

DaVinciA⁺

Un cadre de référence pour les systèmes d'intelligence
artificielle gouvernés, validés et transparents

DaVinciA⁺

THE REFERENCE FRAMEWORK FOR GOVERNED AI

Published by A.Ward Publications

In collaboration with Brehon AI Solutions

**Licence : Creative Commons Attribution-NoDerivatives 4.0 International (CC
BY-ND 4.0)**

ISBN:978-1-918501-01-8

© 2025 A.Ward Publications, in collaboration with Brehon AI Solutions.

This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License (CC BY-ND 4.0).

To view a copy of this license, visit:

<https://creativecommons.org/licenses/by-nd/4.0/>

[Chapitre 1 – Résumé exécutif](#)

[Chapitre 2 – Portée, public et intention](#)

[Chapitre 3 – Présentation du cadre](#)

[Chapitre 4 – Architecture](#)

[Chapitre 5 – Cycle de vie de la validation](#)

[Chapitre 6 – Gouvernance et supervision](#)

[Chapitre 7 – Alignement de conformité](#)

[Chapitre 8 – Modèles de déploiement et d'adoption](#)

[Chapitre 9 – Études de cas](#)

[Chapitre 10 – Annexe technique](#)

[Chapitre 11 – Synthèse et glossaire](#)

[Annexe A — Dossier minimal de preuves pour l'examen de la gouvernance](#)

Chapitre 1 – Résumé exécutif

DaVinciA⁺ est un cadre de référence pour la gouvernance, la validation et la supervision opérationnelle des systèmes d'intelligence artificielle.

Il fournit des principes structurés, des orientations de cycle de vie et des mécanismes de gouvernance afin de soutenir la conception, le déploiement et l'exploitation responsables de l'IA dans des environnements réglementés et d'entreprise.

DaVinciA⁺ est indépendant des fournisseurs, neutre sur le plan technologique et indépendant de toute implémentation, et est destiné à être appliqué en complément des normes existantes en matière de réglementation, de qualité et de gestion des risques.

Le cadre soutient les activités de conformité en traduisant des obligations de haut niveau — telles que la gestion des risques, la supervision humaine, la transparence et la validation — en pratiques opérationnelles répétables et en artefacts de gouvernance.

DaVinciA⁺ n'est pas un dispositif de certification, ne remplace pas les lois ou normes applicables et ne constitue ni une approbation réglementaire ni un avis juridique.

Les systèmes d'intelligence artificielle sont désormais intégrés dans des processus de prise de décision qui affectent la sécurité, la conformité et les résultats réglementaires. Dans de nombreuses organisations, ces systèmes sont déployés sans structure unifiée permettant de garantir la transparence, la traçabilité et l'auditabilité sur l'ensemble de leur cycle de vie.

DaVinciA⁺ définit un cadre de gouvernance et de validation conçu pour imposer de la structure, de la responsabilité et la production de preuves aux systèmes d'IA opérant dans de tels

environnements. DaVinciA⁺ a été développé pour répondre directement à cette lacune. Il présente un cadre de gouvernance et de validation conçu non comme un modèle théorique, mais comme un moyen pratique de structurer l'IA afin que son comportement puisse être compris, surveillé et justifié tout au long de son cycle de vie.

La gouvernance des systèmes d'IA doit être établie dès la phase de conception. Les mécanismes de supervision introduits après le déploiement sont nécessairement incomplets, car le comportement du système reflète déjà des hypothèses, des contraintes et des choix de conception intégrés en amont. DaVinciA⁺ prévoit donc que la gouvernance soit exprimée explicitement au sein de l'architecture du système, avant le déploiement et tout au long de l'exploitation. Une fois déployé, le comportement du système reflète déjà des hypothèses et des choix de conception intégrés bien avant que les mécanismes de supervision n'aient été envisagés. DaVinciA⁺ intervient donc aux fondations. Il prévoit que les systèmes d'IA soient exprimés au travers de trois couches interconnectées — identité et intention, connaissances et logique, et supervision et audit — chacune apportant une forme distincte de contrainte et de responsabilité. Ces couches créent une structure stable autour de laquelle le reste du système peut évoluer, permettant aux organisations de faire évoluer l'IA sans perdre la visibilité ni le contrôle sur ce que le système fait ou pourquoi il le fait. Pour les dirigeants, cette structure réduit l'incertitude opérationnelle, accélère la préparation à l'examen réglementaire et diminue le coût à long terme des reprises en introduisant une gouvernance disciplinée dès les premières phases.

Parallèlement à cette structure architecturale, DaVinciA⁺ introduit une approche de cycle de vie fondée sur des pratiques de qualification historiquement réservées aux industries à haute fiabilité. Les vérifications d'installation, la qualification opérationnelle et la validation des performances forment une séquence progressive garantissant que le système est correct dans sa configuration, correct dans son comportement et correct dans son utilisation en conditions réelles. Une fois déployée, cette même discipline s'étend à la surveillance continue. La dérive est considérée comme un phénomène attendu, et non comme une surprise ; les preuves sont accumulées de manière continue plutôt que ponctuelle ; et les changements sont gérés dans le cadre de revues contrôlées et documentées. Le cycle de vie ne valide pas uniquement un modèle ; il valide l'ensemble de l'environnement opérationnel dans lequel l'IA fonctionne.

La nécessité d'une telle discipline est devenue particulièrement aiguë à mesure que les organisations passent de cas d'usage à modèle unique à des écosystèmes multi-agents. L'IA moderne ne fonctionne généralement pas comme un modèle isolé répondant à des requêtes indépendantes. Elle est de plus en plus un réseau d'agents spécialisés, chacun exécutant une partie d'un flux de travail, chacun dépendant des résultats des autres, et chacun susceptible d'influencer des résultats pertinents pour la conformité. Dans des environnements non structurés, ces agents peuvent dériver, déléguer de manière imprévisible ou agir à contre-sens. DaVinciA⁺ introduit une gouvernance formelle dans ces interactions au moyen de limites explicites, de voies de délégation contrôlées et de mécanismes d'audit qui consignent chaque

échange. Le résultat est un système dans lequel le comportement multi-agents devient reconstructible plutôt qu'émergent ou opaque.

DaVinciA⁺ est délibérément neutre sur le plan technologique. Il ne prescrit ni la manière dont une organisation doit entraîner ses modèles, ni les outils d'orchestration qu'elle doit utiliser. Il définit au contraire les attentes de gouvernance qui doivent s'appliquer indépendamment de la pile technologique, du secteur ou du cas d'usage. Cette neutralité lui permet de s'intégrer aux infrastructures existantes — pipelines de machine learning dans le cloud, clusters de calcul sur site, outils de création d'agents, moteurs de flux de travail — sans imposer de dépendance technologique. Le cadre se situe au-dessus du socle technique et apporte de la cohérence à des systèmes hétérogènes.

Tout aussi important est son alignement conceptuel avec les attentes réglementaires mondiales. Plutôt que de revendiquer la conformité, DaVinciA⁺ reflète, dans des pratiques opérationnelles, les priorités structurelles présentes dans les principaux cadres réglementaires et normatifs — le règlement européen sur l'IA (EU AI Act), ISO 42001, GAMP 5, MDR/IVDR, ISO 13485 et 14971, IEC 62304 et les bonnes pratiques FDA GMLP. Il fournit les types de processus, d'artefacts et de traçabilité couramment attendus dans ces cadres, tout en évitant toute suggestion qu'il s'y substituerait. Il s'agit d'une couche de gouvernance, et non d'un régime de certification. Les organisations qui l'adoptent restent soumises à l'ensemble des évaluations réglementaires requises ; DaVinciA⁺ les prépare simplement en leur fournissant les preuves, la discipline et la documentation que ces évaluations exigent.

À travers ces éléments — architecture, validation, supervision, journalisation et alignement réglementaire — un principe unique traverse le cadre : l'IA doit rester responsable vis-à-vis de l'organisation qui la déploie. La responsabilité, dans ce contexte, n'est pas abstraite. Elle correspond à la capacité de démontrer, preuves à l'appui, comment le système a été conçu, comment il se comporte, comment il est surveillé et comment les risques sont maîtrisés. DaVinciA⁺ rend cette démonstration possible. Il offre aux organisations un moyen d'établir l'ordre avant l'échelle, la clarté avant la complexité et la traçabilité avant le déploiement.

Ce faisant, il requalifie l'IA non comme une capacité volatile à gérer de manière défensive, mais comme un actif opérationnel pouvant être gouverné avec la même rigueur que d'autres systèmes critiques. DaVinciA⁺ permet aux entreprises de faire progresser leurs initiatives en matière d'IA avec confiance, en sachant que la performance, la sécurité et la conformité restent continuellement observables et placées sous autorité humaine. Il ne ralentit pas l'innovation ; il fournit la structure qui rend l'innovation durable.

Chapitre 2 – Portée, public et intention

2.1 Portée

DaVinciA⁺ s'applique aux systèmes d'intelligence artificielle dont les résultats influencent des processus opérationnels, pertinents pour la conformité, sensibles pour la sécurité ou critiques pour la prise de décision. Le cadre est destiné à être utilisé à la fois dans des domaines réglementés et non réglementés où la traçabilité, la responsabilité et la supervision sont requises. Le terme « universel » renvoie à des principes de gouvernance applicables à l'ensemble des secteurs et des juridictions, et non à un traitement réglementaire uniforme ou à une classification uniforme des risques.

Le cadre est applicable aux :

- Systèmes d'IA à modèle unique déployés dans des flux de travail opérationnels
- Systèmes d'IA multi-agents réalisant un raisonnement distribué ou délégué
- Systèmes d'IA intégrés dans des environnements réglementés (y compris les domaines de la santé, du MedTech, du pharmaceutique, de la finance et des infrastructures)
- Déploiements d'IA en entreprise nécessitant une auditabilité et une gouvernance sur l'ensemble du cycle de vie

DaVinciA⁺ est délibérément neutre sur le plan technologique. Il ne prescrit pas de modèles, de plateformes, d'outils d'orchestration ou d'architectures d'infrastructure spécifiques. Les attentes en matière de gouvernance demeurent constantes quelle que soit l'implémentation technique.

2.2 Hors périmètre explicite

DaVinciA⁺ ne :

- Définit ni n'évalue la performance ou la précision des modèles
- Prescrit pas de techniques d'entraînement des modèles ni de construction des jeux de données
- Fournit pas de revendications cliniques, juridiques ou de sécurité concernant les résultats des systèmes
- Remplace ni ne supplante les évaluations réglementaires, certifications ou approbations
- Agit en tant que dispositif d'évaluation de la conformité ou système de certification

Ces exclusions sont délibérées. DaVinciA⁺ gouverne la structure, la supervision et les preuves, et non la capacité des modèles ni la performance des résultats.

2.3 Public visé

Ce document s'adresse à :

- La direction exécutive responsable des risques liés à l'IA, de la responsabilité et de la gouvernance
- Les professionnels des affaires réglementaires, de la qualité et de la conformité

- Les responsables techniques concevant, déployant ou supervisant des systèmes d'IA
- Les auditeurs et les examinateurs de gouvernance évaluant des déploiements d'IA

2.4 Intention

DaVinciA⁺ est publié en tant que cadre de gouvernance de référence. Son objectif est d'établir une structure commune permettant de concevoir, d'examiner et de gouverner les systèmes d'IA de manière cohérente, indépendamment du domaine, du secteur ou de l'implémentation technique.

Déclaration de non-standard

DaVinciA⁺ n'est pas une norme, une spécification ni un dispositif d'évaluation de la conformité. Il s'agit d'un cadre de gouvernance de référence destiné à soutenir, et non à remplacer, les processus réglementaires et normatifs formels.

Ce qu'est DaVinciA⁺

Un cadre de gouvernance de référence pour les systèmes d'IA

Une approche structurée de la validation et de la supervision

Une méthode d'opérationnalisation des obligations réglementaires

Une couche neutre complémentaire aux normes et aux lois existantes

un cadre de référence pour des opérations d'IA prêtes à l'audit

Ce que DaVinciA⁺ n'est pas

Un dispositif de certification ou d'accréditation

Une autorité réglementaire

Une plateforme logicielle propriétaire

Un substitut aux normes ISO, IEC, MDR, FDA ou aux obligations légales

Une garantie d'approbation réglementaire

Chapitre 3 – Présentation du cadre

Les systèmes d'IA fonctionnent comme des instruments opérationnels de prise de décision plutôt que comme des composants techniques isolés. Leurs résultats influencent des processus réglementés, des activités critiques pour la sécurité et la responsabilité organisationnelle. À mesure que leur influence s'étend, un cadre cohérent est nécessaire pour garantir que ces systèmes restent structurés, contraints et gouvernables tout au long de leur cycle de vie. DaVinciA⁺ établit cette structure en définissant la manière dont les systèmes d'IA sont décrits, gouvernés et validés, indépendamment de leur implémentation technique. À mesure que leur influence croît, le besoin d'un cadre cohérent apportant structure, discipline et transparence à la manière dont ces systèmes sont conçus et maintenus s'accroît également. DaVinciA⁺ a été développé pour répondre à ce besoin en proposant un modèle unifié de description, de gouvernance et de validation de l'IA sur l'ensemble de son cycle de vie.

Le cadre repose sur une hypothèse fondamentale : un système d'IA doit être compris non seulement au regard des tâches qu'il exécute, mais aussi des conditions dans lesquelles il les exécute. L'ingénierie logicielle traditionnelle reconnaît depuis longtemps l'importance de la finalité, des préconditions, des contraintes et de la responsabilité. L'IA requiert une structure analogue, adaptée à des systèmes dont le comportement émerge de schémas appris plutôt que de code déterministe. DaVinciA⁺ exprime cette structure au moyen de trois couches interdépendantes qui définissent ce qu'est le système, la manière dont il raisonne et la façon dont il demeure dans les limites qui lui sont autorisées.

La première couche établit l'identité et l'intention du système. L'identité définit ce qu'est le système ; l'intention définit ce que le système est autorisé à faire. La séparation de ces concepts élimine les ambiguïtés et empêche l'extension incontrôlée des responsabilités du système. Elle clarifie le domaine dans lequel l'IA est censée opérer, les responsabilités spécifiques qui lui sont attribuées et les limites dans lesquelles elle doit rester. Cela élimine l'ambiguïté à la source. En formalisant la mission et les contraintes du système, DaVinciA⁺ empêche la dérive de périmètre et permet que chaque choix de conception ultérieur soit évalué au regard de ces engagements initiaux. Cette couche identifie également les parties prenantes humaines responsables des résultats du système, ancrant la gouvernance dans la responsabilité personnelle et organisationnelle.

La deuxième couche concerne les connaissances et la logique — la mécanique interne par laquelle l'IA interprète les entrées, évalue les informations et produit des résultats. Dans les déploiements conventionnels, ces mécanismes sont souvent masqués par l'abstraction. DaVinciA⁺ prévoit qu'ils soient explicités et, lorsque cela est possible, contraints. Il définit les sources de données auxquelles le système peut accéder, les formes de raisonnement qu'il peut employer, les outils qu'il peut invoquer et les garde-fous qui orientent ses décisions. En capturant explicitement ces éléments et en les maintenant sous contrôle de version, DaVinciA⁺

fournit la traçabilité nécessaire à l'investigation, à la surveillance et à l'examen réglementaire. Le raisonnement du système devient un espace gouverné plutôt qu'une boîte noire.

La troisième couche traite de la supervision et de l'audit. Aucun système d'IA ne devrait fonctionner sans mécanisme clair de supervision, d'escalade et de génération continue de preuves. Cette couche introduit des points de contrôle structurés permettant d'identifier quand une intervention humaine est requise, quand des décisions dépassent les limites autorisées et quand des résultats exigent une vérification. Elle impose également un enregistrement d'audit exhaustif retracant les actions, le contexte et la justification. Cet enregistrement n'existe pas pour lui-même ; il crée les conditions permettant aux organisations de démontrer leur responsabilité, d'enquêter sur des anomalies et de satisfaire aux exigences de contrôle externe.

Ces trois couches constituent le cœur de l'architecture DaVinciA⁺, mais le cadre dépasse la simple description structurelle pour s'étendre à la technique opérationnelle. La méthode DaVinciA fournit le fondement philosophique de la manière dont l'IA doit être conçue au sein de cette architecture. Elle met l'accent sur la clarté de la finalité, l'économie de conception, le raisonnement contrôlé et le respect de la supervision humaine. Elle décourage la complexité inutile, la délégation incontrôlée et les comportements ambigus. En pratique, elle vise à réintroduire l'intentionnalité dans un domaine technologique qui évolue souvent plus rapidement que la gouvernance ne peut s'y adapter.

La pertinence du cadre devient particulièrement évidente dans les contextes où plusieurs agents d'IA doivent collaborer. En l'absence de structure, les systèmes multi-agents peuvent créer des chaînes de délégation difficiles à observer ou à reconstituer. DaVinciA⁺ introduit des interfaces explicites entre les agents, définit des voies de communication autorisées et précise que chaque échange est consigné dans une piste d'audit unifiée. Cela transforme un comportement autrement dynamique et faiblement borné en une séquence d'interactions contrôlées pouvant être examinées, testées et justifiées.

Une caractéristique importante de DaVinciA⁺ est sa neutralité. Il n'exige pas que les organisations adoptent des modèles, des plateformes ou des systèmes d'orchestration spécifiques. Il fournit au contraire une couche de gouvernance applicable aux environnements cloud, aux installations HPC sur site et aux outils d'orchestration agentique. L'implémentation technique peut varier ; les principes de gouvernance, eux, demeurent inchangés. Ce choix de conception permet à DaVinciA⁺ de fonctionner comme une référence intégratrice au sein d'architectures d'entreprise complexes, en apportant de la cohérence même lorsque les outils sous-jacents diffèrent.

Supervision et audit
Revue · Escalade ·
Responsabilité

Connaissances et logique
Raisonnement · Contrainte ·
Autorisation

Identité et intention
Définition · Périmètre ·
Exclusion

« Architecture conceptuelle de gouvernance. Ces couches ne constituent ni une certification, ni une approbation réglementaire. Les flèches représentent des relations d'information et de contrainte, et non des pipelines techniques de données. »

Le cadre reflète la trajectoire structurelle des réglementations émergentes en matière d'IA en opérationnalisant des principes qui apparaissent de manière cohérente dans le règlement européen sur l'IA (EU AI Act), ISO 42001, GAMP 5, MDR/IVDR et les lignes directrices de la FDA — sans impliquer de conformité. Bien qu'il ne revendique pas la conformité, ses concepts sont éclairés par les attentes formulées dans l'EU AI Act, ISO 42001, GAMP 5, MDR/IVDR, ISO 13485, ISO 14971 et IEC 62304. Ces instruments partagent une emphase commune sur la gestion documentée des risques, la transparence, la supervision responsable et la discipline du cycle de vie. DaVinciA⁺ traduit opérationnellement ces attentes en mécanismes de gouvernance pratiques que les organisations peuvent adopter précocement, bien avant l'entrée en application d'obligations réglementaires formelles.

Pris dans leur ensemble, ces éléments créent un modèle de gouvernance cohérent pour les systèmes d'IA. DaVinciA⁺ offre un moyen d'articuler ce qu'est un système d'IA, comment il prend des décisions, comment il est contraint et comment son comportement est surveillé dans le temps. Il permet aux organisations de construire des systèmes d'IA qui sont non seulement fonctionnels mais responsables, non seulement capables mais délibérés. Il remplace une conformité réactive par une structure proactive et fournit une base sur laquelle une IA sûre, évolutive et digne de confiance peut être développée.

Chapitre 4 – Architecture

L'architecture du système définit les conditions dans lesquelles l'intelligence est autorisée à opérer. Dans les systèmes d'IA, l'architecture gouverne le comportement en situation d'incertitude, détermine la manière dont le risque se manifeste et établit si la responsabilité peut être démontrée. DaVinciA⁺ traite l'architecture comme un construit de gouvernance plutôt que comme un plan logiciel. Elle définit les conditions dans lesquelles l'intelligence est autorisée à opérer, les limites dans lesquelles les décisions sont formées et les mécanismes par lesquels la supervision est maintenue. Cette perspective déplace l'attention des capacités des modèles individuels vers le système plus large qui les contient et les contraint.

Modélisation des menaces et considérations relatives aux modes de défaillance

À mesure que les systèmes d'IA évoluent en complexité et en autonomie, des modes de défaillance et des vecteurs de menace non évidents peuvent apparaître, qui ne sont pas immédiatement perceptibles par les seuls tests fonctionnels. DaVinciA⁺ reconnaît la modélisation des menaces comme une pratique complémentaire importante pour anticiper et analyser ces risques.

Les organisations peuvent appliquer, le cas échéant, des cadres établis tels que STRIDE pour les menaces de cybersécurité et LINDDUN pour l'analyse d'impact sur la vie privée. Pour les risques spécifiques à l'IA — notamment les boucles de délégation incontrôlées, l'utilisation abusive d'outils, l'omission d'escalade ou l'effondrement des chemins de raisonnement — DaVinciA⁺ fournit un modèle d'audit traçable fondé sur les exécutions et les étapes, permettant à la fois l'analyse *a posteriori* et les tests avant déploiement lors de la Qualification Opérationnelle (OQ).

De futures annexes formaliseront des arbres de menaces représentatifs et des schémas de défaillance dérivés de ce modèle de traçabilité. Ces éléments auront un caractère consultatif et ne prescriront pas de techniques d'atténuation ni de choix d'implémentation spécifiques.

L'architecture repose sur le principe qu'un système d'IA doit être compris à travers les rôles qu'il joue, les informations qu'il utilise et les contrôles qui façonnent son comportement. En termes pratiques, cela exige une clarté quant à la finalité du système et à l'environnement dans lequel il fonctionne. DaVinciA⁺ introduit donc une description formelle de l'identité et de l'intention comme élément architectural fondamental. Cette description précise ce que le système est destiné à accomplir, les limites qu'il doit respecter et les responsabilités qui demeurent exclusivement humaines. Elle permet que les choix de conception puissent être évalués par rapport à une déclaration explicite de finalité, réduisant le risque de dérive de périmètre ou d'extension fonctionnelle non intentionnelle.

S'appuyant sur cette base, l'architecture décrit les connaissances et la logique qui informent les décisions du système. Les modèles d'IA opèrent souvent dans un espace informationnel vaste et faiblement borné, en s'appuyant sur des données et des outils susceptibles d'évoluer dans le temps. Pour contrer cette tendance à l'opacité, DaVinciA⁺ prévoit un ensemble défini de sources de connaissances, de processus de raisonnement et d'actions autorisées. Il impose que les mécanismes par lesquels le système interprète les données, invoque des outils et évalue le contexte soient documentés et versionnés. Cette structure permet aux organisations d'identifier la manière dont les décisions sont formées, d'évaluer si ces décisions demeurent dans les limites des politiques et des contraintes réglementaires, et d'examiner les écarts ou les résultats inattendus.

La supervision et l'audit constituent le troisième composant structurel de l'architecture. Aucun système d'IA, aussi bien conçu soit-il, ne devrait fonctionner hors de portée de la supervision. DaVinciA⁺ intègre donc la supervision directement dans l'architecture, plutôt que de la traiter comme une couche externe ou optionnelle. Il définit quand une revue humaine est nécessaire, comment l'incertitude ou les conflits sont destinés à être escaladés et quelles preuves sont destinés à être produites à chaque étape du fonctionnement. Cela inclut l'enregistrement du raisonnement du système, la documentation de l'utilisation des outils et la capture des éléments contextuels permettant l'analyse a posteriori. En intégrant la supervision au cœur même de l'architecture, DaVinciA⁺ est conçu de manière à ce que la responsabilité ne repose pas sur une reconstruction rétrospective, mais qu'elle soit générée en continu à mesure que le système fonctionne.

Ces éléments architecturaux deviennent particulièrement importants dans les environnements multi-agents. À mesure que les organisations adoptent des flux de travail agentiques, les composants individuels dépendent de plus en plus des résultats des autres. En l'absence de structure, ces interactions peuvent engendrer des comportements difficiles à prévoir ou à vérifier. DaVinciA⁺ répond à cette problématique en définissant des voies de communication contrôlées entre les agents et en exigeant que chaque interaction soit consignée dans la piste d'audit. Il définit un cadre dans lequel la délégation s'effectue dans des limites autorisées, qu'aucun agent ne peut étendre son périmètre de manière autonome et que la supervision humaine est déclenchée lorsque les interactions génèrent de l'incertitude ou du risque. Cela

transforme ce qui pourrait autrement être un comportement émergent ou opaque en une séquence d'étapes responsables. En pratique, cela signifie qu'un agent ne peut déléguer que par des voies autorisées, ne peut pas modifier de manière autonome son périmètre et doit générer des données d'audit pour chaque interaction. Ces contrôles transforment un comportement agentique dynamique en séquences reconstructibles et révisables.

L'architecture intègre également les réalités pratiques du déploiement en entreprise. Les systèmes d'IA sont rarement statiques : les modèles sont remplacés, les jeux de données évoluent, des outils sont ajoutés et les flux de travail changent. DaVinciA⁺ anticipe ce dynamisme en intégrant des mécanismes de changement contrôlé dans la conception architecturale. La configuration et la logique restent sous contrôle de version, permettant aux organisations de suivre l'impact des mises à jour sur le comportement. Des points de contrôle de validation garantissent que les systèmes modifiés continuent de fonctionner dans leur périmètre prévu. Les enregistrements d'audit fournissent les preuves nécessaires pour démontrer que les changements ont été mis en œuvre de manière responsable et avec une supervision appropriée.

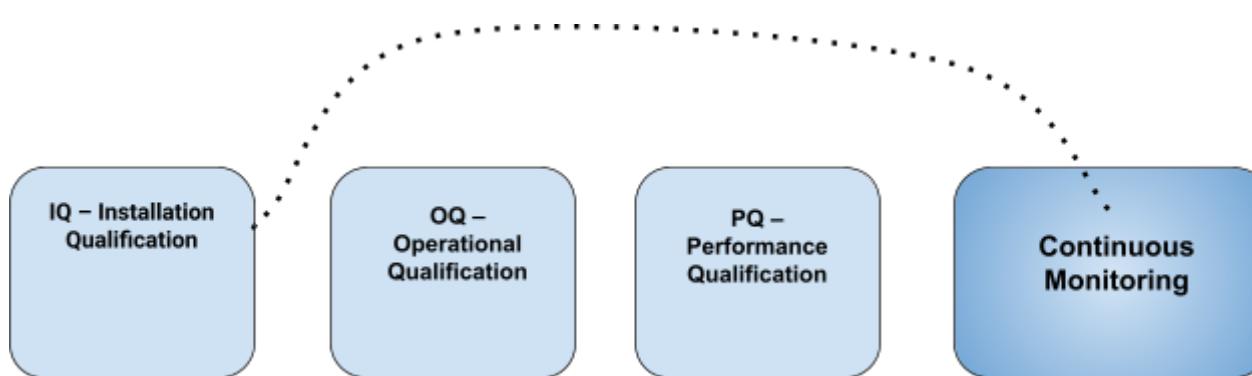
L'une des forces de l'architecture DaVinciA⁺ réside dans son indépendance vis-à-vis de toute pile technologique spécifique. Les principes de gouvernance s'appliquent de manière équivalente aux services cloud, aux plateformes sur site, aux orchestrateurs de flux de travail et aux outils de création d'agents. Cette neutralité permet à l'architecture de fonctionner comme une couche unificatrice à travers des systèmes divers, offrant aux organisations un cadre cohérent même lorsque les composants techniques varient selon les départements ou les projets. L'accent reste mis sur la structure, la traçabilité et le contrôle, plutôt que sur les spécificités techniques du développement des modèles.

Pris dans leur ensemble, l'architecture est conçue pour garantir que les systèmes d'IA demeurent intelligibles, contrôlables et responsables tout au long de leur cycle de vie. Elle fournit les contraintes nécessaires à une exploitation sûre sans freiner l'innovation ni limiter le choix des modèles. En définissant la manière dont l'intention, le raisonnement et la supervision sont destinés à être exprimés, DaVinciA⁺ offre une voie pratique vers un déploiement responsable à grande échelle. Il établit une base stable sur laquelle des capacités d'IA complexes peuvent être construites, intégrées et gouvernées avec confiance.

Chapitre 5 – Cycle de vie de la validation

La validation des systèmes d'IA est une activité continue tout au long du cycle de vie, et non une évaluation ponctuelle. DaVinciA⁺ adopte un modèle de qualification issu des industries à haute fiabilité, exigeant des preuves que les systèmes sont correctement configurés, qu'ils fonctionnent dans des contraintes définies et qu'ils demeurent adaptés à leur finalité dans des conditions réelles d'utilisation. La validation logicielle traditionnelle suppose un comportement déterministe : une fois un système installé et testé, ses résultats restent prévisibles à moins que des modifications explicites ne soient introduites. Les systèmes d'IA remettent en cause cette hypothèse. Leur comportement dépend non seulement du code, mais aussi des paramètres des modèles, des distributions de données, des interactions avec des outils et de l'environnement opérationnel global. Pour cette raison, DaVinciA⁺ adopte une approche de validation fondée sur le cycle de vie, garantissant que le système demeure démontrablement adapté à sa finalité tant au moment du déploiement que dans le temps.

Le cycle de vie commence par l'établissement du fait que le système a été correctement configuré. La qualification d'installation (Installation Qualification – IQ) vérifie que tous les composants — modèles, outils, orchestrateurs, sources de données et garde-fous — sont déployés conformément aux intentions et correspondent aux spécifications documentées. Dans les systèmes conventionnels, cette étape est relativement simple ; dans les déploiements d'IA, elle requiert une vigilance accrue, car des changements dans les versions de modèles, les paramètres d'environnement ou les autorisations des outils peuvent modifier de manière significative le comportement du système. DaVinciA⁺ traite la configuration comme un artefact contrôlé afin de garantir que l'intégrité structurelle du système est préservée dès le départ.



États conceptuels de validation illustrant une qualification continue et conditionnelle.

La séquence est illustrative et ne prescrit ni l'ordre des processus, ni leur calendrier, ni leur mode d'implémentation.

La qualification opérationnelle (Operational Qualification – OQ) examine la manière dont le système se comporte dans des conditions attendues. L'objectif n'est pas seulement de tester la fonctionnalité, mais de comprendre les contours du raisonnement du système et de confirmer que les garde-fous, les voies d'escalade et les mécanismes de supervision réagissent conformément à la conception. Cette étape permet d'examiner si le système respecte son périmètre défini, gère l'incertitude de manière appropriée et produit des résultats qui demeurent dans les limites des politiques et des contraintes réglementaires. C'est à ce stade que la distinction entre un fonctionnement correct et un résultat correct devient critique : un système d'IA peut produire des résultats apparemment plausibles tout en violant des règles internes ou en contournant la supervision. DaVinciA⁺ met l'accent sur la nécessité de valider le comportement, et non uniquement les résultats. Cette distinction est essentielle pour l'IA : un système peut produire des résultats acceptables tout en enfreignant des règles internes, en omittant des étapes de supervision ou en contournant des déclencheurs d'escalade. La validation comportementale soutient la conformité structurelle, et pas seulement la plausibilité des résultats.

La qualification de performance (Performance Qualification – PQ) se concentre sur la capacité du système à fonctionner de manière fiable dans son contexte réel. Contrairement aux étapes précédentes, qui sont menées dans des environnements contrôlés, cette phase examine le système au sein de flux de travail opérationnels, en interaction avec des utilisateurs réels, des données réelles et des contraintes opérationnelles effectives. L'objectif de la PQ n'est pas d'établir la perfection, mais de démontrer que le système maintient sa stabilité, que la supervision reste efficace et que les écarts sont détectés et traités. Les systèmes d'IA opérant dans des contextes réglementés ou critiques pour la sécurité requièrent une attention particulière à ce stade, car des facteurs contextuels peuvent influencer le comportement de manière subtile, au-delà de ce que les tests contrôlés peuvent pleinement anticiper.

La validation ne s'achève pas au déploiement. Les systèmes d'IA évoluent à mesure que les environnements changent, que les outils sont mis à jour et que les modèles sont réentraînés ou remplacés. DaVinciA⁺ intègre donc la surveillance continue comme composant central du cycle de vie. La dérive — qu'elle soit statistique, comportementale ou contextuelle — est présumée plutôt que traitée comme une anomalie. La surveillance recueille des preuves sur l'ensemble des exécutions du système, permettant aux organisations d'identifier les risques émergents, d'évaluer si le comportement reste conforme aux spécifications et de déterminer quand une revalidation est nécessaire. Cette accumulation continue de preuves permet de détecter une éventuelle dégradation de la conformité au fil du temps.

La gestion des changements fournit la structure de gouvernance pour les mises à jour. DaVinciA⁺ considère toute modification apportée aux modèles, aux instructions (prompts), aux flux logiques, aux autorisations des outils ou aux sources de connaissances comme un changement nécessitant une revue documentée. L'objectif n'est pas d'entraver l'itération, mais de garantir que les mises à jour sont mises en œuvre de manière intentionnelle et étayées par

des preuves. Les changements sont destinés à être évalués quant à leur impact potentiel sur le comportement, la sécurité et les obligations de conformité. Le cas échéant, le système est renvoyé à des étapes antérieures du cycle de vie de validation afin de confirmer qu'il continue de fonctionner dans les limites initialement définies.

Tout au long de ce cycle de vie, la documentation joue un rôle central. Les artefacts de validation ne sont pas de simples productions administratives ; ils constituent le moyen par lequel une organisation démontre qu'elle a compris, gouverné et maîtrisé le système de manière responsable. Les enregistrements d'installation, les résultats de tests, les observations de performance, les journaux de surveillance et les historiques de changements forment un ensemble cohérent de preuves soutenant les demandes réglementaires, les enquêtes internes et la responsabilité continue. DaVinciA⁺ fournit la structure nécessaire pour générer ces preuves de manière cohérente et dans une forme alignée avec les attentes des autorités réglementaires et des organismes de normalisation.

Le modèle de cycle de vie vise à ce que les systèmes d'IA demeurent responsables sur l'ensemble de leur horizon opérationnel. Il reconnaît que la validation n'est pas une certification statique, mais un processus vivant qui doit s'adapter à la nature évolutive de l'IA. En intégrant des points de contrôle structurés, une surveillance continue et une gestion disciplinée des changements, DaVinciA⁺ fournit aux organisations les moyens de maintenir la confiance dans leurs systèmes, même lorsque les conditions évoluent. Il propose une approche pratique et rigoureuse pour garantir que l'IA reste sûre, prévisible et alignée sur sa finalité prévue à toutes les étapes de son déploiement.

« Représentation du cycle de vie (IQ → OQ → PQ → Surveillance). La séquence est conceptuelle et peut revenir en arrière lors de la gestion des changements ou de la remédiation de la dérive. »

Chapitre 6 – Gouvernance et supervision

La gouvernance est le mécanisme par lequel une organisation exerce son autorité sur ses systèmes d'IA. Elle décrit un cadre dans lequel les responsabilités sont clairement attribuées, que les processus décisionnels demeurent transparents et que les systèmes fonctionnent dans des limites éthiques, opérationnelles et réglementaires définies. DaVinciA⁺ considère la gouvernance non comme une activité périphérique, mais comme une composante intégrale de la conception du système. La supervision doit être intégrée à l'architecture, exprimée au travers de pratiques opérationnelles et étayée par des preuves capables de résister à un examen interne et externe.

Un principe fondamental du cadre est que les systèmes d'IA ne doivent pas se voir accorder une autonomie implicite. Même des modèles hautement performants doivent opérer dans des contraintes définies, sous la supervision de rôles humains identifiables. La gouvernance commence donc par l'établissement de la responsabilité du comportement du système. Cela inclut les personnes chargées de définir son périmètre, de maintenir sa configuration, de surveiller son fonctionnement et d'approuver toute modification. Dans les environnements réglementés, ces responsabilités s'alignent naturellement sur les fonctions existantes de direction qualité, clinique, réglementaire et technique. DaVinciA⁺ fournit une structure permettant d'exprimer ces responsabilités de manière claire et cohérente. Les rôles de supervision typiques incluent :

- **Propriétaire du système (System Owner)** – responsable de la définition de l'intention et de l'approbation des limites
- **Responsable Qualité/Réglementaire (Quality/Regulatory Lead)** – contribue à l'alignement des processus avec les attentes organisationnelles et réglementaires
- **Examinateur opérationnel (Operational Reviewer)** – réalise des évaluations avec (intervention humaine (HITL) en cas d'incertitude ou d'escalade



Conditions d'escalade illustratives et réponses de gouvernance associées.

Ce schéma ne représente ni une logique exécutable, ni une prise de décision automatisée, ni un flux de contrôle du système.

Ces rôles maintiennent l'autorité humaine tout au long du cycle de vie sans contraindre l'innovation.

La supervision est ensuite appliquée au moyen d'une combinaison de contrôles procéduraux et techniques. Sur le plan procédural, les organisations doivent déterminer quand une revue humaine est requise, comment l'incertitude ou le risque sont escaladés et quelle documentation doit accompagner les décisions automatisées. Sur le plan technique, la supervision est assurée par des garde-fous qui contraignent le comportement du système et par des mécanismes qui consignent chaque action sous une forme exploitable pour l'analyse. Cette structure dual est conçue pour rendre la supervision efficace tant en exploitation que dans des contextes d'audit, où des preuves peuvent être nécessaires pour reconstituer la manière dont une décision a été prise.

Les systèmes d'IA reposant sur plusieurs agents introduisent des défis supplémentaires en matière de gouvernance. Lorsque des agents collaborent ou se délèguent des tâches, le processus décisionnel peut être distribué entre plusieurs composants. En l'absence de structure, cette distribution peut obscurcir la responsabilité ou générer des comportements difficiles à interpréter. DaVinciA⁺ répond à ce défi en définissant des voies contrôlées entre les agents et en exigeant que chaque interaction soit consignée dans le registre d'audit. La délégation ne peut avoir lieu en dehors des routes autorisées, et les agents ne peuvent pas étendre de manière autonome leurs responsabilités ni modifier leurs limites opérationnelles. De cette manière, le cadre préserve la clarté même lorsque les flux de travail deviennent complexes.

Un autre aspect important de la supervision concerne le traitement de l'incertitude. Les systèmes d'IA opèrent fréquemment dans des contextes où les données d'entrée sont incomplètes, ambiguës ou incohérentes. Dans de telles situations, la prise de décision ne doit pas reposer uniquement sur l'inférence algorithmique. DaVinciA⁺ prévoit que les systèmes reconnaissent l'incertitude et procèdent à une escalade appropriée. La supervision humaine n'est pas une garantie théorique, mais un composant opérationnel intégré au comportement du système. Les critères d'escalade sont destinés à être explicites, documentés et testés afin que les examinateurs humains interviennent lorsque leur jugement est requis.

La gouvernance s'étend également à la documentation et aux preuves associées au fonctionnement du système. Les organisations sont destinés à être en mesure de démontrer non seulement qu'un système a produit un résultat acceptable pour une tâche donnée, mais qu'il a fonctionné dans des processus autorisés, en utilisant des raisonnements et des sources de données approuvés, sous une supervision efficace. DaVinciA⁺ introduit que les journaux d'audit consignent le contexte, le raisonnement, l'utilisation des outils et les résultats associés à chaque exécution du système. Cet enregistrement permet d'enquêter sur les anomalies, de soutenir les demandes réglementaires et de constituer la base factuelle de l'amélioration

continue. Il transforme la supervision d'une attente abstraite en un processus pratique et vérifiable.

Il est essentiel que la gouvernance demeure adaptable. À mesure que les cadres réglementaires évoluent, que les modèles changent et que les organisations étendent leur usage de l'IA, les mécanismes de supervision doivent évoluer en conséquence. DaVinciA⁺ fournit une structure suffisamment flexible pour intégrer de nouvelles exigences sans compromettre la stabilité du système. Son accent mis sur la documentation, l'auditabilité et la prise de décision permet que les mises à jour puissent être intégrées de manière méthodique, avec une compréhension claire de leur impact sur les responsabilités et le risque.

Pris dans leur ensemble, ces pratiques créent un environnement de gouvernance dans lequel l'IA demeure sous un contrôle humain délibéré. La responsabilité est explicite plutôt qu'implicite, la supervision est continue plutôt que ponctuelle, et les preuves sont générées de manière organique à mesure que le système fonctionne. DaVinciA⁺ aide les organisations à dépasser une supervision informelle pour adopter un modèle de gouvernance structuré, transparent et défendable, capable de répondre à la fois aux exigences opérationnelles et aux attentes réglementaires. Il transforme la supervision d'une activité réactive en un fondement de déploiement responsable de l'IA.

Matrice des seuils d'escalade (extrait du modèle DMS-GOV-011)

DaVinciA⁺ définit l'escalade non comme une réponse ad hoc, mais comme un mécanisme de gouvernance structuré déclenché par des conditions clairement définies. Le tableau ci-dessous illustre des seuils d'escalade représentatifs et les actions de supervision correspondantes. Ces seuils sont configurables par l'organisation déployante et visent à garantir que l'incertitude, les violations de limites ou les conditions de risque émergentes soient traitées sous une autorité humaine documentée.

Condition de déclenchement

Cible de l'escalade

Rôle de supervision

Action requise

Sensibilité temporelle

Violation de garde-fou (par ex. dépassement d'un seuil de biais ou de politique)

Examinateur HITL

Examinateur opérationnel

Approuver, bloquer ou rediriger la sortie du système

Immédiate

Confiance inférieure au seuil défini

Responsable QA

Qualité / Réglementaire

Exiger une justification, une revue ou un nouveau test

24–48 heures

Délégation en dehors de la voie autorisée

Propriétaire du système

Responsable du risque

Arrêt du système, journalisation et revalidation

Immédiate

Classe d'entrée non observée ou changement de distribution des données

Comité d'escalade

QA & RA

Revoir la lignée des données et signaler la dérive

≤ 72 heures

Cette structure reflète la logique d'escalade définie dans le modèle 2.2 du module de supervision de la gouvernance DaVinciA⁺ et démontre comment l'intention de gouvernance est opérationnalisée en un comportement système auditabile.

Modèle de classification des risques

Niveau 3 — Critique

Supervision humaine systématique

Revalidation requise

Niveau 2 — Modéré

Escalade en cas de dérive ou d'incertitude

Niveau 1 — Minimal

Revue périodique

Paliers de supervision indicatifs fondés sur l'impact potentiel.

Les niveaux de risque ne reflètent pas la capacité du système et n'impliquent ni certification, ni approbation, ni classification réglementaire.

Niveau 1 : Minimal

Résultats non critiques et réversibles

Revue périodique

Niveau 2 : Modéré

Influence indirecte sur la sécurité ou la conformité

Escalade HITL en cas de dérive

Niveau 3 : Critique

Sécurité des patients, risque financier, exposition juridique

HITL systématique + revalidation

Matrice RACI de gouvernance DaVinciA⁺

Cette matrice définit les rôles requis pour mettre en œuvre, superviser et maintenir la gouvernance des systèmes d'IA dans le cadre de DaVinciA⁺. Elle distingue les rôles responsables de l'exécution, redevables des résultats, consultés lors des décisions et informés en continu.

Rôle \ Responsabilité

Définir l'intention

Approuver la logique de supervision

Actions d'escalade

Surveiller la performance

Exécuter les interventions

Rôle \ Responsabilité	Définir l'intention	Approuver la logique de supervision	Actions d'escalade	Surveiller la performance	Exécuter les interventions
Propriétaire du système	A	C	C	I	A
Responsable Qualité / Réglementaire	C	A	C	A	C
Examinateur opérationnel	I	I	R	R	C
Responsable du risque	C	C	A	C	I

Architecte IA	R	R	I	C	C
----------------------	---	---	---	---	---

Développeur	I	I	I	C	R
--------------------	---	---	---	---	---

Logique de gouvernance sous-jacente à la matrice

Le Propriétaire du système est redevable à la fois de la définition de l'intention et de l'exécution des interventions au niveau métier (il est ultimement responsable).

Le Responsable QA/Réglementaire est redevable de la logique de supervision et de l'adéquation de la surveillance, conformément aux attentes réglementaires et aux exigences d'audit.

L'Examinateur opérationnel est responsable des actions d'escalade en temps réel et de la surveillance de la performance — il constitue la première ligne de l'exploitation gouvernée.

Le Responsable du risque est redevable des actions d'escalade du point de vue du risque (arrêter, accepter ou atténuer), mais pas de la surveillance quotidienne.

L'Architecte IA est responsable de la traduction de l'intention en conception technique (intention + logique de supervision) et est consulté sur la performance et les interventions.

Le Développeur est responsable de l'exécution effective des interventions (déploiement des changements, correctifs, retours arrière) une fois les décisions prises.

Propriétaire du système

Assume la responsabilité finale de la finalité du système, de son déploiement et de son utilisation continue.

Responsable Qualité / Réglementaire

Soutient l'alignement avec les réglementations applicables, les normes et les attentes en matière de preuves.

Examinateur opérationnel

Examine le comportement du système au regard des limites opérationnelles et de risque définies.

Responsable du risque

Assume l'évaluation et l'acceptation des risques associés à l'utilisation du système.

Architecte IA

Conçoit la structure du système afin de soutenir la traçabilité, le contrôle et la supervision.

Développeur

Met en œuvre les composants du système conformément aux contraintes de conception approuvées.

Attribution illustrative de l'autorité humaine et de la responsabilité dans le cadre de DaVinciA⁺. Les rôles représentent des responsabilités de gouvernance, et non une séquence de tâches ni un flux opérationnel.

Chapitre 7 – Alignement de conformité

La conformité réglementaire se démontre par des preuves, et non par des déclarations. DaVinciA⁺ ne revendique pas la conformité aux cadres juridiques ou normatifs. Il définit les structures opérationnelles au moyen desquelles l'alignement avec les attentes réglementaires peut être examiné, évalué et maintenu. À mesure que les cadres mondiaux évoluent — notamment le règlement européen sur l'IA (EU AI Act), ISO 42001 et les normes établies régissant les technologies médicales, pharmaceutiques et critiques pour la sécurité — les organisations ont besoin d'un moyen d'interpréter ces obligations en termes opérationnels. DaVinciA⁺ ne remplace pas ces exigences et ne fonctionne pas comme un dispositif de certification. Il fournit au contraire la discipline structurelle par laquelle la conformité peut être soutenue, examinée et maintenue tout au long du cycle de vie du système.

Le défi central auquel sont confrontées les organisations est que les exigences réglementaires sont généralement fondées sur des principes plutôt que prescriptives. Elles définissent des résultats — tels que la transparence, la gestion des risques, la gouvernance des données et la supervision humaine — sans préciser comment ces résultats sont destinés à être mis en œuvre. DaVinciA⁺ répond à ce défi en intégrant ces attentes dans l'architecture et les pratiques de cycle de vie déjà décrites. L'accent mis par le cadre sur l'intention définie, le raisonnement contrôlé, la supervision structurée et la journalisation d'audit exhaustive permet aux organisations de produire les types d'artefacts et de preuves que les autorités réglementaires attendent habituellement. La préparation réglementaire devient ainsi un sous-produit naturel d'une conception de système disciplinée, plutôt qu'un effort rétrospectif visant à justifier des décisions après coup.

De nombreux thèmes présents dans le règlement européen sur l'IA (EU AI Act) reflètent des priorités structurelles également prises en compte par DaVinciA⁺.

Des correspondances structurelles illustratives incluent :

Identité et intention → reflètent les définitions de gouvernance formulées dans ISO 42001

Journalisation d'audit → correspond opérationnellement aux attentes en matière de documentation technique exprimées dans l'EU AI Act

Surveillance de la dérive → reflète les attentes de surveillance post-marché décrites dans les orientations réglementaires

Gestion des changements → reflète les principes de gestion du cycle de vie décrits dans GAMP 5 et les exigences de cycle de vie de la norme ISO 13485

L'accent mis par l'Acte sur la qualité des données, la documentation technique, la surveillance des risques, le contrôle humain, la transparence et la surveillance post-marché reflète les

pratiques de cycle de vie intégrées au cadre. De même, la norme ISO 42001 met l'accent sur les structures de gouvernance, les responsabilités et les systèmes de management garantissant que l'IA est exploitée de manière sûre et responsable. DaVinciA⁺ soutient ces attentes en définissant des rôles redevables, en documentant les limites du système et en exigeant une surveillance continue du comportement et de la performance. Bien qu'il ne revendique pas la conformité, le cadre fournit une base structurée permettant d'examiner les attentes réglementaires.

Des parallèles similaires peuvent être observés dans des industries fortement réglementées. Les normes relatives aux dispositifs médicaux, telles que ISO 13485, ISO 14971 et IEC 62304, exigent des processus de développement maîtrisés, une prise de décision fondée sur le risque et des preuves documentées de validation. GAMP 5, appliquée de longue date aux logiciels dans des environnements réglementés, met l'accent sur la gestion du cycle de vie, la traçabilité et la justification documentée. DaVinciA⁺ renforce ces principes sans chercher à les reproduire ni à les remplacer. En structurant les systèmes d'IA en couches définies, en intégrant des points de contrôle de validation et en garantissant l'auditabilité des décisions, le cadre fournit la discipline opérationnelle requise pour les organisations se préparant à la documentation technique, aux soumissions réglementaires ou aux audits qualité.

L'alignement dépasse les seules législations et normes formelles. Les comités de gouvernance internes, les fonctions de gestion des risques d'entreprise, les instances de supervision clinique et les équipes d'audit ont tous besoin de visibilité sur le comportement des systèmes d'IA. Ils sont destinés à être en mesure de comprendre la justification des décisions, d'évaluer si le système a fonctionné dans les limites des politiques établies et d'identifier si les risques ont été reconnus et escaladés de manière appropriée. DaVinciA⁺ facilite cette visibilité en produisant un enregistrement opérationnel transparent. Les journaux d'audit, les historiques de configuration, les rapports de performance et les dispositifs de gestion des changements sont générés dans le cadre du fonctionnement normal, fournissant aux parties prenantes internes les preuves nécessaires pour prendre des décisions éclairées.

En pratique, la conformité est rarement statique. À mesure que les attentes réglementaires évoluent, les systèmes et les processus qui les soutiennent doivent également évoluer. DaVinciA⁺ est conçu pour s'adapter sans compromettre la stabilité ni accroître le risque. Sa structure permet aux organisations d'intégrer méthodiquement de nouvelles exigences en mettant à jour les règles de supervision, en validant de nouveaux comportements, en introduisant une documentation supplémentaire ou en modifiant les critères d'escalade. L'architecture et le cycle de vie sont suffisamment flexibles pour absorber les évolutions réglementaires tout en maintenant la prévisibilité et le contrôle. Cette adaptabilité est particulièrement importante dans les juridictions où les réglementations spécifiques à l'IA émergent rapidement et où les attentes en matière d'application peuvent se développer au fil du temps.

En traitant la conformité comme une caractéristique opérationnelle plutôt que comme une déclaration formelle, DaVinciA⁺ aide les organisations à se préparer à un avenir où la transparence et la responsabilité seront des exigences fondamentales du déploiement de l'IA. Le cadre ne se contente pas de soutenir l'alignement avec les normes actuelles ; il anticipe les attentes de gouvernance de la prochaine décennie. Il positionne les organisations pour répondre avec confiance aux audits, aux demandes d'information et aux évaluations, et fournit une base disciplinée à partir de laquelle une IA sûre, responsable et traçable peut être déployée à grande échelle.

Chapitre 8 – Modèles de déploiement et d'adoption

Le déploiement d'un cadre de gouvernance de l'IA au sein d'une organisation requiert davantage qu'une intégration technique. Il exige une approche pragmatique qui respecte les

processus existants, reconnaît les contraintes opérationnelles et soutient la maturation progressive des capacités. DaVinciA⁺ fournit une structure suffisamment flexible pour s'adapter aux organisations se trouvant à différents stades de leur parcours en matière d'IA, depuis l'expérimentation précoce jusqu'aux déploiements à grande échelle et réglementés. Ses modèles d'adoption sont conçus pour s'intégrer aux pratiques établies plutôt que pour les remplacer, permettant ainsi aux organisations de renforcer la gouvernance sans perturber les activités en cours.

Au niveau introductif, les organisations commencent souvent par des cas d'usage ciblés où les risques sont limités et l'environnement opérationnel bien compris. Dans ces contextes, **DaVinciA⁺ Light** propose une voie simplifiée fondée sur les principes fondamentaux d'identité, de traçabilité, de supervision et de changement contrôlé. Il fournit une méthode structurée garantissant que même les prototypes précoce ou les déploiements pilotes produisent la documentation et les preuves nécessaires à l'examen interne. Ce modèle allégé est volontairement conservateur : son objectif est d'instaurer la discipline avant l'échelle, en démontrant que la gouvernance peut être appliquée sans entraver l'innovation.

À mesure que les systèmes mûrissent et s'intègrent plus profondément dans les flux de travail opérationnels, les exigences en matière de gouvernance s'accroissent. Les composants d'IA peuvent commencer à influencer des activités réglementées, des décisions sensibles pour la sécurité ou des interactions orientées client. Des systèmes multi-agents peuvent être introduits pour coordonner des tâches ou automatiser des processus complexes. À ce stade, les organisations passent généralement à **DaVinciA⁺ Enterprise**, qui englobe l'ensemble des structures de cycle de vie, d'architecture et de supervision décrites précédemment. Ce modèle fournit une documentation exhaustive, des preuves de validation, des pistes d'audit et des procédures de gestion des changements adaptées aux audits internes et à l'examen réglementaire externe. La transition n'est pas brutale ; elle reflète une progression naturelle à mesure que la dépendance de l'organisation à l'IA augmente.

Le choix entre des modèles de déploiement basés sur le cloud, sur site ou hybrides n'affecte pas de manière significative les principes de gouvernance du cadre. DaVinciA⁺ est conçu pour fonctionner indépendamment de plateformes ou d'outils d'orchestration spécifiques. Les environnements cloud peuvent offrir efficacité et passage à l'échelle, tandis que les déploiements sur site peuvent être privilégiés pour des raisons de confidentialité, de réglementation ou de sécurité. Les modèles hybrides permettent aux organisations de conserver les composants sensibles en interne tout en s'appuyant sur une infrastructure externe pour des tâches moins critiques. Dans chaque scénario, la gouvernance demeure la couche de contrôle : la finalité du système, ses limites, son raisonnement et ses obligations de supervision ne changent pas avec le substrat technique.

L'adoption requiert également une clarté quant aux rôles. Le déploiement réussi dépend de la coopération entre les équipes techniques, les fonctions qualité et réglementaires, la gestion des

risques et la direction opérationnelle. DaVinciA⁺ introduit un modèle de gouvernance qui identifie qui est responsable de la définition de l'intention du système, de la validation du comportement, de la surveillance de la performance et de la supervision des changements. Ces responsabilités s'alignent naturellement sur les structures organisationnelles existantes, permettant une adoption sans nécessiter de réorganisation extensive. En clarifiant les attentes dès le départ, les organisations peuvent éviter des incertitudes ultérieures, en particulier lorsque les systèmes commencent à influencer des décisions pertinentes pour la conformité.

Le passage à l'échelle du cadre sur plusieurs systèmes ou départements nécessite une approche mesurée. DaVinciA⁺ met l'accent sur l'extension progressive des pratiques de gouvernance, soutenue par des modèles, des procédures répétables et une documentation cohérente. Les organisations peuvent commencer par appliquer le cadre à un seul cas d'usage, puis l'étendre à d'autres domaines une fois les bénéfices démontrés. Au fil du temps, cela conduit à un environnement de gouvernance cohérent dans lequel tous les systèmes d'IA sont documentés de manière comparable, partagent des mécanismes de supervision communs et produisent des preuves pouvant être agrégées en une vision unifiée des risques et de la performance.

Il est important de ne pas considérer le déploiement uniquement sous l'angle de la conformité. Les entreprises adoptant DaVinciA⁺ constatent généralement une réduction des ambiguïtés dans le développement, des validations internes plus rapides et une meilleure préparation aux audits. Ces gains opérationnels deviennent plus marqués à mesure que les systèmes d'IA se déploient à l'échelle des fonctions. Les organisations adoptant DaVinciA⁺ constatent souvent que la structure fournie améliore la fiabilité opérationnelle et réduit l'incertitude dans le développement. Des limites claires réduisent les reprises dues à des attentes mal alignées. Une supervision explicite renforce la confiance dans la prise de décision. Une documentation exhaustive rationalise la collaboration entre les équipes. Ces bénéfices sont peut-être plus visibles dans les secteurs réglementés, mais ils s'étendent à tout domaine où l'IA influence des décisions importantes.

DaVinciA⁺ fonctionne ainsi à la fois comme un cadre de gouvernance et comme un facilitateur opérationnel. Il offre aux organisations une voie vers une adoption responsable sans sacrifier l'élan. En fournissant des structures stables capables de passer à l'échelle, de s'adapter et de résister à l'examen, le cadre soutient à la fois l'innovation et la responsabilité. Ses modèles de déploiement reflètent une compréhension pratique des réalités organisationnelles, garantissant que la gouvernance évolue parallèlement aux capacités et aux responsabilités des systèmes qu'elle soutient.

Modèle de maturité DaVinciA⁺

Les organisations adoptant DaVinciA⁺ progressent à travers des niveaux définis de capacité de gouvernance. Ces niveaux reflètent une discipline structurelle croissante, une profondeur de supervision accrue et une préparation renforcée aux audits. Le modèle de maturité ci-dessous fournit une structure de référence pour guider l'adoption, évaluer les progrès et planifier les phases d'implémentation suivantes.

Niveau

Description

Indicateurs opérationnels

Niveau 1 — Pilote

Déploiement initial avec des éléments fondamentaux de gouvernance

- Identité et intention définies
- Journalisation d'audit de base
- Points de contrôle de supervision manuelle

Niveau 2 — Structuré

Systèmes multi-agents et compatibles OQ

- Rôles RACI attribués
- Logique d'escalade formalisée
- Surveillance de la dérive activée

Niveau 3 — Entreprise

Déploiement de niveau réglementé avec supervision du cycle de vie

- Tests PQ réalisés
- Comité de gestion des changements opérationnel
- Revues de gouvernance mensuelles consignées

Niveau 4 — Prêt pour l'audit

Systèmes pleinement matures avec traçabilité complète

- Dossier minimal de preuves produit
- Préparation aux audits externes confirmée
- Preuves de conformité (non déclaratives) disponibles

Chaque niveau s'appuie sur le précédent, augmentant la confiance, la défendabilité et le contrôle. Ce modèle ne prescrit ni vitesse ni calendrier, mais fournit un chemin structuré permettant à la gouvernance DaVinciA⁺ de monter en échelle en cohérence avec la criticité du système et le contexte réglementaire.

Chapitre 9 – Études de cas

Les études de cas sont incluses afin d'illustrer comment DaVinciA⁺ peut être appliqué dans des contextes pratiques pour établir la structure, la traçabilité et la supervision dans des systèmes d'IA. Elles n'ont pas pour objet de démontrer des performances, des caractéristiques de sécurité, une conformité réglementaire ou une aptitude à la certification. Elles fournissent plutôt des exemples de la manière dont des organisations ont appliqué le cadre pour organiser des raisonnements complexes, formaliser des limites et introduire une gouvernance fondée sur le cycle de vie. Les cas suivants reflètent deux projets relevant de domaines distincts, chacun à un stade de maturité différent, dans lesquels DaVinciA⁺ a été adopté afin de renforcer la discipline de gouvernance.

Étude de cas 1 — Système expert à forte intensité de connaissances dans un domaine non réglementé

Une organisation développant un système d'IA pédagogique spécialisé cherchait à transformer un vaste corpus de connaissances expertes en une architecture multi-composants cohérente, interprétable et auditable. Avant l'adoption de DaVinciA⁺, la conception du système reposait sur des modules conceptuels définis de manière lâche, dépourvus de limites documentées, de règles d'interaction ou d'attentes en matière de supervision. Cette situation créait une ambiguïté quant au comportement du système et limitait la capacité à faire évoluer la solution de manière responsable.

DaVinciA⁺ a été introduit afin de fournir une base structurée.

Les activités clés comprenaient :

- La formalisation de l'identité et de l'intention afin de définir clairement la finalité du système, ses contraintes et ses non-objectifs.
- L'établissement de limites au niveau des agents pour garantir que chaque composant de raisonnement opérait dans le cadre de responsabilités approuvées.
- La documentation des sources de connaissances et de la logique de raisonnement, permettant un examen transparent et un contrôle de version.
- La mise en œuvre de règles de supervision et de voies de délégation contrôlées, afin que les interactions entre composants multiples puissent être surveillées et reconstituées.

- La mise en place d'une évolution traçable, garantissant que les extensions ultérieures ne compromettaient pas l'intégrité structurelle du système.

Bien que cet environnement ne soit pas soumis à une supervision réglementaire, l'introduction de DaVinciA⁺ a permis à l'organisation de faire évoluer le système d'un prototype informel vers une structure stable et gouvernable. Le cadre a apporté clarté, auditabilité et croissance maîtrisée, sans contraindre l'innovation.

Étude de cas 2 — Système d'aide à la décision pertinent pour la conformité dans un contexte réglementé

Une autre organisation développant un outil d'aide à la décision assisté par l'IA pour des flux de travail sensibles à la conformité avait besoin d'un modèle de gouvernance capable de soutenir un examen réglementaire futur. Le système devait ingérer des règles spécifiques au domaine, interpréter des informations structurées et non structurées, et assister des examinateurs humains dans des processus fondés sur le jugement. Dès le départ, l'organisation a reconnu que la gouvernance du cycle de vie, la traçabilité et la supervision humaine seraient essentielles pour démontrer une exploitation responsable.

DaVinciA⁺ a été choisi comme cadre de gouvernance interne.

Il a été appliqué pour :

- Définir l'usage prévu du système, ses couches architecturales et ses limites opérationnelles.
- Introduire des processus de raisonnement contrôlés avec une logique versionnée, des outils approuvés et des sources de données documentées.
- Planifier et documenter le cycle de vie de validation (IQ, OQ, PQ) afin de garantir l'aptitude structurelle, comportementale et opérationnelle.
- Mettre en œuvre une journalisation d'audit exhaustive, permettant la reconstitution des décisions, des déclencheurs d'escalade et des interventions de supervision.
- Établir des mécanismes de gestion des changements, garantissant que les mises à jour des modèles, des outils ou des flux de travail s'effectuaient dans le cadre de revues documentées.

En ancrant le système dans DaVinciA⁺, l'organisation a constitué une base de preuves solide bien avant que des soumissions réglementaires ou des activités d'audit externe ne soient envisagées. Le cadre a garanti que l'évolution du système demeurait transparente et maîtrisable, et que la supervision humaine était systématiquement intégrée aux décisions pertinentes pour la conformité.

Synthèse des enseignements des études de cas

Dans les deux exemples — l'un non réglementé et exploratoire, l'autre réglementé et pertinent pour la conformité — le même schéma s'est dégagé :

- La finalité et les limites sont devenues explicites plutôt qu'implicites.
- Les processus de raisonnement sont devenus inspectables et gouvernés plutôt qu'opaques.
- La supervision est devenue structurée, permettant une intervention humaine prévisible (HITL).
- La traçabilité est devenue intrinsèque, soutenant à la fois l'assurance interne et la préparation aux audits externes.
- La croissance du système est demeurée maîtrisée, évitant toute dérive involontaire du périmètre ou du comportement.

Ces exemples démontrent comment DaVinciA⁺ peut ancrer le développement de l'IA dans des principes de gouvernance sans formuler de revendications relatives à la performance, à la sécurité ou à la conformité. Leur objectif est illustratif : montrer comment le cadre peut soutenir la clarté, la responsabilité et une évolution disciplinée dans des environnements d'IA diversifiés.

Chapitre 10 – Annexe technique

Portée de la gouvernance des données, de la protection de la vie privée et de la cybersécurité

DaVinciA⁺ suppose que les contrôles fondamentaux de gouvernance des données — y compris les politiques de confidentialité, la gestion des accès, la minimisation des données, la conservation et les protections de cybersécurité — sont mis en œuvre et gérés au niveau de l'infrastructure, de la plateforme ou du système de management de la qualité (QMS). Ces contrôles sont considérés comme des prérequis et non comme des composants du cadre DaVinciA⁺ lui-même.

Les extensions futures du cadre fourniront des correspondances de référence avec des normes et orientations établies, notamment ISO/IEC 27701, le RGPD et NIST SP 800-53, afin de soutenir les organisations souhaitant intégrer la gouvernance de l'IA à des environnements plus larges de contrôle de la confidentialité et de la sécurité. Ces correspondances resteront non normatives et neutres quant à l'implémentation.

L'objectif de l'annexe technique dans un livre blanc public n'est pas de fournir des détails opérationnels, mais d'offrir aux lecteurs une compréhension plus claire des types d'artefacts et de preuves qui soutiennent un système d'IA gouverné. Dans les environnements réglementés et d'entreprise, les parties prenantes exigent souvent une visibilité sur les structures permettant la supervision, l'auditabilité et la gestion du cycle de vie. DaVinciA⁺ fournit ces structures au moyen d'un ensemble d'éléments conceptuels qui sous-tendent le comportement du système

sans exposer la logique propriétaire ni les détails internes d'implémentation. L'annexe résume ces éléments afin d'illustrer comment la transparence technique est obtenue en pratique.

Un composant central du cadre est le registre d'audit. Les systèmes d'IA génèrent une séquence de décisions, d'invocations d'outils, de chemins de raisonnement et d'interprétations contextuelles qui sont destinés à être consignées sous une forme durable et examinable. DaVinciA⁺ traite la journalisation d'audit comme une activité continue plutôt que comme une fonctionnalité de diagnostic optionnelle. Chaque exécution du système produit un enregistrement structuré permettant aux enquêteurs, auditeurs et équipes de supervision de reconstituer les événements avec clarté. Ces enregistrements reflètent généralement l'intention déclarée du système, les entrées reçues, les limites dans lesquelles il a opéré et les actions entreprises en conséquence. Bien que le format spécifique de ces enregistrements varie selon l'organisation et la plateforme, l'exigence sous-jacente demeure la même : la transparence doit être intégrée au système à un niveau fondamental.

Étroitement liée se trouve la notion de métadonnées. Les systèmes d'IA dépendent de nombreuses variables contextuelles — versions des modèles, paramètres de configuration, identifiants de jeux de données, seuils de décision et conditions environnementales — qui influencent le comportement. En l'absence de métadonnées précises, même des changements mineurs peuvent créer une incertitude quant à la manière ou aux raisons pour lesquelles un système est parvenu à un résultat donné. DaVinciA⁺ prévoit que les métadonnées soient capturées de manière systématique et conservées comme partie intégrante de la piste d'audit. Cette approche permet que les changements puissent être tracés, que le comportement puisse être interprété avec précision et que les preuves demeurent cohérentes tout au long du cycle de vie du système. Les métadonnées servent de tissu conjonctif entre la configuration, le raisonnement et la supervision. Les catégories de métadonnées incluent généralement :

- Métadonnées de modèle (versions, paramètres, fournisseurs)
- Métadonnées de configuration (autorisations des outils, paramètres d'environnement)
- Métadonnées de décision (chemin de raisonnement, activations des garde-fous)
- Métadonnées de supervision (examinateurs, raisons d'escalade, résultats)

Ces catégories assurent la cohérence sans révéler d'éléments internes propriétaires.

Un autre élément important concerne la description formelle des limites du système. DaVinciA⁺ n'exige pas que les organisations publient leur logique interne, mais il encourage une articulation claire du périmètre, des contraintes et des fonctions autorisées. Ces descriptions aident les parties prenantes à comprendre l'usage prévu du système et à évaluer si son comportement reste cohérent avec cette finalité. Dans les environnements multi-agents, ces limites s'étendent aux relations entre agents, en précisant quelles interactions sont autorisées, comment la délégation s'opère et à quels points la supervision humaine doit intervenir. Bien que ces descriptions soient spécifiques à l'implémentation, le cadre soutient qu'elles sont consignées de manière cohérente et examinable.

La gestion des changements fait également partie de l'écosystème technique soutenant la gouvernance. Les systèmes d'IA évoluent au travers de mises à jour des modèles, des outils, des jeux de données et des règles opérationnelles. DaVinciA⁺ structure ces changements au moyen de processus formels de revue qui évaluent l'impact potentiel et déterminent si une revalidation est requise. L'annexe ne prescrit pas de flux de travail spécifiques, mais souligne l'importance de documenter la justification de chaque changement, les preuves qui le soutiennent et les décisions de supervision associées. Cette discipline vise à ce que l'évolution du système demeure délibérée et traçable, plutôt qu'incrémentale et non examinée.

Enfin, l'annexe reconnaît les mécanismes de test et de surveillance qui accompagnent un déploiement responsable de l'IA. Les organisations peuvent mettre en œuvre un ensemble de techniques — tests de vérification, évaluations comportementales, surveillance de la dérive et évaluations périodiques — afin de garantir que le système continue de fonctionner dans des paramètres attendus. DaVinciA⁺ fournit le fondement conceptuel de ces activités en définissant ce qui doit être observé, ce qui doit être enregistré et comment les décisions relatives au comportement du système sont destinés à être prises. Les spécificités de chaque méthode dépendent de l'environnement technique, du contexte réglementaire et des besoins opérationnels de l'organisation.

Pris ensemble, ces éléments illustrent l'infrastructure de soutien requise pour une IA responsable. L'annexe technique ne cherche pas à décrire l'implémentation de manière prescriptive ; elle fournit plutôt une vision cohérente des artefacts et des processus qui permettent la transparence, la supervision et la gouvernance du cycle de vie. Elle renforce l'objectif général du cadre : garantir que les systèmes d'IA demeurent intelligibles et contrôlables tout au long de leur développement, de leur déploiement et de leur évolution.

« Structure de supervision illustrative reflétant les principes de gouvernance DaVinciA⁺. Il ne s'agit pas d'une configuration système requise. »

Note :

« La matrice des seuils d'escalade est illustrative et non normative. Les organisations doivent configurer les seuils conformément à leur processus de gestion des risques. »

Journalisation d'audit et infrastructure de traçabilité

« Exemple illustratif de piste d'audit capturant une interaction multi-étapes, un blocage par garde-fou et un déclencheur d'escalade. Le format est aligné sur le schéma de métadonnées DaVinciA⁺ (DMS-AUD-070). »

{

```
"run_id": "RUN-2025-0415-0371",

"timestamp_utc": "2025-04-15T09:36:18Z",

"agent_id": "AGENT-DECISION-1A",

"user_request": "Generate draft response for regulatory comment letter",

"input_context": {

  "data_sources": ["doc://eu-ai-act-v3.4", "doc://client-guidance-notes"],

  "risk_tier": "Tier-3",

  "governance_mode": "High Oversight"

},

"steps": [

  {

    "step_id": "STEP-001",

    "timestamp_utc": "2025-04-15T09:36:19Z",

    "tool_invoked": "summarisation.agent",

    "input_summary": "Parse key terms from EU AI Act extract",

    "reasoning_snapshot": "Extracting articles relevant to classification scope",

    "guardrail_triggered": false,

    "escalation_triggered": false,

    "output_summary": "Identified Articles 6, 10, and 23 as relevant to request"

  },

  {

    "step_id": "STEP-002",

    "timestamp_utc": "2025-04-15T09:36:24Z",

    "tool_invoked": "response-generator.model-gpt4",
```

```
"input_summary": "Build draft response using regulatory summary",  
"reasoning_snapshot": "Synthesising commentary based on compliance structure",  
"guardrail_triggered": true,  
"guardrail_type": "LegalClaim-Restriction",  
"escalation_triggered": true,  
"escalation_path": "HITL_Review",  
"output_summary": "[BLOCKED] Output contained unverified conformity claim. Routed to System Owner for review."  
}  
,  
"final_output": "[Escalated to Human Reviewer]",  
"reviewer_notes": "Model attempted to assert CE conformity. Blocked and returned for rewrite. Escalation logged as E-2025-0349.",  
"audit_signoff": {  
"reviewed_by": "QA-OVERSIGHT-22",  
"review_timestamp": "2025-04-15T09:41:02Z"  
}  
}
```

Avertissement :

« Exemple non normatif présentant un format possible d'enregistrement d'audit. Les organisations peuvent mettre en œuvre des schémas alternatifs conformes à leur QMS. »

Chapitre 11 – Synthèse et glossaire

Feuille de route prospective

DaVinciA⁺ est destiné à évoluer de manière incrémentale à mesure que les attentes en matière de gouvernance, les environnements réglementaires et les pratiques opérationnelles mûrissent. Les publications complémentaires prévues incluent :

- Des modèles de modélisation des menaces et des bibliothèques de modes de défaillance
- Des orientations relatives à l'intégration de la gouvernance des données et de la protection de la vie privée
- Des bibliothèques étendues de configuration de la supervision HITL
- Des modèles de référence de risques spécifiques à des domaines (par ex. MedTech, Finance)

Ces éléments seront publiés sous forme de compléments optionnels et non normatifs et ne modifieront pas le statut de référence fondamental du cadre DaVinciA⁺.

Synthèse

Le développement et le déploiement de systèmes d'intelligence artificielle exigent un niveau de structure et de responsabilité proportionné à l'importance des décisions que ces systèmes influencent. DaVinciA⁺ fournit une approche pratique et disciplinée pour y parvenir. Il établit un cadre clair fondé sur une intention définie, un raisonnement contrôlé et une supervision continue, garantissant que les systèmes d'IA demeurent transparents, prévisibles et placés sous autorité humaine à toutes les étapes de leur cycle de vie.

L'architecture du cadre exprime le système au travers de trois couches interdépendantes qui clarifient la finalité, contraignent le comportement et soutiennent l'auditabilité. Son cycle de vie de validation étend cette structure au déploiement, en soulignant qu'une exploitation responsable requiert une surveillance et des preuves continues, et non une évaluation ponctuelle. Les pratiques de gouvernance garantissent que la responsabilité est explicite, que la supervision est intégrée aux opérations quotidiennes et que les décisions peuvent être reconstituées et examinées. L'alignement de conformité positionne les organisations pour répondre à l'évolution des attentes réglementaires au moyen de processus démontrables plutôt que de déclarations formelles. Les modèles d'adoption permettent au cadre de s'étendre à travers les domaines et les niveaux de maturité, depuis les pilotes précoce jusqu'au déploiement à l'échelle de l'entreprise. Les études de cas illustrent comment DaVinciA⁺ a déjà apporté de la structure à des projets réels, en soutenant la clarté, la traçabilité et une évolution maîtrisée.

Pris dans leur ensemble, ces composants forment une approche cohérente de la gouvernance de l'IA. DaVinciA⁺ établit une base stable sur laquelle les organisations peuvent innover de manière responsable, en maintenant la confiance dans leurs systèmes tout en naviguant dans les évolutions réglementaires. Il fournit un moyen de garantir que l'IA demeure un instrument

contrôlé, transparent et responsable, capable de soutenir des décisions complexes sans compromettre la supervision ni l'intégrité organisationnelle.

DaVinciA⁺ permet aux organisations d'établir la gouvernance avant l'échelle, les preuves avant l'audit et la clarté avant la complexité. En opérationnalisant des principes de gouvernance couramment référencés par les autorités réglementaires, il permet aux entreprises de faire progresser leurs initiatives en matière d'IA avec confiance tout en maintenant une responsabilité continue.

Glossaire

Responsabilité (Accountability)

L'obligation, pour des rôles humains identifiables, de superviser, d'évaluer et de justifier le comportement, les décisions et les résultats d'un système d'IA.

Agent

Composant spécialisé au sein d'un système d'IA qui exécute des tâches ou des fonctions de raisonnement définies dans des limites documentées et sous supervision.

Journalisation d'audit (Audit Logging)

Enregistrement systématique des actions du système, des étapes de raisonnement et des informations contextuelles afin de permettre la reconstitution et l'examen du comportement.

Gestion des changements (Change Control)

Processus structuré d'évaluation et de documentation des modifications apportées à un système d'IA, incluant l'analyse des impacts et les exigences de revalidation.

Alignement de conformité (Compliance Alignment)

Pratique consistant à structurer les systèmes et les processus de manière à soutenir les attentes des cadres réglementaires sans revendiquer la conformité.

Configuration

Paramètres techniques et opérationnels documentés définissant la manière dont un système d'IA est instancié, y compris les versions des modèles, les outils et les autorisations.

Surveillance continue (Continuous Monitoring)

Évaluation continue du comportement du système afin de détecter des écarts, des risques émergents ou des changements de performance nécessitant une intervention.

Voie de délégation (Delegation Pathway)

Interaction autorisée par laquelle un agent peut demander des informations ou une assistance à un autre agent dans des limites définies.

Dérive (Drift)

Évolution du comportement du système ou des données sous-jacentes affectant les résultats ou le raisonnement, nécessitant une surveillance et une revalidation potentielle.

Escalade

Processus par lequel un système d'IA identifie une incertitude, un risque ou une violation de limites et transfère la prise de décision à la supervision humaine.

Gouvernance

Ensemble des structures, processus et responsabilités garantissant que les systèmes d'IA opèrent dans des limites éthiques, opérationnelles et réglementaires définies.

Identité et intention (Identity and Intent)

Description formelle de la finalité, du périmètre, des contraintes et des non-objectifs du système, servant d'ancrage aux décisions architecturales et opérationnelles.

Cycle de vie (Lifecycle)

Ensemble des activités impliquées dans le développement, la validation, le déploiement, la surveillance et la mise à jour d'un système d'IA.

Métadonnées (Metadata)

Informations contextuelles décrivant le fonctionnement du système, notamment les versions des modèles, les détails de configuration et les conditions environnementales.

Supervision (Oversight)

Supervision humaine intégrée au fonctionnement du système afin d'évaluer les résultats, de traiter l'incertitude et de garantir que les décisions demeurent dans des limites autorisées.

Qualification de performance (Performance Qualification – PQ)

Évaluation du comportement d'un système d'IA dans son environnement opérationnel réel.

Processus de raisonnement (Reasoning Process)

Logique interne par laquelle un système d'IA interprète des entrées et génère des résultats, incluant les chemins de décision et l'utilisation des outils.

Gestion des risques (Risk Management)

Identification, évaluation et atténuation des préjudices potentiels associés au comportement ou à la défaillance d'un système.

Périmètre (Scope)

Ensemble autorisé de tâches, de responsabilités et de domaines dans lesquels un système d'IA peut opérer.

Traçabilité (Traceability)

Capacité à reconstituer le comportement du système au moyen d'un raisonnement documenté, de journaux d'audit et de métadonnées contextuelles.

Validation

Évaluation structurée d'un système d'IA visant à confirmer qu'il fonctionne correctement, en toute sécurité et dans des limites définies tout au long de son cycle de vie.

Annexe A — Dossier minimal de preuves pour l'examen de la gouvernance

Dossier minimal de preuves pour l'examen de la gouvernance

Les évaluations de gouvernance réglementaire et d'entreprise sont fondées sur des preuves. DaVinciA⁺ ne définit ni n'impose d'artefacts spécifiques ; il établit plutôt une structure minimale de preuves au moyen de laquelle la gouvernance, la supervision et la discipline du cycle de vie peuvent être examinées. Le tableau ci-dessous illustre un ensemble minimal de preuves représentatif, couramment attendu lors d'un examen de gouvernance ou d'audit.

Artefact

Modèle source

Description

Enregistrement d'identité et d'intention

DMS-GOV-001

Périmètre du système déclaré, limites et non-objectifs

Règles de supervision et d'escalade

DMS-GOV-011

Conditions dans lesquelles une supervision HITL est requise

Schéma de journal d'audit et exemples

DMS-AUD-070

Exécutions enregistrées, étapes, activations des garde-fous et traçabilité

Matrice RACI

DMS-GOV-010

Cartographie définie des responsabilités et de la redevabilité

Registre de gestion des changements

DMS-CC-061

Modifications du système documentées et évaluations d'impact

Rapports IQ / OQ / PQ

DMS-VAL-021 / 031 / 041

Preuves de validation d'installation, comportementale et en conditions réelles

Journal de surveillance de la dérive

DMS-MON-050

Enregistrements de détection de dérive statistique et comportementale

Ces artefacts correspondent aux structures définies dans le kit de validation et le guide de déploiement DaVinciA⁺ et sont présentés à des fins illustratives de gouvernance uniquement.