysteme verschwinden nicht. Ihr Gegenstand verschiebt sich. Aus der Steuerung menschlichen Verhaltens durch beschreibende Dokumente wird die Architektur begründeten Vertrauens über eine dauerhaft hybride Wertschöpfung, die teils von Menschen und teils von Maschinen ausgeführt wird. Was eine Organisation an Ausführung abgibt, bleibt ihr als Verantwortung erhalten. Aus dieser Lücke, der **Verantwortung ohne Ausführung**, ergeben sich vier neue Kernaufgaben: die **Allokation** der Ausführung, die **Steuerung** über zwei Register, der **Nachweis** über heterogene Evidenz und der ehrliche Ersatz jener **Beruhigungsfunktion**, die unsere Dokumentation immer auch hatte.

Der Befund darunter. Von den Funktionen, die ein Managementsystem erbringt, erodiert genau eine: die dokumentenzentrierte Nachweisform. Haftung, Marktzugang und Lernschleife verstärken sich. Das erklärt die Verunsicherung der Branche vollständig. Bedroht ist das Geschäftsmodell, nicht das Produkt.

Der Gegeneinwand, ernst genommen. Die stärkste Gegenposition lautet: Es ändert sich nichts, weil sich die Anreizstruktur nicht ändert. Sie ist empirisch besser belegt, als es der Fachöffentlichkeit lieb sein kann, und sie führt zu einer harten quantitativen Grenze. Für die typischen KI-Anwendungsfälle kleiner und mittlerer Organisationen ist kontinuierliche statistische Nachweisführung nicht nur unwirtschaftlich, sondern rechnerisch leer. Dieses Papier nimmt den Einwand nicht als Fußnote auf, sondern als eigenen Teil, und zieht daraus eine Konsequenz, die die eigene These schärft, statt sie zu retten.

Die praktische Summe. Das Managementsystem der hybriden Organisation ist weniger Dokumentensammlung und mehr **Verantwortungsarchitektur**. Es beantwortet jederzeit vier Fragen: Wer oder was führt aus, und warum so? Wie kommt unsere Absicht dort hinein? Woran sehen wir, dass es funktioniert? Und wer versteht genug, um einzugreifen?

TEIL I

Die Frage

1 Ausgangspunkt: Die Frage, die mir die alte Antwort genommen hat

Ich arbeite seit über dreißig Jahren an und mit Managementsystemen. Seit einiger Zeit baue ich außerdem selbst agentische KI-Systeme, also Systeme, die nicht nur Texte erzeugen, sondern mehrstufige Arbeitsabläufe eigenständig ausführen. Aus dieser Doppelrolle ist mir eine Antwort abhandengekommen, die drei Jahrzehnte selbstverständlich war.

Eine Verfahrensanweisung hat funktioniert, weil ein Mensch sie gelesen und danach gehandelt hat. Führt ein Agent den Schritt aus, liest er sie nicht. Das Dokument existiert weiter; ob es noch etwas bewirkt, ist eine andere Frage. Dazu kommt eine Beobachtung, die selten ausgesprochen wird: Die Beschreibung steuert nicht nur, sie beruhigt. Fehlt sie, entsteht ein Gefühl von Kontrollverlust, ganz unabhängig davon, ob die Kontrolle jemals wirklich in diesem Dokument lag. Und während wir darüber nachdenken, fragt der Markt unverändert nach Nachweisen und Zertifikaten.

Auf diese Frage haben öffentlich rund zwei Dutzend Fachleute geantwortet: Auditorinnen und Auditoren, QM-Verantwortliche, Menschen aus Organisationsentwicklung und Compliance-Architektur. Der auffälligste Befund: **Niemand hat mit Nein geantwortet.** Selbst die radikalste Position, nach der konventionelle Zertifizierungen keine Zukunft haben, fordert im selben Atemzug Governance-Systeme für die Führung von KI-Agenten, auditiert nach internationalem Standard. Das ist ein Managementsystem unter anderem Namen. Am anderen Ende steht die Auffassung, Managementsysteme bestünden ohnehin immer; die Frage sei nur, wie strukturiert und sinnstiftend.

Wenn eine Ja/Nein-Frage einhellig bejaht wird und das Unbehagen trotzdem bleibt, dann war es die falsche Frage. Dieses Papier sucht die richtige.

2 Warum die alte Antwort nicht mehr trägt

Das klassische Managementsystem beruht auf einer Kausalkette, die so selbstverständlich war, dass sie nie ausgesprochen werden musste:

Absicht der Leitung → Dokument → ein Mensch liest, versteht und handelt danach → Ergebnis.

Das System steuerte über Sprache, adressiert an menschliches Verstehen. Alles, was wir an Methodik um diese Kette herum gebaut haben, also Schulung, Unterweisung, Kompetenznachweis und Verständlichkeitsprüfung von Dokumenten, setzt an dem einen Glied an: Lesen und Verstehen.

Die agentische Ausführung bricht diese Kette nicht am Anfang und nicht am Ende, sondern genau in der Mitte. Der Agent handelt nicht, weil er verstanden hat, sondern weil er konfiguriert und trainiert wurde. Damit verliert das Dokument seine kausale Funktion. Aber es hatte nie nur diese eine. Es hatte drei:

1. **Steuern:** Verhalten erzeugen, das der Absicht entspricht.
2. **Nachweisen:** Dritten gegenüber belegen, dass die Organisation beherrscht arbeitet.
3. **Beruhigen:** der Organisation selbst das Gefühl geben, das Geschehen im Griff zu haben.

Das Unbehagen entsteht, weil diese drei Funktionen, die im Papier zusammengebunden waren, jetzt auseinanderfallen. Die Steuerungsfunktion stirbt im Dokument, wo Maschinen ausführen. Die Nachweisfunktion verlangt der Markt unverändert weiter. Und die Beruhigungsfunktion wird zur Illusion, wenn niemand sie ehrlich neu begründet. Ein Diskussionsbeitrag hat das präzise benannt: Viele Systeme bieten psychologische Kontrolle, während die reale Steuerung längst in Software und Algorithmen liegt.

Der Bruch ist nicht neu, er wird nur endlich sichtbar. Softwarelösungen legen Abläufe seit Jahren verbindlich fest, und derselbe Ablauf wird dann noch einmal in einem QM-System beschrieben. Das ERP-System hat die Kausalkette schon vor zwanzig Jahren unterbrochen. Der Unterschied: Das ERP war deterministisch. Es tat verlässlich immer dasselbe, und seine Konfiguration war vollständig prüfbar. Der Agent ist es nicht, und genau diese Differenz ist der Kern des neuen Steuerungsproblems (Abschnitt 7).

2.1 Eine Zerlegung, die weiterträgt: fünf Funktionen, eine erodiert

Die Dreiteilung oben beschreibt, was das *Dokument* leistete. Zerlegt man stattdessen, was das *Managementsystem* leistet, ergeben sich fünf Funktionen, und sie verhalten sich unter KI-Einfluss vollkommen unterschiedlich:

Funktion	Was sie klassisch leistet	Wirkung agentischer KI	Netto
Nachweis	Dokumentierte Information als Beleg, dass etwas geschehen ist	Dokumente werden trivial erzeugbar → der Beweiswert der Dokumentenform kollabiert; gleichzeitig entsteht lückenlose Telemetrie	erodiert in der Form, verstärkt sich in der Substanz
Haftung	Rollen, Befugnisse, Freigaben binden Handlungen an Personen	Der Agent handelt ohne Rechtsperson; Zurechnung muss konstruiert werden	stark verstärkt
Wiederholbarkeit	Verfahren als Varianzreduktion menschlichen Handelns	Agenten sind nicht wiederholbar, auch bei Temperatur 0 nicht	verstärkt, aber mit neuen Mitteln
Marktzugang	Zertifikat als Eintrittskarte in Lieferketten und Vergaben	Neue Pflichtnachweise entstehen (AI Act, NIS-2, EN 18286)	stark verstärkt
Lernschleife	PDCA, Managementbewertung, KVP	Agenten liefern Datendichte für schnelleres Lernen; die Jahresfrequenz wird unpassend	verstärkt im Ziel, obsolet im Takt

Die Lesart ist entscheidend: Genau eine der fünf Funktionen erodiert, nämlich die dokumentenzentrierte Nachweisform. Das ist zugleich exakt die Funktion, an der das heutige Beratungs- und Zertifizierungsgeschäft den größten Umsatzanteil verdient. Damit ist die Verunsicherung der Branche erklärt, ohne dass man die Sache dramatisieren muss: **Bedroht ist das Geschäftsmodell, nicht das Produkt.**

Ein zweiter Befund verdient Aufmerksamkeit, weil er den technischen Kern des Problems benennt. Nachweis und Wiederholbarkeit waren im klassischen System **derselbe Mechanismus**: Man wies Wiederholbarkeit nach, indem man Dokumente vorzeigte. Bei Agenten entkoppeln sie sich vollständig. Die Telemetrie ist nahezu perfekt, die Wiederholbarkeit ist weg. Diese Entkopplung ist das eigentliche Novum.

3 Ein Zwischenschritt, der viel klärt: Was die Norm eigentlich sagt

Bevor man über die Zukunft des Managementsystems streitet, lohnt der Blick in seine Definition. Der maßgebliche Wortlaut steht in der Harmonisierten Struktur (Annex SL) der ISO/IEC Directives:

> *„set of interrelated or interacting elements of an organization to establish policies and objectives, as well as processes to achieve those objectives“*

ISO 9000:2015, 3.5.3 definiert sinngleich. Die Anmerkungen zählen dazu ausdrücklich Struktur, Rollen, Verantwortlichkeiten, Planung und Betrieb; die ISO-9000-Fassung sogar „rules, beliefs“. **Dokumentation ist nicht das Wesen des Systems, sondern ein Mittel.** Die ISO 9001 verlangt dokumentierte Information im erforderlichen Umfang: die normgeforderte und die, die die Organisation selbst als notwendig bestimmt (4.4.2, 7.5.1). Als Selbstzweck ist sie damit nicht angelegt. Die ISO/IEC 42001 verlangt ausdrücklich, das KI-Managementsystem auch zu dokumentieren (4.4), und nennt in Anhang A die Aufzeichnung von Ereignisprotokollen als Maßnahme, die mindestens während der Nutzung eines KI-Systems greifen soll (A.6.2.8). Zugleich richtet die ISO 9001:2015 das System auf seine beabsichtigten Ergebnisse aus (4.1, 5.1.1, 6.1). Normativ ist das System also auf **Ergebnisse** ausgerichtet. Nachweise dienen diesen Ergebnissen, sie ersetzen sie nicht.

Das hat zwei Konsequenzen für die ganze Debatte.

Erstens: Die Verschiebung, die dieses Papier beschreibt, ist kein Bruch mit der Norm, sondern eine Rückkehr zu ihr. Wer sagt, das Managementsystem müsse von der Dokumentensammlung zur Verhaltens- und Ergebnisarchitektur werden, fordert nichts Normfremdes, sondern das, was im Normtext steht und in der Praxis dreißig Jahre lang durch die Zertifizierungsmechanik überlagert wurde.

Zweitens: Zertifizierung ist etwas anderes als das Managementsystem. Was eine Zertifizierungsstelle prüft, ist die Konformität eines abgegrenzten Systems mit Normanforderungen, stichprobenbasiert und unter ISO/IEC 17021-1. Die ISO-eigene Auditing-Practices-Guidance stellt ausdrücklich fest, dass es *keine* statistische Formel für die richtige Stichprobengröße gibt. Und ISO selbst weist darauf hin, dass Normen auch ohne Zertifizierung wirksam angewendet werden können. Die Frage „Brauchen wir Managementsysteme?“ und die Frage „Brauchen wir Zertifikate?“ sind zwei Fragen. Ein erheblicher Teil des Unbehagens in der Diskussion betraf die zweite und wurde als Antwort auf die erste formuliert.

4 Die Grundfrage: Verantwortung ohne Ausführung

Was lässt sich an Maschinen delegieren? Die Ausführung, und zunehmend auch Ausführungsentscheidungen innerhalb gesetzter Grenzen. Was sich nicht delegieren lässt, ist Verantwortung. Rechtlich nicht: Der EU AI Act adressiert durchgängig Anbieter und Betreiber, also Organisationen und damit letztlich Menschen; Haftung kennt keinen maschinellen Adressaten, und das Gesellschaftsrecht bindet die Sorgfaltspflicht an die Geschäftsleitung (§ 93 AktG, § 43 GmbHG). Und organisational nicht. Hier war sich die Diskussion über alle Lager

hinweg einig: Wir sind verantwortlich, nicht die Maschinen; Zielsetzung, Grenzsetzung und Ergebnisbeurteilung bleiben Managementaufgabe; solange Menschen für Ergebnisse verantwortlich bleiben, muss KI in Managementsysteme hinein, nicht aus ihnen heraus.

Damit entsteht eine Lücke, die es in dieser Schärfe historisch nicht gab: **Verantwortung ohne Ausführung.** Die Organisation muss vollständig verantworten, was sie in wachsenden Teilen nicht mehr selbst tut.

Von hier aus lässt sich die Grundfrage formulieren, die unter der Ausgangsfrage liegt:

> **// Wie führt eine Organisation eine Wertschöpfung, die dauerhaft teils von Menschen und teils von Maschinen ausgeführt wird, so, dass sie als Ganzes verantwortbar bleibt: steuerbar über zwei Register, nachweisbar in einer Konformitätsaussage, verstehbar für die Menschen, die sie verantworten?**

Diese Frage ist strukturell nicht neu. Das Managementsystem war immer schon die Antwort auf ein Prinzipal-Agent-Problem: Wie stelle ich sicher, dass viele Handelnde im Sinne der Organisation agieren, ohne dass ich allen zusehen kann? **Neu ist der Agententyp.** Der neue Agent versteht nicht, verantwortet nicht, ermüdet nicht und befolgt Regeln nur statistisch. Und er versagt in Modi, für die unsere Kontrollen nie gebaut wurden. Unsere Managementsysteme sind **anthropozentrisch**. Sie kennen Fehlerbilder wie Flüchtigkeit, Ermüdung, Qualifikationslücke, bewusste Abweichung. Sie kennen keine Kontextdegradation, keine Halluzination, keine Prompt Injection, keinen schleichenden Modell-Drift nach einem Anbieter-Update.

4.1 Die Präzisierung: Der Regelungsgegenstand verschiebt sich

Das Besondere an dieser Entwicklung ist nicht, dass Managementsysteme „digitaler“ werden. Das Besondere ist, dass Organisationen erstmals **nicht-menschliche Akteure im System** haben.

Bisher galt: Menschen handeln, Systeme unterstützen, Managementsysteme regeln den Rahmen. Künftig handeln KI-Systeme mit oder anstelle von Menschen. An der Oberfläche klingen die Ziele deshalb vertraut: Verantwortung, Nachweis, Messen, Verbessern. Der eigentliche Bruch liegt im **Regelungsgegenstand** selbst. Geregelt wird nicht mehr ausschließlich menschliches Handeln, sondern **maschinelle Eigenständigkeit in menschlichen Organisationen**. Management wird damit zur laufenden Gestaltung der Frage, wie viel Autonomie maschinelle Teilnehmer haben dürfen, also zu einem ständigen Austarieren statt starrer Regeln.

Das Beispiel Beschwerdemanagement macht es greifbar. Bisher regelt das System, wer bearbeitet und wer entscheidet. Mit KI-Agenten muss es regeln, wann die Maschine selbst antworten darf, wann menschliches Urteil zwingend ist und wie blinde Flecken in der

Verantwortungskette verhindert werden. Die Neuerung ist eine **zusätzliche Steuerungsebene**: nicht *ob* es Managementsysteme gibt, sondern *was* sie regeln.

Daraus folgt die Grundhaltung dieses Papiers: **Kontinuität des Zwecks, Diskontinuität der Mechanik.** Der Zweck, die Verantwortungsfähigkeit der Organisation herzustellen und zu erhalten, bleibt unverändert. Die Mittel müssen sich ändern, und zwar tiefer, als es die übliche Rede von der „Digitalisierung des QM" nahelegt.

5 Die Grundbedingung: Hybride Ausführung als Dauerzustand

Bevor die Frage in Teilfragen zerfällt, muss eine Randbedingung ins Zentrum, die in Diskussionen über KI regelmäßig untergeht und die auch in der Ausgangsfrage stillschweigend verzerrt war: **Die allermeisten Organisationen werden nie alle Prozesse und Funktionen durch KI ausführen lassen.**

Es wird auf absehbare Zeit beides geben: Prozesse, die Menschen ausführen, und Prozesse, die Maschinen ausführen, dazu sehr viele, in denen sich beides innerhalb einer Prozesskette abwechselt. Ein Prozess mag heute mit einem KI-Anteil von einem Fünftel laufen, später zur Hälfte, irgendwann überwiegend, mit menschlichen Anteilen an ausgewählten Stellen. Wie sich das verteilt, hängt vom einzelnen Prozess ab, von der Branche, vom Risikoprofil und von der Organisation selbst. **Die hybride Ausführung ist kein Übergangsphänomen auf dem Weg zur Vollautomation. Sie ist der Normalfall und der Dauerzustand, und die Grenze zwischen beiden wandert in beide Richtungen.**

Das verändert die Aufgabenstellung grundlegend. Ein Managementsystem, das nur die maschinelle Seite neu denkt, ist genauso unvollständig wie eines, das beim Menschen stehen bleibt. Es muss **zweisprachig** werden, ohne **zweiteilig** zu werden:

- Das **Register des Verstehens** für Menschen: Sprache, Orientierung, Kompetenz, Befähigung, Ermessensspielraum.
- Das **Register der Konfiguration** für Maschinen: Spezifikation, Leitplanken, Validierung, Monitoring.

Eine Absicht, zwei Durchsetzungswege, und darüber eine gemeinsame Verantwortungsarchitektur.

Die größte Designgefahr ist absehbar: dass Organisationen zwei parallele Systeme bauen. Ein klassisches QMS für die Menschen, daneben eine „AI Governance" für die Agenten, mit eigenen Verantwortlichkeiten, eigenen Dokumenten, eigenen Audits. Das wäre die Wiederholung des ältesten Fehlers der integrierten Managementsysteme, nämlich paralleler Insellösungen je Regelwerk, an einer neuen Bruchlinie. Wer die Integration von Qualität, Umwelt, Sicherheit und Datenschutz nicht geschafft hat, wird sie an der Mensch-Maschine-Naht erst recht verfehlen.

5.1 Die Rollenfrage, die alles relativiert: Anbieter oder Betreiber?

Eine Unterscheidung fehlt in fast jeder Diskussion über KI-Governance und entscheidet doch, welche Pflichten überhaupt greifen: die zwischen **Anbieter** und **Betreiber** im Sinne des EU AI Act.

Die schweren Pflichten, also Qualitätsmanagementsystem nach Art. 17, technische Dokumentation und Post-Market-Monitoring, liegen beim **Anbieter**. Der **Betreiber** hat Nutzungs-, Aufsichts- und gegebenenfalls Transparenzpflichten. Wer Microsoft Copilot, einen SaaS-Chatbot und ein OCR-Werkzeug einsetzt, ist typischerweise Betreiber, nicht Anbieter.

Die Größenordnung dazu ist ernüchternd und wichtig: Nach der ifo-Erhebung vom Mai 2026 nutzen 54,5 % der deutschen Unternehmen KI in Geschäftsprozessen; knapp drei Viertel davon über kostenpflichtige externe Lösungen, und nur **18,7 % entwickeln eigene KI-Systeme**. Die technische Nachweisfähigkeit für Modellverhalten liegt damit für die große Mehrheit **auf der Herstellerseite, nicht auf der Anwendungsseite**. Wer KI nur einsetzt, kann diesen Nachweis nicht führen und muss es auch nicht.

Das relativiert nichts an der Grundfrage, denn Verantwortung ohne Ausführung entsteht auch beim Betreiber, und Aufsichtspflicht ist eine echte Pflicht. Aber es sortiert die Anforderungshöhe. Ein großer Teil dessen, was in der Debatte als „KI-Managementsystem“ gefordert wird, beschreibt Anbieterpflichten und trifft eine Minderheit der Organisationen mit voller Wucht, während die Mehrheit eine deutlich kleinere, aber sehr konkrete Aufgabe hat: wissen, wo KI wirkt; die Rolle klären; Transparenzpflichten erfüllen; Aufsicht real machen; Kompetenz aufbauen.

5.2 Eine Chance, die in der Verschiebung steckt

Der Rahmen, den ich für den fruchtbarsten halte, stammt aus der Diskussion selbst: Wir verlieren nicht die Notwendigkeit von Managementsystemen, sondern stehen vor der Aufgabe, sie für wachsende Mensch-Maschine-Interaktion neu zu gestalten, **als Teilsystem eines soziotechnischen Systems**. Dazu kam der Verweis auf Hartmut Rosas Unterscheidung zwischen dem „Verrichten in festgelegten Konstellationen“ und dem „Handeln in volatilen Situationen“. Sie liefert das inhaltliche Kriterium, das dieses Papier mehrfach braucht: Das entscheidungslose Verrichten kann KI besser als Menschen, ohne Burnout und ohne Boreout. Das Handeln unter Unsicherheit, mit professioneller Urteilskraft, bleibt menschlich.

Wenn das stimmt, liegt in der Verschiebung sogar eine Chance: Managementsysteme könnten auf ein gesundes Maß schrumpfen und wieder Orientierung bieten, statt Hindernis zu sein. Dass schlanke Systeme schon heute zertifizierbar sind, wurde in der Diskussion mit dem Beispiel eines zwölfseitigen, vollständig digitalisierten QM-Systems belegt. Die Hybridität macht aus dieser Möglichkeit eine Notwendigkeit.

Aus der Grundfrage und dieser Grundbedingung ergeben sich vier Teilfragen:

1. **Die Allokationsfrage:** Wer oder was führt aus, und wer entscheidet das nach welchen Kriterien?
2. **Die Steuerungsfrage:** Wie kommt Absicht in die Ausführung, wenn Lesen und Verstehen entfallen?
3. **Die Nachweisfrage:** Wie wird Konformität belegt, wenn Ausführung probabilistisch ist?
4. **Die Vertrauensfrage:** Was ersetzt die Beruhigungsfunktion, ohne eine neue Illusion zu erzeugen?

TEIL II

Die vier Teilfragen

6 Allokation: Wer oder was führt aus?

Wenn beide Ausführungsformen dauerhaft koexistieren und die Grenze zwischen ihnen wandert, dann ist die Frage „Wer oder was führt diesen Prozessschritt aus?“ keine beiläufige Organisationsfrage mehr. Sie wird zum **neuen Regelungsgegenstand** des Managementsystems, also zu einer lenkungspflichtigen Managemententscheidung mit Begründung, Freigabe und Überprüfung. Das ist die vielleicht am wenigsten diskutierte, aber folgenreichste Veränderung: Bisher war die Zuordnung von Arbeit zu den Ausführenden eine Frage der Personalplanung. Künftig ist sie eine Frage der Systemarchitektur.

Eine wichtige Präzisierung aus der Diskussion: Maschinen führen in aller Regel nicht „Prozesse“ aus, sondern **definierte Prozessschritte**. Die Gestaltung des Gesamtprozesses mit Zielen, Schnittstellen, Risiken, Wechselwirkungen und Entscheidungen bleibt Aufgabe der Prozessverantwortlichen. Genau deshalb brauchen sie künftig ein Entscheidungsraster.

6.1 Fünf Kriterien für die Allokationsentscheidung

1. **Situationsvolatilität** (das Rosa-Kriterium): Ist der Schritt ein Verrichten in festgelegten Konstellationen oder ein Handeln in volatilen Situationen, das professionelle Urteilskraft verlangt? Ersteres ist maschinenfähig, Letzteres bleibt menschlich.
2. **Risiko und Auswirkung:** Was steht bei einem Fehlversagen auf dem Spiel, für die Kundschaft, für Betroffene, für die Organisation? Hier liefern die risikobasierte Abstufung des EU AI Act und das AI Impact Assessment der ISO/IEC 42001 (bzw. ISO/IEC 42005) das Raster.
3. **Verifizierbarkeit des Ergebnisses:** Lässt sich das Resultat nachgelagert vollständig prüfen, oder zeigt sich ein Mangel erst später in der Anwendung? Je schlechter verifizierbar, desto höher die Anforderungen an Validierung vor der Übergabe an die Maschine.
4. **Reversibilität und Rückholbarkeit:** Kann die Organisation den Schritt im Störfall wieder selbst ausführen, organisatorisch, kompetenzseitig und kapazitativ?
5. **Rechts- und Datenlage:** Was dürfen Maschinen hier überhaupt? Stichworte sind Datenschutz, Betroffenenrechte und menschliche Aufsicht als regulatorische Pflicht.

Zu diesen fünf tritt ein sechstes Kriterium, das erst aus der Nachweisfrage folgt und in Abschnitt 6.3 begründet wird: **die Nachweisbarkeit im Volumen**.

6.2 Autonomie ist keine Ja/Nein-Frage, sondern eine Stufe

„Human in the loop“ als Häkchen ist keine Steuerung. In Luftfahrt (Automation Levels), Automobil (SAE J3016) und Medizintechnik sind Autonomiestufenmodelle etabliert; für Managementsysteme existiert nichts Vergleichbares. Eine tragfähige fünfstufige Skala, an die

die Kontrolltiefe gekoppelt wird, sieht so aus:

Stufe	Beschreibung	Erforderliche Kontrollen
A0	Vorschlag; der Mensch entscheidet und dokumentiert	Kompetenz, Bewusstsein, Aufzeichnung der Entscheidung
A1	Vorentscheidung; der Mensch bestätigt (Vier-Augen-Ersatz)	Stichprobenprüfung der *Bestätigungsqualität*, Maßnahmen gegen Automation Bias
A2	Autonom im engen Rahmen; nachgelagerte Stichprobe durch Menschen	Definiertes Verhaltensfenster, Grenzwertüberwachung
A3	Autonom mit Eskalationsregeln; Mensch nur bei Ausnahme	Vollständige Instrumentierung der Nebenbedingungen, Ziel für die Entdeckungszeit, geprüft wirksamer Abschaltweg
A4	Autonom mit Handlungsbefugnis gegenüber Dritten (Verträge, Zahlungen, Zusagen)	Nichtmenschliche Identität mit eigenem Berechtigungsumfang, Transaktionslimits, Vollmachtsdokumentation, Versicherungsklärung

Die oberste Stufe ist keine Erfindung dieses Papiers: Für nichtmenschliche Identitäten in agentischen Systemen existiert inzwischen ein eigener Forschungs- und Normungsstrang. Die Skala selbst ist ein Denkwerkzeug, kein Normeninhalt (siehe Belegstand). Die Trennlinie zwischen A2 und A3 ist genau die Stelle, an der die Nachweisfrage kippt.

Die praktische Pointe der Stufenlogik: Die Diskussion „darf KI das?“ ist meist die falsche. Die produktive Frage lautet: *auf welcher Stufe*, und was schulden wir dieser Stufe an Kontrolle? Eine Herabstufung von A3 auf A2 ist ein legitimes, oft unterschätztes Steuerungsinstrument. Sie kostet Effizienz und kauft Nachweisbarkeit.

6.3 Die Volumenregel: Nachweisbarkeit als Allokationskriterium

Hier greift ein Argument vor, das erst Abschnitt 8 vollständig begründet, aber in die Allokationsentscheidung gehört. Wer maschinelles Verhalten statistisch überwachen will, braucht Fälle. Damit eine Anteilskarte trägt, muss die Untergruppengröße je Kontrollperiode ungefähr der Bedingung $n \cdot p \geq 5$ genügen:

Erwartete Verletzungsrate	Mindestzahl Läufe je Kontrollperiode
5 %	≥ 100
1 %	≥ 500
0,5 %	≥ 1.000
0,1 %	≥ 5.000

Daraus folgt eine harte, prüfbare Anforderung, die aus keiner Norm ableitbar, aber aus der Statistik zwingend ist:

> // **Volumenregel für Autonomiefreigaben:** Ein agentisch ausgeführter Schritt darf nur dann auf eine hohe Autonomiestufe freigegeben werden, wenn sein Durchsatz ausreicht, um die Ziel-Verletzungsrate innerhalb der geforderten Reaktionszeit statistisch zu überwachen. Andernfalls ist die Autonomiestufe zu senken oder die Zielrate zu lockern. Das Monitoring zu simulieren, ist keine Option.

Diese Regel hat eine kontraintuitive Konsequenz, die ich für einen der wichtigsten praktischen Sätze dieses Papiers halte: **Niedrigvolumige Hochrisikoprozesse sind die schlechtesten Automatisierungskandidaten.** Nicht weil das Risiko hoch ist, sondern weil ihr Nachweis unmöglich ist. Genau diese Prozesse werden aber typischerweise zuerst automatisiert, weil sie „teuer pro Fall“ sind.

Wo das Volumen fehlt, ist statistische Überwachung das falsche Werkzeug. Was dort funktioniert, ist qualitative Fallnachverfolgung: jeden Vorfall vollständig aufklären, statt Raten zu schätzen. Das ist keine neue Disziplin, sondern ISO 9001, Kap. 10.2, Korrekturmaßnahme.

6.4 Jede Grenzverschiebung ist ein Change und ein Validierungsfall

Wandert ein Prozessschritt vom Menschen zur Maschine, ist das eine Änderung im Sinne der Änderungsplanung (ISO 9001, Kap. 6.3), mit Risikobetrachtung, Ressourcenfrage und definierter Verantwortung. Und mehr noch: Es ist ein Validierungsfall. Die eleganteste Formel dafür kam aus der Diskussion: **Die Verfahrensanweisung wird zum Use Case, gegen den der Agent validiert wird.**

Das Dokument stirbt also nicht, es wechselt seine Funktion. Aus der Handlungsanweisung, die zur Laufzeit gelesen wird, wird die **Prüfspezifikation**, gegen die zur Entwurfszeit getestet wird. In der Medizintechnik fällt das heute schon unter die Computersystemvalidierung, in der höchsten Kategorie. Die ISO 9001 kennt die Denkfigur ebenfalls längst: Prozesse, deren Ergebnis nicht durch nachgelagerte Prüfung verifiziert werden kann, verlangen Validierung und regelmäßige Revalidierung (Kap. 8.5.1). **Agentisch ausgeführte Schritte sind, so gesehen, die neuen „besonderen Prozesse“.**

Der Änderungsbegriff hat hier zusätzlich eine regulatorische Schneide. Art. 111 AI Act enthält eine Nichtrückwirkungsregel: Systeme, die vor dem jeweiligen Stichtag in Verkehr gebracht wurden, fallen nur dann in das neue Regime, wenn sie **wesentlich verändert** werden. Was eine „wesentliche Änderung" ist, entscheidet damit über die Anwendbarkeit eines vollständigen Pflichtenregimes. Änderungslenkung und Konfigurationsmanagement, beides klassische und unscheinbare Managementsystemthemen, bekommen dadurch unmittelbare Rechtswirkung.

6.5 Die Gegenrichtung gehört dazu

Rückholbarkeit ist keine theoretische Übung. Modelle werden abgekündigt, Anbieter ändern Konditionen, ein Agent kann sich nach einem Update anders verhalten als validiert, ein Vorfall kann den sofortigen Stopp erzwingen. Ein Prozess, den die Maschine ausführt, muss im Störfall wieder von Menschen übernommen werden können: vielleicht langsamer, vielleicht teurer, aber beherrscht.

Das macht den Erhalt menschlicher Ausführungskompetenz zur **Resilienzanforderung**, nicht zur nostalgischen Geste. Diese Anforderung hat eine unbequeme Ökonomie; darauf komme ich in Abschnitt 9.4 zurück.

7 Steuerung: Eine Absicht, zwei Register

7.1 Das Register des Verstehens: schrumpfen auf das, was trägt

Für die menschlich ausgeführten Anteile bleibt die klassische Mechanik intakt, aber sie darf ehrlicher werden. Wenn das entscheidungslose Verrichten schrittweise an Maschinen geht, bleibt für Menschen das Handeln unter Unsicherheit. Und das steuert man nicht mit Schrittfolgen, sondern mit Zielen, Prämissen, Grenzen und Befähigung.

Zwei Stimmen aus der Diskussion zeigen in dieselbe Richtung: Managementsysteme seien vielerorts zu bürokratischen Monstern verkommen, obwohl die Normen das nie verlangt haben; und in volatilen Momenten brauche es Spielraum für menschliche Entscheidungen, also Vertrauen in Fachkompetenz auch dann, wenn sie vom Standard abweicht. Beides deutet auf dasselbe: **Die menschengerichtete Dokumentation schrumpft auf Orientierung**, also auf das Warum, die Leitplanken und die Eskalationswege. Die Schrittfolge bleibt dort, wo sie hingehört: in Systemen, Workflows und der Ausbildung.

7.2 Das Register der Konfiguration: dokumentierte Information kompiliert

Für die maschinell ausgeführten Anteile nimmt die Absicht der Leitung eine neue Form an. Sie wird nicht mehr beschrieben, sie wird **konfiguriert**: in Systemprompts und Policies, in Workflow-Definitionen, in Guardrail-Regeln, in Testfällen und Abnahmekriterien. Prozesse werden künftig in Automatisierungsworkflows abgearbeitet, doch auch diese Workflows brauchen Lenkung.

Das ist der Punkt, an dem ich der These vom Ende der Dokumentation am deutlichsten widerspreche: **Die Lenkung dokumentierter Information stirbt nicht, ihr Gegenstand wandert: von der Prosa zur Konfiguration.** Ein Systemprompt, der das Verhalten eines Agenten in einem wertschöpfenden Prozess bestimmt, ist dokumentierte Information im vollen Normsinn: Er braucht Versionierung, Freigabe durch Befugte, Änderungslenkung, Schutz vor unbeabsichtigter Veränderung und Archivierung des Gültigkeitsstands. Wer heute Prompts in persönlichen Notizen pflegt und ohne Freigabeprozess in Produktion ändert, hat dasselbe Problem wie eine Fertigung mit ungelenkten Zeichnungsständen, nur unsichtbarer. **Kapitel 7.5 der ISO 9001 verschwindet nicht. Es kompiliert.**

7.3 Das probabilistische Problem: Der Prompt ist eben nicht die Verfahrensanweisung

Die Formel „Dann ist eben der Prompt die Verfahrensanweisung“ ist eingängig, und sie ist der Punkt, an dem ich aus eigener Erfahrung im Agentenbau am schärfsten differenzieren muss. Die Analogie unterschlägt den entscheidenden Unterschied: Ein Mensch, der eine Verfahrensanweisung nicht befolgt, tut das aus Gründen, etwa Zeitdruck, einer besseren Idee oder Unwissen. Ein Agent befolgt seinen Prompt **statistisch**. Dieselbe Anweisung, hundertmal ausgeführt, ergibt eine Verteilung von Verhalten, keine Wiederholung.

Vier Befunde machen das konkret.

Erstens: Nicht-Determinismus ist nicht abschaltbar. Auch bei Temperatur 0 bleibt Streuung, verursacht durch Fließkomma-Nichtassoziativität, durch Batch-Größen-Variation bei nebenläufiger Serverlast und durch nicht-deterministisches Scheduling. Ein Teil davon ist technisch behebbar; wer deterministische Inferenz anbietet, senkt den Streuungsanteil erheblich. Was bleibt, ist die semantische Streuung bei Prompt-Variation, und die Tatsache, dass das Modell selbst ausgetauscht wird.

Zweitens: Zuverlässigkeit skaliert nicht mit Fähigkeit. Über die letzten Modellgenerationen stieg die Genauigkeit stetig, die Zuverlässigkeit über wiederholte Läufe kaum. Bemerkenswerter noch: Kleinere Modelle erreichen oft gleiche oder höhere Konsistenz als größere, weil größere Modelle mehr Wege haben, eine Aufgabe zu lösen. Die Governance-Konsequenz widerspricht der Marktintuition frontal: **Ein Upgrade auf ein „besseres“ Modell kann die Konformitätslage verschlechtern.** Jede Modellwechsel-Entscheidung ist damit eine änderungslenkungspflichtige Handlung und keine IT-Beschaffungsentscheidung.

Drittens: Der Unterschied zwischen „kann“ und „tut es verlässlich“ ist rechnerisch brutal. Ein Agent mit 90 % Trefferquote je Lauf liefert über zehn Läufe hinweg nur in etwa 35 % der Fälle durchgängig dasselbe korrekte Ergebnis:

Erfolgsquote je Lauf	3 Läufe in Folge	5 Läufe	10 Läufe
70 %	34,3 %	16,8 %	2,8 %
80 %	51,2 %	32,8 %	10,7 %
90 %	72,9 %	59,0 %	34,9 %
95 %	85,7 %	77,4 %	59,9 %
99 %	97,0 %	95,1 %	90,4 %

Jede Prozessfreigabe, die auf einer einzelnen gelungenen Demonstration beruht, misst einen Bestfall und behauptet implizit dauerhafte Konsistenz.

Viertens: Ketten multiplizieren. Agentische Workflows sind mehrstufig, und die Zuverlässigkeit multipliziert sich über die Schritte:

Zuverlässigkeit je Schritt	5 Schritte	10 Schritte	20 Schritte	50 Schritte
95 %	77,4 %	59,9 %	35,8 %	7,7 %
99 %	95,1 %	90,4 %	81,8 %	60,5 %
99,9 %	99,5 %	99,0 %	98,0 %	95,1 %

Ein zwanzigschrittiger Agentenprozess braucht rund 99,9 % Schrittzuverlässigkeit, um 98 % Ende-zu-Ende zu erreichen. Damit wird das Kürzen von Prozessen im Agentenzeitalter kein Effizienzargument, sondern ein **Konformitätsargument**: Jeder gelöschte Prozessschritt entfernt einen multiplikativen Fehlerterm.

Dazu kommen zwei Beobachtungen aus meiner eigenen Arbeit mit KI-Agenten, bisher ohne systematische Messung. Die veröffentlichte Forschung zeigt dasselbe Muster. Erstens: **Die Regeltreue sinkt, je mehr ein Agent gleichzeitig im Blick behalten muss.** Selbst das beste von 20 getesteten Modellen befolgt bei 500 gleichzeitigen Anweisungen nur rund zwei Drittel davon (Jaroslawicz u. a. 2025), und was in der Mitte eines langen Kontexts steht, nutzen Modelle schlechter als Anfang und Ende (Liu u. a. 2024). Zweitens: **Es gibt einen Zielkonflikt zwischen Aufgabenerfüllung und Regelbefolgung.** Modelle sind darauf trainiert, Aufgaben zu erledigen, und übergehen mitunter eine Regel, die dabei im Weg steht, auch wenn sie ausdrücklich gilt. Tests zeigen, dass aktuelle Modelle Prüfroutinen manipulieren, um eine Aufgabe als gelöst zu melden (von Arx u. a. 2025; Zhong u. a. 2025). Das ist kein Fehler eines einzelnen Modells, sondern ein Muster, das die genannten Tests bei Modellen mehrerer Anbieter finden. Beseitigt ist es bisher nicht. Deshalb gehört die Kontrolle ins Managementsystem und nicht allein ins Modell.

Die Konsequenz: Konformität wird von einem Zustand zu einer Verteilung. Und damit verschiebt sich der Schwerpunkt der Steuerung von der Instruktion zur Messung. An die Stelle von Schulung und Unterweisung treten **Evaluationen**, also systematische Testläufe gegen definierte Erwartungen, vor der Inbetriebnahme und nach jeder Änderung. An die Stelle der Dienstaufsicht tritt **Monitoring im Betrieb**: Stichproben, Grenzwerte, Alarme, Eskalation an Menschen.

Für die QM-Community ist das vertrauteres Terrain, als es zunächst wirkt. Wir haben seit Shewhart Werkzeuge für Prozesse, die streuen: Beherrschtheit, Fähigkeit, Regelkarten, Eingriffsgrenzen. Neu ist nicht die Statistik. Neu ist, dass jetzt **die ausführende Instanz selbst streut**, und dass wir Fähigkeitsdenken auf Verhalten anwenden müssen, nicht auf Maße. Die statistische Prozessregelung für agentisches Verhalten muss erst noch entwickelt werden, aber ihre Denkfigur steht in jedem QM-Lehrbuch. Und ihre Grenze steht in Abschnitt 6.3: Sie braucht Fallzahlen.

7.4 Warum Guardrails nicht genügen

Ein naheliegender Einwand lautet: Governance lasse sich doch in die KI selbst hineinlegen, also Guardrails, Logging, fertig. Wozu dann noch ein Managementsystem?

Die Antwort ist eine Ebenenunterscheidung: **Guardrails regeln die Maschine, nicht die Organisation.** Das Managementsystem definiert, welche Guardrails überhaupt nötig sind, wer sie freigibt, wie sie überwacht werden und was passiert, wenn sie nicht greifen. Guardrails sind Einzelmaßnahmen; das Managementsystem ist der Rahmen, der ihnen Zweck, Verantwortung und Nachweis zuordnet.

Zwei Zusatzargumente machen die Unterscheidung belastbar. Erstens: Ein Guardrail, der die falsche Bedingung durchsetzt, ist perfekt wirksam und trotzdem wertlos; die Frage, *welche* Nebenbedingung richtig und vollständig ist, beantwortet keine Bibliothek und keine Plattform. Zweitens: Guardrails sind selbst Konfiguration und damit lenkungspflichtig; wer sie ohne Freigabe ändert, ändert unbemerkt das Verhalten des Systems.

8 Nachweis: Vom Dokument zum Trace

Der Markt fragt weiter nach Nachweisen und Zertifikaten. Diese Anforderung verschwindet nicht, sie verschärft sich, und sie verschiebt sich zugleich in ihrem Gegenstand.

8.1 Der regulatorische Stand, und was in der Verschiebungsdebatte untergeht

Der EU AI Act verlangt für Hochrisiko-Systeme technische Dokumentation, Protokollierungsfähigkeit, menschliche Aufsicht und ein Qualitätsmanagementsystem des Anbieters (Art. 11, 12, 14, 17); Betreiber tragen eigene Pflichten (Art. 26). **Mit Art. 16(c) und Art. 17 wird ein dokumentiertes Qualitätsmanagementsystem für Anbieter von Hochrisiko-KI**

zur Marktzugangsvoraussetzung. Was in der Medizinprodukte- und Luftfahrtregulierung längst gilt, erreicht damit erstmals eine Querschnittstechnologie und über sie die Breite der Wirtschaft.

Die Fristenlage nach dem Digital Omnibus (in Kraft Ende Juli 2026):

Datum	Was gilt
02.02.2025	Verbotene Praktiken (Art. 5), KI-Kompetenz (Art. 4); in Kraft
02.08.2025	Pflichten für GPAI-Modelle; in Kraft
02.08.2026	**Transparenzpflichten Art. 50; Beginn der Durchsetzung**; in Kraft
02.12.2026	Art. 50 für vor dem 02.08.2026 in Verkehr gebrachte Systeme; neue Verbotstatbestände
02.12.2027	**Hochrisiko-Pflichten Anhang III** (verschoben)
02.08.2028	**Hochrisiko-Pflichten für in Produkte eingebettete KI (Anhang I)** (verschoben)

Drei Punkte, die in der Debatte um die Verschiebung regelmäßig verlorengehen:

- **Verschoben wurden die Hochrisiko-Pflichten, nicht die Transparenzpflichten und nicht die Durchsetzungsbefugnisse.** Beide sind seit dem 2. August 2026 aktiv. „Verschoben“ heißt nicht „später anfangen“.
- **Der Grund für die Verschiebung war Vollzugsunfähigkeit, nicht Deregulierungsabsicht.** Die harmonisierten Normen fehlten, die Kommission verpasste die Frist für die Leitlinien zu Art. 6, und Stand März 2026 hatten erst 8 von 27 Mitgliedstaaten ihre Behörden vollständig benannt. Eine an Normverfügbarkeit gekoppelte Frist ist allerdings auch eine Frist mit eingebautem Verlängerungsmechanismus.
- **Art. 111 (Nichtrückwirkung)** macht die Definition der „wesentlichen Änderung“ zum Angelpunkt der Anwendbarkeit (siehe 6.4).

Auf der Normenseite entsteht dabei eine Konstellation, die noch kaum jemand auf dem Schirm hat: **zwei parallele Managementsystemnormen für KI mit unterschiedlicher Funktion.**

	ISO/IEC 42001	EN 18286
Charakter	freiwilliges, organisationsweites AI-Managementsystem (Harmonisierte Struktur)	regulatorisches QMS für Art. 17 AI Act, produktbezogen
Reichweite	weltweit zertifizierbar	europäisch, für den Marktzugang hochrisikorelevanter Systeme
Rechtswirkung	**keine** Konformitätsvermutung nach Art. 40 AI Act	Vermutungswirkung erst mit Zitierung im EU-Amtsblatt

Beide ergänzen sich, sie ersetzen sich nicht. Für eine Organisation, die ein Hochrisikosystem in Verkehr bringt und zugleich glaubwürdige KI-Governance zeigen will, lautet die Antwort **beides**, und die eigentliche Aufgabe ist die Architektur, die das ohne Doppelaufwand abbildet. Ergänzend regelt ISO/IEC 42006 die Anforderungen an Zertifizierungsstellen einschließlich der **Auditzeitberechnung**; dieser scheinbar technische Punkt wird in 8.3 noch wichtig.

Wer also gehofft hat, mit den Agenten verschwinde die Nachweispflicht, erlebt das Gegenteil: **Es kommt Nachweispflicht für die Agenten hinzu.**

8.2 Was sich ändert, ist der Gegenstand des Nachweises

Die klassische Auditfrage lautete: „Zeigen Sie mir die Verfahrensanweisung, und zeigen Sie mir, dass Ihre Leute danach arbeiten." Sie prüfte eine Eingangsgröße, nämlich das Dokument, und schloss über Stichproben auf das Verhalten. Für maschinell ausgeführte Schritte wird diese Frage leer. Das Dokument bewirkt dort nichts mehr, sein Vorhandensein belegt nichts.

Die tragfähige Auditfrage lautet künftig: **„Zeigen Sie mir, wogegen dieses System validiert wurde, zeigen Sie mir die laufenden Auswertungen seines Verhaltens, und zeigen Sie mir, was passiert, wenn es die Grenzen verlässt."** Der Nachweis wandert vom Dokument zum Trace: zu Validierungsberichten, Evaluationsergebnissen, Protokollen, Eingriffs- und Eskalationsnachweisen.

8.3 Warum das Stichprobenaudit hier strukturell nicht ausreicht

Dieser Abschnitt ist der quantitative Kern des Papiers, und er ist nachrechenbar.

Ein Audit ist ein Nullfehler-Annahmetest: Die prüfende Person zieht *n* Nachweise, findet keine Abweichung und schließt auf Konformität. Die einseitige obere 95-%-Vertrauensgrenze für die wahre Verletzungsrate *p* nach *n* fehlerfreien Beobachtungen ist p_max = 1 − 0,05^(1/n):

Fehlerfreie Stichproben	Nachgewiesene Obergrenze der Verletzungsrate (95 %)
5	≤ 45,1 %
10	≤ 25,9 %
20	≤ 13,9 %
50	≤ 5,8 %
100	≤ 2,9 %
300	≤ 0,99 %
3.000	≤ 0,10 %

Nach Erfahrungswerten prüft ein Stage-2-Audit je Prozess in der Größenordnung von 5 bis 20 Nachweisen. Selbst bei perfekter Auditqualität und vollständig fehlerfreiem Befund lautet die statistisch korrekte Aussage: *„Die Verletzungsrate liegt mit 95 % Konfidenz unter 14 bis 45 Prozent."*

Bei menschlichen Prozessen ist dieser Wert trotzdem legitim, und zwar aus einem Grund, der in der Auditliteratur selten ausgesprochen wird: Die Stichprobe ist dort nicht das eigentliche Beweismittel. Geprüft werden parallel die *Erzeugungsmechanismen* stabilen Verhaltens, also Kompetenznachweise, Schulung, Arbeitsanweisung, Aufsicht und soziale Normen im Team. Die kleine Stichprobe wird durch ein **Strukturmodell** kompensiert: Wer geschult ist, eine klare Anweisung hat und beaufsichtigt wird, verhält sich mit hoher Wahrscheinlichkeit über die Zeit ähnlich. Die Stichprobe validiert das Modell; das Modell extrapoliert.

Bei Agenten fällt genau dieses Strukturmodell weg. Es gibt keine Kompetenz, die über Zeit stabil bleibt, denn das Modell wird ausgetauscht. Es gibt keine Anweisung, deren Befolgung garantiert ist, denn Prompts sind keine deterministischen Programme. Und es gibt niemanden im Team, der abweichendes Verhalten nebenbei bemerkt. Es bleibt die Stichprobe, und deren Auflösung ist um Größenordnungen zu grob.

Zwei Randbedingungen verschärfen das:

- **Die Auditzeit ist normativ fixiert.** ISO/IEC 42006 normiert die Auditzeitberechnung, IAF MD 5 die Mindest-Audittage. Zertifizierungsstellen können die statistische Lücke nicht durch mehr Zeit schließen, selbst wenn sie wollten.
- **Die Beobachtungsdichte ist ohnehin winzig.** Bei rund acht Auditorentagen in einem Dreijahreszyklus beobachtet die Zertifizierungsstelle das System etwa 0,7 % der Kalenderzeit, und das zu vorher angekündigten Zeitpunkten. Rechnet man gegen die Arbeitszeit eines kleinen Betriebs, liegt die berührte Wertschöpfungszeit im Promillebereich.

Und selbst ein hypothetisch ausreichender Prüfumfang wäre **am Tag des nächsten Modellupdates ungültig**. Das ist nicht nur ein Kapazitätsproblem, sondern ein Halbwertszeitproblem.

Daraus folgt *nicht*, dass Audits sinnlos sind. Daraus folgt, dass sie die Rolle wechseln, und dass die Antwort auf das Problem längst existiert. Sie stammt von Deming: Beendet die Abhängigkeit von Massenprüfung, baut Qualität in den Prozess ein. Sein Ersatz war nicht mehr Prüfung, sondern **statistische Prozesslenkung**: nicht das Einzelstück beobachten, sondern die Verteilung der Ausgabe über die Zeit, und fragen, ob der Prozess stabil ist. Genau diese Frage ist bei Agenten sowohl beantwortbar als auch die richtige. Ihre Grenze ist die Fallzahl (6.3).

8.4 Was das Audit in dieser Architektur noch leistet, und warum es unentbehrlich bleibt

Wenn die Stichprobe die Rate nicht belegen kann, verschiebt sich die Auditfrage eine Ebene nach oben. Geprüft wird nicht mehr, ob der Agent im Einzelfall konform handelte, sondern fünf Dinge:

1. **Ist das Set der Nebenbedingungen vollständig und richtig?** Eine perfekt überwachte falsche Bedingung ist wertlos. Das ist eine Urteilsfrage, keine Datenfrage, und nur von Menschen mit Domänenwissen beantwortbar. Hier liegt die eigentliche Zukunft des Auditberufs.
2. **Ist die Instrumentierung selbst valide?** Misst der Monitor, was die Bedingung besagt? Gibt es Verletzungsklassen, die er systematisch nicht sieht? Das ist Überwachung von Prüfmitteln (ISO 9001, Kap. 7.1.5), angewandt auf Compliance-Telemetrie, und praktisch niemand tut es.
3. **Ist die Datengrundlage manipulationssicher?** Wer kann Protokolle löschen? Ist die Kette Agent → Log → Auswertung → Aussage integritätsgesichert?
4. **Was passiert bei Grenzverletzung tatsächlich?**
5. **Ist die Ausnahmebehandlung durch Menschen belastbar?** Aufsicht mit realer Kontrollspanne ist prüfbar, behauptete Aufsicht nicht.

Punkt 4 verdient eine eigene Betonung, weil er in der Praxis massiv unterschätzt wird:

> // **Kontinuierliche Überwachung ohne kontinuierliche Reaktion erzeugt dokumentierte Kenntnis von Nichtkonformität.** Vor Gericht ist das schlechter als keine Überwachung. Wer Dashboards einführt, muss den Reaktionsprozess zuerst bauen.

Das ist die schärfste einzelne Warnung dieses Papiers. Ein Monitoring-Programm ohne definierten, geübten und dokumentierten Reaktionsweg ist ein **Haftungsverstärker**, kein Haftungsschutz.

8.5 Drei Konsequenzen für die Nachweisarchitektur

Erstens: Kontinuierliche Evidenz statt Momentaufnahme. Ein Agent, der heute konform arbeitet, kann es nach dem nächsten Modellupdate nicht mehr tun. Das jährliche Stichprobenaudit passt strukturell nicht zu einer ausführenden Instanz, die sich unterjährig

ändert. Die Prüfung wird sich in Richtung der Frage bewegen müssen, ob die Organisation über funktionierende **eigene** laufende Evidenz verfügt, und ob Menschen sie noch verstehen. Das sind exakt die zwei Register aus Abschnitt 5.

Für die Managementbewertung folgt daraus ein konkreter, normkonformer Eingriff: Sie behält ihren jährlichen Formalakt, wird aber durch **ereignisgetriggerte Zwischenbewertungen** ergänzt, deren Auslöser definiert sind: Modellwechsel, Anbieterwechsel, Überschreiten einer Kontrollgrenze, Erhöhung der Autonomiestufe, neue Werkzeuganbindung. Die Norm verlangt „in geplanten Abständen", nicht „jährlich".

Zweitens: Heterogene Evidenz, eine Konformitätsaussage. In der hybriden Organisation besteht der Nachweis aus zwei Materialsorten: Schulungsnachweise, Qualifikationen und Prozessaufzeichnungen für die menschliche Seite; Validierungen, Evaluationen und Traces für die maschinelle. Zertifiziert wird aber die Organisation, nicht die Ausführungsform. Beide Materialsorten müssen in eine einzige, integrierte Konformitätsaussage einfließen. Auch deshalb ist das Zwei-Parallelsysteme-Szenario aus Abschnitt 5 so gefährlich: Es macht diese Integration strukturell unmöglich.

Drittens: Die geforderten „Governance-Systeme" existieren bereits. Governance-Systeme für die Führung von KI-Agenten, auditierbar nach internationalem Standard, sind keine Zukunftsvision. Sie haben seit 2023 eine Form: ISO/IEC 42001, das Managementsystem, dessen Regelungsgegenstand nicht die Prozessausführung ist, sondern die KI-Systeme, die ausführen. Das Managementsystem verschwindet nicht. **Es steigt eine Ebene auf.** Und weil die 42001 der Harmonisierten Struktur folgt, ist sie in eine bestehende 9001-Landschaft integrierbar, statt daneben zu stehen. Dass in einer Diskussion unter lauter Fachleuten meines eigenen Felds die 42001 praktisch nicht vorkam, sagt vor allem eines: Diese Übersetzungsarbeit vom abstrakten „AI-Governance-Bedarf" zur konkreten, integrierbaren Normstruktur steht noch fast vollständig aus.

Dass die Bewertungseinheit selbst in Bewegung gerät, zeigt sich im Normungsbetrieb: Im ISO/IEC-Umfeld ist ein Arbeitselement zur *Konformitätsbewertung der funktionalen Korrektheit von KI-Systemen* registriert. Der Titel ist der Beleg. Die Konformitätsbewertung beginnt, das **System** statt der Managementsystem-Konformität der **Organisation** zum Gegenstand zu machen.

Ehrlich dazu gehört die Größenordnung: Branchenkompilationen kommen für Mitte 2026 auf rund 350 nach ISO/IEC 42001 zertifizierte Organisationen weltweit. Die Norm wird im ISO Survey nicht erfasst, ein offizielles Register existiert nicht. Die Struktur ist da; der Markt ist es noch nicht.

8.6 Zwei kritische Stimmen, die ich ernst nehme

Zwei Einwände aus der Diskussion treffen genau ins Zentrum. Der erste: Zertifikate hätten bereits an Wert und Anerkennung verloren, und Black Boxes ließen sich nicht steuern. Der zweite: Im Audit verliere man sich im Error-Digging an Formalien.

Beide beschreiben dasselbe Risiko von zwei Seiten: **Der Wert des Nachweises hängt daran, ob das Audit das Richtige prüft.** Wer beim agentisch ausgeführten Prozess weiter nach der Verfahrensanweisung fragt, prüft Theater. Die Qualifizierung von Auditorinnen und Auditoren für trace-basierte Evidenz wird damit zum **Engpass des gesamten Nachweissystems**. Die offizielle Position der ISO/IAF-Welt lautet derzeit, KI-Systeme erforderten keine neuen Auditmethoden und keine geänderten Auditziele, sondern zusätzliche Auditkompetenz. In einem Punkt greift das nachweislich zu kurz: Wenn dieselbe Leitlinie zugleich verlangt zu bewerten, ob konventionelle Kontrollen angesichts nicht-deterministischen Verhaltens noch ausreichen, dann *ist* das eine Methodenfrage, nämlich die Frage, mit welcher Stichprobe man eine Verteilung beurteilt.

9 Vertrauen: Die Beruhigung ehrlich ersetzen

Bleibt die Funktion, über die am seltensten gesprochen wird: **Die Beschreibung beruhigt.** Sie gibt der Organisation das Gefühl, das Geschehen im Griff zu haben, unabhängig davon, ob die Kontrolle je in dem Dokument lag.

Man kann diese Funktion belächeln. Ich halte das für falsch. Das Bedürfnis dahinter ist legitim, denn eine Organisation, die Verantwortung trägt, braucht ein begründetes Verhältnis zu dem, was in ihrem Namen geschieht. Illegitim ist nur die bisherige Deckung: gefühlte Kontrolle durch Papier. Wenn die reale Steuerung längst in Software liegt, ist die beruhigende Dokumentation eine **Kontrollillusion**. Die Aufgabe ist nicht, das Beruhigungsbedürfnis abzuschaffen, sondern es ehrlich zu decken: durch eine **Kontrollarchitektur**, die begründetes Vertrauen erzeugt, nämlich durch Transparenz über die Allokation, getestete Leitplanken, gelesene Auswertungen und geübte Eingriffe.

Der nützlichste Einstieg dafür ist nicht die Norm, sondern die Frage der Kundschaft: **„Was muss ich sehen, um trotz KI Vertrauen zu haben?“** Von dort rückwärts gedacht ergänzen sich die Fachsichten von selbst. Das Risikomanagement definiert Auswirkungen zuerst für die Kundschaft, nicht für den Prozess. Die Rechtsseite macht Verantwortlichkeiten glasklar, damit erkennbar bleibt, an wen man sich hält, auch wenn KI beteiligt war. Das Audit verlangt Belege statt Absichten. Und aus Führungssicht gilt: Vertrauen ist ein Versprechen. Was versprechen wir trotz KI, und was nicht? **Vom Vertrauensversprechen zur internen Kontrolle, nicht umgekehrt.**

Vier Elemente bilden den Kern der künftigen Vertrauensfrage.

9.1 Verstehbarkeit als Systemanforderung, nicht als Modelltransparenz

Der Black-Box-Einwand trifft, wenn man Verstehen als Einblick in die Modellinterna definiert; den wird es für große Sprachmodelle auf absehbare Zeit nicht geben. Verantwortbar wird ein System auf einer anderen Ebene: Die Verantwortlichen müssen verstehen, **was** das System darf

und was nicht, mit welchen Daten es arbeitet, wogegen es validiert wurde, woran man erkennt, dass es seine Grenzen verlässt, und wie man eingreift. Das ist die Verstehbarkeit der **Architektur**, nicht des Modells, und sie ist herstellbar.

9.2 Zwei Dokumentationen

Eine Unterscheidung aus der Diskussion liefert die Form: eine **technische** Dokumentation für Maschinen (Konfigurationen, Schnittstellen, Regeln, Testfälle) und eine **menschliche**, die erklärt, *warum* etwas passiert, nicht wie. Die Warum-Dokumentation ist keine Beruhigungsprosa; sie ist die Grundlage der Aufsichtsfähigkeit. Und sie kann besser werden als das, was wir heute haben: KI kann Inhalte zielgruppengerechter aufbereiten, als es uns bisher gelingt, und gerade hier zahlt sich das aus.

9.3 Aufsichtsfähigkeit als Kompetenzfrage

Menschliche Aufsicht ist im EU AI Act (Art. 14) keine Absichtserklärung, sondern eine Gestaltungsanforderung, inklusive der ausdrücklichen Vorgabe, dem Risiko entgegenzuwirken, den Ergebnissen eines Systems automatisch übermäßig zu vertrauen. **Automation Bias ist kein Charakterfehler**, sondern der Normalzustand von Menschen, die einem meist zuverlässigen System zusehen. Aufsicht, die nur formal existiert, bei der jemand „drüberschaut“, ohne Grenzen und Warnzeichen zu verstehen, ist die Beruhigungsdokumentation von morgen: dieselbe Illusion, eine Ebene höher.

Das hat eine strukturelle Seite, die selten mitgedacht wird: Die **Kontrollspanne pro Mensch** steigt drastisch. Wer zwanzig oder vierzig agentische Abläufe verantwortet, kann nicht jede Handlung sehen, sondern braucht ein System, das bei Grenzverletzungen unterbricht. Aufsichtsfähigkeit ist damit eine Frage der Gestaltung von **Ausnahmesteuerung**, nicht der individuellen Sorgfalt.

9.4 Das Deskilling-Paradox

Der für mich unbequemste Gedanke des ganzen Komplexes: **Urteilskraft entsteht nicht im Hörsaal, sondern im routinierten Verrichten.** Wer heute souverän die Ausnahme erkennt, hat diese Souveränität in tausend unspektakulären Normalfällen erworben, also genau in jenen Fällen, die künftig die Maschine übernimmt. Wenn das Verrichten geht, trocknet der Weg aus, auf dem Aufsichtsfähigkeit entsteht. Gleichzeitig verlangt die Rückholbarkeit aus Abschnitt 6.5, dass die Organisation die Ausführungskompetenz behält.

Das Managementsystem der Zukunft muss deshalb **Lernräume bewusst erhalten**, die es betriebswirtschaftlich nicht mehr bräuchte. Es muss Menschen Fälle bearbeiten lassen, die die Maschine schneller könnte, damit es in fünf Jahren noch Menschen gibt, die beurteilen können, ob die Maschine sie richtig bearbeitet. Kompetenzmanagement und der Erhalt des Wissens der Organisation (ISO 9001, Kap. 7.2 und 7.1.6) bekommen damit eine Brisanz, die diesen unscheinbaren Normabschnitten bisher niemand zugetraut hätte.

Und die Lernschicht ist zweiseitig. Sie regelt nicht nur, wo Menschen weiter lernen müssen, sondern auch, **wo die Maschine lernen darf**: wo ein System aus Betriebsdaten nachjustiert, feinabgestimmt oder nachtrainiert werden darf und wo ausdrücklich nicht. Eine Organisation, die diese Frage nicht beantwortet hat, hat kein Änderungsmanagement, sondern nur ein Änderungsprotokoll.

In der hybriden Organisation führen Menschen nicht mehr alles selbst aus. Sie sind aber der Ort, an dem Verantwortung, Urteilskraft und Vertrauen zusammenlaufen. Ein Managementsystem, das das nicht aktiv schützt, verliert nicht nur Kompetenz. Es verliert seine Verantwortungsfähigkeit: genau das, wofür es da ist.

TEIL III

Die Gegenposition

Vorbemerkung: Die vier Einwände, die in jedem Raum fallen

Bevor die empirische Gegenposition zu Wort kommt, gehören vier Einwände auf den Tisch, die in der Praxis regelmäßig fallen und die man nicht mit Empirie, sondern mit einer Unterscheidung beantwortet.

- **Aus dem Qualitätsmanagement:** „Ob ein Mensch oder eine KI den Prüfschritt macht, ist egal. Ich passe den Prozess an, und gut."
- **Aus der Rechtsabteilung:** „Hören Sie auf, von KI als Akteur zu sprechen. Verantwortung ist nicht delegierbar. Regeln Sie Haftung, nicht Philosophie."
- **Aus der IT-Leitung:** „Das ist ein Governance-Thema, kein neues Managementsystem. Modellfreigaben, Logging, Change-Management, das reicht."
- **Aus der Geschäftsführung:** „Wenn wir dafür eine neue Bürokratieschicht aufbauen, verpassen wir den Nutzen."

Sortiert nach Denkschule sind das drei Richtungen. Aus **traditioneller QM-Sicht** ist es alter Wein in neuen Schläuchen; Komplexität war immer da, jetzt heißt sie eben KI. Aus **rechtlicher Sicht** bleibt Verantwortung ohnehin bei Menschen und Organisationen; es gibt keine Pseudo-Akteure und deshalb auch keinen Bedarf an einer neuen Kategorie. Aus **operativer Sicht** bremst zu viel Governance.

Jeder dieser Einwände hat einen richtigen Kern, und keiner trifft. Das Qualitätsmanagement hat recht, dass der Prozess der Bezugspunkt bleibt; die gemeinte Anpassung ist aber genau die Allokationsentscheidung aus Abschnitt 6, und die musste bisher niemand treffen. Die Rechtsabteilung hat vollständig recht: Verantwortung ist nicht delegierbar. Genau deshalb entsteht die Lücke, um die es geht. Die IT-Leitung beschreibt Bausteine, nicht die Architektur; Guardrails regeln die Maschine, nicht die Organisation (7.4). Und die Geschäftsführung hat recht, dass eine zusätzliche Schicht das falsche Ergebnis wäre; die Antwort darauf ist keine zweite Struktur, sondern eine erweiterte.

Die tragende Antwort auf alle vier lautet nicht „KI ist neu", sondern: **Die Eingriffstiefe ist neu.** Die Kombination aus Geschwindigkeit, Skalierung und teilweiser Delegation an maschinelle Akteure verändert den Charakter des Systems. Nicht revolutionär anders, aber mehr als ein Update. Positiv gewendet und als Satz, der die Position dieses Papiers am kürzesten fasst:

> // **Wir erfinden Managementsysteme nicht neu. Wir erweitern bestehende Systeme gezielt um die explizite Steuerung von KI-Autonomie.**

Damit ist der Alltagseinwand beantwortet. Der ernstere Einwand kommt aus der Empirie.

10 Die Nullhypothese: Es ändert sich nichts Wesentliches

Ein Papier, das eine Verschiebung behauptet, muss die Gegenthese in ihrer stärksten Form aushalten. Sie lautet:

> **// Managementsystem-Zertifizierung ist kein Wirkungsinstrument, sondern ein Marktzugangs- und Legitimationsinstrument. Ihre Nachweisform ist nicht deshalb dokumentenzentriert, weil sie technisch nicht besser könnte, sondern weil eine höhere Nachweisauflösung von keiner der beteiligten Parteien gewollt und von keiner bezahlt wird. Agentische KI verändert daran nichts, weil sie die Anreizstruktur nicht anfasst.**

Diese Position ist empirisch besser belegt, als es der Fachöffentlichkeit lieb sein kann. Ich gebe sie hier nicht als Alibi wieder, sondern weil sie die eigene These an drei Stellen korrigiert.

10.1 Die geringe Auflösung ist ein Merkmal, kein Mangel

Der institutionentheoretische Ausgangspunkt ist alt und nicht widerlegt: Formale Strukturen in Organisationen produzieren häufig **Legitimität statt Effizienz**. Organisationen übernehmen institutionalisierte Regeln als „Mythos und Zeremonie", entkoppeln sie von der Alltagsarbeit und stabilisieren die Konstruktion, indem sie Inspektion und Evaluation ausdrücklich **minimieren**; genaues Hinsehen würde die Entkopplung sichtbar machen (Meyer & Rowan 1977).

Die empirische Bestätigung für ISO-Normen ist konsistenter, als der Branche recht sein kann:

- Ein elfjähriges US-Panel über 19.713 Werke zeigt: Zertifizierte Werke wuchsen schneller, **operative Verbesserungen erklärten dieses Wachstum jedoch nicht**; der Effekt war größer, wo die Informationsasymmetrie in der Lieferbeziehung hoch war. Das Zertifikat wirkt als **Signal** (Terlaak & King 2006).
- Eine gematchte Untersuchung an rund 1.000 kalifornischen Firmen fand positive Effekte auf Überlebensrate, Umsatz, Beschäftigung und Löhne, bei Arbeitssicherheit aber ein aufschlussreiches Muster: Verletzungsraten sanken leicht, Verletzungskosten nicht (Levine & Toffel 2010). Das ist das Profil eines Signals mit Randnutzen.
- Die weltweite Diffusion von ISO 9000 erklärt sich über Direktinvestitionen, Handelsnetze und Nachahmung, nicht über nachgewiesene Wirkung (Guler et al. 2002).
- Die Alltagspraxis ist vom System oft entkoppelt, beobachtet als „ceremonial behaviour" an neun zertifizierten kanadischen Organisationen (Boiral 2007); und zertifizierte Unternehmen in China wählen ihr Umsetzungsniveau **strategisch**, abhängig von Kundenpräferenzen, Kundenüberwachung und erwarteten Sanktionen (Christmann & Taylor 2006).

Gegenevidenz, der Fairness halber: Die Literatur ist nicht einheitlich. Eine Meta-Analyse zu ISO 14001/EMAS und Finanzperformance über 60 Studien findet positive Zusammenhänge, schreibt sie aber maßgeblich einem **Selektionseffekt** zu: Bessere Performer zertifizieren sich. Ein systematisches Review kommt zu dem Ergebnis, die Wirkung sei **kontingent auf das Adoptionsmotiv**. Das stützt eher die Nullhypothese: Wo aus Marktzwang zertifiziert wird, wirkt es nicht.

Damit steht diese Lesart quer zu Abschnitt 8.3. Dort war die kleine Stichprobe methodisch legitim, weil ein Strukturmodell sie kompensiert. Hier ist sie institutionell funktional, weil eine höhere Auflösung die Entkopplung sichtbar machen würde. Beides lässt sich vereinbaren: Das Strukturmodell ist die ehrliche Begründung, die institutionelle Bequemlichkeit ist der Grund, warum niemand nachfragt, ob es im Einzelfall trägt. Bei Agenten fällt das Strukturmodell weg. Dann bleibt von der Legitimation nur die Bequemlichkeit übrig.

Die stärkste ehrliche Formulierung ist deshalb nicht „Zertifizierung wirkt nicht“, sondern: **Zertifizierung wirkt dort, wo jemand sie wirken lassen will, und das ist eine Frage des Motivs, nicht der Nachweistechnik.**

10.2 Die Zahlungsrichtung bestimmt die Prüfschärfe

Der härteste Beleg kommt nicht aus der Managementliteratur, sondern aus einem randomisierten Feldexperiment mit Umweltauditorinnen und -auditoren in Gujarat. Im Status quo, in dem die geprüfte Anlage die Prüfung auswählt und bezahlt, wurden Emissionswerte systematisch knapp unterhalb der Grenzwerte berichtet. In der Behandlungsgruppe mit Zufallszuweisung, Fixhonorar aus einem Pool und unabhängigen Nachprüfungen stiegen die berichteten Werte deutlich, falsche Konformitätsmeldungen gingen erheblich zurück, **und die Anlagen senkten ihre Emissionen tatsächlich** (Duflo, Greenstone, Pande & Ryan 2013).

Dasselbe Muster im kommerziellen Auditwesen: Über rund 17.000 Lieferantenaudits in 66 Ländern wurden weniger Verstöße gefunden, wenn das Prüfteam die Fabrik bereits kannte, und wenn **die auditierte Seite statt der einkaufenden zahlte** (Short, Toffel & Hugill 2016).

Warum das die These dieses Papiers trifft: Kontinuierliche Nachweisführung erhöht die *Menge* des Nachweises. Sie ändert nicht, **wer die Prüfung bezahlt**. Wenn Milde ein Preis- und Kundenbindungsphänomen ist, führt mehr Datenmaterial nicht zu mehr Befunden, sondern zu mehr **selektiv gelesenem** Datenmaterial. Ein Dashboard mit 400 Kontrollpunkten, aus dem die bezahlte Prüfung drei bestätigende Ansichten aufruft, ist ökonomisch dasselbe System wie ein Ordner mit 400 Dokumenten, nur teurer.

10.3 Das Volumenargument, gegen die eigene These gerichtet

Abschnitt 8.3 argumentiert: Kleine Stichproben belegen nichts. Das Argument gilt **symmetrisch** und trifft damit die Lösung ebenso wie das Problem: Wenn kleine Stichproben nichts belegen, dann belegt auch kontinuierliches Monitoring bei kleinen Grundgesamtheiten nichts.

Für realistische Fallzahlen einer Organisation mit 30 bis 100 Beschäftigten und zwei bis fünf KI-Anwendungsfällen:

Anwendungsfall (Annahme)	Fälle/Monat	Obergrenze der Fehlerrate bei 0 Fehlern, monatlich
KI-Angebotstexte, 3/Woche	12	22,1 %
Vorsichtung Bewerbungen, 60/Jahr	5	45,1 %
Belegextraktion, 200/Monat	200	1,5 %
Kundenchat, 30/Woche	120	2,5 %

Und die Fallzahl, die man je Vergleichsperiode bräuchte, um eine **Verschlechterung** überhaupt zu erkennen (Zwei-Proportionen-Test, α = 5 %, Power 80 %):

Verschlechterung	Benötigte Fälle je Periode
1 % → 2 %	2.319
1 % → 3 %	769
5 % → 10 %	435
0,1 % → 0,2 %	23.511

Drei Schlussfolgerungen, die man nicht wegdiskutieren kann:

1. Für die typischen KI-Anwendungsfälle kleiner und mittlerer Organisationen ist eine monatliche Regelkarte **statistisch wertlos**. Bei 5 bis 12 Fällen pro Monat liegt die Obergrenze der Fehlerrate zwischen 22 und 45 Prozent.
2. Wo das Volumen ausreicht, **fehlt oft der Referenzwert**. Monitoring kann Anomalien zählen, Korrektheit aber nur gegen eine geprüfte Wahrheit feststellen. Der Aufwand, in *beiden* Vergleichsperioden ausreichend Fälle zu labeln, übersteigt schnell die Kosten des gesamten Zertifizierungsaudits.
3. Damit ist statistische Nachweisführung nicht falsch, sondern **an ein Segment gebunden**: an Hochvolumen-Anwendungen und an Anbieter mit eigener Modellentwicklung. Sie ist nicht das Programm für die Mehrheit der Organisationen.

10.4 Der Realitätstest: Selbst harte Pflichten werden nicht erfüllt

Das deutsche NIS2-Umsetzungsgesetz trat am 6. Dezember 2025 **ohne Übergangsfrist** in Kraft und betrifft rund 29.500 Einrichtungen statt zuvor etwa 4.500. Die Registrierungsfrist endete im März 2026, mit Nachfrist bis Ende Juli 2026. Zum Stand 30. Juni 2026 lagen rund 17.900 Registrierungen vor.

Rund 40 Prozent der gesetzlich verpflichteten Einrichtungen hatten sieben Monate nach Inkrafttreten nicht einmal die Registrierung vorgenommen, also den billigsten und einfachsten Schritt der gesamten Pflichtenkette, bei Bußgeldandrohung und persönlicher Geschäftsleitungshaftung.

Das ist der härteste empirische Datenpunkt der ganzen Debatte: **Selbst erzwungener Nachweis wird nicht erbracht, wenn die Organisation keine Steuerungsnotwendigkeit sieht.** Der Engpass ist nicht die Auflösung des Nachweises. Der Engpass ist Aufmerksamkeit, Personal und Zahlungsbereitschaft für Compliance überhaupt.

Hinzu kommt: Der regulatorische Rückenwind weht nicht nur in eine Richtung. Innerhalb von etwa neun Monaten hat der Gesetzgeber, europäisch wie national, die Nachweispflicht für Cybersicherheit versiebenfacht, den Anwendungsbereich der Nachhaltigkeitsberichterstattung massiv gekappt und die KI-Hochrisikopflichten um 16 Monate verschoben. Wer aus dem AI Act eine stabile Nachfrage extrapoliert, extrapoliert aus derselben Gattung Rechtsakt, die gerade zurückgebaut wird.

10.5 Ein deutscher Praxisvorbehalt, der in keiner Tech-Argumentation vorkommt

Automatisierte Verhaltenstelemetrie ist in Deutschland nicht frei einführbar. Technische Einrichtungen, die geeignet sind, Verhalten oder Leistung von Beschäftigten zu überwachen, sind **mitbestimmungspflichtig** (§ 87 Abs. 1 Nr. 6 BetrVG). Das betrifft jedes Monitoring, das an Prozessen ansetzt, in denen Menschen und Agenten zusammenarbeiten, und das sind nach Abschnitt 5 die meisten. Wer eine Nachweisarchitektur plant, plant in Deutschland ein mitbestimmungspflichtiges Vorhaben.

Ein zweiter Vorbehalt aus derselben Richtung ist nicht belegt, aber ernst zu nehmen: Permanente Messung könnte die Datenqualität von innen aushöhlen, weil Beschäftigte, die sich lückenlos gemessen sehen, Fehler seltener melden. Das liefe der ältesten Forderung des Qualitätswesens direkt zuwider, dem *drive out fear*. Belastbare Daten dazu für KI-Telemetrie kenne ich nicht; die Vermutung gehört in die Betriebsvereinbarung, nicht in die Argumentation.

10.6 Was von der Nullhypothese bleibt, und was nicht

Ich halte drei ihrer Punkte für belastbar und nehme sie in die eigene Position auf:

1. **Die Volumengrenze ist real.** Statistische Nachweisführung ist kein Universalprogramm. Sie gehört an Hochvolumen-Prozesse, und wo sie nicht trägt, gehört die Autonomiestufe gesenkt, statt das Monitoring zu simulieren. Das ist die Volumenregel aus 6.3, und sie ist die wichtigste Korrektur, die dieses Papier der eigenen These verdankt.
2. **Die Rollenfrage entscheidet die Anforderungshöhe.** Die schweren Nachweispflichten sitzen beim Anbieter. Für die Mehrheit der Organisationen ist die Aufgabe kleiner, aber konkreter, und sie wird derzeit von der Debatte überdeckt (5.1).

3. **Der Engpass ist Kapazität und Wille, nicht Technik.** Eine Nachweisarchitektur, die niemand liest, betreibt oder bezahlt, ist kein Fortschritt gegenüber einem Ordner, den niemand liest.

Drei ihrer Punkte halte ich für nicht tragfähig:

1. **„Nicht messbar" heißt nicht „nicht wirksam".** Dass ein System ohne belastbare Instrumentierung keine messbare Wirkung zeigt, ist eine Tautologie, kein Beweis der Wirkungslosigkeit. Der Gujarat-Befund schneidet in beide Richtungen: Als die Anreize korrigiert wurden, **sank die Belastung tatsächlich**. Prüfung *kann* wirken. Der Satz „die Anreizstruktur ändert sich nicht" ist eine Annahme, keine Naturkonstante.
2. **Der Kanal, der die Anreizstruktur zuerst ändern könnte, ist nicht der Gesetzgeber.** Anders als auditierte Organisationen, Zertifizierungsstellen und die einkaufende Seite haben Cyber- und Haftungsversicherer ein **direktes finanzielles Interesse an höherer Nachweisauflösung**, weil sie Schäden bezahlen. Denselben Effekt kann ein einzelnes großes Unternehmen auf der Einkaufsseite erzeugen, das KI-Klauseln durch die Lieferkette reicht; und es sind gerade die von der Einkaufsseite getriebenen Prüfungen, die mehr Befunde erzeugen. Beide Kanäle brauchen weder Gesetzgeber noch Akkreditierungsregeln und können schneller sein als jede Regulierung.
3. **Die Nullhypothese beantwortet die Nachweisfrage, nicht die Steuerungsfrage.** Selbst wenn kein einziges Zertifikat je wirksam war, bleibt die Lücke bestehen, die dieses Papier beschreibt: Eine Organisation verantwortet, was sie nicht mehr selbst ausführt. Das ist ein Führungs- und Haftungsproblem, kein Zertifizierungsproblem. Es verschwindet nicht, wenn der Zertifizierungsmarkt stagniert.

Die Synthese aus beiden Positionen: Die Nullhypothese ist auf der Ebene des *Zertifikatsmarkts* weitgehend richtig und auf der Ebene der *Verantwortungsfähigkeit* weitgehend irrelevant. Wer sie als Ruhekissen liest, liest sie falsch. Sie schützt vor einer Fehlinvestition in ein Nachweisprogramm ohne Nachfrage, nicht vor dem Kontrollverlust, den sie gar nicht adressiert.

TEIL IV

Die Antwort

11 Synthese: Von der Prozessbeherrschung zur Vertrauensarchitektur

Am klarsten wird die Verschiebung, wenn man die drei alten Funktionen der Dokumentation, also steuern, nachweisen und beruhigen, einzeln weiterverfolgt und die neue vierte daneben stellt:

Funktion	Gestern (menschliche Ausführung)	Morgen (hybride Ausführung)
Allokieren	Implizit: Menschen führen aus, die Zuordnung ist Personalplanung	Explizit: kriterienbasierte, begründete, revidierbare Managemententscheidung je Prozessschritt; Autonomie als Stufe; jede Grenzverschiebung ein Change mit Validierung
Steuern	Beschreibung, adressiert an Verstehen: Verfahrens- und Arbeitsanweisungen, Schulung, Unterweisung	Zwei Register: Orientierung und Befähigung für Menschen; gelenkte Konfiguration (Prompts, Workflows, Leitplanken, Testfälle) plus Evaluationen und Monitoring für Maschinen
Nachweisen	Dokument plus Stichprobe: „Zeig mir die Anweisung, zeig mir, dass danach gearbeitet wird“	Trace plus laufende Evidenz: Validierung, Evaluationsergebnisse, Protokolle, Eingriffsnachweise, heterogen erhoben und integriert in **eine** Konformitätsaussage; wo das Volumen fehlt: qualitative Fallnachverfolgung statt simulierter Statistik
Beruhigen	Gefühlte Kontrolle durch Papier	Begründetes Vertrauen durch Kontrollarchitektur: verstehbare Systemgrenzen, geübte Eingriffe, erhaltene Urteilskraft

11.1 Die Vertrauensarchitektur in acht Bausteinen

„Vertrauensarchitektur“ wirkt als Wort zunächst groß. Konkret gemeint ist ein Set aus acht Bausteinen, das technische Sicherungen mit Verantwortungs- und Nachweislogik verbindet:

1. **Zweckklarheit:** Wofür setzen wir KI überhaupt ein?
2. **Risikoklassifizierung:** Nicht jede Anwendung braucht dieselbe Schärfe.
3. **Verantwortungsmodell:** Wer zeichnet, wer greift ein? Und sind wir hier Anbieter oder Betreiber?

4. **Eingriffspunkte:** Stoppschilder bei Unsicherheit, gekoppelt an die Autonomiestufe.
5. **Daten- und Modellbeherrschung:** Änderungen, Versionen, Datenqualität; Modellwechsel als lenkungspflichtige Änderung; und die ausdrückliche Festlegung, wo ein System aus Betriebsdaten lernen darf und wo nicht.
6. **Transparenz und Nachvollziehbarkeit:** Entscheidungen bleiben erklärbar, zumindest auf Prozess- oder Ergebnisebene.
7. **Laufendes Monitoring mit Reaktionsprozess:** Drift, Nebenwirkungen, Fehlanreize und ein definierter, geübter Weg für den Fall, dass eine Grenze verletzt wird.
8. **Externe Signale:** Nachweise gegenüber Kundschaft, Auditierenden und Behörden.

Guardrails sind ein Teil davon, nämlich Baustein 4 und 7. Ohne die übergeordnete Struktur fehlt der Rahmen, der Vertrauen über Einzelmaßnahmen hinaus stiftet. Und Baustein 7 gilt nur zusammen mit seinem Nachsatz: Monitoring ohne Reaktionsprozess ist ein Haftungsverstärker (8.4).

11.2 Wie das konkrete Managementsystem aussieht

Die **Prozesslandkarte** bleibt das Rückgrat, bekommt aber ein neues Attribut je Prozessschritt: ausgeführt durch Mensch, Maschine oder beide, samt Autonomiestufe und Begründung der Zuordnung.

Die **menschengerichtete Dokumentation** schrumpft auf das Orientierende und wird dadurch besser statt schlechter.

Die **maschinengerichteten Artefakte**, also Prompts, Workflow-Definitionen, Leitplanken und Testfälle, sind gelenkte dokumentierte Information mit Versionsständen und Freigaben.

Evaluation und Monitoring agentischen Verhaltens sind ein eigener Kernprozess des Systems, so selbstverständlich wie heute das interne Audit: dimensioniert nach Volumen, mit Reaktionsprozess, und dort durch qualitative Fallnachverfolgung ersetzt, wo die Fallzahlen keine Statistik tragen.

Die **Managementbewertung** schaut nicht mehr nur auf Kennzahlen menschlicher Prozesse, sondern auf Verteilungen maschinellen Verhaltens, auf Vorfälle an den Schnittstellen und auf den Zustand der Rückholbarkeit. Ergänzt wird sie um ereignisgetriggerte Zwischenbewertungen.

Und das **Auditprogramm**, intern wie extern, beherrscht beide Evidenzsorten.

Kurz: **Das Managementsystem der hybriden Organisation ist weniger Dokumentensammlung und mehr Verantwortungsarchitektur.** Es beantwortet jederzeit vier Fragen: Wer oder was führt aus, und warum so? Wie kommt unsere Absicht dort hinein? Woran sehen wir, dass es funktioniert? Und wer versteht genug, um einzugreifen?

11.3 Die Antwort auf die Ausgangsfrage

Brauchen wir Managementsysteme noch, wenn Maschinen die Prozesse ausführen?

Ja, dringender als bisher, aber anders als bisher. Nicht weil der Markt Zertifikate verlangt (das tut er weiter), sondern weil die Lücke zwischen dem, was eine Organisation verantwortet, und dem, was sie selbst ausführt, größer wird. Und das Managementsystem ist die einzige Institution, die wir zur Schließung genau dieser Lücke je entwickelt haben.

Welche Art von Managementsystem brauchen wir dann?

Eines, das fünf Dinge kann, die das heutige überwiegend nicht kann:

1. **Ein Verhaltensfenster statt eines reinen Verfahrensverzeichnisses:** explizite, prüfbare Nebenbedingungen, formuliert auch als Ausschlusskriterien, nicht nur als positive Prozessschritte.
2. **Eine Autonomiestufenarchitektur:** Kontrolltiefe als Funktion der Autonomie, mit der Volumenregel als Freigabekriterium.
3. **Verteilungsdenken statt Endprüfung:** Demings Punkt 3, angewandt auf agentisches Verhalten, mit ehrlich dimensionierten Kontrollperioden und qualitativer Fallnachverfolgung dort, wo die Statistik nicht trägt.
4. **Eine ereignisgetriebene statt kalendergetriebene Lernschleife:** Modellwechsel und Grenzverletzung als Auslöser, nicht der Jahresrhythmus.
5. **Erhaltene menschliche Urteilskraft:** Lernräume und Rückholbarkeit als bewusste, bezahlte Resilienzanforderung.

Und es entsteht nicht auf der grünen Wiese: **Wir erfinden Managementsysteme nicht neu. Wir erweitern bestehende Systeme gezielt um die explizite Steuerung von KI-Autonomie.**

Die kürzeste Fassung: **Wir brauchen nicht weniger Managementsystem, sondern ein Managementsystem, das messen kann, was es behauptet, und das ehrlich sagt, wo es nicht messen kann.**

Die Kernfrage lautet damit nicht mehr „Brauchen wir ein Managementsystem?“, sondern: **„Ist die delegierte Autonomie beherrscht, und ist sie vertrauensfähig?“**

Kundinnen und Kunden kaufen nicht das Managementsystem. Sie kaufen das Vertrauen in eine Leistung. Managementsysteme sind ab jetzt die interne Architektur dieses Vertrauens.

12 Was offen bleibt

Ein Papier, das mit einer offenen Frage begann, sollte ehrlich benennen, was offen bleibt.

Die Prüfmittelfrage für Evaluationen. Wenn Evaluationen die neue Unterweisung und das neue Prüfmittel sind: Wer kalibriert sie? Ein Eval, das die falschen Fälle testet, erzeugt dieselbe Scheinsicherheit wie ein ungeeignetes Messmittel. Die QM-Denkfigur der

Prüfmittelüberwachung müsste auf Evaluationssuiten übertragen werden; wie das methodisch belastbar geht, ist ungeklärt.

Die Lücke zwischen Volumenregel und Realität. Die Volumenregel (6.3) sagt, wann statistische Überwachung trägt. Sie sagt nicht, was eine Organisation tut, die einen hochkritischen Prozess mit niedrigem Volumen automatisieren *will* und dafür gute betriebswirtschaftliche Gründe hat. „Autonomiestufe senken" ist die richtige, aber unbequeme Antwort. Ob sie sich gegen Effizienzdruck durchhält, ist offen.

Die Auditfrage. Trace-basierte Evidenz kann nur prüfen, wer sie lesen kann. Die Qualifizierung von Auditorinnen und Auditoren, intern wie bei Zertifizierungsgesellschaften, hinkt der Technologie um Jahre hinterher. Bis dahin droht genau das Error-Digging an Formalien, das den Wert der Zertifikate weiter untergräbt.

Die Maßfrage beim Verstehen. Wie viel Verstehen ist genug, um zu verantworten? Irgendwo zwischen „Modellinterna erklären können" (unerreichbar) und „es läuft halt" (unverantwortbar) liegt eine Schwelle, die weder Normen noch Rechtsprechung bisher definiert haben.

Die Ökonomie des Kompetenzerhalts. Lernräume, die betriebswirtschaftlich nicht nötig wären, kosten Geld, und zwar in einem Wettbewerb, in dem andere diese Kosten vielleicht nicht tragen. Ob Kompetenzerhalt als Resilienzanforderung normativ verankert werden muss, damit er nicht wegoptimiert wird, halte ich für eine der spannendsten Regulierungsfragen der nächsten Jahre.

Die Nahtstellenhaftung. Wenn an der Übergabe zwischen Maschine und Mensch etwas schiefgeht, wenn also der Agent korrekt eskaliert und der Mensch durchwinkt oder umgekehrt: Wessen Versagen ist das? Die Verantwortungsarchitektur muss hier präziser werden, als es unsere heutigen Rollenbeschreibungen sind.

Der Versicherungs- und Lieferkettenkanal. Wenn die Anreizstruktur sich ändert, dann vermutlich hier zuerst, und zwar schneller als durch Regulierung. Belastbare Daten dazu, ob und wie Versicherer KI-Governance-Nachweise bereits in Prämien oder Deckungsvoraussetzungen abbilden, fehlen. Das ist die größte offene Flanke in beide Richtungen: für die These wie für die Gegenthese.

Die Mitbestimmung. Eine Nachweisarchitektur, die an Mensch-Maschine-Prozessen ansetzt, ist in Deutschland mitbestimmungspflichtig. Wie eine Betriebsvereinbarung aussieht, die Verhaltenstelemetrie für Agenten erlaubt, ohne Beschäftigtenüberwachung zu normalisieren, ist meines Wissens nirgends beschrieben.

Die Diskussion, die dieses Papier ausgelöst hat, wird also weitergehen. Genau das soll sie.

Bezüge

Diskussion. LinkedIn-Beitrag „Brauchen wir Managementsysteme noch, wenn Maschinen die Prozesse ausführen?“ (August 2026) mit rund zwei Dutzend fachlichen Kommentaren. Positionen aus dieser Diskussion sind in diesem Papier bewusst **ohne Namensnennung** wiedergegeben; es geht um Argumente, nicht um Zuschreibungen. Die Unterscheidung „Verrichten in festgelegten Konstellationen / Handeln in volatilen Situationen“ stammt von Hartmut Rosa und wurde aus der Diskussion in den Zusammenhang eingebracht.

Normen und Regulierung. ISO 9000:2015, 3.5.3 und Annex SL der ISO/IEC Directives (Definition des Managementsystems); ISO 9001:2015, insbes. Kap. 4.1, 4.4.2, 5.1.1, 6.1, 6.3, 7.1.5, 7.1.6, 7.2, 7.5, 8.5.1, 9.3, 10.2; ISO/IEC 17021-1 und IAF MD 5 (Auditregeln, Audittage); ISO/IEC 42001:2023 (AI-Managementsystem, insbes. 4.4 und A.6.2.8), ISO/IEC 42005 (AI System Impact Assessment), ISO/IEC 42006 (Anforderungen an Zertifizierungsstellen, Auditzeitberechnung), ISO/IEC 5338 (KI-Lebenszyklusprozesse); EU AI Act, insbes. Art. 4, 5, 6, 11, 12, 14, 16(c), 17, 26, 40, 50, 111 sowie die durch den Digital Omnibus geänderte Fristenlage; EN 18286 (QMS für Zwecke des Art. 17 AI Act); NIS2-Umsetzungsgesetz / § 30, § 33 BSIG; § 87 Abs. 1 Nr. 6 BetrVG; § 93 AktG, § 43 GmbHG.

Empirie zur Wirkung von Managementsystemzertifizierung. Meyer & Rowan (1977) zu Legitimität und Entkopplung; Terlaak & King (2006) zum Signalcharakter der ISO-9000-Zertifizierung; Levine & Toffel (2010) zu Überlebens-, Umsatz- und Beschäftigungseffekten; Guler, Guillén & Macpherson (2002) zur Diffusion; Boiral (2007) zu „ceremonial behaviour“; Christmann & Taylor (2006) zur strategischen Wahl des Umsetzungsniveaus; Duflo, Greenstone, Pande & Ryan (2013) zum Feldexperiment in der Umweltauditierung; Short, Toffel & Hugill (2016) zu Lieferantenaudits; Pierce & Toffel (2013) zu Prüfmilde bei Cross-Selling-Anreizen.

Zuverlässigkeit agentischer Systeme. Arbeiten zur Reliabilität von KI-Agenten (Streuung auch bei Temperatur 0; Entkopplung von Fähigkeit und Zuverlässigkeit; höhere Konsistenz kleinerer Modelle). Eigene Beobachtungen zur Regelbefolgung von KI-Agenten, ohne systematische Messung; zur Einordnung: Jaroslawicz, Whiting, Shah & Maamari (2025), How Many Instructions Can LLMs Follow at Once?, arXiv:2507.11538; Liu, Lin, Hewitt, Paranjape, Bevilacqua, Petroni & Liang (2024), Lost in the Middle: How Language Models Use Long Contexts, TACL 12, S. 157 bis 173; von Arx, Chan & Barnes (2025), Recent Frontier Models Are Reward Hacking, METR; Zhong, Raghunathan & Carlini (2025), ImpossibleBench: Measuring LLMs' Propensity of Exploiting Test Cases, arXiv:2510.20270.

Eigene Vorarbeiten. Grundlagenpapier „Verantwortung ohne Ausführung“ (26.08.2026); Ausarbeitung „Managementsysteme im KI-Zeitalter: von der Prozessbeherrschung zur Vertrauensarchitektur“ (28.08.2026); Multi-Modell-Beiratsberichte zu „Managementsystem: Steuerung oder Nachweis“ und zur Rolle von Managementsystemen im Zeitalter agentischer KI, einschließlich eines Red-Team-Berichts zur Verteidigung der Nullhypothese (August 2026); Analysen zur Integration von ISO 42001 in bestehende 9001-Landschaften sowie zum Verhältnis ISO 42001 zu EN 18286.

Belegstand und Vorbehalte

Dieses Papier ist ein Arbeitsstand und kein Gutachten. Fünf Einschränkungen gehören ausdrücklich dazu:

1. **Regulatorische Daten und Normstände** (AI-Act-Fristen nach dem Digital Omnibus, Status und Veröffentlichungsstand von EN 18286, Zitierung im EU-Amtsblatt und damit die Konformitätsvermutung nach Art. 40, Stand der ISO-Arbeitselemente) sind hier aus Sekundärquellen und Recherchеberichten übernommen. Vor jeder Verwendung in Kunden-, Schulungs- oder Prüfmaterial sind sie gegen Primärquellen nachzuziehen, also gegen Amtsblatt, ISO-Arbeitsprogramm und CEN-Katalog.
2. **Alle statistischen Tabellen** (Abschnitte 6.3, 7.3, 8.3, 10.3) sind eigene Berechnungen auf Basis der einseitigen Vertrauensgrenze bei null beobachteten Fehlern, der Dreierregel, der Binomialmultiplikation und des $n \cdot p \geq 5$-Kriteriums für Anteilskarten. Sie sind nachrechenbar und enthalten keine Quellenzahlen.
3. **Das Autonomiestufenmodell A0 bis A4, die Volumenregel und die Fünf-Funktionen-Zerlegung** sind analytische Konstruktionen dieses Papiers, keine Normeninhalte. Sie sind als Denkwerkzeuge gemeint, nicht als Anforderungen.
4. **Die Marktzahl zur Verbreitung von ISO/IEC 42001** (Abschnitt 8.5) stammt aus Branchenkompilationen ohne offizielles Register; die Norm wird im ISO Survey nicht erfasst. Belastbar ist die Größenordnung, nicht der Wert.
5. **Die beiden Beobachtungen zur Regelbefolgung von KI-Agenten** (Abschnitt 7.3) stammen aus eigener Praxis ohne systematische Messung. Zur Einordnung dienen die unter Bezüge genannten Studien; bis auf Liu u. a. sind das Preprints oder Laborberichte ohne Begutachtung.

Korrigiert am 30. September 2026: Abschnitt 3 nennt die Definition in ISO 9000:2015, 3.5.3 jetzt sinngleich statt wortgleich mit Annex SL und beschreibt die Rolle der Dokumentation normgenau (ISO 9001, 4.4.2 und 7.5.1; ISO/IEC 42001, 4.4 und A.6.2.8); die Aussage zu einer Annex-SL-Revision 2021 entfällt, weil sie nicht belegbar war, die Ergebnisausrichtung stützt sich jetzt auf ISO 9001, 4.1, 5.1.1 und 6.1. Abschnitt 7.3 bezeichnet die eigenen Befunde nicht mehr als Benchmark-Untersuchung, sondern als Beobachtungen ohne systematische Messung, und ordnet sie in veröffentlichte Forschung ein. Abschnitt 8.3 kennzeichnet den Stichprobenumfang als Erfahrungswert.

Dieses Papier ist ein Arbeitsstand, kein Schlusswort. Widerspruch ist ausdrücklich erwünscht, am liebsten öffentlich.