abamix
AI. Applied. Governed.

EXECUTIVE PAPER · NOVEMBER 2026

Urteilskraft

Warum wirksame KI-Aufsicht eine Aufgabe der Geschäftsleitung ist

AUTOR Prof. Dr. Amadou Sienou, abamix GmbH, Stuttgart

VERSION 1.1
DATUM 07.11.2026
STATUS Freigegeben

www.abamix.com

ÜBERSICHT

Inhalt

	Das Wichtigste in Kürze	3
1	**Der Mensch in der Schleife ist noch kein Urteil**	4
1.1	Warum Beteiligung nicht reicht	4
1.2	Zwei Modi des Urteilens	4
1.3	Was bei KI beim Menschen bleibt	5
2	**Wie gut Menschen tatsächlich urteilen**	6
2.1	Verzerrung und Rauschen	6
2.2	Wann Intuition trägt	7
3	**Was das Recht verlangt**	9
3.1	KI-Verordnung: Aufsicht und Fristen	9
3.2	DSGVO und SCHUFA-Urteil	10
3.3	Vier Bedingungen wirksamer Aufsicht	10
4	**Wie Urteilskraft in die Arbeit mit KI kommt**	11
4.1	Der Arbeitszyklus	11
4.2	Was die Wirkung trägt	12
5	**Sechs Entscheidungen der Geschäftsleitung**	13
5.1	Struktur und Aufsicht	13
5.2	Kompetenz und Perspektive	14
5.3	Messung und Kalibrierung	14
6	**Der nächste Schritt**	15
7	**So unterstützt abamix**	16
	Quellen	17

Das Wichtigste in Kürze

Ein Mensch in der Schleife garantiert keine Urteilskraft. Wirksame Aufsicht über KI braucht Kompetenz, Zeit, Unabhängigkeit und eine reale Befugnis zur Ablehnung.

KI übernimmt die Anwendung von Regeln. Die Entscheidung, ob eine Regel auf diesen Fall passt, bleibt bei Ihnen. Diese Fähigkeit heißt Urteilskraft. Sie lässt sich nicht vollständig in Regeln fassen, aber systematisch kultivieren: durch Struktur, Rückmeldung und Perspektivübernahme.

Ohne Struktur ist menschliches Urteil verrauscht. In einem Noise-Audit bei einem Versicherer lag die mediane Abweichung zwischen zwei Underwritern bei identischen Fällen bei 55 %, erwartet hatten Führungskräfte etwa 10 %. Das Recht misst Aufsicht an ihrer Wirksamkeit: Nach dem SCHUFA-Urteil des EuGH schließt eine formale menschliche Mitwirkung, die faktisch dem Score folgt, eine ausschließlich automatisierte Entscheidung nicht aus.

Sechs Entscheidungen liegen bei der Geschäftsleitung:

1. Regeln und Modelle für Routinefälle einsetzen, menschliche Urteilszeit für neuartige Fälle und Wertkonflikte schützen.
2. Für jedes Hochrisiko-System nach Anhang III bis zum 2. Dezember 2027 dokumentieren, wer übersteuern darf, mit welcher Kompetenz und welchem Zeitbudget.
3. Urteilskraft trainieren, nicht nur Wissen vermitteln.
4. Vor Hochrisiko-Einführungen die Standpunkte betroffener Gruppen schriftlich einholen.
5. Das eigene Rauschen messen.
6. Selbstsicherer Intuition in Umwelten geringer Validität misstrauen.

Rechtliche Aussagen geben den Stand vom 07.11.2026 wieder und ersetzen keine Rechtsberatung.

KAPITEL

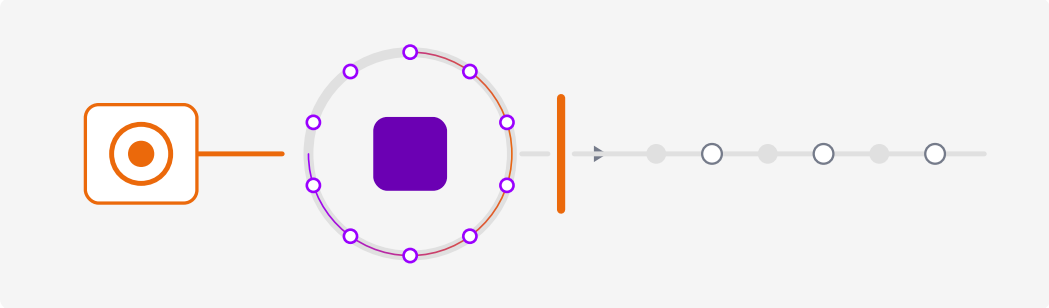
01

Der Mensch in der Schleife ist noch kein Urteil

IN DIESEM KAPITEL

1.1 Warum Beteiligung nicht reicht
1.2 Zwei Modi des Urteilens
1.3 Was bei KI beim Menschen bleibt

Die verbreitete Antwort auf KI-Risiken lautet: Ein Mensch prüft das Ergebnis. Diese Antwort ist notwendig, aber sie reicht nicht.



1.1 Warum Beteiligung nicht reicht

Ein Mensch, der Ausgaben abzeichnet, ohne die Zeit, die Kompetenz oder die Befugnis zu haben, sie abzulehnen, legitimiert eine Entscheidung, ohne sie zu treffen. Regelungen, die menschliche Aufsicht vorschreiben, erzeugen häufig eine bloß formale Legitimation, ohne die Entscheidungsqualität zu verbessern. [2]

1.2 Zwei Modi des Urteilens

Urteilskraft ist die Fähigkeit, die zwischen Wissen und Handeln vermittelt. Nach Kant ist sie das Vermögen, das Besondere als enthalten unter dem Allgemeinen zu denken. Sie wirkt in zwei Modi: [7]

Dimension	Bestimmende Urteilskraft	Reflektierende Urteilskraft
Gegeben	Regel, Gesetz, Begriff	Nur der Fall
Entsprechung in der Organisation	Compliance, Prozessvorgaben, regelbasierte Entscheidungsrechte	Strategie unter Mehrdeutigkeit, Krisenentscheidungen, Bewertung neuer KI-Anwendungsfälle
Wer sie übernehmen kann	Regeln, Modelle, Checklisten	Menschen mit Verantwortung

Wollte man Regeln für die Anwendung von Regeln geben, bräuchte jede dieser Regeln wiederum Urteilskraft. Kants Folgerung: Urteilskraft will „gar nicht belehrt, sondern nur geübt sein“. [6] Gutes Urteilen prüft das eigene Urteil zudem an den möglichen Standpunkten anderer. Mit Arendt heißt das, die eigene Einbildungskraft zu trainieren, „auf Besuch zu gehen“. [14]

1.3 Was bei KI beim Menschen bleibt

Systeme maschinellen Lernens sind leistungsfähige Apparate bestimmender Urteilskraft: Sie ordnen Einzelfälle gelernten Mustern zu. Ein Rest bleibt stets beim Menschen. Jemand muss entscheiden, ob die Kategorie des Modells auf diesen Fall passt, ob der Fall außerhalb der Trainingsverteilung liegt und welcher konfligierende Wert Vorrang hat.

KAPITEL

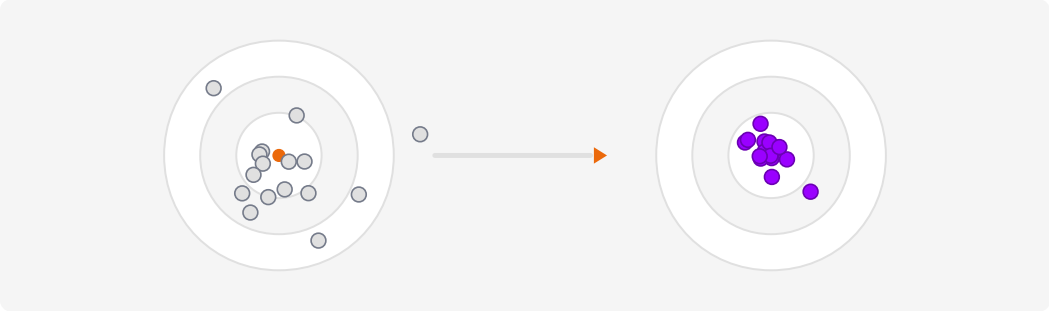
02

Wie gut Menschen tatsächlich urteilen

IN DIESEM KAPITEL

2.1 Verzerrung und Rauschen
2.2 Wann Intuition trägt

Unstrukturiertes Urteilen ist verzerrt und verrauscht. Strukturiertes, rückgekoppeltes Urteilen ist kultivierbar.



2.1 Verzerrung und Rauschen

Intuitives Urteilen beruht auf Faustregeln, die systematische Fehler erzeugen. [16] Hinzu kommt Rauschen: unerwünschte Streuung zwischen Urteilen, die identisch sein sollten. Ein Noise-Audit bei einem Versicherer macht die Größenordnung sichtbar. [4, 5]

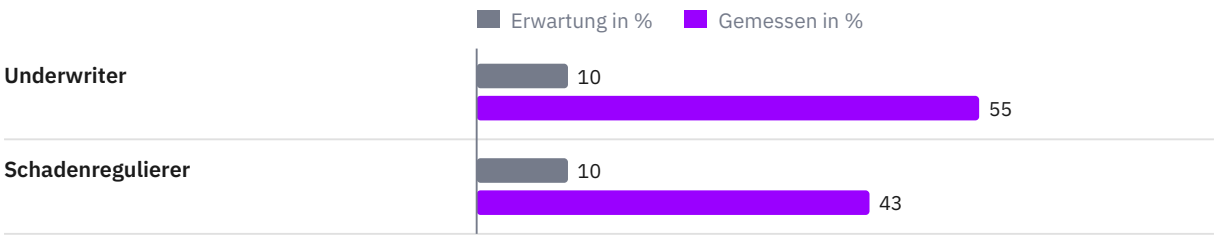


Abbildung 1: Noise-Audit bei einem Versicherer: erwartete und gemessene Abweichung bei identischen Fällen

Die Kennzahlen stammen aus Beratungsmandaten in einzelnen Unternehmen. Sie illustrieren die Größenordnung, sind aber keine universellen Konstanten. Als Gegenmittel empfehlen die Autoren Entscheidungshygiene: Urteile strukturieren, unabhängige Einschätzungen aggregieren, intuitive Gesamturteile verzögern. [5]

2.2 Wann Intuition trägt

Intuition ist nicht wertlos. Erfahrene Einsatzleiter der Feuerwehr verglichen in weniger als 12 % der Fälle Optionen; überwiegend erkannten sie ein vertrautes Muster. [8] Kahneman und Klein zeigen jedoch, dass intuitives Urteil nur unter zwei Bedingungen verlässlich ist: in einer Umwelt mit hoher Validität und bei ausreichender Übung mit schneller, eindeutiger Rückmeldung. Subjektive Gewissheit ist kein Indikator für Treffsicherheit. [3]

Ordnet man Entscheidungen nach Validität der Umwelt und Art der Rückmeldung, ergeben sich vier Felder:

- **Geübte Intuition** (hohe Validität, schnelle Rückmeldung): Intuition kann tragen, wie bei erfahrenen Einsatzleitern der Feuerwehr.
- **Fehlende Lerngelegenheit** (hohe Validität, verzögerte Rückmeldung): Rückmeldung muss künstlich hergestellt werden, etwa durch Prognosen und Entscheidungsjournale.
- **Instabile Hinweisreize** (geringe Validität, schnelle Rückmeldung): Ohne stabile Hinweisreize trägt Intuition nicht; Struktur und Entscheidungshygiene gehen vor.
- **Strategie und KI-Risiko** (geringe Validität, verzögerte Rückmeldung): Hier verdient Intuition das geringste Vertrauen; Basisraten und unabhängige Zweitmeinungen werden Pflicht.

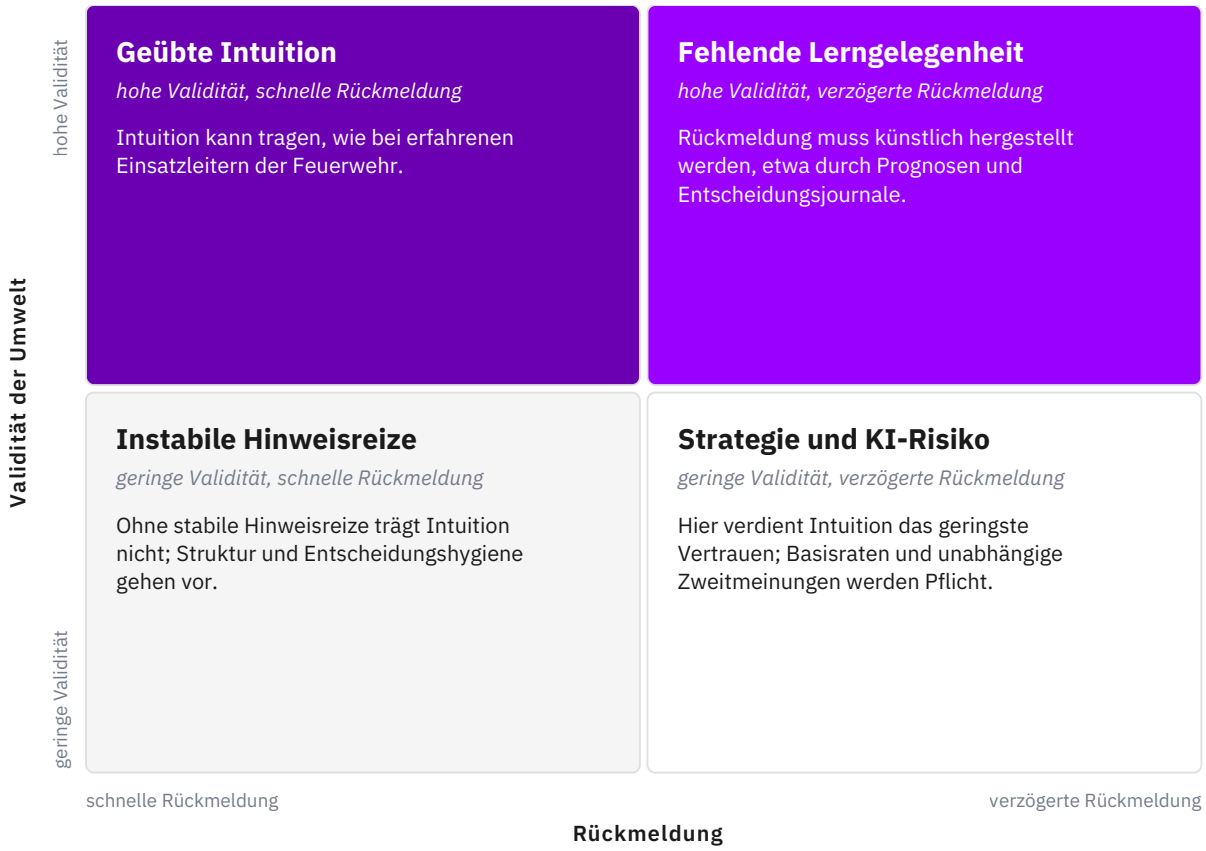


Abbildung 2: Wann Intuition trägt: vier Felder nach Kahneman und Klein

KAPITEL

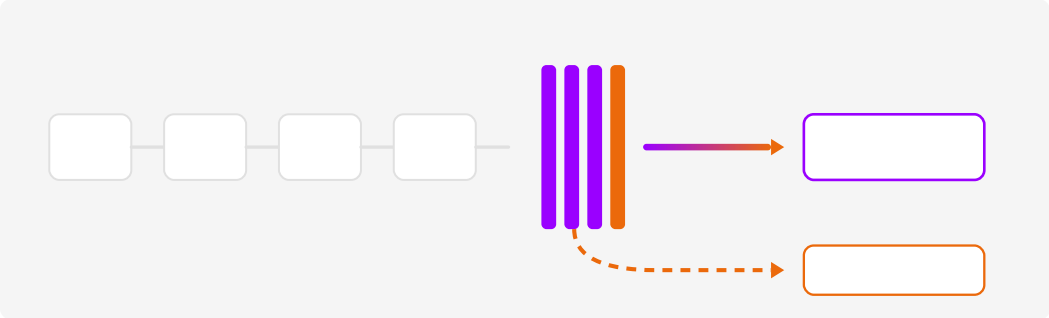
03

Was das Recht verlangt

IN DIESEM KAPITEL

3.1 KI-Verordnung: Aufsicht und Fristen
3.2 DSGVO und SCHUFA-Urteil
3.3 Vier Bedingungen wirksamer Aufsicht

Das Recht macht die Prüfung zur Pflicht und misst sie an ihrer Wirksamkeit, nicht an ihrer Form.



3.1 KI-Verordnung: Aufsicht und Fristen

Hochrisiko-KI-Systeme sind so zu konzipieren, dass natürliche Personen sie wirksam beaufsichtigen können. Aufsichtspersonen müssen Fähigkeiten und Grenzen des Systems verstehen, sich des Automation Bias bewusst bleiben, Ausgaben korrekt interpretieren, Ausgaben übersteuern oder außer Acht lassen und den Betrieb unterbrechen können (Art. 14 Verordnung (EU) 2024/1689). Betreiber übertragen die Aufsicht Personen mit der erforderlichen Kompetenz, Ausbildung und Befugnis (Art. 26 Abs. 2). [12]

Der Digital-Omnibus zur KI-Verordnung (Verordnung (EU) 2026/1744, in Kraft seit 27. Juli 2026) hat die Fristen verschoben, nicht aufgehoben: [13]

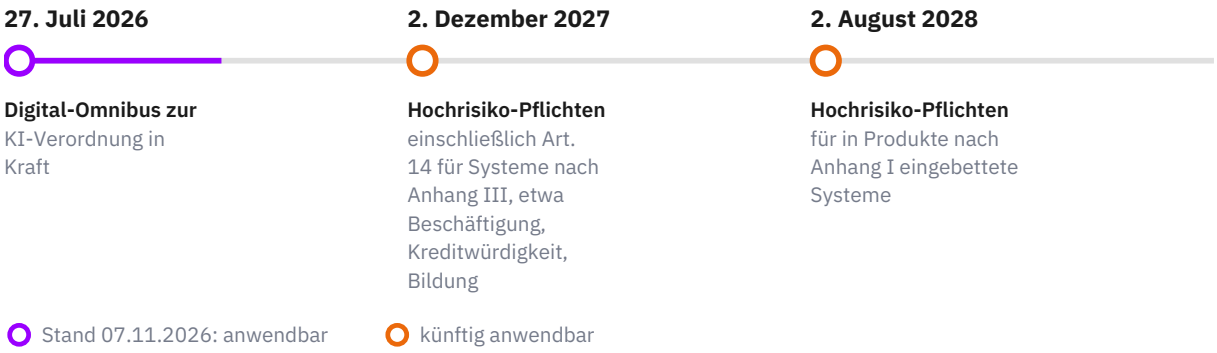


Abbildung 3: Anwendbarkeit der Hochrisiko-Pflichten nach dem Digital-Omnibus zur KI-Verordnung

3.2 DSGVO und SCHUFA-Urteil

Art. 22 Abs. 1 DSGVO gewährt das Recht, keiner ausschließlich automatisierten Entscheidung mit rechtlicher oder ähnlich erheblicher Wirkung unterworfen zu werden. [11] Der EuGH entschied, dass bereits die automatisierte Ermittlung eines Wahrscheinlichkeitswerts eine solche Entscheidung darstellt, wenn ein Dritter diesen Wert maßgeblich seiner Entscheidung über ein Vertragsverhältnis zugrunde legt (Urteil vom 7. Dezember 2023, C-634/21). [1] Für Sie heißt das: Entscheidend ist nicht, ob ein Mensch beteiligt ist, sondern ob er tatsächlich urteilt.

3.3 Vier Bedingungen wirksamer Aufsicht

Erfahrung allein schützt nicht. In einer Studie zur Mammographie verschlechterten fehlerhafte KI-Vorschläge die Treffsicherheit von Radiologinnen und Radiologen aller Erfahrungsstufen deutlich. [10] Aufsicht braucht vier Bedingungen:

KOMPETENZ	Die Aufsichtsperson versteht Fähigkeiten und Grenzen des Systems und kann Ausgaben interpretieren
ZEIT	Für die Prüfung steht ein Zeitbudget je Entscheidung zur Verfügung
UNABHÄNGIGKEIT	Die Prüfung erfolgt nicht durch dieselbe Instanz, die das Ergebnis erzeugt hat
BEFUGNIS	Die Aufsichtsperson darf übersteuern, ablehnen und den Betrieb unterbrechen

Abbildung 4: Vier Bedingungen wirksamer menschlicher Aufsicht

KAPITEL

04

Wie Urteilskraft in die Arbeit mit KI kommt

IN DIESEM KAPITEL

4.1 Der Arbeitszyklus

4.2 Was die Wirkung trägt

Urteilskraft wächst durch viele Fälle, ehrliche Rückmeldung, Perspektivübernahme und Struktur gegen Rauschen. Wirksam wird sie erst, wenn diese Hebel im Arbeitsalltag stattfinden.



4.1 Der Arbeitszyklus

Die Arbeitsmethode Briefing-Oriented Work legt das menschliche Urteil vor und nach die KI-Arbeit:

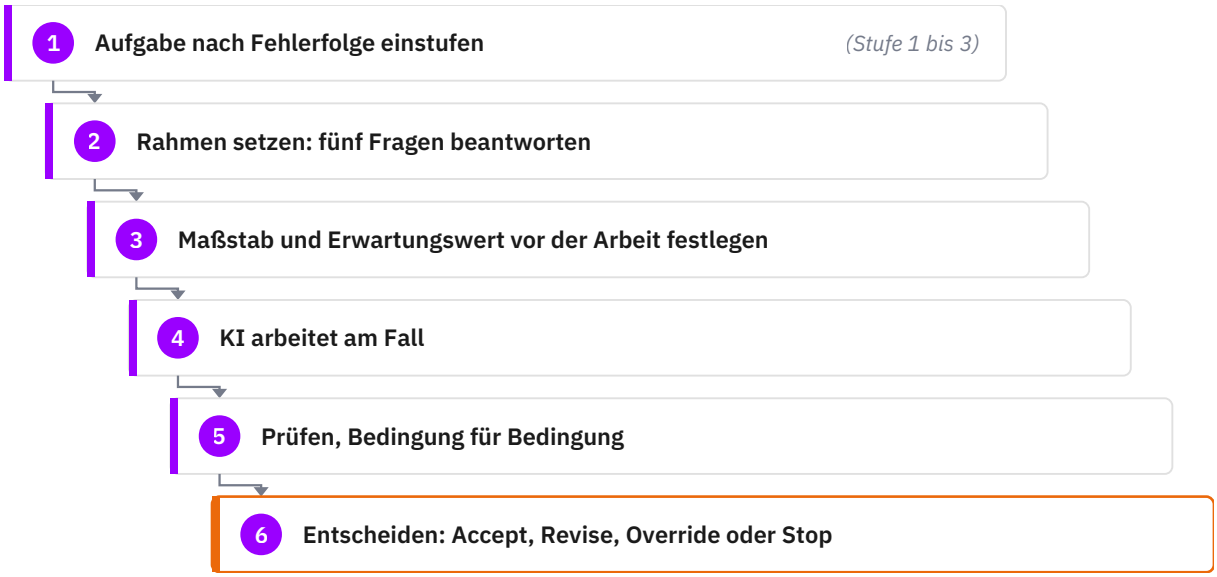


Abbildung 5: Arbeitszyklus mit Urteil vor und nach der KI-Arbeit

4.2 Was die Wirkung trägt

- **Die Rahmung bleibt beim Menschen.** Zweck, Gegenstand und Gütemaßstab werden formuliert, bevor die Maschine am Fall arbeitet.
- **Maßstab und Erwartung vor der Arbeit.** Das Gütekriterium wird vor der Arbeit geschrieben, der erwartete Wert vor der Messung notiert.
- **Drei Urteilsfragen auf der höchsten Stufe.** Was wurde nicht gesagt? Wessen Perspektive fehlt? Würde ich ohne KI genauso entscheiden? Die dritte Frage wirkt wie eine Selbstdistanzierung, und Selbstdistanzierung steigert weises Urteilen experimentell nachweisbar. [9]

Die Wirkung ist eine begründete Hypothese ohne Pilotergebnisse; Automatisierung kann zudem die Prüfkompetenz schwächen. [15]

KAPITEL

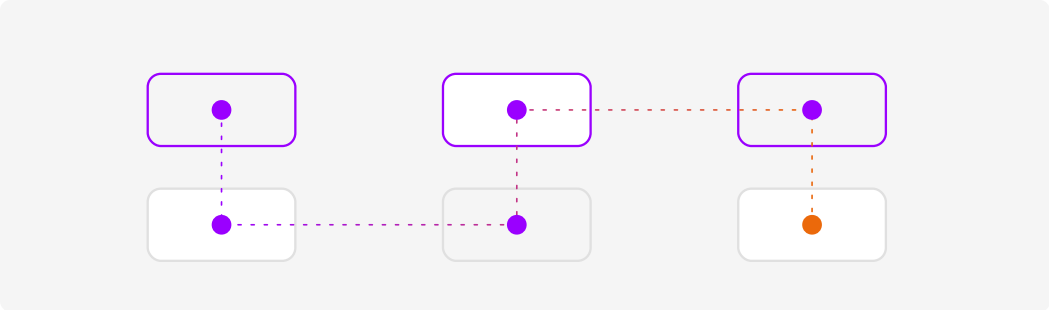
05

Sechs Entscheidungen der Geschäftsleitung

IN DIESEM KAPITEL

5.1 Struktur und Aufsicht
5.2 Kompetenz und Perspektive
5.3 Messung und Kalibrierung

Urteilskraft entsteht nicht durch Appelle, sondern durch Bedingungen, die nur die Führung schaffen kann.



5.1 Struktur und Aufsicht

Entscheidung 1 – Beide Modi im Governance-Design trennen.
Konkret: Für jeden KI-gestützten Entscheidungsprozess festlegen, welche Fälle das System entscheidet und welche ein Mensch.

Entscheidung 2 – Aufsicht substanziell machen.
Konkret: Bis zum 2. Dezember 2027 für jedes Anhang-III-System dokumentieren, wer übersteuern darf, über welche Kompetenz und welches Zeitbudget diese Person verfügt. Übersteuerungs- und Akzeptanzquoten laufend auswerten.

5.2 Kompetenz und Perspektive

Entscheidung 3 – Urteilskraft trainieren, nicht nur Wissen vermitteln.
Konkret: KI-Kompetenz mit wiederholter Fallarbeit und bewerteter Rückmeldung verbinden; geprüft wird die Anwendung auf unbekannte Fälle.

Entscheidung 4 – Erweiterte Denkungsart institutionalisieren.
Konkret: Vor Hochrisiko-Einführungen eine schriftliche Auseinandersetzung mit den Standpunkten betroffener Gruppen verlangen.

5.3 Messung und Kalibrierung

Entscheidung 5 – Eigenes Rauschen messen.
Konkret: Mehrere Führungskräfte dieselben KI-Risikofälle unabhängig bewerten lassen; die Streuung zeigt, wo gemeinsame Maßstäbe fehlen.

Entscheidung 6 – Selbstsicherer Intuition in Umwelten geringer Validität misstrauen.
Konkret: Für strategische Prognosen und KI-Risikobewertungen Basisraten und unabhängige Zweitmeinungen verpflichtend machen.



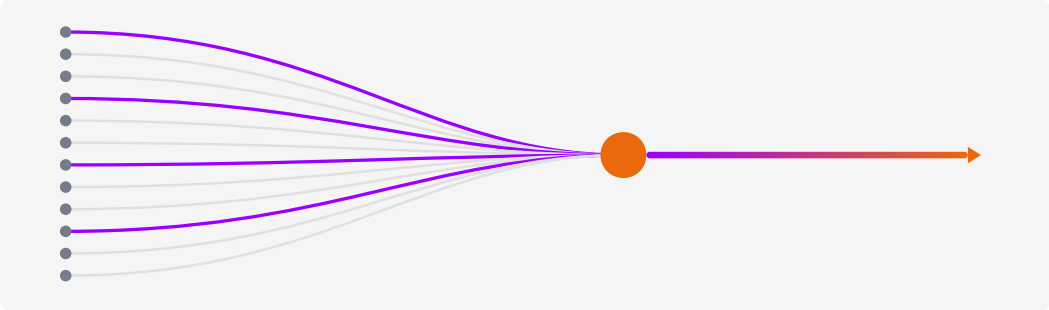
Abbildung 6: Sechs Entscheidungen der Geschäftsleitung

KAPITEL

06

Der nächste Schritt

Die zentrale Gestaltungsaufgabe besteht darin, Bedingungen zu schaffen, unter denen menschliche Urteilskraft tatsächlich ausgeübt werden kann. Es genügt nicht, menschliche Beteiligung bloß zu deklarieren.



Drei Fragen eignen sich als Prüfstein für Ihre nächste Sitzung der Geschäftsleitung:

1. Wissen wir für jedes KI-gestützte Entscheidungsverfahren, welche Fälle ein Mensch beurteilt und warum?
2. Hat jede Aufsichtsperson die Kompetenz, die Zeit und die Befugnis, eine Ausgabe abzulehnen?
3. Wissen wir, wie stark die Urteile unserer Fachleute bei identischen Fällen voneinander abweichen?

Wo eine Antwort fehlt, beginnt der Aufbau von Urteilskraft als Organisationskompetenz. Begriff, Forschung, Recht und Verfahren im Detail beschreibt das Whitepaper „Urteilskraft“ der abamix GmbH.

KAPITEL

07

So unterstützt abamix

Sie wollen menschliche Aufsicht über KI so gestalten, dass in Ihrer Organisation tatsächlich geurteilt und nicht nur abgezeichnet wird.

Executive Briefing Session
Für Sie als Vorstand oder Geschäftsführung, wenn KI-Risiken auf die Agenda des Gremiums kommen. Sie klären die Steuerungsmodi von „in the loop“ bis zur vollständigen Delegation innerhalb von Leitplanken, dazu Entscheidungsrechte und Eskalation, und rahmen Ihre Risikobereitschaft.

AI Enablement Program
Sie verankern Briefing-Kompetenz als Arbeitsmethode in Ihrem Team: BRIEF, WORK, CHECK als wiederholbarer Zyklus und eine Stufenkopplung, die festlegt, welche Entscheidungen welche Stufe menschlicher Prüfung erfordern.

AI Managed Governance
Sie halten Ihre Aufsicht im laufenden Betrieb wirksam. Ihr Agenten- und Modellregister wird gepflegt, der Steuerungsmodus pro Agent überprüft und Ihr Policy Pack gegenüber regulatorischen Änderungen versioniert.

AI Assurance
Sie lassen KI-Systeme im Produktivbetrieb, die andere gebaut haben, unabhängig testen und bewerten und erhalten Nachweise für Revision und Aufsicht.

Quellen

[1] EuGH, Urteil vom 7. Dezember 2023, C-634/21, *SCHUFA Holding (Scoring)*, ECLI:EU:C:2023:957. https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:62021CJ0634

[2] Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, *45*, 105681. https://doi.org/10.1016/j.clsr.2022.105681

[3] Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

[4] Kahneman, D., Rosenfield, A. M., Gandhi, L., & Blaser, T. (2016). Noise: How to overcome the high, hidden cost of inconsistent decision making. *Harvard Business Review*, *94*(10), 38–46.

[5] Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark.

[6] Kant, I. (1904/1911). *Kritik der reinen Vernunft* (1781/1787). In *Kant's gesammelte Schriften* (Bd. III–IV). Reimer. Zitiert: A133/B172.

[7] Kant, I. (1908/1913). *Kritik der Urteilskraft* (1790/1793). In *Kant's gesammelte Schriften* (Bd. V). Reimer. Zitiert: Einleitung IV, AA V, 179.

[8] Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the Human Factors Society Annual Meeting*, *30*(6), 576–580. https://doi.org/10.1177/154193128603000616

[9] Kross, E., & Grossmann, I. (2012). Boosting wisdom: Distance from the self enhances wise reasoning, attitudes, and behavior. *Journal of Experimental Psychology: General*, *141*(1), 43–48. https://doi.org/10.1037/a0024158

[10] Dratsch, T., Chen, X., Rezazade Mehrizi, M., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D., & Pinto dos Santos, D. (2023). Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*, *307*(4), e222176. https://doi.org/10.1148/radiol.222176

[11] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 (Datenschutz-Grundverordnung). ABl. L 119 vom 4.5.2016, S. 1–88.

[12] Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Verordnung über künstliche Intelligenz). ABl. L, 2024/1689, 12.7.2024.

[13] Verordnung (EU) 2026/1744 zur Änderung der Verordnung (EU) 2024/1689 (Digital-Omnibus zur KI). In Kraft seit 27. Juli 2026. https://eur-lex.europa.eu/eli/reg/2026/1744/oj

[14] Arendt, H. (1982). *Lectures on Kant's political philosophy* (R. Beiner, Hrsg.). University of Chicago Press. [Dt.: *Das Urteilen. Texte zu Kants Politischer Philosophie*. Piper, 1985.]

[15] Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

[16] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

KONTAKT

Ansprechpartner

Prof. Dr. Amadou Sienou

abamix GmbH, Stuttgart

info@abamix.com · www.abamix.com

Erstgespräch, 20–30 Minuten: meeting.abamix.com

IMPRESSUM

Herausgeber	abamix GmbH · Lothringerstr. 11 · 70435 Stuttgart
Dokument	Executive Paper — Kurzfassung des Whitepapers „Urteilskraft“ (v1.1) · Urteilskraft
Version	1.1 · 07.11.2026 · Status: Freigegeben
Autor	Prof. Dr. Amadou Sienou, abamix GmbH, Stuttgart
Zitiervorschlag	Sienou, A. (2026). *Urteilskraft: Warum wirksame KI-Aufsicht eine Aufgabe der Geschäftsleitung ist*. Executive Paper, Version 1.1. Stuttgart: abamix GmbH.

© 2026 abamix GmbH. Alle Rechte vorbehalten. Regulatorische Einschätzungen geben den Stand zum Versionsdatum wieder und ersetzen keine Rechtsberatung.



abamix

AI. Applied. Governed.

abamix GmbH · Lothringerstr. 11 · 70435 Stuttgart

info@abamix.com · www.abamix.com