

# „Governance vor Geschwindigkeit“ Strukturelle Verantwortung im Zeitalter agentischer KI

Ein Beitrag zur Frage, wie Ethik in handlungsfähige Entscheidungsarchitekturen übersetzt werden kann.

---



## **Governance vor Geschwindigkeit – Teil 2**

*Wenn Inferenz zur Verpflichtung wird*

---

Andrzej Skulski

Perspektiven – Charta der Menschlichkeit im Zeitalter der KI. Version 1.1 · März 2026

Dieser Beitrag ist ein unabhängiger Perspektivtext und spiegelt die Position des Autors wider.

© 2026 Andrzej Skulski

Lizenz: Creative Commons CC BY-NC 4.0

(Nutzung mit Namensnennung, nicht-kommerziell)

---

# Einleitung

Dieser Text baut auf dem Perspektivbeitrag:  
„**Governance braucht Haltung**“ (Februar 2026) auf.

Der erste Text argumentierte, dass KI-Ethik nur dann wirksam wird, wenn sie in **institutionelle Entscheidungsarchitektur** übersetzt wird.

Ethik kann nicht im Modell verankert werden.

Sie entsteht in den Strukturen, die definieren, **wann ein System handeln darf**.

Der vorliegende Beitrag verschiebt den Fokus auf den kritischsten Moment dieser Architektur:

den Zeitpunkt,  
an dem **Inferenz zur Handlung wird**.

Mit dem Übergang zu agentischen KI-Systemen verändert sich die operative Realität.  
Systeme generieren nicht mehr nur Antworten.

Sie initiieren Prozesse.

Sie kommunizieren im Namen von Organisationen.

Sie lösen Aktionen aus.

In diesem Moment entsteht eine neue Governance-Frage:

Nicht mehr nur

„Welche Regeln gelten?“

sondern

„Wer trägt Verantwortung **im Moment der Ausführung**?“

Viele aktuelle Governance-Modelle reagieren darauf mit der Formel  
**Human-in-the-Loop**.

Doch diese Formel beschreibt nur eine Prozessposition.

Sie beantwortet nicht die entscheidende Frage:

Wer besitzt **Runtime-Autorität**,  
wenn ein System handlungsfähig wird?

Dieser Beitrag entwickelt die These, dass nachhaltige KI-Governance nicht primär im Design von Policies entsteht, sondern in der klar benannten menschlichen Autorität **an der Grenze zwischen Inferenz und Handlung**.

Dort entscheidet sich, ob technische Beschleunigung zu verantworteter Praxis wird –  
oder zu delegierter Verantwortungslosigkeit.

---

# Kapitel 1

## Von Modellqualität zu Entscheidungsarchitektur

Die öffentliche Debatte über KI konzentriert sich häufig auf die Qualität der Modelle: Trainingsdaten, Parameteranzahl, Benchmarks, Genauigkeit, Bias-Reduktion, Interpretierbarkeit.

Diese Fragen sind wichtig.  
Aber sie greifen zu kurz.

Denn je stärker Systeme in operative Prozesse eingebunden werden – in Verwaltung, Medizin, Recht, Personalwesen oder strategische Planung – desto weniger entscheidet die Modellgüte allein über die reale Wirkung.

Entscheidend wird die Architektur, in der das Modell eingebettet ist.

Ein Sprachmodell kann statistisch präzise arbeiten – und dennoch problematische Entscheidungen erzeugen, wenn:

- seine Outputs automatisch in Workflows eingespeist werden,
- seine Vorschläge nicht kritisch hinterfragt werden,
- seine Empfehlungen institutionell privilegiert sind,
- seine Parameter ohne transparente Revision angepasst werden.

Mit dem Übergang zu agentischen Systemen verschiebt sich der Schwerpunkt fundamental. KI antwortet nicht mehr nur. Sie initiiert Prozesse. Sie trifft Vorentscheidungen. Sie löst Aktionen aus. Sie kommuniziert im Namen von Organisationen.

Damit verschiebt sich die ethische Fragestellung:

Nicht mehr:  
„Wie gut ist das Modell?“

Sondern:  
„Unter welchen Bedingungen darf dieses System handeln?“

Diese Verschiebung ist strukturell.

Modellqualität beschreibt interne Leistungsfähigkeit.  
Entscheidungsarchitektur beschreibt externe Wirkungsmacht.

Und Wirkungsmacht ist das, was Institutionen betrifft.

Wer darf Freigaben erteilen?

Welche Schwellenwerte gelten?

Wann greift eine menschliche Überprüfung?

Welche Normen sind als admissible definiert?

Wie werden diese Normen versioniert?

Wer ist berechtigt, sie zu ändern?

Wie wird dokumentiert, dass sie geändert wurden?

Je komplexer Systeme werden, desto weniger genügt es, auf „Human-in-the-Loop“ zu verweisen.

Denn ein Mensch im Prozess ersetzt keine klare Verantwortungsarchitektur.  
Er kann sie höchstens ausführen.

Wenn Normen unklar sind,  
wenn Zuständigkeiten diffus bleiben,  
wenn Governance nur formell existiert,  
dann verschiebt sich Verantwortung ins Unbestimmte.

Und genau dort beginnt strukturelle Erosion.

Die größte Gefahr moderner KI-Systeme ist nicht die spektakuläre Fehlentscheidung.  
Es ist die schleichende Normalisierung von Delegation ohne klare Rechenschaft.

Deshalb muss Governance mehr sein als Compliance.

Sie muss Entscheidungsarchitektur sichtbar machen.

Denn Ethik wird nicht im Modell verankert.

Sie wird in der Struktur verankert, die entscheidet, wann und wie das Modell wirken darf.

---

## Kapitel 2

### **Admissibility, Versionierung und die Frage nach legitimer Normsetzung**

Sobald KI-Systeme nicht nur assistieren, sondern entscheiden oder vorentscheiden, entsteht eine zentrale Frage:

Wer definiert, was zulässig ist?

In technischen Systemen spricht man von „Admissibility“ – also den Bedingungen, unter denen ein Output als gültig, akzeptabel oder ausführbar gilt.

In einfachen Anwendungen ist das trivial.

In komplexen, agentischen Systemen ist es hochpolitisch.

Denn „admissible“ bedeutet nicht nur technisch möglich.

Es bedeutet normativ erlaubt.

Wenn ein System:

- bestimmte Empfehlungen priorisiert,
- bestimmte Risiken toleriert,
- bestimmte Datenquellen bevorzugt,
- bestimmte Handlungsoptionen ausschließt,

dann spiegelt es eine implizite Normsetzung wider.

Und genau hier beginnt die eigentliche Governance-Frage.

Admissibility ist kein Naturgesetz.

Sie ist ein definierter Rahmen.

Doch wer definiert diesen Rahmen?

Die Entwickler?

Das Unternehmen?

Der Gesetzgeber?

Ein Ethik-Board?

Oder – implizit – Marktmechanismen?

In vielen Organisationen wird diese Ebene unterschätzt.

Man diskutiert über Prompt-Design, Safety-Filter oder Monitoring –

aber nicht über die institutionelle Legitimation der zugrunde liegenden Normen.

Hinzu kommt ein weiterer Aspekt:

Normen sind nicht statisch.

Gesellschaftliche Maßstäbe verschieben sich.

Regulatorische Anforderungen ändern sich.

Unternehmen passen Strategien an.

Risikoabwägungen entwickeln sich.

Wenn ein KI-System operative Macht besitzt, muss daher jede Veränderung der Admissibility-Bedingungen:

- explizit definiert,

- versioniert,
- dokumentiert,
- nachvollziehbar gemacht werden.

Nicht nur aus technischer Sauberkeit.  
Sondern aus demokratischer Notwendigkeit.

Unsichtbare Normverschiebung ist gefährlicher als offener Regelbruch.

Wenn sich Entscheidungsgrenzen verändern, ohne dass dies erkennbar ist,  
wird Governance zur Fassade.

Versionierung schafft Transparenz.  
Transparenz schafft Prüfbarkeit.  
Prüfbarkeit schafft die Möglichkeit zur Korrektur.

Aber auch das reicht nicht aus.

Denn selbst eine sauber dokumentierte Norm setzt voraus,  
dass es eine legitime Instanz gibt, die sie festlegt.

Hier liegt der Kern der Debatte:

Architektur kann Transparenz erzwingen.  
Sie kann Versionierung sichern.  
Sie kann Audit-Trails erzeugen.

Aber sie kann Legitimität nicht herstellen.

Legitimität entsteht außerhalb des Systems.

Deshalb ist die Frage nicht nur:  
„Wie implementieren wir admissible Standards?“

Sondern:  
„Wer trägt die Verantwortung für ihre Definition – und wie wird diese Verantwortung überprüft?“

In einer Zeit, in der Systeme immer schneller lernen,  
muss normative Revision langsamer und bewusster erfolgen.

Nicht jede Optimierung ist eine Verbesserung.  
Nicht jede Effizienz ist ein Fortschritt.

Governance bedeutet deshalb nicht nur Kontrolle.  
Sie bedeutet bewusste Verzögerung.

Ein System, das normative Änderungen friktionsfrei internalisiert, mag effizient sein.

Aber es ist nicht reflexiv.

Und ohne Reflexivität wird technische Dynamik zur institutionellen Trägheit.

Die eigentliche Herausforderung liegt daher nicht im Modell.

Sie liegt in der Fähigkeit von Institutionen,  
Normen bewusst zu setzen,  
sichtbar zu ändern  
und öffentlich zu verantworten.

---

## Kapitel 3

### Human-in-the-Loop oder Human-in-Responsibility?

Der Begriff „Human-in-the-Loop“ wird oft als Sicherheitsversprechen verstanden.

Ein Mensch schaut noch einmal drüber.

Ein Mensch bestätigt.

Ein Mensch korrigiert.

Damit scheint Verantwortung gesichert.

Doch in agentischen Systemen verschiebt sich die Realität.

Wenn ein System:

- Vorschläge vorstrukturiert,
- Optionen selektiert,
- Risiken gewichtet,
- Sprache formt,
- Handlungspfade vorzeichnet,

dann trifft es bereits Vorentscheidungen.

Der Mensch bestätigt häufig nur noch das,  
was das System als plausibel vorbereitet hat.

Das ist keine Bosheit.  
Das ist kognitive Ökonomie.

Je komplexer die Welt wird,  
desto mehr verlassen wir uns auf vorverarbeitete Strukturen.

Das Problem ist daher nicht,  
dass der Mensch „im Loop“ fehlt.

Das Problem ist,  
dass der Mensch zunehmend im **Nachvollzug** ist –  
nicht im Ursprung.

Human-in-the-Loop beschreibt eine Position im Prozess.  
Human-in-Responsibility beschreibt eine Haltung.

Und das ist ein Unterschied.

Verantwortung bedeutet:

- Kriterien aktiv hinterfragen
- Maßstäbe regelmäßig prüfen
- Annahmen explizit machen
- Systemlogiken verstehen wollen
- im Zweifel bewusst bremsen

Ein Klick auf „Bestätigen“ ist keine Verantwortungsübernahme.

In Organisationen entsteht hier ein Spannungsfeld:

Wettbewerb erzeugt Druck.  
Druck erzeugt Beschleunigung.  
Beschleunigung reduziert Reflexionszeit.

Und genau dort wird Human-in-the-Loop zur formalen Absicherung,  
während die normative Entscheidung faktisch bereits im Systemdesign lag.

Agentische Systeme verstärken diesen Effekt.

Je autonomer ein System handelt,  
desto größer wird die Versuchung,  
seine Ergebnisse als „technisch neutral“ zu betrachten.

Aber kein Modell ist neutral.  
Es operiert immer innerhalb gesetzter Rahmenbedingungen.

Die Frage lautet daher nicht:

„Ist ein Mensch beteiligt?“

Sondern:

„Hat dieser Mensch reale Entscheidungsautorität –  
oder nur formale Prüfpflicht?“

Echte Verantwortung braucht:

- Informationszugang
- Transparenz über Systemgrenzen
- Zeit zur Bewertung
- institutionellen Rückhalt für abweichende Entscheidungen

Ohne diese Elemente wird Human-in-the-Loop  
zur psychologischen Entlastungskonstruktion.

Es fühlt sich an wie Kontrolle.  
Ist aber oft nur Mitzeichnung.

Gerade deshalb braucht es Training.

Nicht nur technisches Training.  
Sondern Urteilsfähigkeit.

Die Fähigkeit zu sagen:

„Das ist effizient – aber nicht vertretbar.“

„Das ist plausibel – aber nicht legitim.“

„Das ist möglich – aber nicht verantwortbar.“

Human-in-Responsibility heißt nicht,  
gegen Technologie zu arbeiten.

Es heißt,  
nicht von ihr ersetzt zu werden  
in der normativen Entscheidung.

Wenn wir KI-Systeme gestalten,  
müssen wir daher nicht nur Loops definieren,  
sondern Verantwortungsarchitekturen.

Denn die entscheidende Frage bleibt:

Wer trägt die Konsequenzen?

Und ist dieser Mensch oder dieses Gremium  
wirklich in der Lage,  
die normative Richtung zu beeinflussen –  
oder nur das Tempo?

---

## Kapitel 4

### Governance als Struktur – nicht als Symbol

Governance wird häufig als Regelwerk verstanden.

Policies.  
Richtlinien.  
Prinzipien.  
Board-Strukturen.

All das ist notwendig.

Aber es reicht nicht.

Denn Governance kann zwei sehr unterschiedliche Formen annehmen:

**Symbolische Governance**  
und  
**strukturelle Governance.**

Symbolische Governance erzeugt das Gefühl von Ordnung.  
Strukturelle Governance erzeugt tatsächliche Entscheidungsarchitektur.

Der Unterschied liegt nicht im Dokument.  
Er liegt in der Wirksamkeit.

---

### 1. Symbolische Governance

Symbolische Governance erkennt man daran, dass sie:

- auf Prinzipien verweist,
- Verantwortlichkeiten benennt,

- Transparenz fordert,
- aber operative Machtverhältnisse nicht verändert.

Sie beruhigt.

Sie signalisiert Haltung.  
Sie erzeugt Anschlussfähigkeit.

Doch wenn Entscheidungen im Alltag weiterhin durch Effizienz, Kosten, Marktlogik oder Zeitdruck dominiert werden,  
bleibt Governance dekorativ.

Sie ist dann Teil der Kommunikation –  
nicht Teil der Steuerung.

---

## 2. Strukturelle Governance

Strukturelle Governance greift tiefer.

Sie stellt Fragen wie:

- Wer definiert die Admissibility-Schemata?
- Wer versioniert Normen?
- Wer darf sie ändern?
- Wer dokumentiert Abweichungen?
- Wer kann Prozesse stoppen?

Und vor allem:

Ist das Stoppen realistisch möglich –  
oder nur theoretisch vorgesehen?

Governance wird erst strukturell,  
wenn sie Entscheidungswege tatsächlich verändern kann.

Das bedeutet:

- klare Mandate

- nachvollziehbare Zuständigkeiten
- institutionalisierte Review-Zyklen
- transparente Protokollierung von Normverschiebungen
- echte Interventionsrechte

Ohne diese Elemente bleibt Governance performativ.

---

### **3. Der entscheidende Punkt: Friktion**

Effizienz reduziert Reibung.

Governance erzeugt notwendige Reibung.

In komplexen KI-Systemen ist Reibung kein Fehler – sie ist Sicherheitsmechanismus.

Wenn jede normative Verschiebung:

- explizit formuliert,
- dokumentiert,
- versioniert,
- begründet

werden muss,  
entsteht strukturelle Friktion.

Diese Friktion verlangsamt.

Aber sie macht sichtbar.

Und Sichtbarkeit ist die Voraussetzung für Verantwortlichkeit.

---

### **4. Der blinde Fleck: Kultur**

Selbst perfekte Architektur ersetzt keine Haltung.

Governance lebt von:

- Organisationskultur
- Mut zur Korrektur
- Bereitschaft zum Widerspruch
- Schutz von Dissens

Wenn Governance nur auf dem Papier existiert,  
wird sie vom Alltag absorbiert.

Struktur ohne Kultur bleibt mechanisch.  
Kultur ohne Struktur bleibt machtlos.

Beides gehört zusammen.

---

## **5. Warum das jetzt entscheidend wird**

Agentische Systeme verschieben die operative Ebene.

Sie handeln.  
Sie priorisieren.  
Sie skalieren.

Je stärker sie eingebunden sind,  
desto größer wird die Versuchung,  
normative Fragen als technische Fragen zu behandeln.

Doch Legitimität entsteht nicht im Modell.

Sie entsteht im institutionellen Aushandlungsprozess.

Und dieser Prozess braucht Struktur.

Nicht als Feigenblatt.  
Nicht als Marketing.  
Sondern als tragfähige Entscheidungsarchitektur.

---

Governance ist dann gelungen,

wenn sie nicht nur Effizienz ermöglicht,  
sondern Begrenzung.

Wenn sie nicht nur Innovation trägt,  
sondern auch Stoppsignal sein kann.

Wenn sie nicht nur Prinzipien formuliert,  
sondern Verantwortlichkeit erzwingt.

Und wenn sie nicht ersetzt,  
was eigentlich menschliche Aufgabe bleibt:

Maßstäbe zu setzen.

Maßstäbe zu prüfen.

Maßstäbe zu verändern.

---

## Kapitel 5

### Admissibility, Macht und Drift

Der Begriff *Admissibility* klingt technisch.

Er wirkt neutral.

Formal.

Regelgebunden.

Aber er ist es nicht.

Admissibility ist nie nur eine technische Kategorie.

Sie ist immer Ausdruck von Macht.

---

### 1. Was bedeutet Admissibility wirklich?

In agentischen Systemen bezeichnet Admissibility die Frage:

Was darf das System tun?

Welche Antworten gelten als zulässig?

Welche Entscheidungen dürfen ausgeführt werden?

Welche Handlungen werden blockiert?

Auf den ersten Blick ist das eine Spezifikationsfrage.

Auf den zweiten Blick ist es eine normative Festlegung.

Denn jede Definition von „zulässig“  
setzt voraus,

dass jemand festgelegt hat,  
was legitim ist.

Und genau hier beginnt die eigentliche Problematik.

---

## 2. Normative Drift

Normen verändern sich nicht sprunghaft.  
Sie verschieben sich graduell.

- durch neue Geschäftsinteressen
- durch regulatorische Anpassungen
- durch politischen Druck
- durch gesellschaftliche Gewöhnung
- durch Effizienzargumente

Was gestern als „unzulässig“ galt,  
kann morgen als „vertretbar“ erscheinen.

Diese Verschiebung ist menschlich.  
Sie ist historisch normal.  
Sie ist unvermeidbar.

Gefährlich wird sie erst,  
wenn sie unsichtbar geschieht.

---

## 3. Unsichtbare Internalisierung

Wenn normative Verschiebungen  
direkt in Systemparameter einfließen,  
ohne dokumentiert, versioniert oder diskutiert zu werden,  
entsteht eine stille Normalisierung.

Das System verhält sich anders.  
Aber niemand kann präzise sagen, wann sich die Grenze verschoben hat.

Die Drift wird absorbiert.

Und genau hier liegt das strukturelle Risiko:

Nicht der einzelne Fehler.  
Nicht der einzelne Missbrauch.

Sondern die schleichende Verschiebung dessen,  
was als akzeptabel gilt.

---

## 4. Machtfrage statt Technikfrage

Wer definiert Admissibility?

- Der Anbieter?
- Das Unternehmen?
- Ein Governance-Board?
- Der Gesetzgeber?
- Der Markt?

Und noch wichtiger:

Wer überprüft regelmäßig,  
ob diese Definition noch tragfähig ist?

Admissibility ist keine rein technische Variable.  
Sie ist Ausdruck institutioneller Autorität.

Wenn diese Autorität nicht transparent organisiert ist,  
wird Admissibility zur stillen Machtposition.

---

## 5. Drift als strukturelles Phänomen

Drift ist kein Skandal.  
Drift ist Struktur.

In komplexen Systemen verändern sich:

- Datenlagen
- Nutzungskontexte

- Incentives
- gesellschaftliche Erwartungen

Die Frage ist also nicht,  
ob Drift stattfindet.

Die Frage ist,  
ob sie sichtbar gemacht wird.

Architektur kann Drift nicht verhindern.  
Aber sie kann sie protokollieren.

Und genau das ist der Unterschied zwischen:

- stiller Normalisierung  
und
  - nachvollziehbarer Evolution.
- 

## 6. Die eigentliche Herausforderung

Wenn wir Governance ernst nehmen,  
müssen wir akzeptieren:

Legitimität ist kein stabiler Zustand.  
Sie ist ein kontinuierlicher Aushandlungsprozess.

Admissibility darf deshalb nicht als statische Konfiguration verstanden werden,  
sondern als versioniertes, überprüfbares, diskussionsfähiges Schema.

Nur dann bleibt Raum für Korrektur.

Ohne diese Struktur  
verlagert sich normative Entwicklung  
vom öffentlichen Diskurs  
in den technischen Unterbau.

Und dort ist sie kaum noch sichtbar.

---

## 7. Der Kernpunkt

Architektur kann Legitimität nicht erzeugen.

Aber sie kann verhindern,  
dass sich Legitimität unbemerkt verschiebt.

Das ist kein moralischer Anspruch.  
Es ist eine strukturelle Minimalbedingung.

Und vielleicht liegt genau hier  
die Verbindung zwischen Ethik und Governance:

Nicht in moralischen Appellen.  
Nicht in Prinzipienkatalogen.

Sondern in der Frage,  
ob normative Veränderungen nachvollziehbar bleiben.

---

## Kapitel 6

### Human-in-the-Loop – Illusion oder strukturelle Verantwortung?

„Human-in-the-Loop“ klingt beruhigend.

Es suggeriert:

Der Mensch kontrolliert.  
Der Mensch entscheidet.  
Der Mensch bleibt verantwortlich.

Doch diese Formel kann zwei völlig unterschiedliche Bedeutungen haben.

Und genau hier trennt sich Symbolik von Substanz.

---

#### 1. Die beruhigende Version

In vielen Organisationen bedeutet Human-in-the-Loop:

- Ein Mensch klickt auf „Bestätigen“.
- Ein Mensch schaut stichprobenartig drauf.
- Ein Mensch kann im Notfall eingreifen.

Formal ist der Mensch beteiligt.  
Praktisch folgt er der Maschine.

Wenn die KI in 98 % der Fälle „richtig“ liegt,  
wird der Mensch zum Abnicker.

Je höher die Trefferquote,  
desto größer die psychologische Trägheit.

Verantwortung wird formal behalten,  
aber faktisch delegiert.

Das ist keine böse Absicht.  
Das ist kognitive Ökonomie.

---

## 2. Automation Bias

Ein bekanntes Phänomen aus der Luftfahrt und Medizin:

Menschen neigen dazu, automatisierten Systemen zu vertrauen –  
selbst dann, wenn Hinweise auf Fehler bestehen.

Man nennt das Automation Bias.

Je komplexer ein System,  
je höher seine statistische Leistungsfähigkeit,  
desto schwieriger wird aktiver Widerspruch.

Human-in-the-Loop kann deshalb zur Illusion werden,  
wenn die Struktur nicht bewusst gegen diese Dynamik arbeitet.

---

## 3. Verantwortung ist kein Button

Echte Verantwortung entsteht nicht dadurch,  
dass ein Mensch im Prozess vorkommt.

Sie entsteht durch:

- klare Entscheidungsbefugnis
- dokumentierte Begründung

- nachvollziehbare Ablehnungsmöglichkeit
- geschützte Widerspruchsräume

Wenn der Mensch nur eingreift,  
wenn „etwas offensichtlich falsch ist“,  
ist er kein Entscheidungsträger mehr,  
sondern Restabsicherung.

Das reicht für Komfortsysteme.

Nicht für Systeme mit Autoritätswirkung.

---

## **4. Agentische Systeme verschieben die Dynamik**

Bei agentischen Systemen wird es ernst.

Hier geht es nicht mehr nur um Textvorschläge,  
sondern um:

- automatische E-Mails
- operative Entscheidungen
- Prozessauslösungen
- Zugriff auf Systeme

In solchen Umgebungen  
kann Human-in-the-Loop nicht nur eine kosmetische Schicht sein.

Er muss strukturell verankert sein.

Nicht als Ausnahme –  
sondern als bewusst gestaltete Interventionsarchitektur.

---

## **5. Haltung und Training**

Dr. Giehl hat einen entscheidenden Punkt formuliert:

Human-in-the-Loop erfordert Haltung und Training.

Das ist keine Nebenbemerkung.

Wenn Menschen nicht lernen,

- Entscheidungen zu hinterfragen,
- KI-Outputs zu prüfen,
- Begründungen einzufordern,
- Unsicherheit auszuhalten,

wird Human-in-the-Loop zur leeren Formel.

Dann entsteht eine paradoxe Situation:

Je leistungsfähiger das System,  
desto schwächer die menschliche Entscheidungsfähigkeit.

Nicht aus Bosheit.

Sondern aus Bequemlichkeit und Zeitdruck.

---

## **6. Wettbewerb als struktureller Druck**

Christian Fischer hat eine zentrale Frage gestellt:

Wie soll man ethisch entscheiden,  
wenn man im Wettbewerb steht?

Genau hier liegt der Kern.

Wenn Geschwindigkeit belohnt wird  
und Zögern sanktioniert,  
dann wird Human-in-the-Loop ökonomisch unattraktiv.

Und Ethik wird zur Verzögerung.

Deshalb kann Human-in-the-Loop nur funktionieren,  
wenn Organisationen bewusst:

- Reflexionszeit einplanen
- Review-Prozesse institutionalisieren
- Fehlentscheidungen nicht sanktionieren, sondern analysieren

Verantwortung braucht Raum.

Ohne Raum bleibt sie Theorie.

Wettbewerb erzeugt Beschleunigung.  
Beschleunigung reduziert Reflexionszeit.

Genau deshalb reicht moralischer Appell nicht aus.

Organisationen benötigen strukturelle Mechanismen, die auch unter Zeit- und Wettbewerbsdruck verantwortliche Entscheidungen ermöglichen.

Runtime-Autorität ist ein solcher Mechanismus.

Sie verankert Verantwortung nicht im Diskurs, sondern im operativen Moment der Ausführung.

---

## **7. Die eigentliche Frage**

Human-in-the-Loop ist nicht die Antwort.

Er ist die Frage.

Wer übernimmt Verantwortung,  
wenn das System effizienter ist als wir?

Wer widerspricht,  
wenn es unbequem wird?

Wer stoppt,  
wenn alle weiterlaufen wollen?

Wenn wir Human-in-the-Loop nur als Sicherheitsgurt betrachten,  
wird er reißen.

Wenn wir ihn als kulturelle Praxis verstehen,  
kann er tragfähig werden.

---

## **8. Schlussgedanke**

Architektur kann Transparenz schaffen.  
Governance kann Regeln formulieren.  
Modelle können stabilisiert werden.

Aber Verantwortung bleibt menschlich.

Nicht als moralischer Anspruch.

Sondern als strukturelle Notwendigkeit.

Human-in-the-Loop ist keine technische Funktion.

Er ist eine Entscheidung.

Jeden Tag neu.

---

## Kapitel 7

### **Bildung, Kompetenz und die nächste Generation – Warum die eigentliche Arbeit jetzt beginnt**

Wenn wir über Governance, Architektur und Human-in-the-Loop sprechen,  
sprechen wir implizit über Kompetenz.

Und genau hier beginnt das eigentliche Problem.

Nicht im Modell.

Nicht im Prompt.

Nicht im Code.

Sondern im Bildungssystem.

---

### **1. Delegation ohne Urteilskraft**

Agentische Systeme werden in Organisationen eingesetzt,  
lange bevor Menschen systematisch darauf vorbereitet sind.

Schüler nutzen LLMs,  
bevor sie gelernt haben, Argumente sauber zu entwickeln.

Studierende generieren Texte,  
bevor sie gelernt haben, Quellen kritisch zu prüfen.

Mitarbeitende automatisieren Prozesse,  
bevor sie verstanden haben, welche Risiken damit verbunden sind.

Das Problem ist nicht Nutzung.

Das Problem ist Delegation ohne Urteilskraft.

---

## 2. KI als Beschleuniger vorhandener Defizite

KI zerstört kein funktionierendes Bildungssystem.

Aber sie verstärkt ein instabiles.

Wenn grundlegende Fähigkeiten fehlen:

- logisches Denken
- sprachliche Präzision
- Quellenkritik
- Frustrationstoleranz
- Konzentrationsfähigkeit

dann wird KI nicht kompensieren.

Sie wird kaschieren.

Und was kaschiert wird, kann nicht reifen.

Das ist die eigentliche Gefahr.

Nicht Überlegenheit der KI.  
Sondern Verlernen von Urteilskraft.

---

## 3. Der gläserne Schüler

Ein weiteres Risiko:

Wenn Lernsysteme datengetrieben optimieren,  
entsteht ein Profil.

Leistungsprofile.  
Verhaltensprofile.  
Interaktionsprofile.

Die Frage lautet nicht nur:

Was lernt der Schüler?

Sondern:

Wer lernt über den Schüler?

Wenn Governance nicht klar definiert ist,  
wird Bildung zur Datenquelle.

Und der Lernende zur Ressource.

---

## **4. Der Lehrer ist nicht ersetzbar – aber transformierbar**

Ein guter Lehrer ist unersetzbar.

Nicht wegen der Wissensvermittlung.  
Sondern wegen der Haltung.

Ein Lehrer strukturiert:

- Zweifel
- Irrtum
- Widerstand
- Entwicklung

KI kann Inhalte liefern.  
Sie kann Feedback generieren.  
Sie kann Übungsvarianten erstellen.

Aber sie trägt keine Verantwortung für Charakterbildung.

Wenn Bildung zur reinen Effizienzoptimierung wird,  
verlieren wir mehr als wir gewinnen.

---

## **5. Was Schulen wirklich brauchen**

Nicht nur KI-Regeln.

Sondern:

- systematische Schulung im kritischen Umgang mit KI

- Transparenz darüber, wie Modelle funktionieren
- Training im Erkennen von Halluzinationen
- klare Richtlinien zur Attribution
- Räume für eigenständiges Denken ohne KI-Unterstützung

Die zentrale Kompetenz der Zukunft ist nicht Prompting.

Es ist Urteilsfähigkeit.

---

## **6. Wettbewerb und Bequemlichkeit**

Warum ist das schwer?

Weil KI Vorteile bringt.

Zeitgewinn.

Effizienz.

Bequemlichkeit.

Und Schüler stehen im Wettbewerb.

Studierende stehen im Wettbewerb.

Lehrkräfte stehen im Wettbewerb.

Wenn KI 20 % schneller ist,  
wird sie genutzt.

Deshalb reicht moralischer Appell nicht.

Es braucht strukturelle Integration.

---

## **7. Bildung als Ort der Gegenbewegung**

Vielleicht liegt hier die größte Verantwortung.

Schulen könnten der Ort sein,  
an dem:

- Denken verlangsamt wird

- Argumente zerlegt werden
- Quellen überprüft werden
- Unsicherheit ausgehalten wird

Nicht trotz KI.  
Sondern wegen KI.

Wenn alles schneller wird,  
muss Bildung stabilisieren.

---

## 8. Die eigentliche Arbeit beginnt jetzt

Die Frage ist nicht:

„Hat KI menschliches Niveau erreicht?“

Die Frage ist:

Erreichen wir das Niveau,  
das nötig ist, um mit ihr verantwortungsvoll zu arbeiten?

Agentische Systeme werden kommen.  
Governance-Frameworks werden entstehen.  
Regulatorik wird sich entwickeln.

Aber wenn die nächste Generation nicht lernt,

- wie man argumentiert,
- wie man zweifelt,
- wie man Verantwortung übernimmt,

dann wird Human-in-the-Loop nur eine Phrase bleiben.

---

## 9. Schlussgedanke

KI fordert nicht nur Technik heraus.

Sie fordert den Logos heraus.

Nicht weil sie denkt.  
Sondern weil wir entscheiden müssen,  
wie wir denken wollen.

Die eigentliche Frage ist nicht,  
ob KI intelligenter wird.

Sondern ob wir es bleiben.

---

## Kapitel 8

### Von der Theorie zur Praxis – Leitlinien für Organisationen und Bildungseinrichtungen

Wenn wir Governance, Verantwortung und Bildung ernst nehmen,  
dann müssen sie sich in Strukturen übersetzen lassen.

Nicht als Leitbild.  
Sondern als Betriebssystem.

---

#### 1. Trennung von Modell, System und Verantwortung

Erste Klarheit:

Ein LLM ist kein System.  
Ein Agent ist kein Unternehmen.  
Und ein Output ist keine Entscheidung.

Organisationen sollten drei Ebenen explizit trennen:

1. **Modell-Ebene**
  - probabilistischer Kern
  - keine Garantie, nur Wahrscheinlichkeiten
  
2. **System-Ebene**
  - Guardrails
  - Tool-Access
  - Logging
  - Zugriffskontrolle
  - Fail-Closed-Mechanismen

### 3. Verantwortungsebene

- Wer genehmigt Use Cases?
- Wer überprüft Guidelines?
- Wer stoppt Systeme im Zweifel?

Wenn diese Ebenen nicht sauber getrennt sind,  
entsteht Delegation ohne Haftung.

---

## 2. Klare Agent-Owner-Struktur

Jeder agentische Use Case braucht:

- einen fachlichen Owner
- einen technischen Owner
- einen Governance-Verantwortlichen

Nicht „IT macht das schon“.

Nicht „HR nutzt es halt“.

Sondern benannte Zuständigkeit.

Ohne Agent-Owner entsteht Shadow-AI 2.0.

---

## 3. Versionierte Guidelines

Wenn admissibility evolutionär ist,  
dann müssen Guidelines versioniert sein.

- Datum
- verantwortliche Instanz
- dokumentierte Änderung
- Begründung

Normative Drift darf nicht unsichtbar passieren.

Transparenz ersetzt keine Entscheidung.  
Aber sie verhindert stille Verschiebung.

---

## 4. Fail-Closed statt Good-Faith

In High-Risk-Kontexten gilt:

Wenn Unsicherheit besteht → Blockieren.  
Nicht: „Wird schon passen.“

Beispiele:

- Kein Zugriff auf neue Tools ohne Freigabe
- Kein Versand externer Kommunikation ohne Review
- Keine rechtliche Einschätzung ohne Quellenprüfung

Human-in-the-Loop bedeutet nicht nur „überwachen“.

Es bedeutet: Eingreifen können.

---

## 5. Attribution als Pflicht, nicht als Option

Gerade in Bildung und Forschung:

Jede KI-gestützte Arbeit sollte deklarieren:

- Welche Tools wurden genutzt?
- Für welche Teile?
- Wurden Quellen geprüft?

Nicht zur Kontrolle.

Sondern zur epistemischen Ehrlichkeit.

Wenn Attribution verschwindet,  
verliert Erkenntnis ihre Grundlage.

---

## 6. Training statt Verbot

Verbote funktionieren kurzfristig.

Kompetenz wirkt langfristig.

Organisationen sollten verpflichtend schulen:

- Wie funktionieren LLMs?
- Was ist Halluzination?
- Was ist Prompt Injection?
- Wo liegen Haftungsrisiken?
- Was bedeutet probabilistisches Systemverhalten?

Nur wer versteht, kann verantworten.

---

## 7. Bildung: Zwei Modi bewusst trennen

In Schulen und Hochschulen:

- **KI-gestützter Modus**
- **KI-freier Modus**

Beide bewusst, transparent und didaktisch begründet.

Nicht heimliche Nutzung.

Nicht totale Verdrängung.

Sondern strukturierte Integration.

---

## 8. Reflexionsfenster institutionalisieren

Regelmäßige Fragen:

- Welche Use Cases haben wir eingeführt?

- Welche Risiken sind neu entstanden?
- Wo hat sich die Entscheidungsarchitektur verschoben?
- Haben wir noch die Kontrolle über Tool-Access?

Governance ist kein Dokument.  
Sie ist ein Prozess.

---

## 9. Kulturelle Voraussetzung

Technik kann Transparenz schaffen.  
Sie kann Logs erzwingen.  
Sie kann Versionen dokumentieren.

Aber sie kann keine Haltung erzeugen.

Organisationen brauchen:

- Fehlerkultur
- Einspruchsmöglichkeit
- Schutz für kritische Stimmen
- Bereitschaft, Tempo zu reduzieren

Wenn Tempo absolut gesetzt wird,  
verliert Governance.

---

## 10. Schluss

Die eigentliche Frage lautet nicht:

Wie schnell können wir KI integrieren?

Sondern:

Unter welchen Bedingungen dürfen wir sie integrieren?

Ethik ohne Struktur bleibt Symbol.  
Struktur ohne Ethik bleibt Machtinstrument.

Zwischen beiden liegt Verantwortung.

Und die ist nicht delegierbar.

Der erste Wandel erzeugt Aufmerksamkeit.  
Der zweite schafft Ordnung.

Der erste wird diskutiert.  
Der zweite wird gelebt.

Ein Paradigmenwechsel beginnt nicht mit einem Manifest.

Er beginnt in dem Moment,  
in dem Organisationen entscheiden,

dass Geschwindigkeit nicht mehr der einzige Maßstab ist.

Sondern Tragfähigkeit.

Und Tragfähigkeit entsteht dort,

wo technologische Handlungsmacht  
mit klar benannter menschlicher Verantwortung  
verbunden bleibt.

---

## Schlusssektion

### Zwischen Beschleunigung und Verantwortung

Wir erleben derzeit eine Phase technologischer Beschleunigung,  
die viele Institutionen überfordert.

Die Debatte schwankt zwischen Euphorie und Alarmismus.  
Zwischen „Revolution“ und „Untergang“.

Beides greift zu kurz.

KI-Systeme sind weder magisch noch dämonisch.  
Sie sind leistungsfähige probabilistische Werkzeuge,  
die in immer komplexere Entscheidungsarchitekturen eingebettet werden.

Die entscheidende Frage lautet daher nicht:

Können sie denken?

Sondern:

Unter welchen Bedingungen dürfen sie handeln?

Mit jedem Schritt von „Antwort“ zu „Aktion“  
verschiebt sich Verantwortung von der Person zur Struktur.

Wenn Governance nicht klar definiert ist,  
wenn admissibility implizit bleibt,  
wenn normative Maßstäbe unbemerkt drifteten,

dann entsteht Delegation ohne Bewusstsein.

Architektur kann keine Ethik erzeugen.  
Aber sie kann verhindern,  
dass Verantwortung unbemerkt internalisiert wird.

Ethik bleibt menschlich.  
Governance macht sie operativ.  
Bildung macht sie tragfähig.

Die Zukunft entscheidet sich daher nicht im Modell,  
sondern in der Art,  
wie wir mit seinen Möglichkeiten umgehen.

Nicht Geschwindigkeit ist der Maßstab.

Sondern Tragfähigkeit.

---

## **Autoren-Statement für „Perspektiven“**

### **Andrzej Skulski**

Koch, Unternehmer, unabhängiger Publizist im Bereich KI-Governance und Bildungsfragen.

Arbeitet an der Schnittstelle von Verantwortung, Entscheidungsarchitektur und praktischer Umsetzung technologischer Systeme im Alltag von Organisationen. Sein Fokus liegt auf der Frage, wie Governance-Strukturen gestaltet sein müssen, damit Ethik nicht symbolisch bleibt, sondern handlungsfähig wird – insbesondere im Kontext agentischer KI-Systeme und Bildungseinrichtungen.