RM

REZA MONSHIZADEH

AI · TECHNOLOGY · GOVERNANCE

BOARD BRIEFING REPORT

BANKING EUROPEO & MERCATI FINANZIARI

Dove l’AI è autorizzata ad agire senza approvazione umana?

Un Board Briefing Report su agentic AI, human oversight, decision rights e confini di controllo nel banking europeo.

Reza Monshizadeh · Monshizadeh.com · 7 min di lettura

01 SITUATION · UNO SCENARIO IPOTETICO PER IL BOARD

RM

L’AI passa dalla raccomandazione all’azione.
Il punto non è più solo l’output del modello.
Il punto è il confine d’azione.

ASSUMIAMO LA SEGUENTE SITUAZIONE

Un workflow supportato dall’AI non produce più soltanto testi, sintesi o raccomandazioni. Può attivare uno step di workflow, aggiornare un record, escalare un alert, redigere una comunicazione al cliente, prioritizzare un caso, raccomandare una risposta di controllo o preparare un’azione per l’esecuzione.

Questo scenario è generico e ipotetico. Non descrive alcuna istituzione specifica.

LA DOMANDA CHE NE DERIVA

Dove l’AI è autorizzata ad agire?

Dove serve approvazione umana?

Dove il confine è documentato, monitorato e revocabile?

02 MANAGEMENT SUMMARY

RM

Tre punti di orientamento per il board

01

L’agentic AI cambia la domanda di governance

La governance AI tradizionale chiede se un modello sia accurato, spiegabile, monitorato e controllato. L’agentic AI aggiunge un’altra domanda: cosa è autorizzata a fare l’AI?

02

La capacità non è permesso

Un modello può essere tecnicamente capace di agire autonomamente. Questo non decide dove è autorizzato ad agire. Il permesso deve essere determinato da materialità, reversibilità, impatto cliente, effetto legale e risk appetite.

03

La Human Oversight deve essere progettata

Una risposta forte non è “there is a human in the loop”. Una risposta forte mostra dove l’AI propone, dove esegue solo dopo approvazione, dove è permessa autonomia bounded e dove l’azione autonoma non è consentita.

Source: EU AI Act; GDPR Article 22; ESMA AI Statement; IOSCO AI Toolkit.

03 CALIBRAZIONE DELLA QUALITÀ DELLA RISPOSTA

RM

Confine di controllo, non capacità tecnologica

GREEN — DECISIONALE	AMBER — PARZIALMENTE DIMOSTRATO	RED — NON DECISIONALE
Esiste un inventario aggiornato delle azioni AI.	Gli AI use cases sono inventariati, ma i diritti d'azione non sono pienamente mappati.	*"It only recommends."*
Ogni workflow supportato dall'AI ha un livello di autonomia definito.	La human review esiste, ma è documentata in modo inconsistente.	*"A human is somewhere in the loop."*
I punti di approvazione umana sono mappati per ruolo e autorità.	Le permissions tecniche esistono, ma non sono collegate ai business decision rights.	*"The workflow is internal."*
Tool access, API access e workflow permissions sono controllati.	Gli exception logs esistono, ma non sono rivisti a livello governance.	*"The user can override it."*
Le azioni sono loggate con attore, autorizzazione, timestamp ed esito.	Il business ownership esiste, ma non per i confini di autonomia.	*"The model has no final decision right."*
Exception, override e stop mechanisms sono monitorati.	Gli stop mechanisms esistono in design, ma non sono stati testati.	*"The vendor controls the permissions."*
Il board reporting distingue AI assistive, advisory e action-triggering.		*"We will define human approval during rollout."*
		"This is only a productivity tool."

GREEN is not a human-in-the-loop label. **GREEN is an evidenced control boundary.**

04 CHI DEVE RISPONDERE

RM

Technology può abilitare autonomia. Il Business possiede l’impatto dell’azione. **Il Board deve vedere il confine.**

CIO / CTO	Architettura, integrazione workflow, tool access, API permissions, confini di sistema e logging.
Business Owner	Impatto decisionale, finalità del workflow, materialità, reversibilità e conseguenze operative.
CRO / Operational Risk	Classificazione del rischio, tolleranza per azione autonoma, rischio residuo e soglie di escalation.
Compliance / Legal / DPO	AI Act, GDPR, conduct, protezione dati, decisioning automatizzato e interpretazione regolamentare.
CISO / Technology Risk	Identity, access, autorizzazione, cyber risk, scenari di abuso e stop mechanisms tecnici.
COO	Esecuzione operativa, handover di controllo, redesign dei processi e impatto sulla service continuity.
Internal Audit	Revisione indipendente dei confini d’azione, logs, approvals, stop mechanisms ed evidenza di governance.

05 EVIDENZE CHE IL BOARD DOVREBBE RICHIEDERE

RM

Dieci artefatti che rendono governabile l’azione AI

01 AI Action Inventory
Mostra quali workflow supportati dall’AI solo propongono, quali attivano azioni e quali possono agire entro confini definiti.

02 Autonomy Level Classification
Mostra se il use case è assistive, advisory, approval-based, bounded autonomous o monitored autonomous.

03 Action Boundary Map
Mostra cosa l’AI può fare, cosa non può fare e cosa succede al confine.

04 Human Approval Matrix
Mostra quale azione richiede approvazione, quale ruolo approva e se l’approvazione è meaningful.

05 Tool / API Permission Register
Mostra a quali sistemi, dataset, workflow steps, canali di comunicazione o tools esterni l’AI può accedere.

06 Decision Rights Map
Mostra chi può approvare, espandere, restringere, mettere in pausa o terminare i diritti d’azione dell’AI.

07 Exception and Override Log
Mostra dove azioni AI sono state bloccate, overridden, escalated o corrette.

08 Client / Conduct Impact View
Mostra se le azioni AI influenzano clienti, advice, pricing, risk, qualità del servizio, reclami o risultati regolamentari.

09 Testing and Simulation Evidence
Mostra cosa accade quando l’AI classifica male, escala erroneamente, comunica in modo errato o agisce fuori aspettativa.

10 Stop / Rollback Protocol
Mostra chi può restringere, sospendere o disabilitare la capacità AI — e se quel meccanismo è stato testato.

06 THE AUTONOMY LADDER

RM

L’autonomia AI non è un interruttore binario. **Deve essere calibrata secondo materialità e reversibilità.**

01

SUGGEST

L’AI propone. Un umano decide e agisce.

Default per azioni materiali, irreversibili o legalmente sensibili.

02

EXECUTE WITH APPROVAL

L’AI prepara o esegue un’azione solo dopo approvazione umana esplicita.

Adatta quando l’automazione crea efficienza, ma la decisione resta umana.

03

BOUNDED AUTONOMY

L’AI agisce indipendentemente entro uno spazio d’azione stretto e predefinito.

Adatta per attività reversibili, ad alto volume e strettamente monitorate.

04

MONITORED AUTONOMY

L’AI agisce autonomamente mentre il monitoring rileva exceptions, anomalie e boundary breaches.

Adatta solo per contesti low-risk, reversibili e completamente loggati con stop mechanisms testati.

DESIGN RULE Il livello di autonomia segue la materialità e la reversibilità dell’azione. Non segue la capacità del modello.

07 GOVERNANCE CHECK

RM

Segnali di allerta sulla qualità della risposta

Queste risposte possono sembrare rassicuranti in una riunione del board. Tuttavia non sono ancora decision-grade.

“It only recommends.”

Una raccomandazione può comunque plasmare una decisione regolamentata. La domanda è se cambia comportamento, prioritizzazione o risultato.

“A human is in the loop.”

La presenza umana non basta. Il board deve sapere se l’umano ha tempo, competenza, autorità e informazioni per intervenire in modo meaningful.

“The workflow is internal.”

Workflow interni possono comunque influenzare valutazioni di rischio, controlli, reporting, risultati cliente o obblighi regolamentari.

“The user can override it.”

Un’opzione di override non è lo stesso di controllo effettivo. Il board ha bisogno di override frequency, review evidence e escalation logic.

“The model has no final decision right.”

Il diritto di decisione finale è solo una domanda. Triggering, ranking, escalating, suppressing o drafting possono comunque plasmare materialmente la decisione.

“The vendor controls the permissions.”

La gestione tecnica delle permissions può stare presso il vendor. L’accountability per l’azione permessa resta presso l’istituzione.

“We will define human approval during rollout.”

I confini di approvazione sono requisiti di design. Non dovrebbero essere scoperti durante il deployment.

“This is only a productivity tool.”

Un productivity tool può diventare un tema di controllo quando tocca workflow regolamentati, dati cliente, viste di rischio o razionali decisionali.

Nessuna di queste affermazioni è necessariamente falsa. **Semplicemente, non costituiscono evidence sufficiente per board-level reliance.**

08 PATH TO GREEN

RM

Sei condizioni per un confine di controllo dimostrato

01 Diritti d’azione visibili

L’istituzione può mostrare quali workflow supportati dall’AI possono suggerire, triggerare, eseguire, comunicare o modificare record.

02 Autonomia classificata

Ogni AI use case riceve un livello di autonomia basato su materialità, reversibilità e impatto.

03 Approvazione umana meaningful

I punti di approvazione umana sono progettati affinché l’umano possa comprendere, challenge, override e fermare l’azione supportata dall’AI.

04 Permissions controllate

Tool access, API access, workflow permissions e communication rights sono approvati, limitati e monitorati.

05 Azioni ricostruibili

Ogni azione autonoma o semi-autonoma può essere ricostruita: cosa ha agito, sotto quale autorità, su quale input, con quale risultato.

06 Stop mechanism testato

L’istituzione può restringere, mettere in pausa o terminare rapidamente i diritti d’azione AI — e ha testato tale capacità.

DOMANDA SUCCESSIVA PER IL BOARD

Quale workflow supportato dall’AI creerebbe domani il maggiore problema di accountability se agisse senza approvazione umana esplicita — e possiamo dimostrare il confine che lo impedisce?

FONTI SELEZIONATE

EU AI Act, in particolare Articles 12, 14, 15, 19, 26 e Annex III.

GDPR, in particolare Article 22 sulle decisioni automatizzate.

DORA, in particolare requisiti di ICT governance, logging, detection e continuity.

ESMA, Public Statement on AI and Investment Services, maggio 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

Un umano nel processo non è la stessa cosa del controllo umano.

Un modello può essere tecnicamente capace di agire autonomamente. Questo non decide dove è autorizzato ad agire.

www.monshizadeh.com

Reza Monshizadeh · Senior Technology Executive · Frankfurt