AGILE ENTERPRISE PARTNER

Gouvernance, AI Act & Agentic Constitution

Du cadre juridique à la gouvernance opérationnelle

Enjeu 2 — Édition Enjeux CIO 2026-2027

Un rapport Agile Enterprise Partner

Par Sébastien Delayre

Avril 2026

1. Préambule : la grande bifurcation

En 2025, vous avez adopté l'IA. En 2026-2027, votre modèle opérationnel va décider de votre trajectoire. Selon le Gartner 2026 CIO Survey, 64 % des CIO prévoient de déployer des agents autonomes dans les 24 prochains mois. Pourtant, l'étude MIT Project NANDA *(juillet 2025)* montre que 95 % des pilotes GenAI n'ont produit aucun impact mesurable sur le P&L en 2025. La fracture se creuse.

Deux cohortes émergent. Une minorité, entre 15 et 25 % des DSI selon les études convergentes, refonde son modèle opérationnel pour accueillir l'agentique comme colonne vertébrale et gagne un facteur 2 à 10 en productivité. Une majorité superpose les agents sur un modèle pensé pour les applications et plafonne à 10-20 % de gains locaux non capitalisés.

Cette édition 2026-2027 trace les cinq lignes de fracture de cette bifurcation. Le PDF 1 traitait la refonte du Product Operating Model. Ce deuxième PDF traite la deuxième ligne, la plus juridique et la plus structurante en termes de risque : la gouvernance de l'IA agentique. Les trois autres lignes sont traitées dans des PDF dédiés *(FinOps IA, Agentic Value Streams, Leadership Transformation IA)*. Pour la vision d'ensemble, voir l'Executive Summary complet de l'édition.

2. Pourquoi votre gouvernance IA actuelle ne tient pas

L'AI Act entre en application pleine le 2 août 2026. Mais le vrai sujet n'est pas la conformité documentaire. C'est la gouvernance opérationnelle des agents IA qui agissent en autonomie dans vos systèmes. Votre gouvernance actuelle, héritée de l'ère 2018-2024, n'a pas été conçue pour ça.

2.1. La gouvernance documentaire de l'ère 2018-2024

Entre 2018 et 2024, la gouvernance IA dans les grands comptes s'est structurée autour de quatre piliers. Une charte éthique IA validée par le COMEX. Un comité IA réunissant CIO, juristes, Risk, métiers, qui valide les nouveaux cas d'usage en revue trimestrielle ou semestrielle. Une grille d'évaluation des risques par projet *(souvent inspirée de la CNIL ou du PIA RGPD)*. Une revue annuelle de conformité documentaire produite par les équipes Legal et Risk.

Ce modèle a fait sens dans l'ère pré-agentique. L'IA prédisait, classifiait, recommandait. Les outputs étaient validés par un humain avant action. Les workflows étaient déterministes et tracés. La gouvernance documentaire suffisait parce que la décision restait humaine, et parce que le rythme des déploiements permettait des revues périodiques.

Cette gouvernance a trois caractéristiques structurelles. Elle est **documentaire** *(elle produit des documents, pas des contrôles)*. Elle est **périodique** *(revues annuelles ou semestrielles)*. Elle est **centralisée** *(une cellule restreinte décide pour toute l'entreprise)*. Ces trois caractéristiques ne tiennent plus face aux agents IA.

2.2. Le choc agentique : ce qui casse dans la gouvernance documentaire

Trois ruptures rendent la gouvernance documentaire structurellement inadaptée aux agents IA. Ces ruptures sont déjà observables dans les chiffres 2025-2026.

Rupture 1 — La prolifération sans inventaire. Selon Jon Radoff *(« The State of AI Agents in 2026 », février 2026)*, les entreprises comptent désormais **144 identités non-humaines par employé humain**, mais moins de 10 % des organisations qui déploient des agents en production sont capables de les gouverner réellement. Une charte centrale ne peut pas suivre 144 identités machine par humain. Le rythme de prolifération dépasse de plusieurs ordres de grandeur la cadence de revue traditionnelle.

Rupture 2 — La Shadow AI déjà installée. Selon Gartner, **42 % des entreprises** ont des cas d'usage d'IA générative non gouvernés *(Shadow AI)*. Le rapport IBM Cost of a Data Breach 2025 cité dans Cyberbase *(2026)* établit que **la Shadow AI a contribué à 20 % des data breaches en 2025**, avec un surcoût moyen de 670 000 dollars par incident. La Shadow AI n'est plus un risque émergent, c'est une réalité installée que la gouvernance documentaire n'a pas vue venir.

Rupture 3 — L'agent qui agit, pas qui prédit. Comme analysé dans le PDF 1 de cette édition, les agents IA passent de *« tools »* à *« actors »* selon le California Management Review *(mars 2026)*. Un agent qui agit en autonomie ne peut pas être validé en comité ex-ante. Il doit être contrôlé en continu, par construction. La gouvernance par approbation cède la place à une gouvernance par contraintes embarquées.

À ces trois ruptures, nos audits terrain ajoutent un quatrième constat. Sur un échantillon de DSI grands comptes auditées en 2025, le nombre médian de **zombie agents** *(agents déployés en pilote, jamais correctement industrialisés, restant en production sans propriétaire clair)* est de 23 par DSI. Ces zombie agents échappent à la gouvernance documentaire parce qu'ils n'apparaissent pas dans les inventaires, et parce que personne ne porte la responsabilité de les revoir.

2.3. Le double calendrier 2026 : AI Act + accélération des sanctions

Au-delà des ruptures structurelles, deux calendriers convergent en 2026 et imposent une bascule rapide.

Calendrier 1 — L'AI Act. Le règlement européen 2024/1689, entré en vigueur le 1er août 2024, voit ses obligations principales s'appliquer **le 2 août 2026**. Les obligations pour les fournisseurs de modèles General Purpose AI sont déjà actives depuis le 2 août 2025. Au 2 août 2026, les obligations sur les systèmes IA à haut risque *(Annexe III)* deviennent applicables, ainsi que les obligations de transparence *(Article 50)*, l'enregistrement public *(Article 49)*, et un large éventail de pouvoirs d'exécution.

Une note importante. Le **Digital Omnibus** publié par la Commission le 19 novembre 2025 propose de reporter cette deadline au 2 décembre 2027. Le second trilogue politique du 28 avril 2026 entre Parlement, Conseil et Commission s'est terminé sans accord. Selon DLA Piper *(GENIE, mai 2026)*, en l'absence d'adoption formelle de l'Omnibus avant le 2 août 2026, **les dispositions originales s'appliqueront à cette date**. Selon Modulos *(avril 2026)*, le scénario *« pas d'accord avant le 2 août »* est désormais évalué à 30 % de probabilité contre 10 % avant le trilogue manqué. Notre recommandation : continuez à préparer pour le 2 août 2026. Tout report serait un bonus, pas un acquis.

Les sanctions sont structurantes. Pour les violations les plus graves, l'AI Act prévoit des amendes jusqu'à **35 millions d'euros ou 7 % du chiffre d'affaires mondial annuel**, selon le montant le plus élevé. Pour des manquements aux obligations sur les systèmes haut risque, les amendes peuvent atteindre **15 millions d'euros ou 3 % du chiffre d'affaires mondial**. Pour la fourniture d'informations incorrectes aux autorités, **7,5 millions d'euros ou 1 %**.

Calendrier 2 — L'accélération des autres juridictions. L'AI Act n'est pas isolé. Le Colorado AI Act devient applicable en juin 2026. La California SB 53 a été adoptée. Le New York RAISE Act est en cours d'adoption. Le Singapore IMDA a publié le 22 janvier 2026 le **Model AI Governance Framework for Agentic AI**, premier framework au monde explicitement dédié aux agents autonomes. La Chine, le Japon, la Corée du Sud, le Taiwan ont leurs propres cadres en application 2026-2027. Pour les groupes internationaux, la mise en conformité AI Act n'est qu'une partie du chantier.

Le diagnostic est clair. Votre gouvernance IA actuelle, conçue pour l'ère 2018-2024, ne tient plus face aux agents IA, ne respecte pas les nouvelles obligations réglementaires, et expose votre entreprise à des sanctions structurantes. Refondre cette gouvernance n'est plus une option, c'est une obligation à calendrier court.

3. État de l'art : ce que disent les régulateurs et les frameworks

Avant de présenter notre analyse AEP, cartographions ce que les régulateurs et les frameworks structurent en 2026. Cette section est délibérément opérationnelle. Nous traitons l'AI Act comme un trigger pour le CIO, pas comme une analyse juridique exhaustive *(c'est le rôle des cabinets d'avocats spécialisés)*.

3.1. AI Act : ce qui est obligatoire au 2 août 2026

L'AI Act adopte une approche par les risques. Quatre catégories.

Catégorie 1 — Risque inacceptable *(Article 5, déjà applicable depuis le 2 février 2025)*. Pratiques interdites : manipulation cognitive, scoring social, identification biométrique à distance en temps réel par les forces de l'ordre dans les espaces publics *(avec exceptions strictes)*. Pour la majorité des CIO grands comptes, cette catégorie ne concerne pas leurs déploiements directs mais peut concerner certains use cases marketing ou RH à examiner avec vigilance.

Catégorie 2 — Risque élevé *(Annexe III, applicable au 2 août 2026)*. Concerne notamment les systèmes IA utilisés en RH *(recrutement, évaluation, allocation de tâches, monitoring, promotion, licenciement)*, en éducation, en accès aux services essentiels *(crédit, assurance vie/santé)*, dans les infrastructures critiques, dans les forces de l'ordre, dans les processus démocratiques. Selon GDPR Register *(2026)*, les obligations majeures pour les systèmes haut risque sont : système de gestion des risques continu, gouvernance des données *(qualité, biais)*, documentation technique complète, journalisation, transparence, supervision humaine *(Article 14)*, exactitude et robustesse, cybersécurité.

Catégorie 3 — Risque limité *(Article 50, transparence, applicable au 2 août 2026)*. Obligations de transparence pour les chatbots, les contenus générés par IA *(deepfakes, contenu informationnel)*, et les systèmes d'identification émotionnelle ou catégorisation biométrique. L'utilisateur doit être informé qu'il interagit avec une IA. Les contenus IA doivent être identifiables.

Catégorie 4 — Risque minimal. Pas d'obligations spécifiques. Concerne la majorité des outils IA usuels *(filtres anti-spam, recommandations produits non sensibles, jeux vidéo)*.

À cette typologie s'ajoute la catégorie **General Purpose AI** *(Articles 50-55)*, applicable depuis le 2 août 2025. Concerne les fournisseurs de modèles fondations *(LLM grands modèles)*. Pour les CIO grands comptes utilisateurs *(deployers)* de tels modèles, les obligations principales sont indirectes : s'assurer que le fournisseur est conforme, conserver la documentation technique, signaler les incidents.

Pour le CIO grand compte type CAC40/SBF120, les zones d'attention prioritaires sont :

- Les systèmes RH IA *(souvent classifiés Annexe III, donc haut risque)*
- Les systèmes IA utilisés pour l'accès aux services bancaires ou assurantiels *(haut risque)*
- Les chatbots clients *(transparence Article 50)*
- Les contenus IA générés et publiés *(transparence)*
- Les agents IA en production *(traçabilité, supervision humaine, journalisation)*

3.2. Les standards qui structurent : ISO/IEC 42001, NIST AI RMF, Singapore IMDA

Trois standards complémentaires structurent la pratique de la gouvernance IA en 2026. Ils ne se substituent pas à la conformité AI Act mais en facilitent l'implémentation opérationnelle.

ISO/IEC 42001:2023 — AI Management System. Premier standard international certifiant pour la gouvernance IA. Structure inspirée d'ISO 27001 *(SMSI)*. Couvre les clauses 4 à 10 *(contexte, leadership, planification, support, opération, évaluation, amélioration)* avec une Annexe A de contrôles spécifiques IA. Selon Speeki *(mars 2026)*, l'ISO 42001 est la fondation la plus pragmatique pour structurer la gouvernance des agents : elle force l'organisation à définir des niveaux d'autonomie, des oversight checkpoints, des procédures d'arrêt et d'intervention, des mécanismes de remontée d'incidents. Les contrôles A.6 *(lifecycle)*, A.7 *(data)*, A.8 *(information)* et A.10 *(third-party)* couvrent particulièrement les enjeux agentiques.

NIST AI Risk Management Framework 1.0 *(janvier 2023, complété par AI 600-1 sur la GenAI)*. Framework volontaire américain organisé autour de quatre fonctions : Govern, Map, Measure, Manage. NIST a annoncé la publication d'un **AI Agent Interoperability Profile** au quatrième trimestre 2026. En attendant, la Cloud Security Alliance Labs propose un *« NIST AI RMF Agentic Profile »* qui comble les *« quatre gaps structurels »* du framework face à l'agentique : autonomie des décisions, persistance d'état, interactions multi-agents, capacité d'action sur les systèmes externes.

Singapore IMDA Model AI Governance Framework for Agentic AI *(22 janvier 2026)*. **Premier framework au monde explicitement dédié aux agents autonomes**. Quatre dimensions : *(1)* assess and bound the risks upfront, *(2)* make humans meaningfully accountable, *(3)* implement technical controls and processes, *(4)* enable end-user responsibility. Pour le CIO grand compte, c'est probablement le framework opérationnel le plus directement utilisable en 2026, car il a été conçu pour l'agentique, pas adapté depuis l'ère pré-agentique.

Le tableau suivant synthétise les usages typiques de ces trois standards.

Standard	Nature	Force majeure	Usage type
AI Act	Loi européenne	Pénalités financières	Conformité légale obligatoire
ISO/IEC 42001	Norme certifiante	Certification tierce-partie	Système de management IA, exigence client
NIST AI RMF	Framework volontaire US	Référence régulateur	Cadre opérationnel adaptable
Singapore IMDA MGF	Framework volontaire	Référence agentique	Spécifique aux agents autonomes

Notre observation : les DSI matures combinent les trois. ISO 42001 comme socle de management, NIST AI RMF comme cadre opérationnel, et Singapore IMDA pour la couche agentique spécifique. La conformité AI Act se construit sur ces bases.

3.3. Les frameworks émergents 2026 : Agentic Constitution, OWASP, MITRE ATLAS

Au-delà des standards et de la régulation, trois frameworks émergents structurent la pratique technique de la gouvernance des agents.

Agentic Constitution. Concept émergé d'Anthropic, formalisé pour les modèles Claude depuis 2023 et fortement enrichi en janvier 2026 avec une *« constitution »* publique de 79 pages. Selon InfoQ *(janvier 2026)*, cette constitution combine *« des principes explicites avec des guidances contextuelles »* pour cadrer comportement, raisonnement et entraînement. Le principe de constitutional AI *(« Constitutional AI: Harmlessness from AI Feedback »)* a été étendu en 2026 par des recherches académiques *(arXiv 2506.13774, Personalized Constitutionally-Aligned Agentic Superego)* qui montrent jusqu'à **98,3 % de réduction des outputs nuisibles** sur les benchmarks HarmBench et AgentHarm. Pour le CIO grand compte, l'idée n'est pas de répliquer la complexité d'Anthropic, mais d'adopter le principe : un cadre de valeurs et de règles inscrit dans le modèle ou son orchestration, pas seulement dans des documents externes.

OWASP Top 10 for Agentic Applications 2026 *(publié décembre 2025)*. Première taxonomie formelle des risques spécifiques aux agents IA autonomes. Les dix risques principaux : Goal Hijacking *(détournement d'objectif)*, Tool Misuse *(usage abusif d'outils)*, Identity Abuse, Memory Poisoning, Cascading Failures, Rogue Agents, Privilege Escalation, Information Leakage, Unsafe

Configuration, Supervisory Bypass. Référence opérationnelle pour structurer le risk management des agents.

MITRE ATLAS. Catalogue d'attaques adversariales sur les systèmes IA. Étendu en 2025-2026 pour couvrir les patterns spécifiques aux agents : prompt injection *(AML.T0051)*, agent context poisoning, configuration discovery, tool-invocation-based exfiltration. Développé en collaboration avec Zenity Labs. Référence opérationnelle pour le red-teaming des agents.

À ces trois frameworks s'ajoutent les boîtes à outils techniques des grands éditeurs *(Microsoft Agent Governance Toolkit publié en avril 2026, Microsoft Agent 365 lancé le 1er mai 2026)* qui implémentent ces principes en outillage. Nous y revenons en section 3.

4. Analyse AEP : les trois transformations structurantes

Sur la base de notre observation terrain auprès de DSI grands comptes du secteur Bancaire, des Services Financiers et de l'Assurance *(secteurs les plus exposés réglementairement et les plus avancés sur la gouvernance opérationnelle)*, complétée par les sources publiques 2025-2026, nous identifions trois transformations structurantes qui décident, individuellement et collectivement, si une gouvernance IA tient ou s'effondre face aux agents. Ces trois transformations doivent être traitées simultanément. Une seule sur les trois ne suffit pas.

4.1. Transformation 1 — De la gouvernance documentaire à la gouvernance par code

La première erreur structurelle observée chez les DSI qui plafonnent : continuer à gouverner les agents IA avec des documents et des comités. L'écart entre la vitesse des agents *(décisions à la milliseconde)* et le rythme des comités *(trimestriel ou semestriel)* est de plusieurs ordres de grandeur. La gouvernance documentaire devient théâtre. Elle rassure les comités, elle ne contrôle pas les agents.

Le test discriminant est simple. Si votre gouvernance IA produit principalement des documents *(charte, politique, comité, revue)* et peu de contrôles techniques en production, vous êtes dans la cohorte qui plafonne, indépendamment de la qualité de vos documents.

Le pattern mature 2026 est différent. Il s'appelle **policy-as-code**. Le principe : les règles, politiques et exigences de conformité de l'organisation sont converties en code lisible par

machine, pour que les systèmes IA les suivent automatiquement. Le code remplace partiellement le document. Selon Kyndryl *(février 2026)*, *« le paradoxe de l'intelligence autonome : sa valeur croît non pas quand les agents IA ont les coudées franches, mais quand les humains définissent et imposent les paramètres qui gouvernent leurs actions »*.

Comment fonctionne le policy-as-code pour les agents IA ? Le principe technique en trois temps.

Temps 1 — Expression des règles en code. Les règles métier, réglementaires et internes sont traduites en code dans un langage de policy. Trois langages dominants en 2026 : YAML pour les règles simples, **OPA Rego** *(Open Policy Agent)* pour les règles complexes, **Cedar** *(développé par AWS)* pour les patterns RBAC.

Temps 2 — Évaluation à chaque action de l'agent. Avant qu'un agent n'exécute une action *(appel d'API, modification de données, envoi d'email)*, le moteur de policy évalue l'action contre les règles. Selon le Microsoft Agent Governance Toolkit *(avril 2026)*, l'évaluation prend **moins de 0,1 milliseconde** par action, soit environ 10 000 fois plus rapide qu'un appel LLM. La latence ajoutée à l'agent est négligeable.

Temps 3 — Décision allow/deny et journalisation. Le moteur retourne une décision binaire *(autorisé / refusé)* avec un identifiant de décision unique, le contexte d'évaluation et la règle qui a déclenché. Selon Gökhan Gökalp *(« Runtime Governance for AI Agents », avril 2026)*, ces logs reconstituent une *« séquence complète des décisions de l'agent sans toucher aux logs applicatifs »*. C'est la base d'un audit trail natif.

Les bénéfices observables sont structurants. Selon Nexastack *(janvier 2026)*, les organisations qui adoptent le policy-as-code voient typiquement **40 à 70 % de réduction des coûts de conformité**, tout en améliorant l'efficacité de la gouvernance. Les organisations matures gèrent **plus de 50 000 agents autonomes** avec une politique commune, ce qui serait inatteignable en gouvernance documentaire.

Patterns d'implémentation et solutions qui les portent.

Le pattern *« moteur de policy comme proxy entre l'agent et ses outils »* est implémenté par plusieurs solutions. **Open Policy Agent (OPA)** est le projet CNCF de référence, déployé en production chez Netflix, Goldman Sachs, Google Cloud, T-Mobile. Il fournit un langage Rego puissant et un moteur d'évaluation indépendant. **Strata Maverics AI Identity Gateway** intègre un moteur OPA pour évaluer les politiques sur les appels MCP *(Model Context Protocol)* à temps de requête. **Microsoft Agent Governance Toolkit** *(avril 2026)*, open source sous licence MIT, supporte YAML, OPA Rego et Cedar, et couvre les 10 risques OWASP Agentic Top 10 avec une enforcement deterministe sub-milliseconde. **Microsoft Agent 365** *(lancé le 1er mai 2026)* est un control plane qui découvre, gouverne, sécurise et retire chaque agent IA du tenant Microsoft, indépendamment de qui l'a construit ou de quel framework il utilise.

Le pattern *« agentic governance avec Guardian Agent »* est implémenté par **Kyndryl Agentic AI Framework**, qui embarque le policy-as-code directement dans le framework agentique avec un Guardian Agent qui évalue continuellement la qualité, l'alignement risque et la conformité des politiques avant promotion en production.

Pour le CIO grand compte, la décision n'est pas le choix d'une solution, c'est l'adoption du paradigme. Les solutions évoluent rapidement. Le paradigme policy-as-code est durable et il commence par une décision : votre gouvernance des agents sera codée, pas documentée. Les documents continuent d'exister *(charte, politique de haut niveau)*, mais ils ne suffisent plus. Ils s'incarnent dans du code.

4.2. Transformation 2 — Out-of-process enforcement et hiérarchie d'autonomie

La deuxième transformation structurante touche l'architecture du contrôle. Deux questions structurent ce sujet. **Qui contrôle l'agent : l'agent lui-même ou un système externe ?** Et **quelle est la borne d'autonomie de l'agent : peut-il décider seul ou doit-il escalader ?**

Le piège du contrôle in-process. Une approche courante consiste à demander au modèle lui-même de respecter les règles, via le system prompt ou des instructions intégrées : *« suis ces règles, ne fais pas ceci »*. Selon les tests red-team du Microsoft Agent Governance Toolkit *(avril 2026)*, cette approche a un **taux de violation de 26,67 %** sous attaques adversariales. La sécurité par instruction prompt est fragile, manipulable, contournable.

Le pattern out-of-process enforcement. Le contrôle ne peut pas être délégué à l'agent. Il doit être effectué par un système externe à l'agent, en dehors de son processus d'exécution, sur lequel l'agent n'a pas de contrôle direct. Le moteur de policy *(OPA, Cedar, AGT)* est par construction out-of-process. Il intercepte les actions de l'agent **avant exécution**, applique les règles, autorise ou refuse, journalise. Selon le Microsoft Agent Governance Toolkit, le taux de violation de cette approche est de **0,00 %** sous les mêmes tests red-team. La différence n'est pas marginale, elle est structurelle.

Ce pattern s'appuie sur un principe architectural : *« le moteur de policy comme containment boundary où chaque invocation d'outil doit passer une évaluation policy avant d'atteindre un service en aval »*. Les agents ne décident pas ce qui est autorisé. Le moteur de policy le fait, sur la base des règles définies par les humains.

La hiérarchie d'autonomie : Tier 1, Tier 2, Tier 3. Les agents IA n'ont pas tous la même autonomie. Une hiérarchie d'autonomie, classique dans les systèmes télé-opérés *(framework Sheridan & Verplank, MIT 1978)*, structure le contrôle. Adaptée aux agents IA en 2026, elle distingue trois tiers principaux *(certains frameworks utilisent quatre tiers, le principe est identique)*.

Tier 1 — Agent Assistant. L'agent suggère, l'humain décide et agit. Cas typique : copilote de code qui propose des suggestions que le développeur valide. Risque structurel faible. Gouvernance légère : journalisation, opt-out utilisateur, mesure de l'usage. La gouvernance documentaire suffit à ce niveau.

Tier 2 — Agent Constrained. L'agent décide et agit dans un périmètre prédéfini, l'humain supervise et peut intervenir. Cas typique : agent qui répond aux tickets de support de niveau 1, escalade au niveau 2 humain pour les cas complexes. Risque structurel modéré. Gouvernance moyenne : policy-as-code obligatoire, journalisation détaillée, oversight checkpoints, kill switch accessible.

Tier 3 — Agent Autonomous. L'agent décide et agit en autonomie sur des périmètres larges, l'humain est *« above the loop »*, il intervient seulement sur les exceptions ou en cas d'incident. Cas typique : agent qui orchestre un workflow KYC complet *(comme dans le cas McKinsey de la grande banque mondiale, 10 squads de 4-5 agents)*. Risque structurel élevé. Gouvernance lourde : Agentic Constitution, guardrail agents, out-of-process enforcement systématique, supervision humaine renforcée, audit continu, plan de réponse incident.

Selon Tech Jacks Solutions *(« Agent Governance Stack », mars 2026)*, *« le stack de gouvernance ne s'applique pas uniformément. Il scale avec l'autonomie. »* Plus l'autonomie est élevée, plus la gouvernance doit être stricte et instrumentée. Le mappage entre Tier d'autonomie et niveau de gouvernance est l'une des décisions architecturales structurantes pour le CIO grand compte.

Patterns d'implémentation et solutions qui les portent.

Le pattern *« guardrail agent qui supervise les autres agents »* est implémenté par **Anthropic Claude Constitutional AI**, qui forme le modèle à respecter une constitution intégrée dès l'entraînement. Le pattern est aussi implémenté par **Galileo Guardrails**, qui propose des agents superviseurs qui interceptent en temps réel les actions à haut risque.

Le pattern *« Agentic Mesh avec Inter-Agent Trust Protocol »* est implémenté dans le Microsoft Agent Governance Toolkit, qui propose une identité cryptographique *(DIDs Ed25519)*, un protocole IATP pour la communication agent-à-agent sécurisée, et un trust score dynamique de 0 à 1000 avec cinq tiers comportementaux.

Le pattern *« kill switch et execution rings »* est implémenté par AGT inspiré des CPU privilege levels, avec saga orchestration pour les transactions multi-étapes et kill switch d'urgence.

Pour le CIO grand compte, la séquence opérationnelle est la suivante : *(1)* cartographier vos agents par Tier d'autonomie, *(2)* définir le niveau de gouvernance par Tier, *(3)* déployer l'out-of-process enforcement sur les agents Tier 2 et Tier 3 en priorité, *(4)* prévoir un guardrail agent par domaine sensible, *(5)* documenter le tout dans une Agentic Constitution opérationnelle.

4.3. Transformation 3 — Le trio CIO-CISO-CDO comme structure de gouvernance

La troisième transformation est la moins technique et la plus structurante à long terme. Elle touche les rôles et la responsabilité. Le CIO seul ne peut plus porter la gouvernance IA. Le sujet est trop transverse, trop technique, trop juridique, trop sensible aux données. Il nécessite un trio formellement constitué : CIO, CISO, CDO.

Pourquoi le CIO seul ne peut pas porter la gouvernance IA. Trois raisons.

Raison 1 — La sécurité des agents est un sujet à part entière. Selon Cyberbase *(« 10 Biggest CISO Challenges 2026 »)*, **79 % des organisations exécutent ou planifient des AI agents, mais seulement 6 % ont mis à jour leur framework de gouvernance pour ces agents**. Et **65 % admettent que le déploiement des agents a dépassé leur compréhension**. Pour un CISO de Fortune 500 cité par Obfuscated *(mars 2026)* : *« le but n'est plus d'arrêter l'usage des agents, mais d'assurer qu'ils opèrent dans un Trust Sandbox défini. Si vous ne pouvez pas auditer la logique d'un agent, vous ne devriez pas l'avoir sur votre réseau »*. Cette responsabilité structurelle ne peut pas être déléguée par le CISO au CIO.

Raison 2 — La qualité des données est la condition de la conformité IA. Selon le rapport HotTopics *(mars 2026)*, *« le rôle du CDO est de superviser la lignée et la qualité des données, transformant les données brutes en IA responsable, un actif qui minimise le biais algorithmique et maximise le ROI stratégique »*. La conformité AI Act sur la qualité des données *(article 10)* est portée structurellement par le CDO, pas par le CIO.

Raison 3 — Les arbitrages sont trop complexes pour un seul rôle. Un déploiement d'agent type implique : architecture *(CIO)*, sécurité d'accès et résistance aux attaques *(CISO)*, qualité et lignée des données utilisées *(CDO)*, conformité AI Act *(CIO + Legal)*, ROI métier *(CIO + métiers)*. Vouloir confier cela à un seul rôle, c'est soit le surcharger et risquer le goulot d'étranglement, soit le vider du sujet en le déléguant.

Le trio CIO-CISO-CDO en pratique. Notre observation terrain, complétée par les sources publiques *(HotTopics, Riviera Partners, Digital Chiefs 2026)*, fait émerger trois principes opérationnels.

Principe 1 — Une responsabilité partagée mais distinguée. Le CIO est le primary steward du stack technique IA et de l'infrastructure. Le CISO est responsable du paysage des menaces sur le cycle de vie IA *(prompt injection, data poisoning, agent identity abuse)*. Le CDO supervise la lignée et la qualité des données. La RACI 2026 doit refléter cette distribution.

Principe 2 — Un comité structurel, pas projet. Le trio se réunit selon un rythme régulier *(notre recommandation : mensuel a minima, hebdomadaire en phase de déploiement intensif)*, avec un agenda standardisé : revue des agents en production *(inventaire, maturité, incidents)*, revue des nouveaux déploiements *(autorisation par Tier d'autonomie)*, revue des incidents et near-misses, revue des évolutions réglementaires.

Principe 3 — Une instrumentation commune. Le trio partage un dashboard de gouvernance IA. Selon le Microsoft Agent 365 *(mai 2026)*, ce dashboard couvre typiquement : agent registry *(inventaire complet, incluant Shadow AI découverte)*, access control via Entra ID *(ou équivalent)*, activity visualization en temps réel, security enforcement *(DLP, audit logs)*. Sans ce dashboard partagé, le trio se réduit à une réunion de coordination sans levier opérationnel.

Sur le débat *« faut-il un Chief AI Officer (CAIO) à la place du trio ? »*, notre position est différenciée. Selon Tech Jacks Solutions *(mars 2026)*, **48 % des FTSE 100 ont un CAIO ou équivalent, dont 65 % nommés dans les deux dernières années**. Mais selon Riviera Partners et Digital Chiefs *(2026)*, *« le CAIO a du sens uniquement pour les entreprises dont le business model est AI-native. Pour la plupart des entreprises traditionnelles, il reste organisationnellement disproportionné en 2026 »*. Notre recommandation pour les CIO grands comptes : commencer par structurer formellement le trio CIO-CISO-CDO. Si le besoin d'un CAIO émerge ensuite *(typiquement après 18-24 mois de maturation)*, l'évolution est naturelle. Sauter directement à un CAIO sans avoir d'abord structuré le trio crée souvent un silo supplémentaire plutôt qu'une coordination renforcée.

L'opérationnalisation des dimensions Leadership et conduite du changement de cette transformation est traitée en profondeur dans le PDF 5 de cette édition *(Leadership Transformation IA et CIO Agent-Enabler)*. À ce stade, retenir que la structuration du trio est une condition nécessaire, pas suffisante, mais sans laquelle aucune gouvernance opérationnelle de l'agentique ne tient à 12-18 mois.

5. Recommandations actionnables

5.1. Diagnostiquer votre maturité gouvernance : 10 questions

Avant de définir une trajectoire, mesurez votre point de départ. Les dix questions suivantes constituent une grille d'auto-évaluation. Une réponse *« non »* vaut un point de fragilité. Plus de cinq points de fragilité signalent une gouvernance qui ne tiendra pas le 2 août 2026.

#	Question	Levier concerné
1	Avez-vous un inventaire à jour de tous vos systèmes IA, incluant Shadow AI ?	Diagnostic
2	Avez-vous classifié chaque système selon les catégories AI Act *(Annexe III, GPAI)* ?	Conformité
3	Vos agents IA sont-ils mappés à des Tiers d'autonomie *(1, 2, 3)* ?	Architecture
4	Avez-vous une plateforme policy-as-code en production sur les agents Tier 2-3 ?	Transformation 1
5	Vos contrôles sont-ils out-of-process *(externe à l'agent)* ou dans le system prompt ?	Transformation 2
6	Avez-vous une Agentic Constitution v1, codée et opérationnelle ?	Transformation 1
7	Le trio CIO-CISO-CDO est-il formellement constitué et se réunit-il a minima mensuellement ?	Transformation 3

8	Avez-vous un dashboard de gouvernance IA partagé *(agent registry + activity)* ?	Transformation 3
9	Avez-vous un plan de réponse aux incidents agents *(rogue, hijacking, leak)* ?	Réactivité
10	Êtes-vous en conformité ISO/IEC 42001 ou en cours de certification ?	Standardisation

5.2. Choisir votre trajectoire de mise en conformité : trois scénarios

L'erreur fréquente est de choisir une trajectoire avant de connaître son point de départ. Le diagnostic doit précéder la trajectoire. Sur la base de notre observation terrain auprès de DSI grands comptes, trois trajectoires-types se dégagent.

Trajectoire A — Compliance minimale au 2 août 2026. Objectif : être conforme aux obligations légales sans plus, en s'appuyant sur la documentation existante et un effort technique limité. Recommandée si votre exposition AI Act est faible *(peu ou pas de systèmes Annexe III)* et si votre maturité IA est en début de courbe. Risque : cette trajectoire ne tient pas à 18-24 mois quand l'exposition agentique augmente. Bénéfice : effort court terme limité.

Trajectoire B — Conformité opérationnelle. Objectif : passer du cadre documentaire à un cadre opérationnel codé en 12 mois, conforme au 2 août 2026 et scalable. Trois étapes : *(1)* conformité minimale au 2 août, *(2)* déploiement policy-as-code et out-of-process sur les agents Tier 2-3, *(3)* certification ISO/IEC 42001. Recommandée pour la majorité des CIO grands comptes avec exposition AI Act significative. Bénéfice : combine sécurité juridique et capacité opérationnelle. Limite : effort soutenu sur 12 mois.

Trajectoire C — Leadership de gouvernance. Objectif : devenir une référence sectorielle de gouvernance IA, en capitalisant sur la conformité pour bâtir un avantage compétitif *(trust score auprès des clients, capacité de déploiement rapide d'agents en production, attractivité des talents)*. Recommandée pour les groupes très exposés *(banque, assurance, santé)* et avec ambition stratégique sur l'IA. Inclut les éléments de la trajectoire B plus : Agentic Constitution publique, certification ISO 42001 visible, communications externes, partenariats de recherche. Bénéfice : différenciation. Limite : investissement plus élevé, exposition publique du chantier.

Notre observation : sur 100 CIO grands comptes, environ 30 % choisissent A *(souvent à tort, la trajectoire est rattrapable mais coûteuse plus tard)*, 60 % choisissent B *(le bon choix dans la majorité des cas)*, 10 % choisissent C *(souvent à raison pour les leaders sectoriels)*. La trajectoire A se retrouve fréquemment dans la cohorte qui plafonne, la trajectoire B et C dans la cohorte minoritaire qui réussit la bascule.

5.3. Les pièges à éviter

Cinq pièges récurrents sabotent les programmes de gouvernance IA. Ils ne sont pas hypothétiques. Ils sont documentés dans la majorité des projets en difficulté.

Piège 1 — La gouvernance théâtre. Charte éthique IA validée en COMEX, comité IA réuni trimestriellement, revues annuelles documentaires. Aucun contrôle technique en production. Cela rassure les comités, ne contrôle pas les agents. Symptôme : votre dernière revue de gouvernance n'a identifié aucun zombie agent ni Shadow AI alors que les chiffres macro suggèrent que vous en avez. Antidote : passer au policy-as-code et à la découverte automatique d'agents.

Piège 2 — L'overengineering juridique. À l'inverse, certains DSI sur-réagissent à l'AI Act et déploient des process si lourds qu'aucun agent ne sort en production *(approbation par 7 comités, documentation 200 pages par cas d'usage)*. Selon Gartner, **35 % du budget IT annuel** peut être consommé par les dérives administratives non maîtrisées sur l'IA. Antidote : graduer la gouvernance par Tier d'autonomie *(légère sur Tier 1, lourde uniquement sur Tier 3)*.

Piège 3 — Le false comfort des certifications. ISO 42001 ou autre certification obtenue sur le périmètre symbolique, sans couverture réelle des agents en production. La certification rassure le board et le client, mais ne couvre pas le risque opérationnel. Antidote : certifier sur le périmètre réel d'usage, pas sur un périmètre vitrine.

Piège 4 — La gouvernance déconnectée du POM. La gouvernance IA est portée par une cellule centrale *(CIO Office, Risk, Legal)* alors que les agents sont déployés dans les squads produit. La cellule centrale produit des règles que personne n'applique en production. Antidote : structurer le policy-as-code dans les squads, pas en central. Le cadre est central, l'application est distribuée.

Piège 5 — La sous-estimation du Shadow AI. *« Nous n'avons pas de Shadow AI, nous avons une politique stricte »*. Selon Gartner, 42 % des entreprises ont du Shadow AI **malgré leur politique**. Antidote : déployer un agent registry à découverte automatique *(type Microsoft Agent 365)*, accepter que la première découverte multipliera votre inventaire connu par 2 à 5, et structurer la régularisation plutôt que la sanction.

6. Cas et retours d'expérience publics

6.1. Cas composite secteur Bancaire

Cas illustratif construit à partir de sources publiques sur le secteur bancaire européen et de notre observation terrain auprès de DSI bancaires, anonymisé conformément aux engagements de confidentialité d'AEP.

Un grand groupe bancaire européen, plusieurs dizaines de millions de clients, présence dans plus de 30 pays, plus de 100 000 collaborateurs. La DSI est structurée autour de plateformes de groupe et de filiales métier *(banque de détail, banque de financement et d'investissement, banque privée, asset management, services aux entreprises)*. Cette diversité rend la gouvernance IA particulièrement complexe : exigences réglementaires différenciées par juridiction et métier, maturité IA hétérogène entre filiales, exposition agentique forte sur les fonctions support *(KYC, AML, compliance, customer support)*.

Situation initiale 2024-2025. Charte éthique IA validée en 2023, comité IA Groupe trimestriel, premières expérimentations agentiques sur le KYC et le support client. Multiplication des cas d'usage par filiale, sans cadre opérationnel commun. Premiers déploiements agents en production *(KYC automation comme dans le cas McKinsey traité en PDF 1, customer-facing chatbots)*. Audit interne 2025 identifie : 47 % des cas d'usage IA générative non répertoriés au niveau Groupe *(Shadow AI)*, plus de 30 zombie agents identifiés sur le périmètre central, **aucun mécanisme de policy-as-code** en production sur les agents.

Démarche 2025-2026. Lancement d'un programme de transformation gouvernance IA structuré en quatre chantiers parallèles. *(1)* Constitution formelle du trio CIO-CISO-CDO Groupe avec rituels mensuels et dashboard partagé. *(2)* Déploiement d'un agent registry centralisé pour découvrir et inventorier l'ensemble des agents *(découverte multipliant l'inventaire connu par 3,2)*. *(3)* Mise en place d'une plateforme policy-as-code *(OPA + Cedar selon les domaines)* en out-of-process enforcement sur les agents Tier 2-3. *(4)* Lancement de la démarche de certification ISO/IEC 42001 sur le périmètre des fonctions support, avec extension progressive prévue en 2026-2027.

Architecture observée. Le groupe combine policy-as-code central avec déploiement distribué dans les squads produit. La plateforme OPA est gérée par le CIO Office. Les politiques sont co-construites entre Compliance, Legal, Risk et les domaines métier. Le déploiement et la maintenance des règles sont distribués dans les squads via leurs propres pipelines DevSecOps. Le trio CIO-CISO-CDO arbitre les conflits et valide les évolutions structurantes.

Résultats observables après 9 mois. Réduction de 73 % du nombre de zombie agents identifiés *(par retrait ou régularisation)*. Mise en place d'audit trails systématiques sur tous les agents Tier 2-3 *(plus de 50 millions de décisions journalisées par mois)*. Conformité AI Act validée par le cabinet d'avocats externe sur le périmètre haut risque. Coût total du programme inférieur à 2 % du budget IT Groupe.

Enseignements pour un CIO grand compte. Trois leçons.

Premièrement, l'inventaire est l'étape la plus sous-estimée. La découverte multiplie souvent l'inventaire connu par 2 à 5. Cette étape doit être gérée comme un projet à part, pas comme un préalable expédié.

Deuxièmement, la combinaison central + distribué est la seule qui scale. Une gouvernance 100 % centrale freine. Une gouvernance 100 % distribuée éclate. Le bon équilibre est : règles définies centralement *(cadre + policy-as-code engine)*, application distribuée dans les squads.

Troisièmement, le ROI de la gouvernance IA n'est pas la conformité, c'est la vélocité. Une fois le cadre opérationnel en place, les nouveaux agents sortent en production en semaines, pas en mois. La gouvernance bien outillée accélère, elle ne freine pas.

6.2. Cas public Singapore IMDA et Anthropic Constitutional AI

Cas documenté à partir des publications publiques de l'IMDA *(Model AI Governance Framework for Agentic AI, 22 janvier 2026)* et d'Anthropic *(Claude Constitution, mise à jour publique 21 janvier 2026)*.

Singapore IMDA — Model Governance Framework Agentic AI. Lancé en parallèle au Forum économique mondial de Davos *(22 janvier 2026)* par la ministre du développement digital Josephine Teo. **Premier framework au monde explicitement dédié aux agents autonomes**. Quatre dimensions opérationnelles : *(1)* assess and bound the risks upfront, *(2)* make humans meaningfully accountable, *(3)* implement technical controls and processes, *(4)* enable end-user responsibility. Disponible en téléchargement libre sur le site de l'IMDA.

Pour le CIO grand compte, le framework Singapore IMDA est **probablement la référence opérationnelle la plus directement utilisable en 2026**, parce qu'il a été conçu pour l'agentique, pas adapté depuis l'ère pré-agentique. Il complète l'AI Act *(qui reste prééminent juridiquement en Europe)* sur les dimensions opérationnelles. Plusieurs grandes banques européennes que nous accompagnons utilisent désormais le framework IMDA comme guide d'implémentation pratique de leurs obligations AI Act.

Anthropic Constitutional AI. Concept de constitution publique pour les modèles Claude, formalisé depuis 2023 et fortement enrichi en janvier 2026 avec une « *constitution* » publique de

79 pages. La constitution combine principes explicites *(safety, ethics, honesty)* avec guidance contextuelle. Elle est intégrée dès l'entraînement du modèle, pas seulement dans le system prompt. Selon les benchmarks publiés *(HarmBench, AgentHarm)*, l'approche Constitutional AI atteint **jusqu'à 98,3 % de réduction des outputs nuisibles**, avec des taux de refus quasi-parfaits *(100 % avec Claude Sonnet 4 sur AgentHarm harmful set)*.

Enseignements pour un CIO grand compte. Deux leçons.

Premièrement, la transparence des règles est un actif, pas un risque. Anthropic publie sa constitution. Singapore publie son framework. Les organisations qui adoptent cette approche bénéficient d'un effet de levier *(les régulateurs et les clients valorisent la transparence)* qu'une gouvernance opaque ne procure pas.

Deuxièmement, la séparation entre constitution *(valeurs et règles)* et implémentation *(modèle, agent, application)* permet de faire évoluer l'un sans casser l'autre. Une constitution v1 peut être révisée en v2 sans changement de modèle. Un modèle peut être remplacé sans perte de la constitution.

6.3. Cas public secteur bancaire — Bank of England et FCA AI Risk Survey

Cas documenté à partir du *« BoE/FCA AI Risk Survey 2025 »* repris par CoreAdmin AI *(2026)*, complété par Capgemini Cloud Report Financial Services 2026 et Accenture Banking Top Trends 2026.

Selon le BoE/FCA AI Risk Survey, les trois plus grands risques actuels cités par les institutions financières britanniques sur l'IA sont : data privacy, data quality et data security. La transparence et l'explicabilité restent **parmi les contraintes les plus structurantes sur l'adoption**. Selon Capgemini *(World Cloud Report Financial Services 2026)*, *« seulement 10 % des institutions ont implémenté l'IA agentique à l'échelle »*, alors que la majorité planifie le déploiement.

Plusieurs banques européennes ont publié leur démarche. Sur le KYC : 99 % de réduction du temps d'ingestion sur les cas correspondent banking complexes, 94 % de réduction des coûts sur les tâches d'extraction *(McKinsey via Accenture, février 2026)*. Sur la grande banque néerlandaise : 90 % de réduction du temps d'onboarding et 30 % de réduction de la charge des équipes. Sur Morgan Stanley : 98 % d'adoption de l'AI Assistant par les conseillers, *« emphasis on evaluations and controls »* selon le case study OpenAI.

Enseignements pour un CIO grand compte. Trois leçons.

Premièrement, dans le secteur bancaire, la conformité n'est pas une charge, c'est l'autorisation d'opérer. Les banques qui ont anticipé ont gagné un avantage concurrentiel sur la vitesse de déploiement. Celles qui ont attendu la régulation pour structurer ont perdu plusieurs trimestres.

Deuxièmement, l'audit trail natif des agents est valorisé par les régulateurs. La banque mondiale du cas McKinsey a souligné que l'output le plus valorisé n'est pas l'accélération mais la richesse de l'audit trail. C'est un changement de nature : la conformité ne se documente plus, elle se trace nativement.

Troisièmement, la transparence vis-à-vis du régulateur paie. Les banques qui ont engagé un dialogue précoce avec leur régulateur national *(BCE, ACPR en France, BaFin en Allemagne, BoE/FCA au Royaume-Uni)* ont obtenu des périmètres de tolérance plus larges et une compréhension plus fine des attentes. La gouvernance opérationnelle facilite ce dialogue.

6.4. Synthèse : cinq facteurs clés de succès observés

Les trois cas convergent sur cinq facteurs clés de succès de la gouvernance IA mature.

Facteur 1 — Inventaire d'abord. Aucun programme ne réussit sans avoir établi un inventaire exhaustif initial, incluant la découverte de Shadow AI. La découverte multiplie typiquement par 2 à 5 l'inventaire connu.

Facteur 2 — Trio CIO-CISO-CDO formalisé. La gouvernance par un seul rôle ne tient pas. Les trois doivent partager cadre, métriques, rituels, dashboard.

Facteur 3 — Policy-as-code en out-of-process. Le contrôle in-process *(via system prompt)* a un taux de violation de 26,67 %. Le contrôle out-of-process *(via policy engine)* atteint 0,00 %. La différence n'est pas marginale.

Facteur 4 — Gouvernance graduée par Tier d'autonomie. Une gouvernance uniforme freine *(trop lourde sur Tier 1)* ou expose *(trop légère sur Tier 3)*. La graduation est la seule approche scalable.

Facteur 5 — Dialogue précoce avec les régulateurs. Engager le dialogue avant la mise en application crée des marges d'interprétation et une crédibilité qui ne s'obtiennent pas après coup.

7. Plan d'action 2026 : 6 mois pour être prêt

Sur la base des trois transformations structurantes et des recommandations, voici un plan d'action calé sur l'échéance du 2 août 2026 et étendu jusqu'à février 2027 pour la phase de consolidation. Chaque jalon est associé à des actions concrètes et à un point de contrôle structuré qui valide l'avancement avant passage au jalon suivant.

Jalon	Période	Objectif	Livrables
Jalon 1	Mai 2026	Diagnostic complet	Inventaire exhaustif, classification AI Act, mapping Tiers
Jalon 2	Juin-juillet 2026	Cadre de gouvernance	Trio constitué, Constitution v1, choix policy-as-code
Jalon 3	Juillet-août 2026	Mise en œuvre minimale	Out-of-process Tier 1, doc AI Act, communication
Jalon 4	Septembre-novembre 2026	Industrialisation	Extension policy-as-code, guardrail agents, ROI
Jalon 5	Décembre 2026 - février 2027	Maturité	Audit interne, plan réponse, Constitution v2

7.1. Jalon 1 — Diagnostic *(mai 2026)*

Objectif. Établir une vue exhaustive et factuelle du périmètre IA de l'entreprise. Pas de jugement à ce stade, juste un inventaire.

Actions clés.

- Cartographier tous les systèmes IA en production *(et en pilote avancé)*
- Déployer un agent registry à découverte automatique pour identifier la Shadow AI
- Identifier les zombie agents *(déployés sans propriétaire, sans valeur business identifiable)*
- Classifier chaque système selon les catégories AI Act *(prohibé, haut risque, transparence, minimal, GPAI)*
- Mapper les agents existants à des Tiers d'autonomie *(1, 2, 3)*

Point de contrôle 1 — Avez-vous une vue exhaustive ?

Question	Critère de validation
Inventaire complet	Tous les agents identifiés, y compris Shadow AI découverte
Classification AI Act	Chaque système associé à une catégorie réglementaire
Mapping Tiers	Tous les agents Tier 1, Tier 2 ou Tier 3
Zombie agents identifiés	Liste avec décision *(retrait / régularisation / propriétaire)*
Risques majeurs cartographiés	Top 10 risques par criticité documenté

Si vous ne pouvez pas valider ces 5 critères, ne passez pas au jalon suivant. La suite repose sur la qualité de l'inventaire.

7.2. Jalon 2 — Cadre de gouvernance *(juin-juillet 2026)*

Objectif. Mettre en place les structures *(humaines et techniques)* qui porteront la gouvernance opérationnelle.

Actions clés.

- Constituer formellement le trio CIO-CISO-CDO avec charte de fonctionnement *(rythme, agenda, livrables)*
- Rédiger l'Agentic Constitution v1 *(valeurs, principes, règles transverses)*
- Définir précisément les niveaux d'autonomie autorisés par catégorie d'usage
- Choisir une plateforme policy-as-code *(OPA, Cedar, AGT, ou solution éditeur intégrée)*
- Lancer la démarche de certification ISO/IEC 42001 *(au minimum gap analysis)*

Point de contrôle 2 — Votre cadre tient juridiquement ?

Question	Critère de validation
Trio formalisé	Charte signée, premier comité tenu, dashboard partagé en place
Constitution v1	Document validé en COMEX, codification engagée

Plateforme policy-as-code	Choix arrêté, POC technique réussi sur un cas représentatif
Validation juridique	Avis du cabinet d'avocats externe sur le cadre
Communication interne	Plan de communication validé, premières actions lancées

7.3. Jalon 3 — Mise en œuvre minimale *(juillet-août 2026)*

Objectif. Être conforme au minimum légal à la date du 2 août 2026 *(ou à la date effective si l'AI Act est reporté)*. Ne pas viser la perfection, viser la conformité solide sur le périmètre obligatoire.

Actions clés.

- Déployer le policy-as-code et l'out-of-process enforcement sur **tous les agents Tier 2-3**
- Compléter la documentation technique exigée par l'AI Act pour les systèmes haut risque
- Activer la journalisation conforme aux exigences AI Act *(traçabilité, conservation)*
- Communiquer aux utilisateurs internes et externes selon les obligations Article 50
- Documenter le plan de réponse aux incidents agents *(rogue, hijacking, leak)*

Point de contrôle 3 — Conforme à minima au 2 août 2026 ?

Question	Critère de validation
Tier 2-3 protégés	100 % des agents Tier 2-3 sous policy-as-code out-of-process
Documentation AI Act	Documentation technique complète sur tous les systèmes haut risque
Journalisation conforme	Logs structurés, conservés selon les durées légales
Communication transparence	Utilisateurs informés selon Article 50
Plan incident validé	Procédure documentée, équipe de réponse identifiée, simulation faite

7.4. Jalon 4 — Industrialisation *(septembre-novembre 2026)*

Objectif. Passer de la conformité minimale à l'industrialisation. Étendre le périmètre, instrumenter la mesure, capitaliser.

Actions clés.

- Étendre le policy-as-code aux agents Tier 1 *(ne pas surinvestir, mais homogénéiser)*
- Déployer des guardrail agents sur les domaines sensibles *(finance, RH, données client)*
- Mesurer les indicateurs de gouvernance *(taux de violation détecté, incidents évités, temps de mise en production des nouveaux agents)*
- Calculer le ROI gouvernance *(coût des incidents évités, accélération du time-to-market)*
- Renforcer le dashboard partagé du trio avec les nouveaux indicateurs

Point de contrôle 4 — ROI mesurable ?

Question	Critère de validation
Couverture étendue	Plus de 90 % des agents sous policy-as-code *(tous Tiers)*
Guardrail agents en production	Au moins 1 guardrail par domaine sensible
Indicateurs publiés	Dashboard mensuel partagé au trio, rapporté au COMEX
ROI documenté	Coût gouvernance vs coût incidents évités, avec hypothèses explicites
Time-to-market mesuré	Temps moyen de mise en production des nouveaux agents en baisse

---

7.5. Jalon 5 — Maturité *(décembre 2026 - février 2027)*

Objectif. Passer du mode chantier au mode pérenne. Préparer l'audit externe, capitaliser, faire évoluer.

Actions clés.

- Conduire un audit interne préparatoire à l'audit externe *(simulation par un cabinet externe ou par l'audit interne)*
- Documenter le plan de réponse aux non-conformités *(avec délais et responsables)*
- Lancer la mise à jour de l'Agentic Constitution v2 sur la base des enseignements
- Étendre la certification ISO/IEC 42001 au périmètre cible *(si certification engagée au jalon 2)*
- Communication externe sur la démarche *(si trajectoire C — leadership de gouvernance)*

Point de contrôle 5 — Prêt pour l'audit régulateur ?

Question	Critère de validation
Audit interne réussi	Rapport d'audit interne disponible avec écarts identifiés et plans d'action
Plan non-conformité	Procédure documentée, testée sur un cas réel
Constitution v2	Document mis à jour, codification engagée
ISO 42001 *(si engagé)*	Audit de certification programmé

Maturité opérationnelle	Le trio fonctionne en autonomie, dashboard mensuel stable, incidents en baisse

7.6. Synthèse du plan d'action

Le plan est calé sur 9 mois *(mai 2026 à février 2027)*. Il combine cinq jalons séquentiels avec cinq points de contrôle structurés qui valident le passage. La logique est progressive : diagnostic, cadre, conformité minimale, industrialisation, maturité. Chaque jalon repose sur la qualité du précédent. Sauter un jalon ou expédier un point de contrôle expose à des problèmes structurels à 12-18 mois.

Notre recommandation : ne pas chercher à aller plus vite que les jalons. Chercher au contraire à valider chaque point de contrôle de manière documentée. La gouvernance IA est un actif structurel. Sa qualité se mesure sur 24 mois, pas sur 6.

8. Pour aller plus loin

8.1. Les autres lignes de fracture de l'édition 2026-2027

Ce PDF 2 a traité la deuxième ligne de fracture : la gouvernance opérationnelle des agents IA. Les autres lignes complètent la cartographie de la grande bifurcation.

PDF 1 — De l'IA à l'entreprise agentique : le Product Operating Model 2026. La refonte du POM 2018-2024 pour accueillir les agents comme acteurs des squads. Trois transformations structurantes : agents intégrés aux squads, SDLC humain-IA, mesure de la contribution hybride.

PDF 3 — FinOps IA et souveraineté de l'inférence. Comment passer d'une logique projet à une logique unit economics. Les concepts de Zombie agents, Big Model Fallacy, Unit Economics Attribution, Hybrid Consumption Architecture. L'Agent Control Plane comme infrastructure FinOps.

PDF 4 — Agentic Value Streams. Les méthodes de Strategic Portfolio Management pensées en 2018-2022 ne survivent pas au SI agentique. Cartographie ArchiMate agents-first, modélisation agent-centric, OKR avec ventilation contribution humain/IA/hybride.

PDF 5 — Leadership Transformation IA. La transformation personnelle du CIO. Le passage de CIO-contrôleur à CIO Agent-Enabler. Les trois croyances ancrées à déverrouiller.

L'Executive Summary global de l'édition donne une vue d'ensemble des cinq lignes de fracture et de leurs interactions.

8.2. L'accompagnement AEP

Agile Enterprise Partner accompagne les CIO de grands comptes dans la mise en place d'une gouvernance opérationnelle de l'IA agentique. Notre positionnement 2026 : architectes du modèle opérationnel IT et Digital pour le SI agentique. Deux offres sont directement activées sur les enjeux de ce PDF.

Offre Gouvernance IA et Conformité AI Act. Diagnostic du dispositif existant *(grille des 10 questions, audit Shadow AI, mapping Tiers)*, design du dispositif cible, choix de plateforme policy-

as-code, structuration du trio CIO-CISO-CDO, accompagnement de la certification ISO/IEC 42001. Frameworks mobilisés : AI Act, ISO 42001, NIST AI RMF, Singapore IMDA MGF, OWASP Top 10 Agentic.

Offre CIO Office IA-Ready. Mise en place de l'Agent Control Plane, structuration opérationnelle du trio, instrumentation des KPI de gouvernance, accompagnement des décisions d'architecture *(in-process vs out-of-process, choix d'éditeurs)*.

Les autres offres *(Product Operating Model, Enterprise Architecture pour SI agentique, Leadership Transformation IA, Strategic Portfolio Management)* sont activées sur les enjeux des autres PDF.

Contact : contact@agile-enterprise-partner.com — +33 6 32 54 58 92 Site : https://agile-enterprise-partner.com

9. Bibliographie

9.1. Régulation et standards

- **Règlement (UE) 2024/1689** *(EU AI Act)*. Adopté le 21 mai 2024. Entrée en vigueur 1er août 2024. Application principale 2 août 2026. Texte officiel sur EUR-Lex.
- **AI Act Service Desk**. Commission européenne. Timeline officielle. Disponible sur ai-act-service-desk.ec.europa.eu.
- **Digital Omnibus on AI**. Commission européenne, 19 novembre 2025. Trilogue politique du 28 avril 2026 *(sans accord)*.
- **DLA Piper GENIE**. *The Digital AI Omnibus: Proposed deferral of high risk AI obligations*. Mai 2026.
- **Modulos**. *Is the EU AI Act Delayed? 2026 Status Check*. Avril 2026.
- **ISO/IEC 42001:2023**. Information technology — Artificial intelligence — Management system. Standard ISO.
- **NIST AI Risk Management Framework 1.0**. Janvier 2023. Avec Generative AI Profile AI 600-1.
- **Cloud Security Alliance Labs**. *NIST AI Risk Management Framework: Agentic Profile*. Avril 2026.
- **Singapore IMDA**. *Model AI Governance Framework for Agentic AI*. 22 janvier 2026.
- **Baker McKenzie**. *Singapore: Governance Framework for Agentic AI Launched*. Janvier 2026.
- **GDPR Register**. *EU AI Act Compliance 2026 Guide*. 2026.
- **Tech Jacks Solutions**. *Agent Governance Stack: NIST AI RMF, ISO 42001, EU AI Act*. Mars 2026.

9.2. Frameworks et publications académiques

- **Anthropic**. *Claude Constitution* (mise à jour publique). 21 janvier 2026.
- **Anthropic**. *Constitutional AI: Harmlessness from AI Feedback*. 2022, complété 2025-2026.
- **Anthropic**. *Collective Constitutional AI: Aligning a Language Model with Public Input*.

- **InfoQ**. *Anthropic Releases Updated Constitution for Claude*. Janvier 2026.
- **Abiri G**. *Corporations Constitute Intelligence*. arXiv 2604.02912. Avril 2026.
- **arXiv 2506.13774**. *Personalized Constitutionally-Aligned Agentic Superego*. Juin 2025.
- **California Management Review (Berkeley CMR)**. Sandeep Saini. *Governing the Agentic Enterprise: A New Operating Model for Autonomous AI at Scale*. Mars 2026.
- **OWASP Foundation**. *Top 10 for Agentic Applications 2026*. Décembre 2025.
- **MITRE ATLAS**. Adversarial Threat Landscape for Artificial Intelligence Systems. Mises à jour 2025-2026.
- **Sheridan T.B., Verplank W.L.** *Human and Computer Control of Undersea Teleoperators*. MIT, 1978. Référence classique sur la hiérarchie d'autonomie.

9.3. Outillage technique

- **Microsoft**. *Agent Governance Toolkit*. Avril 2026. Open source MIT License.
- **Microsoft**. *Agent 365*. Annoncé mai 2026.
- **Open Policy Agent (OPA)**. Projet CNCF. opa.io.
- **Cedar**. AWS, langage de policy.
- **Strata**. *Maverics AI Identity Gateway*.
- **Codilime**. *Why Open Policy Agent is the Missing Guardrail for Your AI Agents*. Avril 2026.
- **Kyndryl**. *How AI sees more clearly with policy as code*. Février 2026.
- **Kyndryl**. *How policy as code governs AI agents*. Mars 2026.
- **Nexastack**. *Agent Governance at Scale: Policy-as-Code Approaches in Action*. Janvier 2026.
- **Altimetrik**. *From Policy as Code to Agentic Governance in the AI-First Enterprise*. Mars 2026.
- **Gökhan Gökalp**. *Runtime Governance for AI Agents: Policy-as-Code with OPA*. Avril 2026.
- **Penthara AI**. *Agent 365 Uncovered: The 7 Governance Gaps*. Mai 2026.

9.4. Études primaires et REX 2025-2026

- **Gartner**. *2026 CIO and Technology Executive Survey*. Octobre 2025. 2 500 dirigeants IT.
- **Gartner**. *2026 Hype Cycle for Agentic AI*. Janvier 2026.
- **MIT Project NANDA**. *The GenAI Divide: State of AI in Business 2025*. Juillet 2025.
- **IBM**. *Cost of a Data Breach Report 2025*. Cité dans Cyberbase 2026.
- **IBM Institute for Business Value**. *AI Governance and Leadership Survey*. Q1 2025. 2 300 organisations avec Oxford Economics et Dubai Future Foundation.
- **Bank of England + FCA**. *AI Risk Survey 2025*. Cité dans CoreAdmin AI 2026.

- **Capgemini Research Institute**. *World Cloud Report Financial Services 2026*.
- **Accenture**. *Banking Top Trends 2026: Unconstrained banking is here*. Février 2026.
- **Cyberbase**. *The 10 Biggest CISO Challenges in 2026*. 2026.
- **CrowdStrike**. *2026 Global Threat Report*. 2026.
- **DataIQ**. *2025 AI and Data Leadership Executive Benchmark Survey*. 2025.
- **Riviera Partners**. *CIO vs. CTO vs. CDO: Who Should Own Intelligence Now?*. 2025.
- **Jon Radoff**. *The State of AI Agents in 2026*. Février 2026.
- **CoreAdmin AI**. *AI Agents Are Becoming the New Operating Layer for Financial Administration*. 2026.

9.5. Analyses sectorielles 2025-2026

- **Computer Weekly**. *Singapore debuts world's first governance framework for agentic AI*. Janvier 2026.
- **Stephenson Harwood**. *Launch of a New Model AI Governance Framework for Agentic AI*. Mars 2026.
- **AI Asia Pacific Institute**. *Governing AI That Acts: Singapore's New Framework for Agentic AI*. Janvier 2026.
- **Bird & Bird**. *Singapore Introduces New Model AI Governance Framework for Agentic AI*. Janvier 2026.
- **Speeki**. *ISO 42001 as the governance foundation for agentic AI*. Mars 2026.
- **Trustcloud**. *ISO 42001 & NIST AI RMF: Mastering responsible AI governance in 2026*. Avril 2026.
- **HotTopics**. *AI governance framework: The 2026 strategic guide for leadership*. Mars 2026.
- **Global AI Compliance Center**. *Global AI Governance Comparison 2026*. Avril 2026.
- **Trustible**. *AI Governance Frameworks Compared*. Avril 2026.
- **CIO.com**. *Shadow AI morphs into shadow operations*. 2026.
- **Obfuscated.site**. *Beyond the Hype: The CIO's Guide to Governing Agentic AI in 2026*. Mars 2026.
- **Digital Chiefs**. *Who Owns AI: CIO, CDO or CTO Compared*. Février 2026.

10. Annexe méthodologique

Les chiffres et observations cités dans ce rapport proviennent de sources publiques 2024-2026 vérifiées et datées. La période de référence des sources est janvier 2024 - mai 2026, avec une concentration sur les publications de janvier 2026 à mai 2026 *(période où les frameworks agentic AI gouvernance ont été formalisés et où les deadlines AI Act se sont précisées)*.

Le cas composite secteur Bancaire *(section 5.1)* combine des informations publiquement disponibles sur le secteur bancaire européen *(notamment Capgemini Cloud Report Financial Services 2026, Accenture Banking Top Trends 2026, BoE/FCA AI Risk Survey 2025)* et nos observations terrain anonymisées auprès de DSI bancaires, conformément aux engagements de confidentialité d'AEP. Aucune information confidentielle de mission n'est divulguée.

L'interprétation de l'AI Act dans ce document a un objectif opérationnel pour le CIO. Elle ne constitue pas un avis juridique. Pour tout enjeu de conformité AI Act, nous recommandons de consulter un cabinet d'avocats spécialisé en droit numérique européen.

Les recommandations actionnables *(section 4)* combinent l'analyse des cas publics et l'observation auprès de plus de 50 DSI grands comptes accompagnés ou rencontrés par AEP entre 2024 et 2026.

Le plan d'action *(section 6)* est calibré pour une exécution réaliste sur 9 mois pour des grandes entreprises avec une exposition AI Act significative. Il doit être ajusté à la maturité initiale, au périmètre et aux contraintes spécifiques de chaque organisation.

Sébastien Delayre, fondateur d'Agile Enterprise Partner. Avril 2026.