

— INSIGHT · KI-OPERATING-MODEL & GOVERNANCE

# Die Autonomie kosten.

Unterschiedliche Autonomiestufen erzeugen unterschiedliche Kostenstrukturen. Die meisten Business Cases bepreisen nur die Technologie — deshalb wirken Piloten günstig und Scale-ups sprengen das Budget.

## MANAGEMENT SUMMARY

# Unternehmen kaufen Autonomie. Sie bepreisen sie nicht.

Fünf Erkenntnisse, die Vorstände abwägen sollten, bevor sie das nächste KI-Skalierungsbudget genehmigen.

**01****KI-Ausgaben verdoppeln sich 2026 — aber 70 % des Werts entstehen außerhalb der Algorithmen.**

KI-Ausgaben steigen 2026 von **0,8 % auf 1,7 % des Umsatzes** (BCG AI Radar, 2.400 Führungskräfte). 70 % des KI-Werts entstehen aus Menschen und Prozessen, nur 10 % aus den Algorithmen selbst — doch die Organisationsinvestition zieht selten im Gleichschritt mit.

**02****Vier Autonomiestufen erzeugen vier verschiedene Kostenstrukturen.**

Stufe 1 (Human-in-the-Loop), Stufe 2 (KI entscheidet, Mensch prüft), Stufe 3 (KI entscheidet, Mensch wird informiert) und Stufe 4 (Human-in-the-Exception). Jede erzwingt eine andere Governance-, Monitoring- und Kompetenzlast. Piloten auf Stufe 1 bepreisen nicht das Skalieren auf Stufe 3.

**03****Stufe 2 ist die Falle — und der häufigste Ort für KI-Vorfälle.**

Stufe 2 sieht günstig aus, weil ein Mensch formal beteiligt ist. Tatsächlich finanziert die Organisation drei parallele Systeme: KI-Ausführung, menschliche Prüfung und das Interface dazwischen. Ohne feste Prüfquote, Ausnahmebehandlung und Übersteuerungsbefugnis wird die Aufsicht dekorativ.

**04****EU AI Act Artikel 14 misst Aufsicht an Kompetenz, Ausbildung und Befugnis.**

Eine Warteschlange markierter Fälle, geprüft von Personen ohne Ablehnungsbefugnis, erfüllt die Anforderung nicht. **62 %** der Führungskräfte sagen, KI beeinflusst bereits die Mehrheit ihrer Entscheidungen — nur **5 %** berichten von echtem Fortschritt bei der Governance dieser Entscheidungen (Deloitte, 2025).

**05****Drei Entscheidungen machen aus Autonomiekosten eine echte Zeile im Business Case.**

Klassifizierung jedes KI-Systems nach Ziel-Autonomiestufe; explizite Bepreisung der Organisationslast je Stufe; Klassifizierungs-Governance auf einer Seite. Keine davon braucht neue Technologie. Alle drei sind heute günstiger zu treffen als nach der ersten Skalierungsinitiative.

## INHALT

# Was Sie erwartet.

---

<b>01</b>	<b>Technologie-Kosten sind nicht gleich Gesamtkosten</b>	<b>05</b>
	Warum die organisatorische Belastung im Business Case fehlt — und wo sie auftaucht.	
<b>02</b>	<b>Vier Autonomiestufen, vier Kostenprofile</b>	<b>08</b>
	Das HandsOn AI Operating Model und die organisatorische Last je Stufe.	
<b>03</b>	<b>Die Stufe-2-Falle</b>	<b>11</b>
	Warum die Stufe, die wie der „logische nächste Schritt“ wirkt, oft die teuerste ist.	
<b>04</b>	<b>Drei Entscheidungen, ein Quartal</b>	<b>14</b>
	Wie Vorstände Autonomie bepreisen, bevor sie sie kaufen.	
<b>05</b>	<b>Über den Autor &amp; Quellen</b>	<b>17</b>
	Methodik, Quellen und wie Sie HandsOn erreichen.	

---

KAPITEL

# 01

## **Technologie-Kosten sind nicht gleich Gesamtkosten.**

Der teuerste KI-Fehler in deutschen Unternehmen liegt derzeit in der Budgetierung. Unternehmen kaufen Autonomie, ohne sie realistisch zu bepreisen.

## Wenn ein CFO einen KI-Business-Case unterschreibt, fehlen Kostenbausteine.

**D**er teuerste KI-Fehler in deutschen Unternehmen liegt derzeit in der Budgetierung. Unternehmen kaufen Autonomie, ohne sie realistisch zu bepreisen. Die Technologie-Rechnung zeigt meistens nur eine Kostendimension. Das Organisationssystem, das die neue Autonomie tragen muss — Governance, Monitoring, Verantwortung, Kompetenz — taucht im Business Case selten auf.

Die BCG-AI-Radar-Studie 2026, basierend auf einer Befragung von 2.400 Führungskräften, liefert die Zahlen. KI-Ausgaben sollen 2026 von 0,8 % auf **1,7 % des Umsatzes** steigen — eine Verdopplung über den Status quo. In demselben Datensatz wird gezeigt, dass **70 % des KI-Werts** aus Menschen und Prozessen und nur 10 % aus den Algorithmen selbst entstehen.

Wenn sich das Budget verdoppelt und sieben Zehntel der Wertschöpfung in der Organisation liegen, müssten die Organisationsinvestitionen theoretisch im Gleichschritt mitziehen. In den Unternehmen, mit denen wir arbeiten, tut sie das selten. Wir nennen diese Lücke die „*Cost of Autonomy*“ — und sie wird selten ernsthaft betrachtet.

Wenn ein CFO heute einen KI-Business-Case unterschreibt, enthält das Modell üblicherweise drei Zeilen: Lizenzen, Implementierung und ein großzügiges Einsparungspotenzial, das die Initiative auf dem eigentlichen Prozess verspricht. Was fehlt, sind die strukturellen Kosten, die entstehen, sobald das System live ist und die Autonomie tatsächlich übernimmt.

Diese Kosten skalieren nicht-linear mit der Autonomiestufe, auf der das System läuft. Ein System, das ausschließlich Empfehlungen ausspricht, ist einfach zu steuern. Ein System, das durchgehend innerhalb einer Policy operiert, ist teuer zu steuern — weil die Organisation eine Monitoring-, Eskalations- und Rekalibrierungsarchitektur vorhalten muss.

### DEFINITION · COST OF AUTONOMY

**Die organisatorischen Kosten, die entstehen, sobald ein KI-System live ist und Autonomie tatsächlich übernimmt.**

Governance-Gremien, benannte Verantwortung, Monitoring-Kadenz, Übersteuerungsbefugnis, Kompetenzaufbau, Prüf-FTEs, Eskalationspfade. Diese Posten skalieren nicht-linear mit der Autonomiestufe — und stehen meistens nicht im initialen Budget.

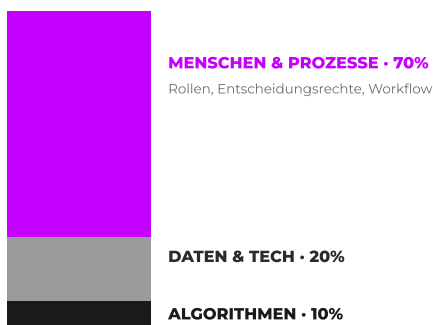
## Sieben Zehntel der Wertschöpfung liegen in der Organisation.

Die BCG-Daten sind eindeutig: Die Algorithmen sind *nicht* der Engpass. Die Engpass-Domäne ist organisatorisch — die Workflows, Rollen, Entscheidungsrechte und Kompetenzen, die KI in Wert übersetzen. Der Business Case muss in dieselbe Richtung zeigen wie die Wertschöpfung.

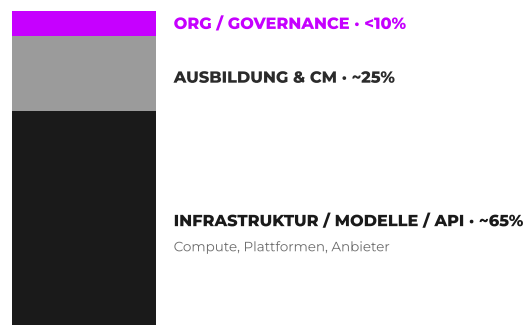
### Abbildung 01 · Wo der KI-Wert entsteht — und wohin das Budget fließt.

Zwei Säulen sollten übereinstimmen. Sie tun es nicht.

QUELLE DES KI-WERTS



TYPISCHE BUDGETVERTEILUNG



QUELLE · BCG AI RADAR 2026 · HANDSON-ANALYSE ZUR BUDGETSTRUKTUR IM EU-MITTELSTAND

**„Die menschliche Aufsicht wird einer natürlichen Person übertragen, die über die erforderliche Kompetenz, Ausbildung und Befugnis verfügt, um die Aufsicht über das Hochrisiko-KI-System wirksam auszuüben.“**

ARTIKEL 14 · EU AI ACT · VERORDNUNG (EU) 2024/1689

## Das Fehlermuster ist vorhersehbar — und teuer.

Ein Pilot als Human-In-The Loop Set Up liefert einen schmeichelhaft günstigen Proof Point. Der Business Case für die Skalierung übernimmt im Folgenden die Kostenstruktur des Piloten. Dann wird das System auf eine höhere Autonomiestufe gehoben — weil dort der tatsächliche ROI liegt — und die organisatorische Last kommt überraschend. Als Vorfälle, gestoppte Rollouts, regulatorische Exposition.

**„Unternehmen kaufen Autonomie. Das Organisationssystem, das die Autonomie tragen muss, taucht im Business Case aber selten auf.“**

KAPITEL 01 · DIE AUTONOMIEKOSTEN

Artikel 14 EU AI Act ist sehr klar: Die menschliche Aufsicht muss einer Person zugewiesen sein, die über die erforderliche Kompetenz, Ausbildung und Befugnis verfügt, ein Hochrisiko-KI-System zu übersteuern oder zu stoppen. Diese Befugnis kostet FTEs, spezifisches Rollendesign, Ausbildungsstunden, Governance-Gremien und klare Berichtslinien. Meistens stehen keine dieser Posten im initialen Budget.

### DIE DIAGNOSE-FRAGE

**Auf welcher Autonomiestufe ist Ihr Pilot bepreist — und auf welcher soll er skalieren?**

Wenn die Antworten unterschiedlich sind, fehlt im Business Case die organisatorische Last für die höhere Stufe. Das ist der häufigste Grund, warum KI-Programme „auf dem Papier“ funktionieren und in der Produktion scheitern. Um diese Frage vorab zu beantworten, braucht das Unternehmen ein Vokabular für Autonomie.

## KAPITEL

# 02

## **Vier Autonomiestufen, vier Kostenprofile.**

Das HandsOn AI Operating Model definiert das Human-AI Interface als das zentrale Designobjekt einer KI-gestützten Organisation. Es drückt sich in einer von vier Autonomiestufen aus — und jede erzwingt eine andere Kostenstruktur.

# Jeder KI-gestützte Entscheidungstyp läuft auf genau einer Stufe.

Das Human-AI Interface ist die organisatorische Architektur für Entscheidungsfindung und Verantwortung, wenn Menschen und KI gemeinsam Verantwortung tragen. Es drückt sich in vier Autonomiestufen aus.



## STUFE 01

### Human-in-the-Loop.

KI empfiehlt; ein Mensch gibt jeden Output frei. Per Design sicher. Governance-Last gering, Kompetenzlast moderat, Monitoring vernachlässigbar. In hochvolumigen Prozessen nicht skalierbar — und genau deshalb selten der Endzustand eines Business Case.

**Kostentreiber · Prüfkapazität pro Output, Trainingszeit der Reviewer.**

## STUFE 02

### KI entscheidet, Mensch prüft.

KI führt innerhalb eines definierten Rahmens aus; ein Mensch prüft Stichproben und bearbeitet Ausnahmen. Wirkt auf dem Papier am günstigsten. Trägt die **höchsten versteckten Kosten** im Modell — weil die Organisation drei parallele Systeme gleichzeitig finanziert.

**Kostentreiber · Prüfquote, Ausnahmebehandlung, Übersteuerungsbefugnis — alles drei muss dokumentiert sein.**

## STUFE 03

### KI entscheidet, Mensch wird informiert.

KI läuft durchgehend innerhalb dokumentierter Policy-Grenzen. Menschen überwachen auf Systemebene. Benannte Verantwortung — KI-Owner, KI-Steward, Übersteuerungsbefugnis — wird tragend. Hier liegt der tatsächliche ROI vieler Programme.

**Kostentreiber · Drift-Tooling, Monitoring-Dashboards, Eskalationspfade, Rekalibrierungsauslöser.**

## STUFE 04

### Human-in-the-Exception.

KI orchestriert mehrstufige Workflows. Menschen setzen Ziele und greifen nur bei Ausnahmen ein. Schwerste Governance-Last. EU-AI-Act-Dokumentationspflichten skalieren hier. Ohne explizit ausgestaltete Ausnahme-Engineering kein produktives System.

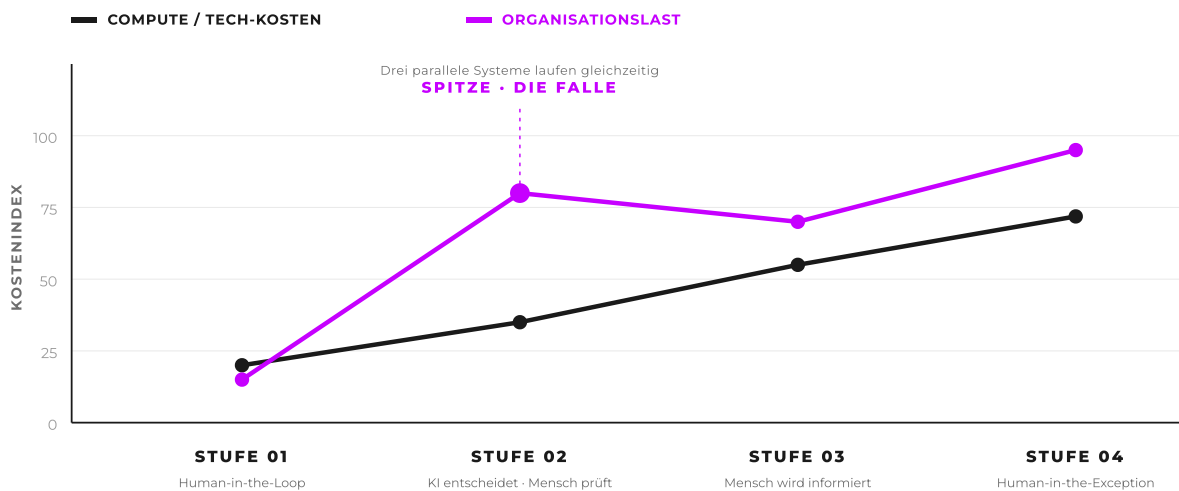
**Kostentreiber · Ausnahme-Engineering, Audit-Dokumentation, Compliance-Reporting.**

## Die Organisationslast skaliert nicht-linear — und Stufe 2 ist die Spitze.

Stufe 1, 3 und 4 erzwingen Klarheit. Auf Stufe 1 gibt es keine Autonomie zu steuern. Stufe 3 und 4 sind offensichtlich autonom — ohne Stichprobengrößen, Monitoring-Schwellen, Kill-Switch und Eskalationspfad geht kein produktives System live. Auf Stufe 2 läuft die Organisation Gefahr, alle drei Posten zu übersehen.

### Abbildung 02 · Autonomiekosten je Stufe.

Indizierte Kosten pro KI-gestütztem Entscheidungstyp je Autonomiestufe. Compute (grau) vs. Organisationslast (lila).



QUELLE · HANDSON-ENGAGEMENT-DATEN · INDIZIERTE KOSTEN JE KI-GESTÜTZTEM ENTSCHEIDUNGSTYP

Der Fehler, den die meisten Organisationen machen: Sie bauen die *technologische Infrastruktur* für Stufe 3 und die *Governance-Infrastruktur* für Stufe 1 — und wundern sich dann, warum das nicht die gewünschten Ergebnisse liefert.

# 03

## **Die Stufe, die wie der logische nächste Schritt aussieht, ist die teuerste.**

Stufe 2 sieht aus wie ein Human-in-the-Loop-Modell, weil formal ein Mensch beteiligt ist. In der Realität ist das Modell aber nur wirksam, wenn drei Voraussetzungen dokumentiert und eingehalten werden — sonst wird die Prüfung zum reinen Abnicken.

## Auf Stufe 2 bezahlt eine Organisation drei parallele Systeme gleichzeitig.

Stufe 2 ist das Risiko im Business Case. Sie sieht aus wie ein Human-in-the-Loop-Modell, weil formal ein Mensch beteiligt ist. Tatsächlich aber finanziert die Organisation die KI-Ausführungsschicht, die menschliche Prüfschicht und das Interface, das beide verbindet. Die Stichprobengröße ist meist nicht definiert, der Umgang mit Ausnahmen nicht ausgestaltet, die Übersteuerungsbefugnis nicht zugewiesen.

### Abbildung 03 · Die Stufe-2-Lücke.

Führungskräfte, die KI-Einfluss auf Entscheidungen melden, vs. Führungskräfte, die Fortschritt bei der Entscheidungs-Governance melden.



57 Punkte Abstand zwischen KI-geprägten Entscheidungen und der Governance dafür.

QUELLE · DELOITTE 2025 GLOBAL HUMAN CAPITAL TRENDS · EXECUTIVE-KOHORTE

#### DEFINITION · DEKORATIVE AUFSICHT

**Eine Warteschlange markierter Fälle, geprüft von Personen, denen Zeit, Fähigkeiten und Mandat fehlen.**

Artikel 14 EU AI Act behandelt dekorative Prüfung nicht als real. Kompetenz, Ausbildung und Befugnis sind der Maßstab, an dem die Aufsichtsbehörde messen wird — und eine Warteschlange ohne Person mit Ablehnungsbefugnis erfüllt ihn nicht. Genau dort entstehen die meisten KI-Vorfälle im Mittelstand: KI, die nominell von Menschen geprüft wird, denen aber Zeit, Fähigkeiten und Mandat fehlen.

Die Lücke zwischen den 62 % und den 5 % liegt fast vollständig auf Stufe 2 — der Zone, die Organisationen für sicher halten, weil jemand formal den Output prüft.

## Zwei Variablen entscheiden, ob die Aufsicht durch den Human in the Loop echt ist.

Beides sind organisatorische Variablen: die Befugnis und Ressourcen der prüfenden Person und der Umgang mit Ausnahmen.

Befugnis heißt Entscheidungsrechte. Wenn die prüfende Person das Modell nicht übersteuern, kein Retraining anfordern oder das System nicht aussetzen kann, wird die Prüfschleife zu einer Durchsatzfunktion, die Aufsicht nur simuliert.

Das HandsOn Entscheidungsrechte-Register behandelt das als Designartefakt: Für jeden KI-gestützten Entscheidungstyp gibt es eine benannte Autorität, einen Evidenzstandard und einen Rekalibrierungsauslöser. Ein Unternehmen, das dieses Register für seine fünf wichtigsten KI-Anwendungsfälle nicht vorlegen kann, betreibt auf Stufe 2 eine Warteschlange mit Prüfeticket, keine belastbare Aufsicht.

Der Umgang mit Ausnahmen ist die zweite Variable. Ein Prüfgate, das 1 % der Fälle markiert und für 100 % Prüfung ausgestattet ist, ist teuer, aber funktional. Ein Prüfgate, das 20 % der Fälle markiert und für 5 % Prüfung ausgestattet ist, ist eine Compliance-Exposition — denn 15 % der Outputs laufen ungeprüft durch, während die Organisation sich erzählt, ein Mensch sei im Prozess involviert.

McKinsey-Daten zum State of AI zeigen: Rund 80 % der KI-einsetzenden Organisationen haben keinen einzigen Workflow um ihre KI-Einführung neu gestaltet. Das ist eine direkte Formulierung desselben Befunds: Die Prüfarchitektur existiert in Folien, nicht im Operating Model. Solange das so bleibt, tauchen die Autonomiekosten als Vorfälle, Nacharbeit und regulatorische Exposition auf — und nicht als Zeile im Business Case.

### Abbildung 04 · Drei Prüfarchitekturen.

100 Outputs pro Szenario. Markiert (lila) vs. tatsächlich geprüft (dunkel) vs. ungeprüft durchgelaufen (grau).

#### SZENARIO A · FUNKTIONAL

Auslöser 1 % · Kapazität 100 % — teuer pro Fall, aber belastbar.



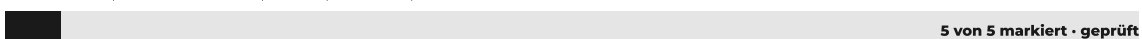
#### SZENARIO B · COMPLIANCE-EXPOSITION

Auslöser 20 % · Kapazität 5 % — 15 % der Outputs laufen ungeprüft durch.



#### SZENARIO C · KALIBRIERT

Auslöser 5 % · Kapazität 5 % — Auslöserquote entspricht Prüfkapazität.



QUELLE · HANDSON-ENGAGEMENT-DATEN · ILLUSTRATIVES MODELL ZUR AUSLÖSER-/PRÜFKAPAZITÄTSLÜCKE

## KAPITEL

# 04

## **Wie ihr Autonomie bepreist, bevor ihr sie kauft.**

Drei Entscheidungen machen aus den Autonomiekosten eine echte Zeile im Business Case und vermeiden nachträgliche Überraschungen. Keine davon braucht neue Tools. Alle drei sind in einem Quartal umsetzbar.

---

## Drei Entscheidungen, ein Quartal — und keine neuen Tools.

Die praktische Frage ist, was ein COO oder CFO tun kann, um die notwendigen Anpassungen vorzunehmen. Drei Entscheidungen liegen in der Entscheidungsmacht jedes Vorstands oder Executive Committees und können noch in diesem Quartal umgesetzt werden.

---

### ENTSCHEIDUNG 01

#### Jedes KI-System nach Ziel-Autonomiestufe klassifizieren.

Nicht die Stufe, auf der es heute läuft. Die Stufe, die der Business Case *braucht*. Ein Forecast-System mit menschlicher Freigabe pro Output ist Stufe 1; wenn der ROI davon abhängt, 10.000 Outputs pro Woche freizugeben, ist das reale Ziel Stufe 2 oder 3. Allein diese Übung fördert **5-15 Systeme** pro mittelständischem Unternehmen zutage, die auf einer Stufe geplant und auf einer anderen budgetiert sind.

---

### ENTSCHEIDUNG 02

#### Die Organisationslast je Stufe explizit bepreisen.

Für jede Zielstufe Governance-, Monitoring-, Kompetenz- und Verantwortungskosten als *explizite Zeilen* schätzen. Prüf-Köpfe auf Stufe 2, Dashboard- und Drift-Tooling auf Stufe 3, Ausnahme-Engineering auf Stufe 4. Ausbildungsprogramme, kalibriert auf die Zielstufe — nicht generische KI-Literacy. Die Übung zwingt das Gespräch aus dem Compute-Budget heraus und in das Betriebskostenbudget, wo die echten Kosten liegen.

---

### ENTSCHEIDUNG 03

#### Klassifizierungs-Governance installieren.

Wer ist berechtigt, ein System von einer Autonomiestufe in die nächste zu heben, und unter welchem Evidenzstandard? Der am häufigsten übersprungene Schritt — und der Punkt, an dem sich die meiste regulatorische Exposition aufbaut. Eine Seite reicht meist — aber sie muss eine **Autorität**, eine **Evidenzschwelle** und eine **Kadenz** benennen.

## Autonomiekosten sind planbar — keine Überraschung.

Die Autonomiekosten entscheiden, ob euer KI-Portfolio den Business Case einlöst oder ihn sprengt. Vier Autonomiestufen; vier organisatorische Kostenstrukturen. Ein Pilot, der auf Stufe 1 bepreist und auf Stufe 2 oder 3 skaliert wird, ohne das Kostenmodell neu zu bauen, ist die häufigste Ursache für KI-Business-Cases, die auf dem Papier stark aussehen und in der Produktion scheitern.

Wer ein KI-Portfolio verantwortet: Nehmt die fünf größten Initiativen, klassifiziert jede nach Ziel-Autonomiestufe und bittet euren CFO, die Organisationslast je Stufe zu bewerten.

Wer auf Vorstands- oder Aufsichtsratsebene sitzt: Bringt drei Entscheidungen auf die Agenda — Klassifizierung, Kosten der Organisationslast, Klassifizierungs-Governance.

Keine davon braucht neue Technologie. Alle drei sind *jetzt* günstiger zu treffen als nach der ersten Implementierungsinitiative.

**„Die Autonomiestufe, die wie der logische nächste Schritt wirkt, ist oft die teuerste — wenn sie nicht ernsthaft umgesetzt wird.“**

SCHLUSSPLÄDOYER · DIE AUTONOMIEKOSTEN

### ENGAGEMENT

**Wenn Sie KI-Entscheidungsrechte klassifizieren oder die Autonomiekosten für Ihr Portfolio bepreisen, melden Sie sich direkt.**

HandsOn arbeitet mit Vorständen und Executive Committees genau an diesem Übergang — vom „Autonomie kaufen“ zum „Autonomie bepreisen“: Ziel-Autonomiestufen klassifizieren, das Entscheidungsrechte-Register aufbauen und die Governance-Kadenz definieren, die EU-AI-Act-konform ist und operativ trägt.

## ÜBER DEN AUTOR

### HANDSON

# Maximilian Stein.

Founder & Managing Partner, HandsOn

Maximilian Stein berät Vorstände und Executive Committees beim Operating-Model-Redesign, das erforderlich ist, um KI-Investitionen in messbare EBIT-Wirkung zu übersetzen. Seine Arbeit fokussiert sich auf die strukturellen Entscheidungen — Entscheidungsarchitektur, Workflow-Design, Kompetenzaufbau und System-Governance — die zwischen einem Pilot und einem skalierten KI-Portfolio liegen.

HandsOn ist eine unabhängige Beratung, die mit europäischen Mittelstands- und Enterprise-Führungsteams am organisatorischen Design für KI arbeitet. Die Insight Series veröffentlicht originale Forschung und feldgeprüfte Frameworks zu Strategie, Struktur und Governance von KI in großen Organisationen.

## QUELLEN

- **BCG** · AI Radar 2026, n = 2.400 Führungskräfte.
- **McKinsey** · The State of AI 2025 (QuantumBlack).
- **Deloitte** · Human Capital Trends, 2025.
- **EU AI Act** · Verordnung (EU) 2024/1689, Artikel 14.
- **HandsOn** · AI Operating Model · Entscheidungsrechte-Register.
- **HandsOn** · Insight Series · Pillar 03 · KI-Governance.

### KONTAKT

info@wearehandson.de  
wearehandson.de

### SERIE

HandsOn Insight  
Pillar 03 · KI-Governance

### AUSGABE

Vol. 01 / Nr. 05  
April 2026

© 2026 HandsOn. Alle Rechte vorbehalten. Diese Publikation enthält ausschließlich allgemeine Informationen. HandsOn erbringt mit dieser Publikation keine geschäftliche, finanzielle, investitionsbezogene, rechtliche, steuerliche oder sonstige professionelle Beratung. Statistiken stammen aus den oben zitierten öffentlich verfügbaren Quellen; Zahlen behalten die Rechte der ursprünglichen Herausgeber. Auszug Artikel 14 EU AI Act · Verordnung (EU) 2024/1689.