Trilogue Annexe I : le rail signe le dossier, pas le système

Le faux débat

Le débat européen sur les dispositifs médicaux intégrant de l'IA est en train de confondre deux objets distincts : le cadre qui certifie le dossier et l'architecture qui gouverne la décision. Cette confusion structure aujourd'hui le trilogue.

Le trilogue du 28 avril 2026 sur l'Annexe I s'est enlisé sur une question désormais centrale : quelle architecture réglementaire retenir pour les dispositifs médicaux intégrant de l'intelligence artificielle.
Le débat oppose deux positions connues.

Le Parlement européen, plusieurs États membres et une partie des fédérations sectorielles défendent un rail sectoriel unique. Leur argument est cohérent : le couple MDR/IVDR, complété par IEC 62304, ISO 14971, ISO 13485, les protocoles de validation clinique et la surveillance post-marché, constitue déjà un dispositif dense de maîtrise du risque logiciel sur le cycle de vie. Ajouter une couche horizontale supplémentaire introduirait redondances documentaires, friction réglementaire et ralentissement industriel.
La Commission maintient au contraire une logique de rail croisé MDR + AI Act, au motif que certains risques propres aux systèmes IA ne sont pas correctement couverts par les cadres sectoriels existants.

Les deux positions sont techniquement défendables.

Elles partagent pourtant une même prémisse, rarement explicitée : qu'un rail réglementaire puisse produire à lui seul la gouvernabilité d'un système d'IA en production.

C'est cette prémisse qui ne tient pas.

Le choix du rail détermine le dossier à signer. Il ne détermine pas si le système reste gouvernable au moment exact où il décide.

Une propriété d'exécution, pas de cycle de vie

Le problème n'est pas documentaire. Il est architectural.

Un système d'IA clinique n'engage pas sa responsabilité lorsqu'il est développé, documenté ou certifié. Il l'engage lorsqu'il produit une décision opposable sur un patient réel, sous distribution ouverte, dans des conditions opérationnelles imparfaitement contrôlées. C'est à ce moment précis que se joue la gouvernabilité.

La gouvernance d'un système probabiliste n'est donc pas une propriété de cycle de vie. C'est une propriété d'exécution à l'inférence.

Cette distinction n'est pas académique. Un système peut être simultanément validé, certifié, traçable, surveillé et conforme au cadre applicable … et rester incapable de déterminer, au moment exact où il produit une décision, si cette décision doit être *acceptée, refusée, dégradée, transférée vers une autre voie de calcul* ou *escaladée vers un humain*. Or c'est précisément cette capacité qui décide si un système est gouvernable.

Le débat actuel organise principalement la conformité du cycle de vie. Il ne spécifie pas les propriétés runtime minimales qui rendent une décision IA juridiquement et opérationnellement gouvernable.

Le point aveugle commun aux deux rails

Les deux architectures réglementaires actuellement débattues régulent principalement des processus, des obligations documentaires, des mécanismes de gestion du risque et des procédures de surveillance ex post. Aucune n'impose explicitement des propriétés d'exécution auditables à l'inférence.

Cette nuance est déterminante. Une propriété dérivable n'est pas une propriété opposable.

Un texte peut permettre d'inférer qu'un fabricant devrait surveiller la dérive d'un modèle, contrôler certaines entrées ou limiter certains usages. Tant que ces propriétés ne sont ni explicitement spécifiées, ni auditables en production, ni vérifiées à l'exécution, elles restent interprétables, hétérogènes, difficilement auditables et juridiquement fragiles. Le système peut alors rester conforme tout en prenant des décisions non gouvernables.

Le mécanisme réel de défaillance

Le mécanisme de défaillance observé dans les systèmes probabilistes est relativement simple.

Une entrée arrive hors de la distribution sur laquelle le modèle a été calibré : nouvelle population, capteur dégradé, protocole clinique différent, combinaison biologique rare, données incomplètes ou dérive temporelle silencieuse.
Le système produit néanmoins une inférence.

- ⇨ La probabilité calculée est interprétée comme une décision exploitable.

Aucune qualification du signal n'intervient ; aucun filtrage d'admission ne bloque l'exécution ; aucune route de repli ne s'active. Une décision clinique, prudentielle ou économique est prise sur un signal dont la fiabilité réelle n'est pas qualifiée.

L'échec ne provient pas seulement d'une erreur algorithmique. Il provient de l'absence de propriétés architecturales capables d'interrompre la chaîne de défaillance avant production d'une décision opposable.
Le système n'a pas besoin de savoir qu'il est en erreur. Il doit savoir qu'il pourrait l'être.
C'est cette capacité, distincte de la justesse, qui sépare un actif logiciel gouvernable d'une dette opérationnelle probabiliste.

Les signaux empiriques

Les données disponibles ne contredisent pas ce diagnostic.

Une étude publiée en 2025 dans *JMIR Medical Informatics* (Chen, Teng, Kuo *et al.*, 2025[1]), portant sur vingt-sept années de rappels AI/ML aux États-Unis à partir de la base openFDA, documente que les dispositifs IA cleared FDA présentent un taux de rappel de l'ordre de 5,8 %, avec une prédominance des causes liées à la conception logicielle et au design des dispositifs, ces dernières représentant approximativement 50 % des causes-racines identifiées sur la cohorte AI/ML.
Une analyse complémentaire sur la FDA AI-Enabled Medical Devices List[2] confirme par ailleurs qu'une part minoritaire des dispositifs listés documente publiquement des évaluations de sécurité prémarché détaillées.
Ces données n'établissent pas une causalité unique. Elles restent néanmoins compatibles avec un constat plus structurel : les cadres actuels savent largement auditer le cycle de vie logiciel ; ils spécifient encore insuffisamment les propriétés runtime qui gouvernent la décision probabiliste en production.

Le critère architectural minimal

À quoi ressemble une gouvernance effectivement inscrite dans l'architecture du système ?
À une spécification minimale de trois contrôles exécutables à l'inférence : *contrôle de validité, contrôle d'admission, contrôle de transfert*. Ces trois fonctions ne constituent pas des préférences d'ingénierie. Elles correspondent aux trois points où la chaîne de défaillance peut être interrompue avant production d'une décision opposable. En supprimer une suffit à réintroduire le risque systémique.

Contrôle de validité

Un système gouvernable doit déterminer explicitement si le signal reçu appartient encore à la zone pour laquelle sa performance a été démontrée. Cela suppose une validation sans fuite méthodologique, un split respectant les conditions réelles d'apparition des données futures, une calibration vérifiée sur jeu indépendant, et la déclaration explicite d'un domaine d'applicabilité. En dehors de ce domaine, l'inférence doit être considérée comme non fiable par défaut.
Le dispositif technique correspondant, la déclaration explicite de la zone dans laquelle la performance reste revendiquée, constitue le *port de validité* du système. La grammaire des ports a été développée dans l'Article V de la série hexagonale ; le port de validité en est l'instance la plus immédiatement opérationnelle pour les déploiements en environnement régulé.
La preuve minimale exigible ne réside pas dans une courbe de performance supplémentaire. Elle se compose d'un protocole de validation publié, d'un domaine d'applicabilité documenté, et des conditions de dégradation attendues hors de ce domaine.
Un système incapable de qualifier son propre périmètre de validité ne peut pas opposer sa décision.

Contrôle d'admission

Un système gouvernable doit décider si une requête est admissible avant même l'appel du modèle. La décision d'admission ne peut être ni implicite, ni statistique, ni laissée à l'interprétation de l'utilisateur final : elle doit produire explicitement l'une des trois sorties :

1. *acceptation,*
2. *rejet,*
3. *reroutage*

sur la base de règles publiables et instrumentées.

Le point critique n'est pas la capacité du modèle à répondre. Le point critique est sa capacité à refuser.
La preuve minimale exigible se compose des règles d'admission publiées, de la typologie des rejets, et du taux observé de refus et de reroutage.

Sans contrôle d'admission, le système ne gouverne pas son espace d'action. Il le subit.

Contrôle de transfert

Un système gouvernable ne peut pas reposer sur une route de calcul unique. Il doit comporter au minimum une route nominale, une route de repli vers un modèle calibré de référence, et une route d'escalade vers un humain ou un système spécialisé. Les critères de bascule entre ces routes doivent être explicitement définis, journalisés et activables.
L'escalade ne peut pas être une formule réglementaire abstraite du type « validation humaine requise ». Elle doit correspondre à une mécanique d'exécution effectivement activable. Une supervision humaine non routée n'est pas une garantie. C'est un récit de conformité.
Le Predetermined Change Control Plan (PCCP) imposé par la FDA pour les SaMD constitue probablement le premier déplacement réglementaire explicite vers une gouvernance pré-spécifiée des comportements évolutifs du système : la route d'évolution du modèle y est pré-spécifiée, journalisée et activable selon des critères publiés en amont du déploiement. Le PCCP ne formalise toutefois que la gouvernance des modifications du modèle ; il ne couvre pas, à lui seul, le mécanisme d'escalade décisionnelle à l'inférence. Il en marque cependant la direction. La question n'est plus de savoir si un humain regarde. C'est de savoir si la machine sait quand le déranger.
La preuve minimale exigible se compose des critères de bascule publiés, du taux d'escalade observé et des délais de prise en charge des routes de repli.

Sans contrôle de transfert, un système probabiliste ne possède aucun mécanisme structurel de confinement de ses propres conditions de défaillance.

Une doctrine industrialisable

Cette architecture n'est pas théorique. Elle existe déjà dans certains systèmes multimodèles industriels où la route nominale est conditionnée par qualification préalable du signal, où un fallback calibré prend le relais en cas d'incertitude, et où une escalade humaine est déclenchée hors domaine d'applicabilité.

Sur PREDICARE, programme territorial de médecine prédictive structuré pour le GHT Aube, l'architecture de référence implémente cette séparation par construction. Chaque signal entrant, qu'il provienne d'un capteur connecté, d'un système d'information hospitalier ou d'un acteur ambulatoire, est qualifié selon une typologie *Bronze / Silver / Gold* avant toute inférence. Un signal Bronze (tensiomètre relevé à 14h37 sans contexte clinique vérifié, par exemple) ne franchit pas la couche d'admission. Un signal Silver autorise une inférence sur route nominale. Un signal dégradé en cours d'exécution bascule vers une route de repli sur modèle calibré de référence. Un signal hors domaine d'applicabilité signé déclenche une escalade vers le praticien clinique de la CPTS de référence, dans un délai défini, traçable et auditable.
Cette instance ne démontre pas que la doctrine est universellement satisfaite. Elle démontre qu'un système portant explicitement les trois contrôles à l'inférence est techniquement industrialisable, et que son architecture de référence peut être spécifiée dans un environnement régulé réel, et non pas seulement dans un démonstrateur de laboratoire.

Une lacune symétrique UE / US / CN

La comparaison internationale ne modifie pas le diagnostic.
Le cadre européen, qu'il soit sectoriel ou croisé, n'impose pas explicitement les trois contrôles d'exécution. Le QMSR américain, entré en vigueur le 2 février 2026, nomme la dérive des modèles et la gouvernance des données sans exiger formellement la délimitation du domaine d'applicabilité, le filtrage en ligne des entrées ni le routage multi-chemins à l'inférence. Côté chinois, la NMPA simplifie les procédures de change registration tant que l'algorithme central reste considéré comme stable ; là encore, le contrôle porte principalement sur l'identité du modèle et son cycle de modification, non sur les propriétés d'exécution de la décision probabiliste.
La lacune est donc symétrique. Elle ne relève pas d'un retard européen. Elle relève de l'état actuel de la régulation mondiale des systèmes probabilistes temps réel.

Coût assumé et zones de friction

Les trois contrôles ne sont pas gratuits.

Le contrôle de validité impose un protocole de validation plus exigeant et la maintenance continue d'un domaine d'applicabilité explicite (déclaration qui doit être révisée à chaque évolution de la population cible, du protocole clinique ou du capteur amont).
Le contrôle d'admission ajoute une latence à chaque inférence, qui devient critique sur les pipelines à fort volume ou sur les contextes d'urgence.
Le contrôle de transfert exige une journalisation continue et une infrastructure d'escalade (humaine ou algorithmique) qui doit elle-même être disponible, calibrée et auditable.

Trois zones de friction méritent d'être nommées sans les minimiser.

1. Première friction : la déclaration d'un domaine d'applicabilité reste techniquement délicate pour les modèles fondationnels généralistes, dont la zone de performance est par construction floue et évolutive. La doctrine ne supprime pas cette difficulté ; elle l'oblige à être explicitée plutôt que masquée. Le bénéfice est d'ordre épistémique, pas calculatoire.
2. Deuxième friction : le contrôle d'admission peut entrer en tension avec les exigences de réactivité clinique en situation d'urgence. La doctrine n'impose pas une admission bloquante

par défaut ; elle impose une décision typée explicite, qui peut inclure un mode dégradé prioritaire activable et journalisé. La gouvernabilité n'est pas synonyme de lenteur, mais elle est synonyme de traçabilité de l'arbitrage.

3. Troisième friction : le coût opérationnel total est mesurable, en heures de calcul, en lignes de code et en équivalents temps plein. Le coût alternatif, premier contentieux, premier audit contradictoire, premier incident critique sur signal mal qualifié, n'est ni mesurable a priori ni bornable a posteriori. Le choix n'est pas entre gouvernabilité et performance. Il est entre coût d'architecture connu et coût d'incident inconnu.

Le véritable enjeu du trilogue

Le trilogue reste utile. Il organise des responsabilités, définit des périmètres, clarifie des obligations et structure l'industrialisation réglementaire européenne. La doctrine ici défendue n'entend pas s'y substituer ; elle propose d'en étendre le périmètre de spécification du cycle de vie vers les propriétés runtime. Mais un cadre ne peut pas résoudre une propriété qu'il n'a pas explicitement spécifiée.

La thèse défendue ici reste falsifiable. Il suffirait qu'un dispositif conforme à un cadre existant publie une spécification explicite de ses propriétés d'exécution, des métriques runtime auditables, des mécanismes vérifiables de qualification, de rejet et d'escalade, et un audit indépendant en conditions réelles, pour démontrer qu'un cadre actuel suffit déjà à produire cette gouvernabilité. À défaut de tels dossiers publics, la conclusion tient.

Le rail signe le dossier.

La gouvernabilité se joue à l'inférence.

Sans contrôle de validité, sans contrôle d'admission, sans contrôle de transfert, aucune des deux architectures réglementaires actuellement débattues ne garantit qu'un système reste gouvernable au moment exact où il décide.

Cette note traite formellement la gouvernance des dispositifs médicaux IA dans le contexte du trilogue européen. Elle pose en filigrane une question plus large, qui sera l'objet de la 4/6 : à quelles conditions une décision reste-t-elle opposable lorsque le moteur qui la produit est probabiliste et partiellement non déterministe ? C'est à cette question, et non à celle de la conformité, que ***la gouvernance opérationnelle*** devra répondre.

[Série : Gouvernance = architecture - 3/6]
Voir également :
L'architecture hexagonale n'est pas seulement un pattern de développement. C'est une condition structurelle de la gouvernabilité. · Twingital Institute

Le port de promotion contractualisé : ce que la FDA a construit avant que les architectes ne le nomment · Twingital Institute

Le Framework RAISE · Twingital Institute