RM

REZA MONSHIZADEH

AI · TECHNOLOGY · GOVERNANCE

BOARD BRIEFING REPORT

BANKING EUROPÉEN & MARCHÉS FINANCIERS

Où l’IA est-elle autorisée à agir sans approbation humaine?

Un Board Briefing Report sur l’agentic AI, le human oversight, les decision rights et les frontières de contrôle dans le banking européen.

Reza Monshizadeh · Monshizadeh.com · 7 min de lecture

01 SITUATION · UN SCÉNARIO HYPOTHÉTIQUE DE BOARD

RM

L’IA passe de la recommandation à l’action.
La question n’est plus seulement l’output du modèle.
La question est la frontière d’action.

SUPPOSONS LA SITUATION SUIVANTE

Un workflow soutenu par l’IA ne produit plus seulement des textes, des résumés ou des recommandations. Il peut déclencher une étape de workflow, mettre à jour un enregistrement, escalader une alerte, rédiger une communication client, prioriser un cas, recommander une réponse de contrôle ou préparer une action pour exécution.

Ce scénario est générique et hypothétique. Il ne décrit aucune institution spécifique.

LA QUESTION QUI SUIT

Où l’IA est-elle autorisée à agir?

Où une approbation humaine est-elle nécessaire?

Où la frontière est-elle documentée, monitorée et révocable?

02 MANAGEMENT SUMMARY

RM

Trois points d’orientation pour le board

01

L’agentic AI change la question de gouvernance

La gouvernance traditionnelle de l’IA demande si un modèle est exact, explicable, monitoré et contrôlé. L’agentic AI ajoute une autre question: que l’IA est-elle autorisée à faire?

02

La capacité n’est pas une permission

Un modèle peut être techniquement capable d’agir de façon autonome. Cela ne décide pas où il est autorisé à agir. La permission doit être déterminée par la matérialité, la réversibilité, l’impact client, l’effet juridique et l’appétit au risque.

03

Le Human Oversight doit être conçu

Une réponse forte n’est pas “there is a human in the loop”. Une réponse forte montre où l’IA propose, où elle exécute uniquement après approbation, où une autonomie bornée est permise et où l’action autonome n’est pas autorisée.

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

03 CALIBRATION DE LA QUALITÉ DE RÉPONSE

RM

Frontière de contrôle, pas capacité technologique

GREEN	DÉCISIONNEL
Un inventaire actuel des actions IA existe.	
Chaque workflow soutenu par l’IA a un niveau d’autonomie défini.	
Les points d’approbation humaine sont mappés par rôle et autorité.	
Les accès aux tools, APIs et permissions de workflow sont contrôlés.	
Les actions sont loggées avec acteur, autorisation, horodatage et résultat.	
Les mécanismes d’exception, d’override et d’arrêt sont monitorés.	
Le board reporting distingue l’IA assistive, advisory et action-triggering.	

AMBER	PARTIELLEMENT DÉMONTRÉ
Les AI use cases sont inventoriés, mais les droits d’action ne sont pas entièrement mappés.	
La human review existe, mais elle est documentée de manière inconsistante.	
Les permissions techniques existent, mais ne sont pas reliées aux decision rights business.	
Les exception logs existent, mais ne sont pas revus au niveau gouvernance.	
Le business ownership existe, mais pas pour les frontières d’autonomie.	
Les mécanismes d’arrêt existent en design, mais n’ont pas été testés.	

RED	NON DÉCISIONNEL
“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 QUI DOIT RÉPONDRE

RM

Technology peut permettre l’autonomie. Le Business détient l’impact de l’action. **Le Board doit voir la frontière.**

CIO / CTO	Architecture, intégration workflow, accès tool, permissions API, frontières système et logging.
Business Owner	Impact décisionnel, finalité du workflow, matérialité, réversibilité et conséquences opérationnelles.
CRO / Operational Risk	Classification du risque, tolérance à l’action autonome, risque résiduel et seuils d’escalade.
Compliance / Legal / DPO	AI Act, GDPR, conduct, protection des données, décisions automatisées et interprétation réglementaire.
CISO / Technology Risk	Identity, access, autorisation, cyber risk, scénarios d’abus et mécanismes techniques d’arrêt.
COO	Exécution opérationnelle, transferts de contrôle, redesign des processus et impact sur la continuité de service.
Internal Audit	Revue indépendante des frontières d’action, logs, approbations, mécanismes d’arrêt et evidence de gouvernance.

05 PREUVES QUE LE BOARD DEVRAIT DEMANDER

RM

Dix artefacts qui rendent l’action IA gouvernable

01 AI Action Inventory
Montre quels workflows soutenus par l’IA se limitent à proposer, lesquels déclenchent des actions et lesquels peuvent agir dans des frontières définies.

02 Autonomy Level Classification
Montre si le use case est assistive, advisory, approval-based, bounded autonomous ou monitored autonomous.

03 Action Boundary Map
Montre ce que l’IA peut faire, ce qu’elle ne peut pas faire et ce qui se passe à la frontière.

04 Human Approval Matrix
Montre quelle action requiert une approbation, quel rôle l’approuve et si l’approbation est meaningful.

05 Tool / API Permission Register
Montre à quels systèmes, datasets, étapes de workflow, canaux de communication ou tools externes l’IA peut accéder.

06 Decision Rights Map
Montre qui peut approuver, étendre, restreindre, mettre en pause ou terminer les droits d’action de l’IA.

07 Exception and Override Log
Montre où des actions IA ont été bloquées, overridden, escaladées ou corrigées.

08 Client / Conduct Impact View
Montre si les actions IA affectent clients, advice, pricing, risk, qualité de service, réclamations ou résultats réglementaires.

09 Testing and Simulation Evidence
Montre ce qui se passe lorsque l’IA classifie mal, escalade incorrectement, communique incorrectement ou agit hors attente.

10 Stop / Rollback Protocol
Montre qui peut restreindre, suspendre ou désactiver la capacité IA — et si ce mécanisme a été testé.

06 THE AUTONOMY LADDER

RM

L’autonomie IA n’est pas un interrupteur binaire. **Elle doit être calibrée selon la matérialité et la réversibilité.**

01

SUGGEST

L’IA propose. Un humain décide et agit.

Par défaut pour les actions matérielles, irréversibles ou juridiquement sensibles.

02

EXECUTE WITH APPROVAL

L’IA prépare ou exécute une action uniquement après approbation humaine explicite.

Approprié lorsque l’automatisation crée de l’efficacité, mais que la décision reste humaine.

03

BOUNDED AUTONOMY

L’IA agit indépendamment dans un espace d’action étroit et prédéfini.

Approprié pour des activités réversibles, à volume élevé et fortement monitorées.

04

MONITORED AUTONOMY

L’IA agit de façon autonome tandis que le monitoring détecte exceptions, anomalies et breaches de frontière.

Uniquement approprié pour des contextes low-risk, réversibles et entièrement loggés avec mécanismes d’arrêt testés.

DESIGN RULE Le niveau d’autonomie suit la matérialité et la réversibilité de l’action. Il ne suit pas la capacité du modèle.

07 GOVERNANCE CHECK

RM

Signaux d’alerte sur la qualité de réponse

Ces réponses peuvent sembler rassurantes en réunion de board. Elles ne sont toutefois pas encore décisionnelles.

“It only recommends.”

Une recommandation peut tout de même orienter une décision réglementée. La question est de savoir si elle modifie comportement, priorisation ou résultat.

“A human is in the loop.”

La présence humaine ne suffit pas. Le board doit savoir si l’humain a le temps, la compétence, l’autorité et l’information pour intervenir meaningful.

“The workflow is internal.”

Les workflows internes peuvent tout de même affecter évaluations de risque, contrôles, reporting, résultats clients ou obligations réglementaires.

“The user can override it.”

Une option d’override n’est pas un contrôle effectif. Le board a besoin de fréquence d’override, evidence de review et logique d’escalade.

“The model has no final decision right.”

Le droit de décision final n’est qu’une question. Triggering, ranking, escalating, suppressing ou drafting peuvent tout de même façonner matériellement la décision.

“The vendor controls the permissions.”

La gestion technique des permissions peut être chez le vendor. L’accountability pour l’action autorisée reste auprès de l’institution.

“We will define human approval during rollout.”

Les frontières d’approbation sont des exigences de design. Elles ne devraient pas être découvertes pendant le deployment.

“This is only a productivity tool.”

Un productivity tool peut devenir un sujet de contrôle lorsqu’il touche des workflows réglementés, des données client, des vues de risque ou des justifications décisionnelles.

Aucune de ces déclarations n’est nécessairement fausse. **Elles ne constituent simplement pas une evidence suffisante pour une board-level reliance.**

08 PATH TO GREEN

RM

Six conditions pour une frontière de contrôle démontrée

01 Droits d’action visibles

L’institution peut montrer quels workflows soutenus par l’IA peuvent suggérer, déclencher, exécuter, communiquer ou modifier des enregistrements.

02 Autonomie classifiée

Chaque AI use case reçoit un niveau d’autonomie basé sur matérialité, réversibilité et impact.

03 Approbation humaine meaningful

Les points d’approbation sont conçus pour que l’humain puisse comprendre, challenger, override et stopper l’action soutenue par l’IA.

04 Permissions contrôlées

L’accès tool, l’accès API, les permissions workflow et les droits de communication sont approuvés, limités et monitorés.

05 Actions reconstructibles

Chaque action autonome ou semi-autonome peut être reconstruite: ce qui a agi, sous quelle autorité, sur quelle input, avec quel résultat.

06 Mécanisme d’arrêt testé

L’institution peut restreindre, mettre en pause ou terminer rapidement les droits d’action IA — et a testé cette capacité.

QUESTION SUIVANTE POUR LE BOARD

Quel workflow soutenu par l’IA créerait demain le plus grand problème d’accountability s’il agissait sans approbation humaine explicite — et pouvons-nous démontrer la frontière qui l’empêche?

SOURCES SÉLECTIONNÉES

EU AI Act, notamment Articles 12, 14, 15, 19, 26 et Annex III.

GDPR, notamment Article 22 sur les décisions automatisées.

DORA, notamment exigences d’ICT governance, logging, detection et continuity.

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

Un humain dans le processus n’est pas la même chose qu’un contrôle humain.

Un modèle peut être techniquement capable d’agir de manière autonome. Cela ne décide pas où il est autorisé à agir.

www.monshizadeh.com

Reza Monshizadeh · Senior Technology Executive · Frankfurt