RM

REZA MONSHIZADEH

AI · TECHNOLOGY · GOVERNANCE

BOARD BRIEFING REPORT

BANCA EUROPEA & MERCADOS FINANCIEROS

¿Dónde puede actuar la AI sin aprobación humana?

Un Board Briefing Report sobre agentic AI, human oversight, decision rights y límites de control en la banca europea.

Reza Monshizadeh · Monshizadeh.com · 7 min de lectura

01 SITUATION · UN ESCENARIO HIPOTÉTICO PARA EL BOARD

RM

La AI pasa de la recomendación a la acción.
La cuestión ya no es solo el output del modelo.
La cuestión es el límite de acción.

SUPONGAMOS LA SIGUIENTE SITUACIÓN

Un workflow apoyado por AI ya no produce solamente textos, resúmenes o recomendaciones. Puede activar un paso de workflow, actualizar un registro, escalar una alerta, redactar una comunicación al cliente, priorizar un caso, recomendar una respuesta de control o preparar una acción para ejecución.

Este escenario es genérico e hipotético. No describe ninguna institución específica.

LA PREGUNTA QUE SIGUE

¿Dónde puede actuar la AI?

¿Dónde debe aprobar un humano?

¿Dónde está el límite documentado, monitorizado y revocable?

02 MANAGEMENT SUMMARY

RM

Tres puntos de orientación para el board

01

Agentic AI cambia la pregunta de governance

La governance tradicional de AI pregunta si un modelo es preciso, explicable, monitorizado y controlado. Agentic AI añade otra pregunta: ¿qué está autorizada a hacer la AI?

02

Capacidad no es permiso

Un modelo puede ser técnicamente capaz de actuar de forma autónoma. Eso no decide dónde está autorizado a actuar. El permiso debe determinarse por materialidad, reversibilidad, impacto cliente, efecto legal y risk appetite.

03

Human Oversight debe diseñarse

Una respuesta fuerte no es “there is a human in the loop”. Una respuesta fuerte muestra dónde la AI propone, dónde ejecuta solo tras aprobación, dónde se permite autonomía bounded y dónde la acción autónoma no está permitida.

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

03 CALIBRACIÓN DE LA CALIDAD DE RESPUESTA

RM

Límite de control, no capacidad tecnológica

GREEN — DECISION-GRADE

- Existe un inventario actualizado de acciones AI.
- Cada workflow apoyado por AI tiene un nivel de autonomía definido.
- Los puntos de aprobación humana están mapeados por rol y autoridad.
- Tool access, API access y workflow permissions están controlados.
- Las acciones se registran con actor, autorización, timestamp y resultado.
- Los mecanismos de exception, override y stop están monitorizados.
- El board reporting distingue AI assistive, advisory y action-triggering.

AMBER — PARCIALMENTE DEMOSTRADO

- Los AI use cases están inventariados, pero los derechos de acción no están plenamente mapeados.
- La human review existe, pero está documentada de forma inconsistente.
- Los permisos técnicos existen, pero no están vinculados a business decision rights.
- Los exception logs existen, pero no se revisan a nivel governance.
- El business ownership existe, pero no para los límites de autonomía.
- Los stop mechanisms existen en diseño, pero no han sido probados.

RED — NO DECISION-GRADE

- *“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.”*

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

04 QUIÉN DEBE RESPONDER

RM

Technology puede habilitar autonomía. El Business posee el impacto de la acción. **El Board necesita ver el límite.**

CIO / CTO	Arquitectura, integración de workflow, tool access, API permissions, límites de sistema y logging.
Business Owner	Impacto decisional, finalidad del workflow, materialidad, reversibilidad y consecuencias operativas.
CRO / Operational Risk	Clasificación de riesgo, tolerancia a acción autónoma, riesgo residual y umbrales de escalación.
Compliance / Legal / DPO	AI Act, GDPR, conduct, protección de datos, decisioning automatizado e interpretación regulatoria.
CISO / Technology Risk	Identity, access, autorización, cyber risk, escenarios de uso indebido y stop mechanisms técnicos.
COO	Ejecución operativa, handovers de control, rediseño de procesos e impacto en service continuity.
Internal Audit	Revisión independiente de límites de acción, logs, aprobaciones, stop mechanisms y evidencia de governance.

05 EVIDENCIA QUE EL BOARD DEBERÍA SOLICITAR

RM

Diez artefactos que hacen gobernable la acción AI

01 AI Action Inventory
Muestra qué workflows apoyados por AI solo sugieren, cuáles activan acciones y cuáles pueden actuar dentro de límites definidos.

02 Autonomy Level Classification
Muestra si el use case es assistive, advisory, approval-based, bounded autonomous o monitored autonomous.

03 Action Boundary Map
Muestra qué puede hacer la AI, qué no puede hacer y qué ocurre en el límite.

04 Human Approval Matrix
Muestra qué acción requiere aprobación, qué rol la aprueba y si la aprobación es meaningful.

05 Tool / API Permission Register
Muestra a qué sistemas, datasets, workflow steps, canales de comunicación o tools externos puede acceder la AI.

06 Decision Rights Map
Muestra quién puede aprobar, ampliar, restringir, pausar o terminar derechos de acción de la AI.

07 Exception and Override Log
Muestra dónde acciones AI fueron bloqueadas, overridden, escaladas o corregidas.

08 Client / Conduct Impact View
Muestra si las acciones AI afectan clientes, advice, pricing, risk, calidad del servicio, quejas o resultados regulatorios.

09 Testing and Simulation Evidence
Muestra qué ocurre cuando la AI clasifica mal, escala incorrectamente, comunica incorrectamente o actúa fuera de expectativa.

10 Stop / Rollback Protocol
Muestra quién puede restringir, suspender o deshabilitar la capacidad AI — y si ese mecanismo ha sido probado.

06 THE AUTONOMY LADDER

RM

La autonomía AI no es un interruptor binario. **Debe calibrarse por materialidad y reversibilidad.**

01

SUGGEST

La AI propone. Un humano decide y actúa.

Default para acciones materiales, irreversibles o legalmente sensibles.

02

EXECUTE WITH APPROVAL

La AI prepara o ejecuta una acción solo tras aprobación humana explícita.

Adecuado cuando la automatización crea eficiencia, pero la decisión sigue siendo humana.

03

BOUNDED AUTONOMY

La AI actúa de forma independiente dentro de un espacio de acción estrecho y predefinido.

Adecuado para actividades reversibles, de alto volumen y estrechamente monitorizadas.

04

MONITORED AUTONOMY

La AI actúa autónomamente mientras el monitoring detecta exceptions, anomalías y boundary breaches.

Solo adecuado para contextos low-risk, reversibles y completamente registrados con stop mechanisms probados.

DESIGN RULE El nivel de autonomía sigue la materialidad y reversibilidad de la acción. No sigue la capacidad del modelo.

07 GOVERNANCE CHECK

RM

Señales de alerta sobre la calidad de respuesta

Estas respuestas pueden sonar tranquilizadoras en una reunión del board. Todavía no son decision-grade.

“It only recommends.”

Una recomendación puede igualmente moldear una decisión regulada. La pregunta es si cambia comportamiento, priorización o resultado.

“A human is in the loop.”

La presencia humana no basta. El board debe saber si el humano tiene tiempo, competencia, autoridad e información para intervenir meaningful.

“The workflow is internal.”

Los workflows internos pueden afectar evaluaciones de riesgo, controles, reporting, resultados cliente u obligaciones regulatorias.

“The user can override it.”

Una opción de override no equivale a control efectivo. El board necesita override frequency, review evidence y escalation logic.

“The model has no final decision right.”

El derecho de decisión final es solo una pregunta. Triggering, ranking, escalating, suppressing o drafting pueden moldear materialmente la decisión.

“The vendor controls the permissions.”

La gestión técnica de permisos puede estar en el vendor. La accountability por la acción permitida permanece en la institución.

“We will define human approval during rollout.”

Los límites de aprobación son requisitos de diseño. No deberían descubrirse durante el deployment.

“This is only a productivity tool.”

Un productivity tool puede convertirse en un asunto de control cuando toca workflows regulados, datos de cliente, vistas de riesgo o racionales de decisión.

Ninguna de estas afirmaciones es necesariamente falsa. **Simplemente no constituyen evidencia suficiente para board-level reliance.**

08 PATH TO GREEN

RM

Seis condiciones para un límite de control evidenciado

01 Derechos de acción visibles

La institución puede mostrar qué workflows apoyados por AI pueden sugerir, activar, ejecutar, comunicar o modificar registros.

02 Autonomía clasificada

Cada AI use case recibe un nivel de autonomía basado en materialidad, reversibilidad e impacto.

03 Aprobación humana meaningful

Los puntos de aprobación humana están diseñados para que el humano pueda comprender, cuestionar, override y detener la acción apoyada por AI.

04 Permisos controlados

Tool access, API access, workflow permissions y communication rights están aprobados, limitados y monitorizados.

05 Acciones reconstruibles

Cada acción autónoma o semi-autónoma puede reconstruirse: qué actuó, bajo qué autoridad, sobre qué input, con qué resultado.

06 Stop mechanism probado

La institución puede restringir, pausar o terminar rápidamente derechos de acción AI — y ha probado esa capacidad.

SIGUIENTE PREGUNTA PARA EL BOARD

¿Qué workflow apoyado por AI crearía mañana el mayor problema de accountability si actuara sin aprobación humana explícita — y podemos evidenciar el límite que lo impide?

FUENTES SELECCIONADAS

EU AI Act, especialmente Articles 12, 14, 15, 19, 26 y Annex III.

GDPR, especialmente Article 22 sobre decisiones automatizadas.

DORA, especialmente requisitos de ICT governance, logging, detection y continuity.

ESMA, Public Statement on AI and Investment Services, mayo de 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 humano en el proceso no es lo mismo que control humano.

Un modelo puede ser técnicamente capaz de actuar de forma autónoma. Eso no decide dónde está autorizado a actuar.

www.monshizadeh.com

Reza Monshizadeh · Senior Technology Executive · Frankfurt