RM

REZA MONSHIZADEH

AI · TECHNOLOGY · GOVERNANCE

BOARD BRIEFING REPORT

EUROPÄISCHES BANKING & FINANZMÄRKTE

# Wo darf KI ohne menschliche Freigabe handeln?

Ein Board Briefing Report zu Agentic AI, Human Oversight, Entscheidungsrechten und Kontrollgrenzen im europäischen Banking.

**Reza Monshizadeh** · Monshizadeh.com · 7 Min. Lesezeit

# 01 SITUATION · EIN HYPOTHETISCHES BOARD-SZENARIO

RM

KI bewegt sich von Empfehlung zu Handlung.
Es geht nicht mehr nur um den Modell-Output.
**Es geht um die Handlungsgrenze.**

## NEHMEN WIR FOLGENDE SITUATION AN

Ein KI-gestützter Workflow erstellt nicht mehr nur Texte, Zusammenfassungen oder Empfehlungen. Er kann einen Prozessschritt auslösen, einen Datensatz aktualisieren, einen Alert eskalieren, eine Kundenkommunikation entwerfen, einen Fall priorisieren, eine Kontrollreaktion empfehlen oder eine Handlung zur Ausführung vorbereiten.

*Dieses Szenario ist generisch und hypothetisch. Es beschreibt kein konkretes Institut.*

## DIE ANSCHLIESSENDE FRAGE

Wo darf KI handeln?

Wo muss ein Mensch freigeben?

Wo ist die Grenze dokumentiert, überwacht und widerrufbar?

# 02 MANAGEMENT SUMMARY

RM

## Drei Orientierungspunkte für den Vorstand

01

### Agentic AI verändert die Governance-Frage

Klassische KI-Governance fragt, ob ein Modell akkurat, erklärbar, überwacht und kontrolliert ist. Agentic AI ergänzt eine weitere Frage: Was darf die KI tun?

02

### Fähigkeit ist keine Erlaubnis

Ein Modell kann technisch in der Lage sein, autonom zu handeln. Das entscheidet aber nicht, wo es handeln darf. Erlaubnis muss sich nach Materialität, Reversibilität, Kundenauswirkung, rechtlicher Wirkung und Risikoappetit richten.

03

### Human Oversight muss gestaltet werden

Eine starke Antwort lautet nicht: „Ein Mensch ist im Prozess.“ Eine starke Antwort zeigt, wo KI vorschlägt, wo KI nur nach Freigabe ausführt, wo begrenzte Autonomie erlaubt ist und wo autonome Handlung nicht zulässig ist.

Quelle: EU AI Act; GDPR Artikel 22; ESMA AI Statement; IOSCO AI Toolkit.

# 03 KALIBRIERUNG DER ANTWORTQUALITÄT

RM

## Kontrollgrenze, nicht technische Fähigkeit

### GRÜN — ENTSCHEIDUNGSFÄHIG

- Ein aktuelles KI-Handlungsinventar existiert.
- Jeder KI-gestützte Workflow hat ein definiertes Autonomieniveau.
- Menschliche Freigabepunkte sind nach Rolle und Autorität gemappt.
- Tool-Zugriff, API-Zugriff und Workflow-Berechtigungen sind kontrolliert.
- Handlungen werden mit Akteur, Autorisierung, Zeitstempel und Ergebnis geloggt.
- Exception-, Override- und Stop-Mechanismen werden überwacht.
- Board Reporting unterscheidet assistive, advisory und action-triggering AI.

### GELB — TEILWEISE BELEGT

- KI-Use-Cases sind inventarisiert, aber Handlungsrechte sind nicht vollständig gemappt.
- Human Review existiert, ist aber uneinheitlich belegt.
- Technische Berechtigungen existieren, sind aber nicht mit Business-Entscheidungsrechten verbunden.
- Exception Logs existieren, werden aber nicht auf Governance-Ebene überprüft.
- Business Ownership existiert, aber nicht für Autonomiegrenzen.
- Stop-Mechanismen existieren im Design, wurden aber nicht getestet.

### ROT — NICHT ENTSCHEIDUNGSFÄHIG

- *“It only recommends.”*
- *“A human is somewhere in the loop.”*
- *“The workflow is internal.”*
- *“The user can override it.”*
- *“The model has no final decision right.”*
- *“The vendor controls the permissions.”*
- *“We will define human approval during rollout.”*
- *“This is only a productivity tool.”*

GRÜN ist kein Human-in-the-Loop-Label. **GRÜN ist eine belegte Kontrollgrenze.**

# 04 WER ANTWORTEN MUSS

RM

Technology kann Autonomie ermöglichen. Das Business verantwortet die Handlungswirkung. **Das Board muss die Grenze sehen.**

| | |
|---|---|
| **CIO / CTO** | Architektur, Workflow-Integration, Tool-Zugriff, API-Berechtigungen, Systemgrenzen und Logging. |
| **Business Owner** | Entscheidungswirkung, Workflow-Zweck, Materialität, Reversibilität und operative Konsequenzen. |
| **CRO / Operational Risk** | Risikoklassifizierung, Toleranz für autonome Handlung, Residual Risk und Eskalationsschwellen. |
| **Compliance / Legal / DPO** | AI Act, GDPR, Conduct, Datenschutz, automatisierte Entscheidungen und regulatorische Auslegung. |
| **CISO / Technology Risk** | Identity, Access, Autorisierung, Cyber-Risiko, Missbrauchsszenarien und technische Stop-Mechanismen. |
| **COO** | Operative Ausführung, Kontrollübergaben, Prozessdesign und Auswirkung auf Service Continuity. |
| **Internal Audit** | Unabhängige Prüfung von Handlungsgrenzen, Logs, Freigaben, Stop-Mechanismen und Governance-Evidenz. |

# 05 EVIDENZ, DIE DER VORSTAND ANFORDERN SOLLTE

RM

## Zehn Artefakte, die KI-Handlung steuerbar machen

**01 AI Action Inventory**
Zeigt, welche KI-gestützten Workflows nur vorschlagen, welche Handlungen auslösen und welche innerhalb definierter Grenzen handeln dürfen.

**02 Autonomy Level Classification**
Zeigt, ob der Use Case assistive, advisory, approval-based, bounded autonomous oder monitored autonomous ist.

**03 Action Boundary Map**
Zeigt, was KI tun darf, was sie nicht tun darf und was an der Grenze passiert.

**04 Human Approval Matrix**
Zeigt, welche Handlung eine Freigabe benötigt, welche Rolle freigibt und ob die Freigabe meaningful ist.

**05 Tool / API Permission Register**
Zeigt, auf welche Systeme, Datensätze, Workflow-Schritte, Kommunikationskanäle oder externen Tools KI zugreifen kann.

**06 Decision Rights Map**
Zeigt, wer KI-Handlungsrechte genehmigen, erweitern, einschränken, pausieren oder beenden darf.

**07 Exception and Override Log**
Zeigt, wo KI-Handlungen blockiert, überschrieben, eskaliert oder korrigiert wurden.

**08 Client / Conduct Impact View**
Zeigt, ob KI-Handlungen Kunden, Beratung, Pricing, Risiko, Servicequalität, Beschwerden oder regulatorische Ergebnisse beeinflussen.

**09 Testing and Simulation Evidence**
Zeigt, was passiert, wenn KI falsch klassifiziert, falsch eskaliert, falsch kommuniziert oder außerhalb der Erwartung handelt.

**10 Stop / Rollback Protocol**
Zeigt, wer die KI-Fähigkeit einschränken, aussetzen oder deaktivieren kann — und ob dieser Mechanismus getestet wurde.

# 06 THE AUTONOMY LADDER

RM

KI-Autonomie ist kein binärer Schalter. **Sie muss nach Materialität und Reversibilität kalibriert werden.**

01

### SUGGEST

KI schlägt vor. Ein Mensch entscheidet und handelt.

Standard für materielle, irreversible oder rechtlich sensible Handlungen.

02

### EXECUTE WITH APPROVAL

KI bereitet eine Handlung vor oder führt sie nur nach ausdrücklicher menschlicher Freigabe aus.

Geeignet, wenn Automatisierung Effizienz schafft, die Entscheidung aber menschlich bleibt.

03

### BOUNDED AUTONOMY

KI handelt eigenständig innerhalb eines engen, vordefinierten Handlungsspielraums.

Geeignet für reversible, hochvolumige und eng überwachte Aktivitäten.

04

### MONITORED AUTONOMY

KI handelt autonom, während Monitoring Exceptions, Anomalien und Grenzverletzungen erkennt.

Nur geeignet für risikoarme, reversible und vollständig geloggte Kontexte mit getesteten Stop-Mechanismen.

**DESIGN RULE** Das Autonomieniveau folgt der Materialität und Reversibilität der Handlung. Es folgt nicht der Fähigkeit des Modells.

# 07 GOVERNANCE CHECK

RM

## Warnsignale der Antwortqualität

Diese Antworten können in einer Vorstandssitzung beruhigend klingen. Sie sind aber noch nicht entscheidungsfähig.

***“It only recommends.”***

Eine Empfehlung kann trotzdem eine regulierte Entscheidung prägen. Entscheidend ist, ob sie Verhalten, Priorisierung oder Ergebnis verändert.

***“A human is in the loop.”***

Menschliche Präsenz reicht nicht aus. Das Board muss wissen, ob der Mensch Zeit, Kompetenz, Autorität und Informationen hat, um meaningful einzugreifen.

***“The workflow is internal.”***

Interne Workflows können trotzdem Risikobewertungen, Kontrollen, Reporting, Kundenergebnisse oder regulatorische Pflichten beeinflussen.

***“The user can override it.”***

Eine Override-Option ist nicht dasselbe wie wirksame Kontrolle. Das Board braucht Override-Frequenz, Review-Evidenz und Eskalationslogik.

***“The model has no final decision right.”***

Das finale Entscheidungsrecht ist nur eine Frage. Triggering, Ranking, Escalating, Suppressing oder Drafting können die Entscheidung trotzdem materiell prägen.

***“The vendor controls the permissions.”***

Technisches Permission Management kann beim Vendor liegen. Accountability für erlaubte Handlung bleibt beim Institut.

***“We will define human approval during rollout.”***

Freigabegrenzen sind Designanforderungen. Sie sollten nicht während des Deployments entdeckt werden.

***“This is only a productivity tool.”***

Ein Productivity Tool kann zum Kontrollthema werden, wenn es regulierte Workflows, Kundendaten, Risikosichten oder Entscheidungsbegründungen berührt.

Keine dieser Aussagen ist notwendigerweise falsch. **Sie sind nur keine ausreichende Evidenz für Board-Level Reliance.**

# 08 PATH TO GREEN

RM

## Sechs Bedingungen für eine belegte Kontrollgrenze

### 01 Handlungsrechte sichtbar

Das Institut kann zeigen, welche KI-gestützten Workflows vorschlagen, auslösen, ausführen, kommunizieren oder Datensätze verändern können.

### 02 Autonomie klassifiziert

Jeder KI-Use-Case erhält ein Autonomieniveau auf Basis von Materialität, Reversibilität und Wirkung.

### 03 Menschliche Freigabe meaningful

Freigabepunkte sind so gestaltet, dass der Mensch die KI-gestützte Handlung verstehen, hinterfragen, überschreiben und stoppen kann.

### 04 Berechtigungen kontrolliert

Tool-Zugriff, API-Zugriff, Workflow-Berechtigungen und Kommunikationsrechte sind genehmigt, begrenzt und überwacht.

### 05 Handlungen rekonstruierbar

Jede autonome oder semi-autonome Handlung kann rekonstruiert werden: was gehandelt hat, unter wessen Autorität, auf welcher Eingabe, mit welchem Ergebnis.

### 06 Stop-Mechanismus getestet

Das Institut kann KI-Handlungsrechte schnell einschränken, pausieren oder beenden — und hat diese Fähigkeit getestet.

**NÄCHSTE FRAGE FÜR DEN VORSTAND**

Welcher KI-gestützte Workflow würde morgen das größte Accountability-Problem erzeugen, wenn er ohne ausdrückliche menschliche Freigabe handelt — und können wir die Grenze belegen, die das verhindert?

## AUSGEWÄHLTE QUELLEN


EU AI Act, insbesondere Artikel 12, 14, 15, 19, 26 und Annex III.

GDPR, insbesondere Artikel 22 zu automatisierten Entscheidungen.

DORA, insbesondere ICT Governance, Logging, Detection und Continuity Requirements.

ESMA, Public Statement on AI and Investment Services, Mai 2024.

IOSCO, Supervisory Toolkit for AI Use in Capital Markets, 2026.

NIST AI Risk Management Framework and Generative AI Profile.


BOARD BRIEFINGS SERIES

# Executive Questions for European Banking & Financial Markets

## Ein Mensch im Prozess ist nicht dasselbe wie menschliche Kontrolle.

Ein Modell kann technisch in der Lage sein, autonom zu handeln. Das entscheidet aber nicht, wo es handeln darf.


www.monshizadeh.com

Reza Monshizadeh · Senior Technology Executive · Frankfurt