

Implémenter un agent IA en banque et en finance

Méthodologie de décision, de contrôle et de passage à l'échelle

CONVICTION DIRECTRICE Un agent IA bancaire ne se gouverne pas comme un agent conversationnel : la capacité d'agir, la mémoire et les droits d'accès deviennent des objets de produit, de risque et d'audit à part entière.

À destination des directeurs produit, product owners, product managers et directions de business unit

L'IA est déjà matérielle dans la banque européenne ; la question porte désormais sur la gouvernance à l'échelle

Le sujet s'est déplacé de l'expérimentation vers l'industrialisation, et ce déplacement change la nature du problème posé aux équipes produit. Un agent agit au nom de l'institution ou d'un collaborateur, ce qui appelle des contrôles propres à l'action, distincts de ceux qui garantissent la qualité d'une réponse.

SITUATION

L'usage est généralisé

La Banque centrale européenne indique que plus de 85 % des banques qu'elle supervise utilisent l'intelligence artificielle. L'Autorité bancaire européenne observait en 2025 que 55 % des banques interrogées l'utilisaient déjà dans des processus en contact avec la clientèle.

COMPLICATION

Agir engage l'institution

La rupture tient au passage d'une réponse à une action effectuée au nom de l'institution. Les contrôles conçus pour un moteur de réponse ne couvrent ni la mémoire, ni les droits d'accès, ni la capacité d'agir.

RÉOLUTION

Gouverner l'action

Traiter la capacité d'agir, la mémoire et les habilitations comme des objets de produit, de risque et d'audit. La méthode qui suit organise cette gouvernance en six phases et six points de décision.

La fenêtre ouverte par le calendrier européen doit servir à industrialiser la preuve de conformité, et non pour différer le travail.

Huit décisions à prendre, du choix du processus jusqu'à la préparation du retrait

La recommandation d'ensemble tient en une phrase : lancer deux à trois agents bornés sur des processus internes ou semi-opérationnels, en construisant dès le premier pilote les briques de contrôle communes. Une banque optimise un portefeuille, là où une collection de preuves de concept disperse la capacité.

- 1 Concevoir depuis le processus.** Le bon objet de conception est la chaîne de valeur : étapes, exceptions, décisions, données, contrôles et recours.
- 2 Fixer un plafond d'autonomie explicite.** Le niveau maximal se code dans les permissions, jamais dans la consigne donnée au modèle.
- 3 Classer le risque avant le prototype.** Statut au titre du règlement européen, traitement des données, criticité opérationnelle, réversibilité, portée financière.
- 4 Financer une plateforme de contrôle minimale.** Identité propre de l'agent, passerelle d'outils, journal des actions, mémoire cloisonnée, évaluation continue, arrêt d'urgence.
- 5 Exiger des preuves reproductibles.** Jeux d'évaluation représentatifs, tests adverses, scénarios d'incident, exploitation en observation avant tout pouvoir d'action.
- 6 Nommer un propriétaire métier responsable.** Le product owner pilote le produit ; un dirigeant de business unit répond du résultat et du risque accepté.
- 7 Séparer productivité et valeur réalisée.** La valeur se constate après déploiement : adoption, capacité libérée, délai client, qualité et pertes évitées.
- 8 Préparer le retrait dès la conception.** Versionnement, réversibilité, fournisseur alternatif, export des traces, procédures manuelles et arrêt d'urgence.

La quatrième décision conditionne les sept autres. C'est la seule qui ne puisse être rattrapée après coup, parce qu'elle détermine ce que l'on aura le droit de mesurer.

Le calendrier européen fixe les échéances ; la capacité de validation fixe le rythme réel

L'Autorité de contrôle prudentiel et de résolution a inscrit la préparation de la supervision de l'intelligence artificielle à son programme de travail 2026, et confirme l'application au 2 décembre 2027 des règles de haut risque aux systèmes évaluant la solvabilité des personnes physiques.

Echéance	Ce qui change	Implication pour le produit
2 février 2025	Définitions, pratiques interdites et obligation de maîtrise de l'intelligence artificielle	Formation ciblée, inventaire et politique d'usage
2 août 2025	Obligations relatives aux modèles à usage général et gouvernance européenne	Diligence sur les modèles et documentation fournisseur
2 août 2026	Règles de transparence et mise en œuvre de la majorité des dispositions	Information des utilisateurs, contenu synthétique, gouvernance
2 décembre 2027	Règles pour les systèmes à haut risque de l'annexe III	Crédit aux personnes physiques notamment ; conformité complète
2 août 2028	Haut risque intégré aux produits réglementés de l'annexe I	Le cas échéant, filière produit réglementé

Seize mois séparent ce document de l'échéance du 2 décembre 2027. Une conformité conçue après cette date se conçoit sous contrainte, ce qui est la manière la plus coûteuse de la concevoir.

Ces huit décisions se logent dans les cinq leviers de SCALE-6

Le cadre Acadewie SCALE-6 organise la captation de valeur en cinq leviers et une cible. Ce document en approfondit un cas particulier, celui des agents en banque et en assurance : chaque levier renvoie à la partie qui le traite, et à la preuve qui le soutient.

Levier	Ce que le levier demande	Où ce document le traite
S • Sélectionner	Concentrer l’effort sur trois à quatre cas d’usage à fort impact	Partie II : la grille à quatre dimensions et les cinq cas de départ
C • Concentrer	Porter l’essentiel de l’effort sur les personnes et les processus	Partie VII : le coût complet, dont 79,9 % de charge humaine et organisationnelle
A • Accompagner	Former, avec une modalité qui compte autant que le volume	Partie V : le superviseur d’agent, son temps et sa formation
L • Lier	S’appuyer sur des solutions et des partenaires établis	Partie VIII : construire, acheter, codévelopper, et les sept axes de diligence
É • Évaluer	Mesurer en indicateurs de compte de résultat	Partie VII : l’échelle de preuve à trois niveaux et l’équation de valeur
La cible, 6 %	Attribuer au moins 5 % de son résultat d’exploitation à l’IA	Annexe E : le taux d’automatisation qui produit ces points de résultat

SCALE-6 donne la colonne, les agents en donnent le cas particulier. Le cadre établit que 6 % des entreprises captent une valeur significative. Ce document explique comment être dans les 6 % sur le cas précis des agents.



Ce qu'est un agent, et ce qu'il engage

Un agent combine un modèle capable de raisonner, un ensemble d'outils, un état de travail et des règles d'orchestration. La rupture porte sur l'action effectuée au nom de l'institution.

· Définition opérationnelle

· Cinq niveaux d'autonomie

· Autonomie et criticité

· Cinq illusions

Un agent est une délégation sous contrainte, et c'est la contrainte qui se conçoit

DÉFINITION OPÉRATIONNELLE Système logiciel qui reçoit un objectif, choisit une séquence bornée d'étapes, mobilise des données et des outils autorisés, maintient un état ou une mémoire, puis recommande ou exécute des actions avec un niveau de supervision déterminé.

Quatre thèses de valeur, dans l'ordre de solidité décroissante

LA PLUS ÉTABLIE

Compression du temps de traversée

L'agent coordonne la collecte, la vérification, la rédaction et les passages de relais. La valeur vient du cycle réduit bien plus que du temps de frappe économisé.

SOUS CONDITION

Qualité et homogénéité

Politiques, justificatifs et listes de contrôle s'appliquent de façon plus constante, à la condition que les sources et les règles soient versionnées.

EFFET INDIRECT

Capacité de contrôle augmentée

La préconstitution des dossiers, la relecture et la détection d'anomalies rendent du temps aux spécialistes pour les cas complexes.

LA PLUS EXIGEANTE

Nouveaux services

Un service plus proactif et plus personnalisé devient possible, mais les règles de conduite et la protection de la clientèle imposent une validation plus stricte.

Les trois premières thèses portent sur le processus ; la quatrième porte sur l'offre. C'est cette dernière qui se mesure le plus mal et qui se vend le mieux, ce qui justifie de la placer en dernier.

Le niveau d'autonomie se fixe outil par outil, et se code dans les permissions

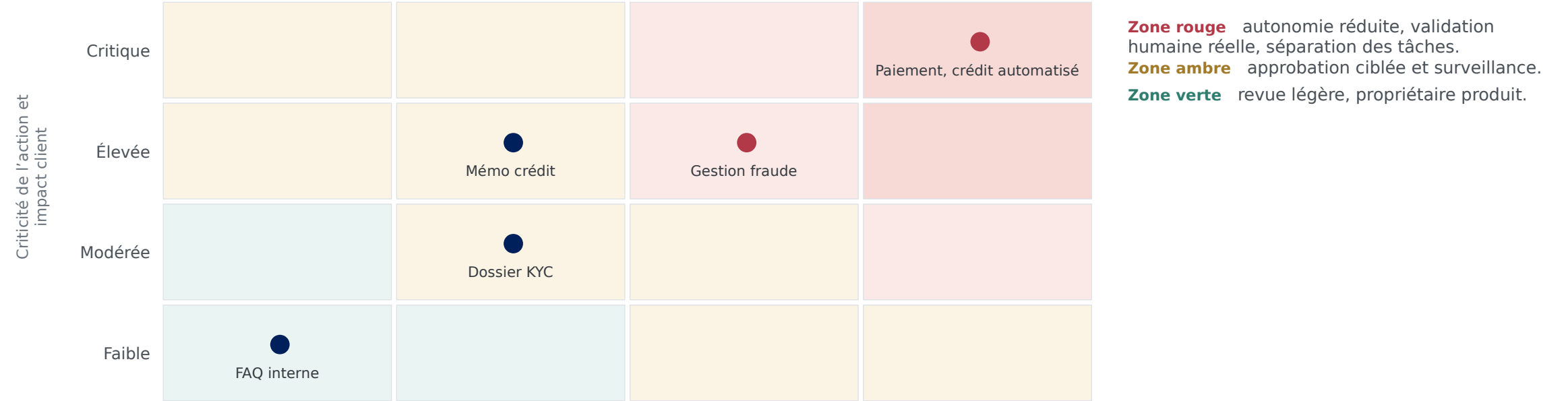
Un même agent peut être limité au conseil sur le crédit, à l'action préparée sur la rédaction d'un courriel, et à l'action bornée sur la création d'une tâche interne. Raisonner par agent conduit à surdimensionner les contrôles là où ils sont inutiles, et à les sous-dimensionner là où ils comptent.

Niveau	Rôle	Capacité	Contrôle dominant
A0	Recherche et génération	Répond, extrait, résume. Aucun outil d'action.	Validation du contenu
A1	Conseil	Produit une recommandation, un score ou un brouillon.	Décideur humain explicite
A2	Action préparée	Prépare une action ; exécution après approbation ciblée.	Aperçu, approbation, séparation des tâches
A3	Action bornée	Exécute des actions réversibles dans une enveloppe prédéfinie.	Moteur de politique, limites, surveillance, retour arrière
A4	Orchestration adaptative	Décompose l'objectif, choisit outils et agents, agit avec latitude.	Exception ; contrôle renforcé et preuve formelle

A4 relève de l'exception et suit une filière d'autorisation distincte. Le niveau maximal se code dans les permissions ; seule cette inscription technique vaut contrôle.

L'autonomie acceptable décroît à mesure que la criticité et l'impact client augmentent

La matrice ci-dessous donne la règle de contrôle par défaut. Elle se lit comme une diagonale : plus un acte engage le client ou l'établissement, plus le niveau d'autonomie admissible est bas, et plus la validation humaine doit être réelle plutôt que formelle.



Le règlement européen classe en haut risque l'évaluation de la solvabilité des personnes physiques, à l'exception de la détection de fraude financière, ce qui n'exonère ni des autres cadres ni d'un contrôle interne robuste.

Cinq points de vigilance qui décident du réalisme d'un projet d'agent

Chacun de ces cinq énoncés est défendable en apparence et faux en pratique. Ils reviennent avec assez de régularité dans les comités pour mériter d’être nommés et réfutés une fois pour toutes.

Le raisonnement	Ce que l'expérience lui oppose
« Le modèle est performant, donc le produit est sûr »	La majorité des défaillances viennent de la chaîne complète : données, outils, règles, interfaces, utilisateurs et changements.
« Nous avons prévu une validation humaine »	Une validation sans temps, sans compétence ni droit réel de contredire peut rester une décision exclusivement automatisée au sens du règlement sur les données personnelles.
« Le retour sur investissement tient au temps économisé »	Sans adoption, reconfiguration des équipes et retrait effectif de coûts, le gain reste virtuel.
« Un agent généraliste connecté à tout ira plus vite »	La polyvalence augmente la surface d'attaque, l'opacité des finalités et la concentration des permissions.
« La conformité s'ajoutera en fin de projet »	Finalité, mémoire, journalisation, droits d'accès et supervision sont des choix d'architecture, arrêtés à la conception.
La deuxième illusion est la plus lourde de conséquences juridiques : sur laquelle un contrôleur reviendra, et le premier à en tirer les conséquences dans son architecture gagne du temps sur tous les suivants.	



Quoi lancer, et dans quel ordre

La première vague doit maximiser trois apprentissages : l'intégration à un processus réel, le contrôle d'une action, et la mesure de la valeur après déploiement. Les cas purement conversationnels sont utiles ; la capacité agentique de l'organisation se teste sur une action.

· Carte des opportunités

· Score de priorisation

· Trois vagues

· Ce qu'il faut exclure

Huit domaines portent des cas agentiques, et leur niveau de risque commande la séquence

Cette carte hiérarchise par apprentissage, en tenant le coût de risque sous contrôle. Les deux dernières lignes, qui concentrent le risque le plus élevé, appellent une vague ultérieure.

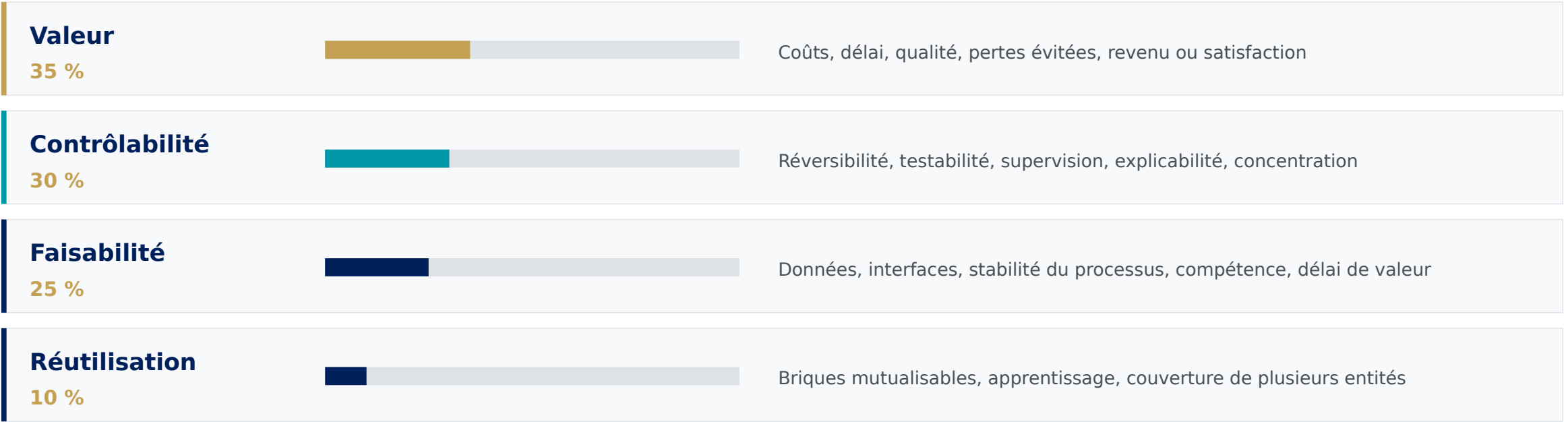
Domaine	Exemples	Autonomie cible	Risque
Front et conseil	Préparation d’entretien, action suivante explicable, compte rendu, suivi	A1 à A2	R2 à R3
Entrée en relation	Complétude, collecte guidée, relances, préanalyse des pièces	A1 à A2	R2 à R3
Crédit	Mémo, revue annuelle, engagements, recherche d’éléments manquants	A1 à A2	R3
Fraude et lutte anti-blanchiment	Constitution de dossier, chronologie, recherche de liens, brouillon	A1 à A2	R3
Opérations	Réconciliations, exceptions, investigations, instructions de résolution	A2 à A3	R2 à R3
Finance et risques	Commentaire de variance, preuve de contrôle, analyse de scénarios	A1 à A2	R2 à R3
Informatique et cybersécurité	Diagnostic, procédure d’exploitation, changement préparé, réponse à incident	A2 à A3	R3
Marchés	Recherche, contrôle avant négociation, surveillance, exécution	A1 à A4	R3 à R4

Les opérations et l’informatique offrent le meilleur rapport entre apprentissage et exposition, parce qu’ils combinent un volume élevé, une vérité terrain observable et une réversibilité forte.

Étude, partie 2.1. Statut : cartographie proposée par l’auteur ; les niveaux d’autonomie et de risque sont des points de départ à confronter à votre appétence.

La contrôlabilité pèse presque autant que la valeur, et c’est ce qui distingue ce score des autres

Chaque dimension se note de un à cinq, les hypothèses se documentent, et la note de contrôlabilité se fait contester par la deuxième ligne. Un cas classé au niveau de risque le plus élevé suit une filière d’autorisation distincte, quelle que soit sa valeur.



Les quatre pondérations et le seuil de 3,5 sont des ordres de grandeur posés par l’auteur, non des valeurs mesurées. Reprenez-les et ajustez-les à votre appétence avant de les opposer à quiconque.

Étude, partie 2.2. Statut : pondérations et seuil proposés par l’auteur, non mesurés, à calibrer par votre comité des risques.

Trois vagues, de l'apprentissage sous faible exposition aux processus matériels

La séquence gradue la réversibilité. Chaque vague apporte la preuve de maîtrise qui autorise la suivante.

VAGUE 1

Apprendre sous faible exposition

Assistants internes, préparation de dossiers, actions au brouillon, données gouvernées, validation humaine aisée.

VAGUE 2

Agir dans une enveloppe

Mises à jour réversibles, notifications internes, orchestration d'exceptions, limites de volume et de valeur.

VAGUE 3

Processus matériels

Usages client ou prudentiels, actions difficiles à inverser, décisions ou recommandations fortement influentes.

Cinq cas à exclure du premier pilote

- Décision autonome d'octroi ou de refus de crédit à une personne physique
- Initiation ou libération autonome d'un paiement, d'un virement, d'une opération de marché ou d'un changement de limite
- Dépôt autonome d'une déclaration réglementaire ou d'un signalement de soupçon sans revue habilitée
- Agent disposant de privilèges d'administration, d'un accès transversal aux boîtes de courriel ou d'une mémoire persistante non cloisonnée
- Cas dont la vérité terrain reste inobservable, processus encore instable, ou performance testable en production seulement

Ces cinq exclusions se signent avant le lancement, au même titre que le périmètre. Une exclusion levée en cours de pilote transforme le pilote en déploiement, et le fait sans son dossier d'autorisation.

Cinq cas de départ, retenus pour la vitesse d'apprentissage qu'ils procurent

Lancer deux à trois de ces cinq cas suffit à construire les briques de contrôle communes, à éprouver la filière d'autorisation et à produire une première mesure de valeur. Les lancer tous les cinq disperse la capacité de validation, qui est la ressource rare.

Cas d'usage	Autonomie	Risque	Pourquoi maintenant
Assistant de politiques internes	A1 · conseille	R1	Faible exposition ; améliore la trouvabilité et la cohérence
Préparation de rendez-vous chargé d'affaires	A1	R2	Valeur commerciale ; sortie vérifiable par le conseiller
Complétude d'entrée en relation et relances	A2 · propose	R2 à R3	Processus volumique ; actions validées par l'analyste
Dossier d'investigation de fraude	A2	R3	Réduit le temps de collecte sans déléguer la décision
Diagnostic d'incident informatique	A2	R3	Résilience et délai de rétablissement ; périmètre technique contrôlable

Retenez au moins un cas de chaque nature. Les deux derniers apportent la preuve d'action, les trois premiers apportent l'adoption.



Le cadre réglementaire, lu par obligations cumulatives

Le règlement sur l'intelligence artificielle s'ajoute à cinq autres cadres de décision. Un agent situé hors de la catégorie haut risque peut rester matériel au regard de la protection des données, de la résilience opérationnelle, des règles prudentielles, de la lutte anti-blanchiment ou de la protection de la clientèle.

· Six cadres cumulatifs

· Cinq tensions

· Taxonomie des risques

· Quatre filières

Six cadres s’appliquent simultanément, et chacun pose ses propres questions à documenter

La méthode doit produire une cartographie cumulative, usage par usage. Répondre au seul règlement sur l’intelligence artificielle laisse ouvertes cinq familles d’obligations dont plusieurs sont, en pratique, plus contraignantes.

Cadre	Questions à documenter
Règlement sur l’IA	Le système entre-t-il dans la définition ? Quel rôle : fournisseur, déployeur, importateur ? Usage interdit, transparence, modèle à usage général ou haut risque ?
Protection des données	Finalité, base légale, minimisation, données sensibles, profilage, article 22, information, droits, analyse d’impact, transferts, sous-traitants.
Résilience opérationnelle	Fonction critique ou importante, service technologique tiers, incident, tests de résilience, continuité, concentration, registre et stratégie de sortie.
Prudentiel et modèles	Impact sur le risque de crédit, de marché, opérationnel, sur le capital, données de risque, validation indépendante, performance et explicabilité.
Conduite et clientèle	Information claire, adéquation, intérêts du client, non-discrimination, réclamation, recours, enregistrement et traçabilité.
Chacun de ces six cadres s’applique pour son propre compte. Un agent hors haut risque peut donc rester matériel, et c’est cette lecture cumulative qui doit figurer dans le dossier de décision.	

Cinq tensions structurent la conformité d'un agent au regard des données personnelles

La note exploratoire publiée par la Commission nationale de l'informatique et des libertés avec le Conseil de l'intelligence artificielle et du numérique, en juillet 2026, recommande le filtrage continu, le cloisonnement des mémoires et des environnements, la supervision humaine des actions critiques, une classification des actions par risque et un dispositif d'arrêt d'urgence.

1	Finalité	Un agent généraliste peut déclencher des traitements non prévus ; chaque outil et chaque mémoire se rattachent à une finalité autorisée.
2	Minimisation	L'accès à une messagerie ou à un espace documentaire entier est rarement nécessaire ; la récupération doit être ciblée, filtrée et justifiée.
3	Mémoire	Les mémoires multiples créent rétention, incohérence et difficulté d'effacement ; préférer le contexte éphémère, une durée de vie et le cloisonnement par tâche.
4	Décision automatisée	Une approbation humaine vaut contrôle lorsque le relecteur dispose de temps, de compétence, d'informations et du pouvoir réel de contredire.
5	Responsabilité et tracabilité	Il faut pouvoir reconstruire les données mobilisées, les agents et services appelés, les décisions intermédiaires, la chronologie, les approbations et les effets.

La quatrième tension est celle qui se règle le moins par la technique. Un relecteur sans temps ni pouvoir de contredire laisse le traitement dans le champ de l'article 22, avec les obligations qui s'y attachent.

Sept familles de risques agentiques, chacune avec le contrôle qui la structure

Cette taxonomie sert à deux choses : construire le jeu de tests adverses de la phase de démonstration, et vérifier que chaque famille dispose d'un contrôle nommé dans l'architecture.

Domaine	Scénarios	Contrôles structurants
Objectif et raisonnement	Mauvaise interprétation, hallucination en cascade, détournement de l'objectif	Objectifs bornés, limites d'étapes, règles déterministes, évaluations par scénario
Outils et actions	Mauvais usage d'un outil, action irréversible, dépassement de mandat	Liste blanche, moindre privilège, budget, approbation, idempotence, retour arrière
Données et mémoire	Fuite, sur-collecte, donnée obsolète, contamination entre agents	Minimisation, listes d'accès, durée de vie, séparation, lignage, purge et réconciliation
Cybersécurité	Injection de consigne, exfiltration, identité usurpée, chaîne d'approvisionnement	Filtrage, bac à sable, identité de l'agent, signature, secrets courts, équipe adverse
Client et conduite	Biais, conseil inadapté, pratiques trompeuses, excès de confiance	Tests d'équité, transparence, recours, scripts de conduite, suivi par segment
Opérations	Dérive, indisponibilité, boucle, coût non borné, incident chez un tiers	Objectifs de service, quotas, coupe-circuit, modèle de repli, continuité, sortie fournisseur
Gouvernance	Responsabilité fragmentée, intelligence artificielle non déclarée, changement non approuvé	Inventaire, fiche d'agent, matrice de responsabilité, contrôle du changement, attestation périodique
La quatrième famille est la seule sans précédent dans les dispositifs existants. C'est aussi celle que la sécurité applicative traditionnelle couvre le moins bien, faute d'équivalent hors du domaine agentique.		

Quatre niveaux de risque interne, et le niveau retenu est le maximum des dimensions

La règle du maximum est ce qui distingue cette classification d’une notation moyenne. Un seul déclencheur critique suffit à imposer la filière la plus exigeante, quelle que soit la valeur du reste du dossier.

Niveau	Déclencheurs indicatifs	Filière d’autorisation
R1 · limité	Interne, lecture seule, données non sensibles, aucun effet client	Propriétaire produit ; sécurité standard ; revue légère
R2 · modéré	Données internes sensibles ou action réversible sans effet significatif	Revue de protection des données et de sécurité ; tests ; approbations ciblées
R3 · élevé	Client, conformité, fonction importante, décision influente ou action matérielle	Deuxième ligne, validation indépendante, exploitation en observation puis canari, comité
R4 · critique	Haut risque au titre du règlement, fonds, crédit défavorable, marché, données très sensibles, irréversible	Appétence arrêtée par la direction ; conformité complète ; autonomie fortement limitée

Le niveau retenu est le maximum des dimensions, non leur moyenne, et cette règle se pose avant le premier prototype, faute de quoi elle sera contestée au moment où elle bloquera un projet.

IV

La méthode 6D : du mandat à la preuve en production

Six phases et six points de décision. Ces points sont des décisions fondées sur des preuves construites dans le déroulement du travail, ce qui les rend compatibles avec une livraison itérative.

- Définir et diagnostiquer
- Concevoir et démontrer
- Déployer et diriger
- Six points de décision

Six phases, six livrables de preuve, et chaque phase autorise la suivante

La méthode se lit comme une boucle et non comme un tunnel : la phase de pilotage renvoie vers la conception dès qu'une preuve de valeur ou de maîtrise disparaît. Le dossier de preuves attendu à chaque point de décision est proportionné au risque et à l'autonomie.



BOUCLE DE PILOTAGE Renouveler, réduire l'autonomie ou retirer l'agent lorsque la preuve de valeur ou de maîtrise disparaît. Tout changement de modèle, de consigne système, d'outil, de permission, de donnée, de mémoire ou de fournisseur se traite comme un changement contrôlé.

Les six points de décision s'appuient sur des preuves construites dans le déroulement du travail. C'est ce qui rend la méthode compatible avec une livraison itérative plutôt qu'opposée à elle.

Six points de décision, chacun avec sa question et ses décideurs

Un point de décision se tranche sur une question fermée. Il se présente lorsque cette question reçoit une réponse par oui ou par non, appuyée sur une pièce.

Point	Question de décision	Décideurs
G0 · Mandat	Le problème, la ligne de base, le propriétaire et le plafond d'autonomie sont-ils explicites ?	Sponsor de business unit et directeur produit
G1 · Admissibilité	Le cas est-il licite, proportionné, testable et dans l'appétence au risque ?	Métier, juridique, conformité, délégué à la protection des données, risque
G2 · Conception	Permissions, données, mémoire, expérience humaine, architecture et contrôles réduisent-ils le risque ?	Architecture, sécurité des systèmes d'information, contrôles
G3 · Preuve	Les seuils d'évaluation, de sécurité, d'équité et de résilience sont-ils atteints et reproductibles ?	Validation indépendante et risque de modèle
G4 · Production	Procédure d'exploitation, surveillance, support, retour arrière, formation et risque résiduel sont-ils acceptés ?	Autorité de changement et dirigeant responsable
RÈGLE DE GOUVERNANCE Toute extension d'autonomie, d'outil, de type de donnée ou de population constitue un nouveau périmètre. Elle repasse au minimum par G1, G2 et G3, et ne se traite jamais comme un réglage de consigne.		



L'organisation : fédérer l'exécution, centraliser les garde-fous

Le modèle recommandé est fédéré. Les entités possèdent les résultats, les processus et les risques de premier niveau ; une plateforme centrale fournit les composants de contrôle et l'observabilité ; les fonctions de deuxième ligne définissent les politiques et valident proportionnellement ; l'audit interne fournit l'assurance indépendante.

· Cinq instances

· Responsabilités

· Trois lignes de défense

· Modèle de financement

Cinq instances, et la question propre que chacune tranche

La supervision bancaire européenne insiste sur une responsabilité non ambiguë, une supervision effective du management et un défi robuste porté par le risque, la conformité et l'audit interne. Le dispositif ci-dessous traduit cette exigence en instances.

ARBITRE

Comité de portefeuille

Arbitre les investissements, la séquence des cas, la capacité de la plateforme, les risques agrégés et les décisions de passage à l'échelle ou de retrait.

DOCTRINE

Conseil des risques liés aux agents

Maintient la taxonomie, les interdits, les seuils de contrôle, les exceptions et la lecture réglementaire commune.

EXÉCUTION

Équipe produit pluridisciplinaire

Product owner et product manager, métier, conception, données et ingénierie, architecture, sécurité, opérations, contrôle embarqué selon le niveau.

SOCLE

Équipe de plateforme

Fournit identité, passerelle d'outils, récupération documentaire, mémoire, évaluations, observabilité, modèles approuvés et motifs de déploiement.

ASSURANCE

Validation indépendante

Évalue les preuves sans dépendre de l'équipe de livraison ; profondeur proportionnée au niveau de risque.

Chacune de ces cinq instances doit pouvoir dire non sur un objet précis, faute de quoi elle devient un point de passage administratif que les équipes apprendront à contourner.

Dix rôles, et un seul propriétaire nommé par agent sur l'ensemble de son cycle de vie

Rôle	Ce dont il répond
Sponsor de business unit	Résultat, capacité opérationnelle et risque résiduel
Propriétaire de l'agent	Cycle de vie, permissions, incidents, revalidation et retrait
Directeur produit	Portefeuille, standards produit, mutualisation, capacité et cohérence de l'expérience
Product owner et product manager	Exploration, fiche d'agent, ligne de base, backlog, métriques, preuves, adoption et exploitation
Ingénierie et plateforme	Architecture, intégrations, passerelle de politique et d'outils, observabilité, objectifs de service et retour arrière
Données	Qualité, lignage, accès, rétention, données de référence, corpus documentaire et propriété
Sécurité des systèmes d'information	Modèle de menace, identité, secrets, tests, détection, réponse à incident et chaîne d'approvisionnement
Juridique, conformité et protection des données	Classification, base légale, conduite, analyse d'impact, transparence, droits et obligations sectorielles
Risque et validation de modèle	Niveau, contestation indépendante, modèle de risque, limites, seuils et surveillance
Audit interne	Assurance indépendante sur la gouvernance, la conception, l'exécution et l'efficacité réelle des contrôles

Le propriétaire de l'agent est le rôle le plus souvent absent des organisations observées. Sans lui, la responsabilité se dilue entre l'équipe qui construit et la fonction qui contrôle, ce qui est exactement ce que la supervision reproche.

Trois lignes de défense adaptées, et quatre enveloppes qui rendent la mutualisation possible

Le modèle de financement est ce qui décide, en pratique, si la plateforme de contrôle existe. Faire porter les briques communes par le premier projet revient à lui faire supporter seul un investissement dont profiteront tous les suivants.

Ligne	Acteurs	Mission sur l'agent
1re ligne	Business unit, produit, opérations, ingénierie	Conçoit, opère, surveille, corrige et démontre la maîtrise
2e ligne	Risque, conformité, protection des données, sécurité	Définit les politiques, conteste, approuve ou valide selon le niveau, agrège les risques
3e ligne	Audit interne	Évalue indépendamment le dispositif et l'efficacité réelle des contrôles

Quatre enveloppes de financement

ENVELOPPE 1

Exploitation

Chaque business unit finance le produit, le changement de processus, les opérations et le support de son

ENVELOPPE 2

Commun

L'entreprise finance la plateforme de contrôle, les modèles approuvés, l'évaluation, l'observabilité et la

ENVELOPPE 3

Renfort de maîtrise

Une enveloppe transversale finance les exigences nouvelles, la validation, l'équipe adverse et

ENVELOPPE 4

Refacturation interne

Coût par tâche réussie et contrôlée, consommation de modèles et d'outils, coût de supervision,

La deuxième enveloppe se décide au niveau de l'entreprise, sous peine de voir chaque équipe reconstruire ses propres contrôles, ce qui multiplie le coût et fragmente la preuve.

Le superviseur d’agent est un rôle nouveau, avec sa charge et ses compétences

Un agent déplace le travail plutôt qu’il ne le supprime : le collaborateur cesse de produire le dossier et se met à juger une proposition. Ce geste est différent, il se forme, et sa charge se mesure dès le pilote.

CE QU’IL FAIT Il juge une proposition et il tranche Il lit les pièces citées, vérifie la règle appliquée, modifie, rejette ou escalade, et motive chaque écart. La motivation devient la matière première de l’évaluation continue.	CE QU’IL LUI FAUT Un écran qui montre le raisonnement Les pièces examinées, la règle citée avec sa version, les éléments restés sans interprétation, et une séparation nette entre la proposition de l’agent et la décision humaine.	CE QUI SE MESURE Le temps de revue et la tendance à approuver Un relecteur qui approuve sans lire produit un dispositif conforme en apparence. Le temps passé et le taux de modification se suivent dès la première semaine d’exposition.
48 % PRÉVOIENT CE RÔLE Part des dirigeants prévoyant de créer des postes de superviseur ou de coordinateur d’agents	92 % CITENT LE MANQUE DE COMPÉTENCES Part des dirigeants financiers citant les compétences en intelligence artificielle comme un défi critique	

Cinq à dix heures par personne, en présentiel, avec un accompagnateur

Le levier A de SCALE-6 porte une règle que peu d'établissements appliquent : la modalité de formation pèse autant que son volume. Les quatre chiffres ci-dessous viennent de la même enquête et se lisent ensemble.

5 à 10 h PAR UTILISATEUR Seuil au-delà duquel l'adoption progresse nettement	+12 pts PRÉSENTIEL Écart d'adoption entre une formation en présentiel et une formation à distance	+14 pts ACCOMPAGNATEUR Écart d'adoption lorsqu'un référent accompagne l'équipe dans le flux de travail	36 % SE SENTENT FORMÉS Part des répondants s'estimant correctement formés à ce jour
---	--	---	--

CE QUE CELA DONNE DANS LE PILOTE

- Semaine 6 : le référent est désigné dans l'équipe métier, et il participe à la conception de l'écran de supervision.
- Semaine 9 : les relecteurs suivent cinq à dix heures en présentiel sur des dossiers réels du jeu de référence.
- Semaine 11 : le référent siège à côté des relecteurs pendant la période à blanc, et consigne les motifs de correction.
- Semaine 13 : le temps de revue et le taux de modification figurent au dossier de décision, au même titre que la qualité.

Le levier le moins financé est aussi le mieux documenté. Un budget de formation à 16 % de l'enveloppe, en baisse de 8 points sur un an, dit où se trouve la marge de progression la moins chère du dispositif.

VI

L'architecture : une ceinture de contrôle autour de l'agent

L'architecture doit pouvoir changer de modèle sans reconstruire la gouvernance. Le modèle est un composant ; le contrôle de l'action réside dans l'identité, la passerelle d'outils, les politiques déterministes, la mémoire, l'observabilité et les mécanismes humains.

· Cinq couches

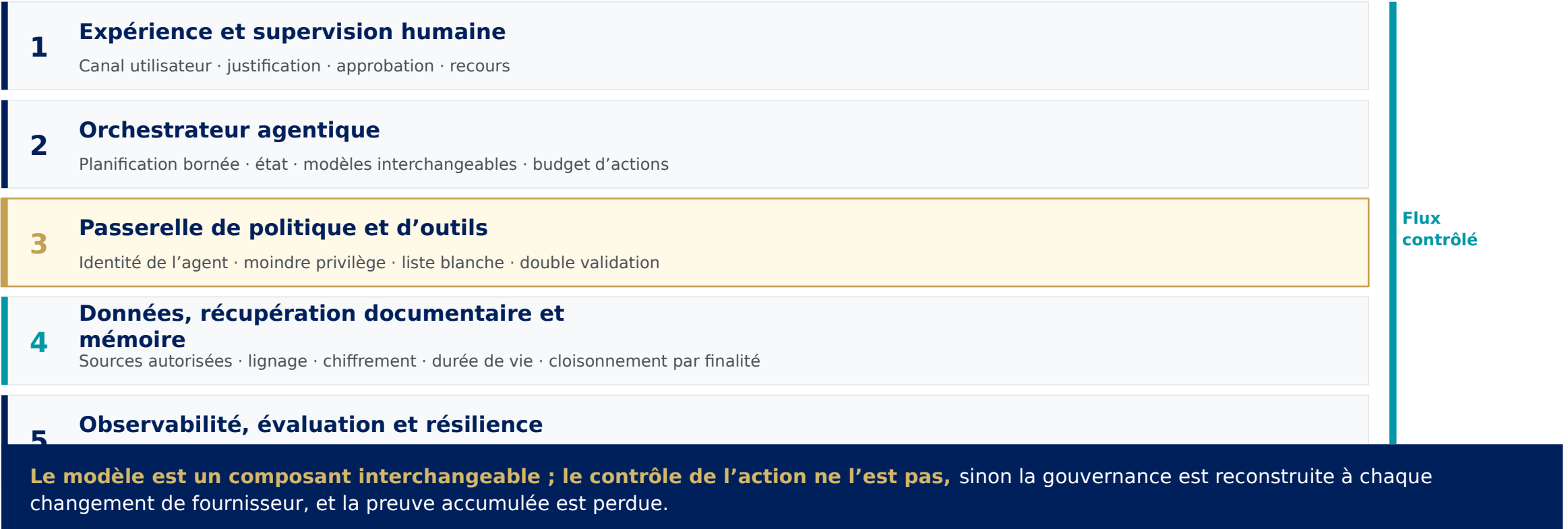
· Dix exigences

· Contrôles par action

· Sécurité agentique

Cinq couches, et c’est la troisième qui décide de ce que l’agent a le droit de faire

Cette architecture est indépendante du fournisseur de modèle. Changer de modèle doit être une opération de la couche deux, sans effet sur les politiques de la couche trois ni sur les preuves de la couche cinq.



Étude, partie 6, figure 3. Statut : architecture proposée par l’auteur, indépendante du fournisseur.

Dix exigences non négociables, à inscrire au contrat comme au dossier d'architecture

- 1 **Identité d'agent distincte.** Aucun compte utilisateur partagé ; authentification, attributs, propriétaire et cycle de vie propres.
- 2 **Moindre privilège par tâche.** Jeton court, portée dynamique, accès limité à la donnée et à l'action nécessaires.
- 3 **Passerelle de politique et d'outils.** Toutes les actions passent par des règles déterministes, des quotas, des validations et des contrôles de contexte.
- 4 **Liste blanche d'outils et schémas stricts.** Entrées typées, sorties validées, destinations autorisées, aucun appel libre à une interface générale.
- 5 **Séparation lecture et écriture.** Les droits d'observer n'impliquent jamais les droits de modifier ou d'envoyer.
- 6 **Mémoire minimale et cloisonnée.** Contextes éphémères par défaut, durée de vie, finalité, chiffrement, purge vérifiable, aucun réentraînement implicite.
- 7 **Traçabilité de bout en bout.** Utilisateur, version de modèle, consigne système, données, outils, paramètres, décisions, approbations et résultat.
- 8 **Supervision proportionnée.** Aperçu avant action, raison et preuves, incertitude, droit de modifier ou refuser, escalade et recours.
- 9 **Arrêt et récupération.** Arrêt d'urgence, coupe-circuit, limites de coût et d'étapes, idempotence, annulation, retour arrière et procédure manuelle.
- 10 **Évaluation continue.** Tests avant production, échantillonnage en ligne, canaris, détection de dérive et revalidation des changements.

Ces dix exigences se vérifient avant la signature. La première et la sixième sont celles qu'un fournisseur accepte le moins volontiers, et ce refus est en soi une information.

Six classes d'action, et le motif de contrôle par défaut de chacune

Raisonner par classe d'action plutôt que par cas d'usage permet de réutiliser les contrôles d'un agent à l'autre. C'est aussi ce qui rend la revue de deuxième ligne rapide : elle porte sur des motifs déjà connus d'elle.

Action	Risques à couvrir	Motif de contrôle par défaut
Lire	Données autorisées, minimisation, masquage, journal des accès	Filtrage par attributs et requête ciblée
Recommander	Sources, incertitude, limites, non-automatisation du jugement	Citations, vue explicative, possibilité de passer outre
Créer ou modifier	Brouillon, différence, validation de schéma, bac à sable, retour arrière	Prévisualisation puis approbation au niveau A2
Communiquer	Destinataire, secret, contenu, ton, obligation d'information	Liste blanche, prévention de fuite, double contrôle externe
Engager	Valeur, limite, séparation des tâches, authentification forte	Règle déterministe et humain habilité
SÉCURITÉ AGENTIQUE Tout contenu récupéré, document, courriel, page web ou sortie d'outil, se traite comme non fiable : il peut porter des instructions malveillantes. Instructions système, données et commandes se séparent, et le modèle ne produit jamais directement une requête privilégiée exécutable sans validation de schéma et de politique.		

VII

Prouver une tâche sûre et une valeur réalisée

L'évaluation d'un agent ne se résume ni à l'exactitude textuelle ni à un taux de satisfaction. Elle porte sur la trajectoire complète : choix du plan, sélection des outils, exactitude des arguments, action, supervision, résultat métier et résilience.

· Six dimensions

· Jeu d'évaluation et seuils

· Tableau de bord

· Dossier économique

Six dimensions d'évaluation, jusqu'à l'exploitation et à l'expérience humaine

Les quatre premières dimensions sont couvertes par les bancs d'essai usuels. Les deux dernières, l'exploitation et l'humain, relèvent de votre propre mesure, et ce sont elles qui déterminent le coût réel et l'acceptation du dispositif.

Dimension	Exemples de métriques
Résultat métier	Délai, capacité, qualité, pertes évitées, conversion, satisfaction, niveau de service
Qualité de tâche	Succès de bout en bout, exactitude, complétude, ancrage documentaire, gestion des exceptions
Action	Bon outil, bons paramètres, respect du mandat, réversibilité, absence d'effet non voulu
Risque et conformité	Hallucination matérielle, données personnelles, biais, conduite, sécurité, preuves, recours
Exploitation	Disponibilité, latence, coût, dérive, incidents, délai de rétablissement, dépendance fournisseur
Humain	Adoption, temps de retour, qualité des décisions de passer outre, biais d'automatisation, charge cognitive, confiance calibrée

La sixième dimension est celle qui prédit le mieux l'échec en production. Un dispositif que les équipes contournent, ou qu'elles approuvent sans lire, produit d'excellentes métriques techniques pour une valeur nulle.

Six composants dans le jeu d'évaluation, des cas de référence aux seuils durs

COMPOSANT 1

Jeu de référence

Cas réels anonymisés ou synthétiques représentatifs, avec résultat attendu et critères de réussite vérifiables.

COMPOSANT 2

Tranches

Segments de clientèle, canal, langue, type de document, complexité, valeur, équipe, données rares et cas de bord.

COMPOSANT 3

Jeu adverse

Injection directe et indirecte, données contradictoires, outil indisponible, identité insuffisante, boucle, sortie trompeuse.

COMPOSANT 4

Seuils durs

Zéro action interdite, zéro fuite critique, zéro dépassement de limite. Les moyennes ne compensent pas un événement intolérable.

COMPOSANT 5

Seuils statistiques

Intervalle de confiance, comparaison à la ligne de base humaine ou système, test de non-régression avant tout changement matériel.

COMPOSANT 6

Échantillonnage en production

Revue ciblée des cas à forte incertitude et revue aléatoire des cas ordinaires, avec boucle de correction contrôlée.

Le quatrième composant est celui qui distingue une évaluation de banque d'une évaluation de laboratoire : et c'est la raison pour laquelle une mise en production s'autorise sur des seuils durs, au-delà des moyennes.

Huit indicateurs de production, chacun avec sa cadence et son propriétaire nommé

Un indicateur devient un instrument de pilotage lorsqu’il porte un propriétaire et une cadence. Les trois indicateurs en temps réel sont ceux qui déclenchent une action immédiate ; les autres alimentent une revue.

Indicateur	Défini tion	Cadence	Propriétaire
Taux de tâches réussies	Part des tâches atteignant le résultat métier sans correction matérielle	Hebdomadaire	Product owner
Taux d’actions sûres	Part des actions conformes au mandat, aux règles et aux limites	Temps réel	Propriétaire de l’agent
Qualité des passages outre	Part des décisions de passer outre justifiées et des erreurs rattrapées	Hebdomadaire	Opérations et risque
Précision des escalades	Part des escalades réellement nécessaires	Hebdomadaire	Opérations
Taux d’erreur matérielle	Erreurs à impact client, financier, prudentiel ou de conformité	Temps réel	Risque
Valeur réalisée	Euros nets, capacité libérée, délai, qualité et volume effectivement absorbé	Mensuelle	Sponsor de business unit
Coût par tâche réussie	Modèle, outils, infrastructure, supervision et reprise	Mensuelle	Produit et finance
Santé des changements	Régressions, dérive et incidents après changement	À chaque livraison	Ingénierie

Le coût par tâche réussie est le seul indicateur qui relie la performance technique à la décision de poursuivre. Les six autres expliquent ; celui-là décide, et c’est le seul que le sponsor doit pouvoir citer de mémoire.

Une équation de valeur, et les cinq conditions qui la rendent démontrable

ÉQUATION Valeur nette annualisée = capacité productive libérée + revenu incrémental + pertes évitées + coût de qualité évité – coûts récurrents – coût de supervision – reprise – amortissement de la mise en conformité.

- 1 Ligne de base avant pilote.** Temps actif, temps d'attente, taux de reprise, qualité, volume, incidents et coût complet, mesurés avant tout déploiement.
- 2 Mesure contrefactuelle.** Cohorte témoin, comparaison avant et après sur périmètre comparable, ou déploiement par vagues, pour isoler la saisonnalité et le changement de processus.
- 3 Taux de capture.** Distinguer les minutes gagnées, la capacité réellement réaffectée, et le coût effectivement retiré des comptes.
- 4 Scénarios.** Défavorable, central et favorable, sur l'adoption, la performance, le prix du modèle et la charge de contrôle.
- 5 Financement progressif.** N'engager l'industrialisation que si la preuve de valeur et les points de contrôle progressent ensemble.

La première condition est celle que l'on saute le plus souvent, parce qu'elle retarde le lancement de deux semaines. Elle conditionne les quatre autres : c'est elle qui transforme une intention en chiffre défendable.

VIII

Conserver la maîtrise des contrôles et de la donnée

Le choix ne porte pas seulement sur un modèle. Il couvre l'orchestration, les connecteurs, la mémoire, l'évaluation, l'observabilité, l'hébergement, le support et la capacité de sortie. Une banque peut acheter une technologie ; elle ne peut pas externaliser sa responsabilité.

· Construire, acheter, codévelopper

· Sept axes de diligence

· Critères de sélection

· Règle de rejet

Construire ce qui vous différencie et vous engage, acheter ce qui est interchangeable

Le critère de partage n'est ni le coût ni le délai, mais la réversibilité. Ce qui porte votre responsabilité et ne peut être repris par un tiers se construit ; ce qui se remplace en un trimestre s'achète.

Option	Périmètre indicatif	Raison
Construire	Passerelle de politique et d'outils, identité et permissions, règles de contrôle, intégration au cœur, traces et propriété	Différenciation et responsabilité
Acheter	Modèles, outils d'orchestration, observabilité ou évaluation, à condition qu'ils soient interchangeables et auditables	Vitesse et accès aux capacités
Codévelopper	Cas matériels ou spécialisés nécessitant une expertise sectorielle et un transfert de compétence	Réduire le délai de valeur sans boîte noire
Évaluer	Plateforme formelle contrôlant le fragement des fonctionnalités, identité, responsabilité	Construction et gestion immortelles

RÈGLE DE REJET Une note globale élevée ne compense jamais l'absence de droit d'audit, l'impossibilité de désactiver l'entraînement sur vos données, une mémoire non effaçable, des permissions opaques ou une stratégie de sortie non testable. Ces cinq points sont éliminatoires, quels que soient le prix et la démonstration.

La diligence porte sur le dispositif entier, non sur le seul modèle. Sept axes se documentent : données, modèle, agent, exploitation, conformité, économie et sortie.

IX

Quatre-vingt-dix jours pour prouver, dix-huit mois pour industrialiser

Le calendrier suppose un sponsor engagé, une plateforme de données existante et l'accès aux fonctions de contrôle. Il privilégie un agent au niveau conseil ou action préparée sur un processus mesurable, sans effet client défavorable autonome.

· Pilote 90 jours

· Feuille de route 0-18 mois

· Conditions du passage

· Trente prochains jours

Treize semaines, six phases, et un point de décision atteint à chaque étape

Ce calendrier est un ordre de grandeur, non un engagement. Ce qui compte est la séquence des points de décision : chaque semaine se franchit avec la pièce qui autorise la suivante.

Semaines	Phase	Sorties attendues	Point
S1 à S2	Définir	Charte, ligne de base, carte de processus, propriétaire, plafond d'autonomie, critères de succès	G0
S2 à S4	Diagnostiquer	Classification, niveau, carte des données, modèle de menace initial, jeu de référence version zéro	G1
S4 à S6	Concevoir	Prototype intégré, fiche d'agent, permissions, expérience d'approbation, journaux, retour arrière	G2
S6 à S9	Démontrer	Évaluations, sécurité, équipe adverse, tests utilisateurs, validation indépendante	G3
S9 à S11	Déployer	Mode observation puis cohorte pilote, procédure d'exploitation, support, formation, tableau de bord	G4
S12 à S13	Décider	Valeur, risque, incidents, reprise, adoption ; arrêter, corriger ou étendre	G5
Les trois issues de la semaine treize ont la même dignité : La décision d'arrêter est la sortie normale d'un pilote qui a rempli son office, non l'aveu d'un échec.			

Étude, partie 9.1. Statut : calendrier proposé par l'auteur, sous les hypothèses énoncées.

Dix-huit mois, quatre horizons, et une capacité qui se construit avant de se dépenser



Sept conditions de passage à l'échelle

Inventaire unique des agents, modèles, outils, permissions, données, propriétaires, tiers, versions et dates de revue · catalogue de motifs approuvés A0 à A3, A4 sous exception · passerelle commune · identité d'agent · gestion des secrets · traces et arrêt d'urgence éprouvés · jeux d'évaluation versionnés et capacité de rejouer un incident · matrice de responsabilité et capacité

Ces sept éléments se construisent en amont du passage à l'échelle. La capacité de deuxième ligne, en particulier, se réserve des trimestres à l'avance et ne s'achète pas dans l'urgence.

Sept décisions à prendre dans les trente prochains jours, à budget d'investissement constant

- 1 Nommer un sponsor exécutif et un propriétaire d'agent par pilote
- 2 Choisir deux cas au niveau conseil ou action préparée, et un cas d'action bornée à explorer, avec une ligne de base disponible
- 3 Adopter la taxonomie d'autonomie A0 à A4 et la classification de risque R1 à R4
- 4 Fixer cinq interdits d'entreprise et la filière d'exception qui leur correspond
- 5 Financer la passerelle d'outils, l'identité d'agent, les journaux et le banc d'évaluation en version un
- 6 Bloquer les créneaux du juridique, de la protection des données, de la sécurité, du risque et des opérations dès l'exploration
- 7 Définir les seuils du point de décision G3 et la règle « arrêter, corriger ou étendre » avant le lancement du pilote

La sixième décision est celle qui glisse le plus souvent. Les fonctions de contrôle sollicitées en fin de conception deviennent un obstacle ; sollicitées à l'exploration, elles deviennent des alliées du dossier.



Le toolkit : transformer le contrôle en exigences produit

Les outils qui suivent traduisent la méthode en objets que manipule une équipe produit : un canevas en douze champs, deux définitions de prêt et de terminé, un backlog de contrôle rédigé en récits d'utilisateur, et les questions à poser en comité de pilotage.

· Agent Canvas

· Prêt et terminé

· Backlog de contrôle

· Questions de comité

Douze champs obligatoires : le canevas qui remplace la note de cadrage

Les douze champs se remplissent avant le point de décision G1, qui les exige tous. Le canevas se remplit par l'équipe produit, se fait relire par les fonctions de contrôle, et se met à jour à chaque changement matériel.

1 Résultat Résultat métier, ligne de base, valeur et seuil de succès	2 Utilisateurs Qui mandate, qui supervise, qui subit l'effet, qui peut recourir	3 Périmètre Tâches incluses, non-objectifs, populations, canaux et géographies	4 Autonomie Niveau A0 à A4 par action et par outil ; conditions d'escalade
5 Risque Niveau R1 à R4, cadres applicables, impacts et risque résiduel	6 Données Sources, sensibilité, finalité, base légale, qualité, lignage et rétention	7 Outils Interfaces, droits, schémas, liste blanche, limites, réversibilité et propriétaire	8 Mémoire Contenu, portée, durée de vie, cloisonnement, correction, effacement et réutilisation
9 Humain Moment, preuve, temps, compétence, pouvoir de contredire et recours	10 Évaluation Jeu de référence, tranches, équipe adverse, seuils durs et non-régression	11 Exploitation Objectifs de service, surveillance, incident, retour arrière, modèle de repli et coûts	12 Cycle de vie Propriétaire, versions, changements matériels, attestations, sortie et retrait

Les champs 8, 9 et 12, la mémoire, la supervision humaine et le cycle de vie, sont les trois que les équipes produit remplissent le moins spontanément, et les trois que la deuxième ligne lit en premier.

Ce qui autorise à explorer, puis ce qui autorise à agir sur un client

DEFINITION OF READY

Prêt à explorer

- Problème et ligne de base mesurés ; alternative sans intelligence artificielle considérée
- Sponsor, propriétaire d'agent, équipe produit et partenaires de contrôle identifiés
- Processus et exceptions cartographiés ; résultat et non-objectifs explicites
- Autonomie maximale, outils, données, mémoire et interdits préqualifiés
- Classification initiale et niveau R1 à R4 disponibles ; accès au jeu de référence confirmé
- Critères de succès, seuils de risque et règle d'arrêt du pilote acceptés

DEFINITION OF DONE

Prêt pour la production

- Fiche d'agent, architecture, classification, analyses requises et décisions de points archivées
- Permissions testées ; aucune action hors passerelle ; secrets éphémères ; mémoire et purge vérifiées
- Jeux de référence et adverses passés ; seuils durs respectés ; validation indépendante close
- Mode observation concluant ; supervision humaine testée en conditions réelles ; formation effectuée
- Journaux, alertes, objectifs de service, procédure d'exploitation, arrêt d'urgence, retour arrière et support opérationnels
- Tableau de bord de valeur et de risque actif ; budget de coût et de latence ; cadence de revue et propriétaire de retrait définis

La colonne de gauche autorise à explorer ; seule celle de droite autorise à agir sur un client. et c'est ce qui permet de commencer vite sans promettre ce qui n'a pas encore été démontré.

Huit questions qui suffisent à instruire un dossier en vingt minutes

Ces huit questions constituent une revue complète en vingt minutes. Elles se posent dans cet ordre, et l'absence de réponse sur l'une d'elles suffit à reporter la décision sans qu'il soit nécessaire d'en débattre.

1 Quel résultat métier change, et quelle ligne de base le prouvera ?

3 Quelle est la portée maximale d'un incident par action, par heure et par client ?

5 Quels tests échoués arrêteraient immédiatement le pilote ou le service ?

7 Quel coût complet par tâche réussie, en incluant la supervision et la reprise ?

2 Quelles actions l'agent peut-il accomplir sans humain, et pourquoi ce niveau est-il nécessaire ?

4 Comment un humain comprend-il, contredit-il et annule-t-il l'action ?

6 Quelle donnée ou permission nouvelle déclencherait une requalification ?

8 Comment revient-on au processus manuel, et comment sort-on du fournisseur ?

Les questions 1 et 7 encadrent le dossier économique ; et les questions 3, 5 et 8 sont celles qui provoquent le plus souvent un report, parce qu'elles portent sur ce qui se passe quand le dispositif échoue.

La banque étend aux agents la gouvernance qu'elle possède déjà

DÉCISION PROPOSÉE Autoriser un programme de quatre-vingt-dix jours sur deux à trois agents de niveau conseil ou action préparée, assorti d'un chantier de plateforme de contrôle minimale et d'une gouvernance de portefeuille. La montée en autonomie se gagne par la preuve, agent par agent et action par action.

Les actifs déjà présents dans la maison

- Culture de contrôle
- Validation de modèles
- Séparation des tâches
- Gestion des accès
- Gestion des tiers
- Résilience opérationnelle

L'enjeu consiste à étendre ces mécanismes aux propriétés spécifiques des agents, à savoir la planification, la mémoire, les permissions dynamiques et l'action autonome, puis à les rendre utilisables au rythme du produit.

A

Annexes opérationnelles

Quatre pièces à reprendre telles quelles : le dossier de preuves minimal attendu à chaque point de décision, le catalogue de contrôles prioritaires, les seuils indicatifs d'un pilote de niveau A2 et R2, et le référentiel de standards mobilisable.

· Dossier par point

· Catalogue de contrôles

· Seuils indicatifs

· Standards

Dossier de preuves minimal attendu à chaque point de décision

Point	Pièces minimales
G0	Charte, ligne de base du processus, arbre de valeur, propriétaire responsable, périmètre et hors-périmètre
G1	Classification cumulative, niveau de risque, carte des données, interdits, analyse d'impact et analyses requises, risque initial
G2	Fiche d'agent, architecture, permissions, mémoire, expérience de supervision humaine, modèle de menace, contrôle par action
G3	Jeux de référence et adverses, rapport d'évaluation, sécurité et équité, validation, non-conformités et remédiation
G4	Résultats d'observation et de pilote, objectifs de service, surveillance, procédure d'exploitation, support, formation, retour arrière, risque résiduel accepté
G5	Valeur réalisée, incidents, décisions de passer outre, dérive, coûts, changements, dette de contrôle, plan de passage à l'échelle ou de retrait

Étude, annexe A. Statut : liste proposée par l'auteur ; à ajuster à votre référentiel de gouvernance de projet.

Catalogue de contrôles prioritaires, avec leur identifiant

GOV-01	Propriétaire et matrice de responsabilité	Propriétaire d’agent unique, délégations et suppléance
GOV-02	Inventaire	Agent, modèle, outils, données, versions, niveau et statut
GOV-03	Contrôle du changement	Changements matériels et revalidation proportionnée
IAM-01	Identité de l’agent	Compte distinct, attesté, durée et propriétaire
IAM-02	Moindre privilège	Portées courtes, contexte et tâche, réattestation
ACT-01	Passerelle d’outils	Liste blanche, schéma, politique, quotas et journal
ACT-02	Approbation	Validation ciblée, séparation des tâches, preuve
ACT-03	Portée maximale	Limites de valeur, de volume et de temps, coupe-circuit
DAT-01	Minimisation	Récupération ciblée, filtres et justification
DAT-02	Mémoire	Durée de vie, séparation, correction, purge, aucun entraînement implicite
SEC-01	Injection	Données non fiables, séparation instructions et données, tests
SEC-02	Secrets	Hors contexte du modèle, délivrance éphémère, rotation
SEC-03	Bac à sable	Exécution isolée, réseau et système de fichiers limités
EVA-01	Jeu de référence	Cas représentatifs, vérité terrain et tranches
EVA-02	Seuils durs	Zéro action interdite, zéro fuite critique, zéro dépassement
EVA-03	Non-régression	Changements de modèle, de consigne, d’outil et de donnée testés
OPS-01	Observabilité	Trace de bout en bout, métriques et capacité de rejouer
OPS-02	Arrêt d’urgence	Arrêt total ou par classe d’action, testé
OPS-03	Continuité	Repli, procédure manuelle, objectifs de reprise et sortie
HUM-01	Supervision	Temps, compétence, information et pouvoir réel

Étude, annexe B. Statut : catalogue proposé par l’auteur ; les identifiants se raccordent à votre référentiel de contrôle interne.

Seuils indicatifs pour un pilote de niveau A2 et R2, à recalibrer chez vous

Ces valeurs sont des points de départ posés par l’auteur, non des mesures. Le métier, le risque et les fonctions de contrôle fixent les valeurs finales selon la gravité des erreurs, la population, la volumétrie et la capacité de rattrapage.

Métrique	Seuil indicatif	Règle associée
Action interdite	0	Arrêt immédiat
Fuite de donnée sensible	0	Arrêt immédiat et incident
Succès de tâche de bout en bout	Supérieur à 90 %	À adapter au processus
Erreur matérielle	Inférieure à la ligne de base humaine et au seuil métier	Par tranche
Action bloquée à juste titre	Supérieure à 95 % sur le jeu adverse	Mesurer les faux positifs
Temps de revue	Gain net par rapport à la ligne de base	Inclure la reprise
Disponibilité et latence	Objectif de service du processus	Inclure les dépendances
Coût par tâche réussie	Sous le scénario défavorable du dossier économique	Inclure la supervision

Les deux premiers seuils sont les seuls qui ne se négocient pas, et le premier réflexe utile consiste à demander sur quelle base ils ont été retenus dans votre établissement.

Cinq mesures publiques établissent le gain de tâche ; le gain économique reste à établir chez soi

Cette annexe dit pourquoi le document avance une méthode plutôt que des résultats de terrain : la littérature accessible mesure le premier niveau et documente la dépendance à la tâche. Chaque ligne porte son dispositif et sa clause de transposition.

Niveau	Résultat mesuré	Dispositif	Transposition
Tâche	+15 % de problèmes résolus par heure, gain concentré sur les moins expérimentés	Brynjolfsson, Li et Raymond, QJE, avril 2025 ; 5 172 conseillers de support	Hors banque ; tâche conversationnelle à faible enjeu réglementaire
Tâche	+17 % de qualité et –34 % de temps en compréhension documentaire	Banque centrale d’Irlande, essai préenregistré, 143 collaborateurs, états Solvabilité II	Secteur financier public ; prépublication
Tâche	–12 % de qualité et aucun gain de temps en analyse de données	Même essai, même semaine, même établissement, seconde tâche	Le même outil produit deux résultats opposés selon la tâche
Tâche	19 points de réponses justes en moins hors du domaine de compétence du modèle	Essai contrôlé, 758 consultants	Hors banque ; borne le périmètre d’emploi plutôt que l’outil
Fiabilité	Précision de 16,7 % à 80,0 %, score de fiabilité de 0,69 à 0,87 en vingt-quatre mois	Équipe Princeton HAL ; 15 modèles, 12 métriques, 2 bancs d’essai	Banc d’essai ; la précision progresse plus vite que la régularité
Économie	« financial quantification of realised benefits remains a challenge »	BCE, supervision bancaire, novembre 2025 ; 107 puis 110 établissements importants	Constat d’autorité sur le secteur ; verbatim d’un texte publié

Le premier niveau est mesuré, le troisième reste à mesurer. La dernière ligne justifie la forme de ce document : le superviseur lui-même constate que la quantification financière des bénéfices réalisés reste un exercice ouvert.

Chaque terme de l'équation porte un nombre sur un cas type de 120 000 dossiers

Cette annexe répond à l'objection d'un comité d'investissement : une équation sans terme instancié ne se discute pas. Les valeurs proviennent du modèle de coût Acadewie, dont les hypothèses sont publiées et recomptables dans le classeur joint.

Terme de l'équation	Valeur sur le cas type	Statut
Volume traité	120 000 dossiers par an	Hypothèse de cas type
Coût complet de première année	1 824 442 €, dont 630 000 € de charges fixes	Modèle Acadewie, recomptable
Coût par dossier traité	15,20 €	Modèle Acadewie, recomptable
Part humaine et organisationnelle	79,9 % du coût complet, contre 20,1 % de charge technique	Modèle Acadewie, recomptable
Capacité productive libérée	56 équivalents temps plein ramenés à 36	Modèle Acadewie, recomptable
Seuil d'équilibre	22,8 % d'automatisation sur le premier parcours	Modèle Acadewie, recomptable
Flux net de première année	887 294 €, après un abattement de prudence de 25 %	Choix de prudence de l'auteur

EFFET DE PORTEFEUILLE Le premier parcours porte la plateforme de contrôle entière. En amortissant les charges réutilisables — 50 % de l'évaluation, 80 % de la gouvernance, 60 % du risque, 40 % de l'intégration — le seuil d'équilibre passe de 22,8 % sur un parcours à 17,2 % sur deux, 15,3 % sur trois et 13,4 % sur six.

Modèle publié, hypothèses ouvertes, résultat non mesuré en production. Ces valeurs se recomptent depuis le classeur joint et se remplacent par les vôtres en changeant six paramètres.

Standards mobilisables, et les vingt références sur lesquelles ce document s’appuie

Référentiel de standards

ISO/IEC 42001:2023	Système de management de l’intelligence artificielle
ISO/IEC 23894:2023	Management des risques spécifiques à l’intelligence artificielle
ISO/IEC 42005:2025	Évaluation des impacts humains et sociétaux
ISO/IEC 27001:2022	Système de management de la sécurité de l’information
NIST AI RMF 1.0 et profil	Gouverner, cartographier, mesurer, gérer
DORA	Risque informatique, incidents, résilience et risque tiers

La norme ISO/IEC 42001 fournit le système de management d’entreprise ; la méthode 6D en constitue la déclinaison produit et agentique, avec un accent sur l’action, les permissions et la preuve.

PRÉCAUTION DE LECTURE Les sources retenues sont en priorité des textes, des autorités et des organismes de normalisation. Les orientations doivent être vérifiées au moment de chaque décision de conformité, le cadre restant évolutif. Les scores, pondérations et seuils proposés dans ce document relèvent de l’appréciation de l’auteur et ne sont pas des valeurs mesurées.

Sources officielles

Commission européenne, calendrier de mise en œuvre du règlement sur l’IA · ACPR, règlement européen sur l’IA, page sectorielle, 2 juillet 2026 · ACPR, programme de travail 2026 · BCE, supervision bancaire, « Technology is neutral, governance is not », février 2026 · Autorité bancaire européenne, application croissante de l’IA dans la banque et les paiements, septembre 2025 · CNIL et Conseil de l’IA et du numérique, IA agentique et protection des données, juillet 2026 · Règlement (UE) 2016/679 · CJUE, arrêt C-634/21 SCHUFA · Règlement (UE) 2022/2554 DORA · BCE, guide sur l’externalisation en nuage, juillet 2025 · ACPR, gouvernance des algorithmes d’IA dans le secteur financier · BIS, Financial Stability Institute, réguler l’IA dans le secteur financier · NIST, cadre de gestion des risques, profil IA générative · NIST NCCoE, identité et autorisation des agents logiciels, février 2026 · Comité européen de la protection des données, avis 28/2024 · ISO/IEC 42001:2023 · ISO/IEC 23894:2023 · Règlement (UE) 2024/1689 consolidé au 27 juillet 2026 · ACPR, équité algorithmique dans le secteur financier, juillet 2026 · BCE, priorités prudentielles 2026-2028.

À P R O P O S D E L ' A U T E U R

Mounir Fassouane

Praticien du produit, il accompagne les directions et équipes produit sur l'intelligence artificielle et publie les études Acadewie, dont *GenAI for Leaders*. Ce guide est né d'un constat de terrain : les équipes ne manquent ni de technologie, ni d'ambition, elles manquent du chemin de preuves qui fait voter un comité.

Besoin d'accompagnement comme Product Owner, Product Manager ou en conseil en transformation IA ?

Me contacter sur LinkedIn

linkedin.com/in/mounirfassouane

Diffusé sous licence Creative Commons BY-ND 4.0 · partage libre avec attribution, sans modification

Mounir Fassouane · Acadewie · www.acadewie.com/agents/