Hors doctrine, avant doctrine

Pourquoi l'agentique en santé sera d'abord régulée par les assureurs, les contrats et les juges

1. Quatre événements régulatoires, un acte négatif structurant

Entre janvier et avril 2026, quatre actes régulatoires ont été produits dont la mise en série n'a pas encore reçu de lecture commune.

1. Le 12 mars 2026, le Centers for Disease Control and Prevention publie *Considerations for Agentic Research in Public Health*, document fédéral américain qui distingue explicitement l'agentique de la générativité conversationnelle et publie des principes opératoires pour son usage en santé publique. Le document est suivi le lendemain par un *Considerations for Generative AI in Public Health* qui en marque, par contraste, la spécificité.
2. Le 2 avril 2026, la Food and Drug Administration adresse à Purolea Cosmetics Lab, fabricant homéopathique OTC à Livonia (Michigan) sous CGMP en vertu de 21 CFR parties 210 et 211, une warning letter dont une section dédiée s'intitule *Inappropriate Use of Artificial Intelligence in Pharmaceutical Manufacturing*. Le grief est inscrit dans 21 CFR 211.22(c) : la Quality Unit avait délégué à des agents IA la rédaction de spécifications, de procédures et de master production records sans revue compétente. La firme a depuis cessé sa production pharmaceutique.
3. En janvier 2026, Mindgard divulgue à Doctronic, assistant clinique IA déployé dans un sandbox réglementaire en Utah, une chaîne de vulnérabilités exploitable par injection de contexte. La publication publique du rapport intervient en mars 2026 : un attaquant peut faire ingérer à l'agent un faux bulletin réglementaire, modifier ses recommandations, et contaminer durablement le patient context via les SOAP notes générés. Une dose d'OxyContin triplée peut être routée vers un clinicien humain pour validation, présentée comme document clinique structuré.
4. Le 2 février 2026, la Commission européenne manque pour la seconde fois la deadline statutaire fixée par l'Article 6 §5 de l'AI Act : les guidelines sur la classification des systèmes haut-risque ne sont pas publiées. La proposition Digital Omnibus du 19 novembre 2025 prévoit d'aligner l'application des règles haut-risque sur la disponibilité effective des standards harmonisés CEN-CENELEC, dont la première deadline avait elle-même été manquée à l'automne 2025.

À ces quatre actes positifs s'ajoute un acte négatif structurant : ***la Commission européenne a retiré, en février 2025, sa proposition d'AI Liability Directive***. Le cadre civil de responsabilité spécifique à l'IA, qui aurait pu qualifier les comportements émergents distribués, est abandonné. La Product Liability Directive révisée 2024/2853 demeure ; sa portée et ses limites sont discutées en §6.

Pris isolément, aucun de ces événements n'établit une doctrine régulatoire stabilisée. Leur mise en série, en revanche, révèle un fait que les commentaires usuels manquent : la doctrine régulatoire ne se déplace pas vers le risque ; elle est en train d'être contournée par d'autres forces, économiques, contractuelles et jurisprudentielles, qui décident à sa place. Le présent article tient que cette substitution est en cours, qu'elle est structurelle, et qu'elle déterminera la régulation effective de l'agentique en santé pour les années qui viennent.

2. Trois locus, pas deux

Les commentaires usuels lisent l'inadéquation régulatoire comme une dissociation à deux termes : ce que la régulation nomme, et ce que les systèmes font. Cette lecture fusionne implicitement la production du comportement et le risque qu'il porte. Elle est trop simple pour saisir un système qui ne pardonne pas les simplifications.

Trois locus distincts doivent être nommés.

1. *Le locus de production du comportement* : l'orchestration. C'est l'endroit où le comportement effectif du système est composé, à l'exécution, à partir de modèles, de règles, de mémoires et de données. Il n'est pas réductible à un composant.
2. *Le locus du risque* : l'émergence distribuée. C'est l'endroit où survient un comportement non aligné avec la finalité régulée. Ce locus ne coïncide pas avec le précédent : certaines orchestrations produisent du comportement sans produire de risque critique ; certains risques apparaissent hors orchestration (par dérive de données, par défaut d'intégration humaine, par exposition non qualifiée à des sources externes). La production est nécessaire mais non suffisante au risque.
3. *Le locus de régulation* : les points d'ancrage institutionnels. C'est l'endroit où la régulation peut effectivement intervenir, contrôler, sanctionner. Il est défini par des objets : dispositifs, opérateurs identifiables, processus auditables.

Ces trois locus ne sont pas équivalents. Leur non-coïncidence est le fait structurel qu'aucun cadre régulatoire actuel n'opère explicitement. La doctrine confond production et risque, comme si gouverner la première suffisait à contenir le second. Elle confond aussi régulation et locus du risque, comme si les ancrer ensemble suffisait à les faire

coïncider. Aucune de ces deux confusions ne tient à l'épreuve d'une chaîne agentique persistante.

3. Antécédents conceptuels

La propriété centrale défendue dans ce texte ne s'invente pas *ex nihilo*. Elle s'inscrit dans une lignée d'analyses qui ont, à différents niveaux, identifié l'inadéquation des cadres d'imputabilité classiques face à des systèmes complexes.

- La théorie de l'accident normal (Perrow, 1984) établit que dans les systèmes à couplage serré et interactions complexes, certains accidents sont structurellement non attribuables à une cause unique : ils émergent de la combinaison de défaillances individuellement tolérables. Cette ligne thématise l'irréductibilité de la causalité dans les systèmes complexes.
- L'*accountability gap*, largement développé depuis Mittelstadt et al. (2016) en éthique des données et de l'IA, désigne l'écart entre la complexité des systèmes algorithmiques et la capacité des cadres juridiques et organisationnels à identifier des responsables. Cette ligne thématise la question juridique.
- La *distributed responsibility* (Floridi & Sanders, 2004 ; développée dans les travaux ultérieurs sur l'agentivité morale artificielle) reconnaît que la responsabilité, dans des systèmes multi-agents, peut être structurellement distribuée plutôt qu'attribuable à un agent unique. Cette ligne thématise la question éthique.
- La *causal opacity* et le *traceability gap*, dans les littératures techniques sur l'IA explicable et l'audit algorithmique, désignent l'impossibilité pratique de suivre la chaîne causale interne à un modèle ou à un système.

Ces lignes se croisent sans coïncider strictement. Elles partagent l'idée que certains comportements ne sont pas attribuables sans perte d'information pertinente. La présente analyse ne prétend pas refonder cette idée. Elle prétend la spécifier pour une classe particulière de systèmes, les chaînes agentiques persistantes, où trois caractéristiques (composition dynamique, persistance structurée traversant les sessions, exposition à des sources externes au périmètre certifié) produisent une non-localisabilité spécifique, observable, et juridiquement conséquente.

4. Non-localisabilité causale : spécialisation aux chaînes agentiques persistantes

La non-localisabilité causale ne désigne pas l'absence de causalité. Elle désigne l'impossibilité pratique et juridique d'attribuer le comportement observable à une transition unique sans perdre l'information pertinente sur la trajectoire. Elle prolonge les notions d'*accountability gap* et de *distributed responsibility*, mais les spécifie dans un cas

particulier : les chaînes agentiques persistantes où mémoire, contexte externe et composition dynamique modifient la sortie au fil des sessions.

Énoncé sous sa forme structurelle :

Pour une chaîne d'orchestration qui maintient une mémoire structurée traversant les sessions, la défaillance observable n'est pas, en général, attribuable à une étape identifiable de la chaîne sans perte d'information pertinente sur la trajectoire. Cette non-attribution est une propriété de la chaîne, pas un défaut de l'observation.

Trois caractéristiques constitutives la produisent :

1. La composition dynamique (la sélection des modèles dépend du contexte d'exécution),
2. La persistance structurée (l'information transite et se reformule entre sessions via des artefacts typés),
3. Et l'exposition à des sources externes au périmètre certifié (le contexte d'inférence est partiellement non qualifié).

Ces trois caractéristiques sont précisément celles qui font la valeur clinique de l'agentique. On ne peut donc ni les supprimer pour faciliter la régulation, ni espérer une chaîne agentique utile qui ne les manifesterait pas.

La conséquence pour les cadres régulatoires classiques est d'ordre structurel, non contingent. Ces cadres (SaMD, AI Act, CGMP) opèrent sous une hypothèse implicite d'attribution causale : pour qu'un opérateur soit tenu responsable d'un comportement, il faut que le comportement soit imputable à une étape identifiable. Quand cette hypothèse cesse de tenir, le cadre ne saisit plus son objet comme objet régulatoire autonome ; il continue à saisir des effets et des responsabilités via des prises indirectes, mais il cesse de formaliser le générateur du comportement.

5. Doctronic : preuve d'existence technique documentée

Le cas Doctronic n'est pas une illustration du phénomène. Il en est une preuve d'existence technique documentée, au sens où il établit, sur un système qu'il est légitime de déployer dans un cadre régulatoire formellement valide, que la propriété énoncée au paragraphe précédent est techniquement observable. Le matériau provient d'un rapport de sécurité publié, non d'un acte régulatoire ni d'une décision juridictionnelle ; cette nature même du matériau participe de l'argument.

L'attaque ne porte pas sur les poids du modèle. Elle ne porte pas sur une session unique. Elle exploite la persistance des SOAP notes (résumés cliniques structurés que l'assistant génère pour le clinicien) comme vecteur de propagation. ***Un faux bulletin réglementaire injecté dans une session compromise contamine le patient context, qui est ré-***

injecté lors des sessions suivantes, présenté comme contexte clinique légitime, et finit validé par un clinicien humain.

La trajectoire de la défaillance traverse trois moments : la fenêtre d'injection (session N), la transduction du contenu manipulé en SOAP note structuré (frontière session/persistance), la lecture humaine confiante (session N + k). Aucun de ces trois moments, pris isolément, ne suffit à produire le comportement observé. Aucun, en conséquence, ne suffit à le qualifier juridiquement. La cause se distribue dans la trajectoire, exactement comme la spécialisation énoncée plus haut le prédit.

Trois conséquences en découlent.

1. La frontière de confiance n'est pas inscrite dans le système : le SOAP note est traité par le clinicien comme appartenant à un registre d'autorité (celui du dossier médical) alors qu'il provient d'une frontière où le contenu utilisateur a été promu en contexte régulé sans qualification explicite.
2. la supervision humaine est techniquement présent et substantiellement absent : le clinicien valide, mais ce qu'il valide n'est plus reconstructible à partir de la sortie qu'il observe.
3. L'imputabilité, enfin, est non assignable sans convention : le déployeur peut invoquer l'attaquant, l'attaquant n'est pas qualifié juridiquement, le clinicien validateur a respecté la procédure, et le sandbox régulateur de l'Utah ne couvre pas ce cas. Toute attribution sera donc *décidée*, par convention contractuelle ou jurisprudentielle, et non *établie* par le cadre régulatoire.

Doctronic, en l'état des éléments publics disponibles, établit ainsi que la non-localisabilité causale n'est pas une abstraction conceptuelle, mais une propriété techniquement observable sur des systèmes en exploitation. La généralisation de cette observation à l'ensemble des architectures du même type est une question empirique ouverte ; sa pertinence comme preuve d'existence ne l'est pas.

6. AI Act et PLD : prises indirectes solides, pas formalisation explicite

Une critique honnête doit nommer ce que la doctrine européenne fait mieux que la doctrine américaine. Quatre points d'appui méritent d'être mobilisés sans sous-estimation.

- L'**Article 25** définit la responsabilité distribuée le long de la chaîne de valeur : tout opérateur qui modifie substantiellement un système à haut risque devient provider.
- L'**Article 14** impose un oversight humain effectif.

- L'**Article 15** exige robustesse, exactitude et cybersécurité, incluant la résilience aux manipulations.
- Les **Articles 56 et 95** structurent les Codes of Practice et codes de conduite, instruments que la Commission a mobilisés pour traiter, en aval de la doctrine, ce qu'elle ne pouvait pas formaliser en amont.

Ces prises sont solides. Elles ne sont pas la formalisation de l'orchestration agentique comme objet régulatoire autonome. La distinction est précise et mérite d'être tenue.

L'Article 25 traite la composition statique : qui devient responsable de quoi quand un acte d'opérateur identifiable modifie le système. Il ne formalise pas la composition dynamique, où la modification du comportement résulte d'une trajectoire d'exécution sans acte d'opérateur qualifiable. La *substantial modification* qu'il exige est une catégorie juridique, pas une catégorie comportementale.

L'Article 14 traite la supervision comme exigence formelle : le système doit *permettre* la supervision. Il ne formalise pas la supervision comme propriété effective, c'est-à-dire le fait que la sortie soit *réellement* analysable par un humain placé dans des conditions cognitives ordinaires. Doctronic satisfait l'Article 14 dans sa lettre : le clinicien valide. Doctronic le viole dans sa fonction : la validation est captive d'une présentation qui en court-circuite la possibilité.

L'Article 15 traite la robustesse du *système tel que défini*, c'est-à-dire le dispositif au sens MDR ou le système d'IA au sens de l'AI Act. Il ne formalise pas la robustesse de la *trajectoire d'exécution* qui traverse plusieurs systèmes, plusieurs sessions, et plusieurs frontières de confiance. La robustesse Article 15 est une propriété du périmètre certifié. Le risque, lui, sort du périmètre.

Les Articles 56 et 95, par leur existence même, signalent ce que les articles précédents ne peuvent accommoder : la Commission renvoie aux fournisseurs de GPAI le soin de structurer les exigences applicables, c'est-à-dire reconnaît implicitement qu'elle ne peut formaliser ce qu'elle régule. Cette reconnaissance est doctrinalement importante. Elle est précisément le point d'entrée du troisième régime régulatoire (§8).

La PLD 2024/2853 mérite la même précision. Elle modernise effectivement le régime de responsabilité produit en incluant les logiciels, les services numériques liés au produit, les mises à jour, et les défauts de cybersécurité parmi les éléments susceptibles de constituer un défaut. Elle peut donc saisir le défaut d'un composant logiciel (y compris IA) ou d'un service numérique intégré, pris isolément. Elle ne saisit pas en revanche la causalité trajectorielle multi-acteurs, multi-sessions, multi-frontières comme objet autonome, c'est-à-dire la défaillance qui émerge de la composition plutôt que de l'un de ses composants. Le retrait de l'AI Liability Directive aggrave cette limite : la proposition initiale visait des présomptions de causalité spécifiques à l'IA, capables de saisir des

comportements distribués sans cause unique. Sa disparition replace l'imputabilité dans un régime qui sait traiter les composants (et leurs défauts), mais qui ne sait pas traiter les trajectoires comme telles.

La doctrine européenne, sur l'agentique, n'est pas naïve. Elle est partiellement inadéquate, ce qui n'est pas le pire défaut.

7. Purolea : la divergence cumulative

L'événement Purolea ne signale pas un déplacement régulatoire. Il signale l'inverse : la stabilité de la régulation, doctrinalement maintenue, face à un risque qui dérive sans elle.

La FDA n'a pas régulé l'agentique. ***Elle a appliqué 21 CFR 211.22(c), disposition existante sur la responsabilité de la Quality Unit, à un cas où un agent IA avait remplacé la supervision humaine***. L'événement est doctrinalement orthodoxe : la FDA défend son périmètre régulatoire en refusant que l'agent IA déplace la responsabilité. Le message implicite est exact : ***la régulation reste où elle est ; à l'opérateur de compenser en organisation ce que l'agent ne fournit pas en conformité***.

Cette stabilité doctrinale a un coût qui n'est pas immédiatement visible. À chaque cycle d'innovation agentique, une nouvelle classe de comportements distribués apparaît dans des architectures déployées. À chaque cycle, la régulation maintient son périmètre. La surface du risque effectif s'éloigne donc, à chaque cycle, du périmètre régulé. Ce n'est pas seulement une asymétrie statique entre locus du risque et locus de régulation. C'est une **divergence cumulative** : l'écart se creuse à chaque innovation, et le rattrapage doctrinal devient progressivement infaisable sans refonte explicite.

La doctrine régulatoire actuelle est dans la situation suivante : elle peut rester stable longtemps, et chaque jour de stabilité accroît l'écart entre ce qu'elle gouverne et ce qui décide effectivement.

8. Trois régimes régulatoires, pas deux

La binaire « régulation suit le risque vs. régulation reste où elle peut tenir » est utile mais incomplète. Un troisième régime existe, opérant, et largement non thématisé par la doctrine.

1. *Premier régime : l'extension formelle.* Le périmètre régulé est étendu pour absorber l'orchestration : refonte du SaMD (MDR Article 2(1), MDCG 2019-11), qualification de la trajectoire d'exécution comme objet certifiable, infrastructure d'audit dynamique. Cette voie est concevable mais coûteuse, lente, et politiquement difficile. Elle n'est pas, en l'état, étayée par les actes régulatoires observés.

2. *Deuxième régime : le maintien et la stratification fonctionnelle*. La régulation préserve son point d'ancrage et impose des compensations procédurales aux opérateurs : AI literacy obligatoire (AI Act Article 4), responsabilité de la Quality Unit (CGMP), data governance distribuée. Purolea relève de ce régime. Il est doctrinalement conservateur et opérationnellement courant.
3. *Troisième régime : l'auto-régulation industrielle contrainte par responsabilité ex post*. Les plateformes définissent et imposent leurs propres règles d'orchestration : guardrails de modèle, tool policies, context filters, logging interne, terms of service contractualisant le périmètre d'usage. Cette régulation n'est pas optionnelle. Elle est rendue obligatoire par la pression conjointe de la responsabilité civile (PLD 2024/2853), de la responsabilité contractuelle envers les déployeurs institutionnels, et de la responsabilité réputationnelle face aux incidents publiquement instruits.

Ce troisième régime est largement déjà institutionnalisé, à un endroit que la doctrine n'a pas su lire correctement : les Codes of Practice et codes de conduite des Articles 56 et 95 de l'AI Act lui-même. Quand l'AI Act renvoie aux fournisseurs de GPAI le soin de structurer les exigences applicables aux modèles à risque systémique, il reconnaît implicitement que la doctrine ne peut pas formaliser ce qu'elle régule. Le régime trois n'est pas un contournement du cadre européen ; il en est partiellement une instanciation que celui-ci n'assume pas comme telle.

L'auto-régulation industrielle n'attend pas la doctrine pour exister. Elle est déjà la doctrine, sur les zones que la doctrine ne formalise pas. Elle a vocation à être dominante sur l'orchestration agentique en santé avant toute évolution formelle.

9. L'économie comme moteur, pas comme contexte

La trajectoire régulatoire effective ne sera pas déterminée par la doctrine. Elle sera déterminée par trois forces économiques qui n'attendent pas la doctrine pour décider.

1. *Les assureurs*. La couverture cyber et responsabilité civile professionnelle des opérateurs déployant des systèmes agentiques en santé est en cours de redéfinition. Les premières exclusions explicites pour systèmes IA non auditables apparaissent en 2025 sur le marché américain. La réécriture des périmètres assurables (primes différentielles selon présence de mécanismes d'audit, retrait progressif de couvertures sur architectures non qualifiées) précède la régulation. Elle la *définit*, par effet de structure : un système non assurable n'est pas déployable, indépendamment de sa conformité régulatoire formelle.
2. *Les tribunaux*. La responsabilité civile dans le cadre de la PLD 2024/2853 va connaître ses premiers cas instruits sur des architectures agentiques en santé. La question est connue : qui paie quand un patient subit un dommage suite à une

recommandation produite par une chaîne dont la cause se distribue ? Elle sera tranchée *cas par cas*, par jurisprudence, avant qu'aucune doctrine régulatoire n'en ait formalisé les conditions. La forme effective de la responsabilité émergera de quelques décisions structurantes, agrégées par effet précédent. Ces décisions imposeront aux déployeurs des contraintes opérationnelles que la régulation reprendra ensuite, peut-être, par voie d'implémentation.

3. *Les industriels.* Face à ces deux forces, les arbitrages des opérateurs ne se font pas entre conformité et non-conformité. Ils se font entre trois positions : conformité stricte (qui ralentit le time-to-market et augmente le coût marginal), externalisation contractuelle (qui transfère le risque résiduel à l'utilisateur via terms of service et disclaimers), et auto-régulation préventive (qui investit dans des socles d'audit pour offrir des garanties contractuelles aux clients institutionnels). Le troisième terme est, pour les acteurs ayant une exposition réputationnelle, le plus rationnel. Il alimente directement le régime trois.

Cette dynamique n'est pas inédite. Dans la finance post-2010, l'encadrement effectif du trading algorithmique s'est largement formé par enforcement actions de la SEC, standards d'industrie de risk management et conventions FINRA, adoptés en réaction au flash crash du 6 mai 2010 et à la défaillance Knight Capital de 2012, plutôt que par anticipation doctrinale ; la Regulation SCI de 2014 codifie *après coup* des pratiques que les acteurs avaient déjà dû imposer. Dans l'aéronautique automatisée, la certification FAA a été réformée par l'Aircraft Certification, Safety, and Accountability Act de 2020 *en réponse* aux investigations NTSB et aux litiges consécutifs aux accidents 737 MAX (2018-2019), pas en anticipation. Dans les deux cas, la doctrine publique a codifié, après coup, les compromis que des acteurs économiques et juridictionnels avaient dû imposer au cours du cycle. Aucune raison structurelle ne donne à l'agentique en santé une trajectoire différente.

Ces trois forces ne se coordonnent pas. Elles convergent par effet de réseau. Et leur convergence définit la régulation effective : l'ensemble des contraintes opérationnelles auxquelles un système agentique en santé est *réellement* soumis, bien avant que la doctrine n'en ait codifié les principes. L'économie n'est pas le contexte de la régulation. Elle en est le moteur.

10. Cas-limites conceptuels : PREDICARE et TweenMe

Le présent texte ne soutient pas que ces deux instances valident la thèse comme preuves générales. Elles servent ici de cas-limites conceptuels : terrains d'épreuve où la non-localisabilité causale devient observable et où les régimes décrits ci-dessus deviennent opératoires.

- PREDICARE, architecture de médecine prédictive séquentielle (ingestion multi-source, inférence contextuelle, recommandation graduée, suivi longitudinal), admet une qualification SaMD composant par composant, mais le comportement global dépend de la trajectoire.
- TweenMe, chaînage de modèles hétérogènes (LLM contextuel, modèles tabulaires de risque, modèles de survie type Fine & Gray, oracles toxicologiques) avec sélection dynamique et mémoire longue, produit une valeur émergente non localisable dans un sous-module.

Sur ces deux terrains, la non-localisabilité causale est observable, le régime de stratification fonctionnelle est insuffisant, et les arbitrages économiques décrits plus haut sont déjà en jeu. Cela suffit à invalider la thèse contraire, selon laquelle le cadre actuel saisit comme objet ce qu'il a vocation à régir.

11. Le socle d'orchestration auditable comme primitive opératoire

La proposition d'un *socle d'orchestration auditable*, défini comme condition de gouvernabilité, exige d'être formulée à un niveau de précision qui la distingue strictement des cadres existants (NIST AI RMF, ENISA AI Threat Landscape, ISO/IEC 42001). À défaut, elle ressemble à du bon sens rebaptisé, ce qu'elle ne doit pas être. La présente section trace le seuil discriminant.

Le socle est défini par une primitive unique : la **falsifiabilité comportementale**.

Énoncé. Soit T une trajectoire d'exécution d'une chaîne agentique persistante, produisant un comportement observable β. Le socle exige qu'il existe une partition causale {E_k} de T telle que :

- (CN, condition nécessaire) chaque transition E_k est associée à un rôle régulatoire identifié et conservé à l'exécution ;
- (CS, condition suffisante) la composition des transitions {E_k} permet de reconstruire β avec conservation des dépendances causales pertinentes, et chaque transition est inspectable à l'exécution sans interruption du système.

Dit autrement, le socle exige qu'un comportement produit par une chaîne agentique puisse être reconstruit comme une suite intelligible de transformations attribuables. Non pas seulement après coup, par interprétation humaine, mais pendant l'exécution elle-même. Chaque étape importante doit laisser une trace causalement exploitable : quel composant a transformé quelle information, sous quelle règle, avec quelle responsabilité et dans quel contexte d'exécution.

L'exigence est proche de celle d'un enregistreur de vol aéronautique, mais appliquée à la décision logicielle : lorsqu'un comportement produit un effet clinique, juridique ou prudentiel, il doit être possible de remonter la trajectoire effective qui l'a généré sans dépendre d'une reconstruction narrative ex post.

Une chaîne peut donc être techniquement performante, conforme à ses processus qualité et même auditée réglementairement, tout en restant non falsifiable comportementalement si, au moment où elle agit, les transformations critiques de contexte ne restent ni typées, ni attribuables, ni reconstructibles.

- *Critère d'échec observable.* Le socle est en défaut s'il existe au moins un comportement β observé pour lequel aucune partition {E_k} satisfaisant (CN) et (CS) ne peut être reconstruite à l'exécution. Doctronic en est l'illustration directe : le SOAP note manipulé, lu par le clinicien, ne permet pas de remonter à la transition qui a transformé le contenu utilisateur en contexte clinique régulé. La généalogie a été effacée, non par défaut technique, mais par absence de typage régulatoire des transitions.
- *Différence opérationnelle avec l'audit classique.* Un audit ISO/IEC 42001 vérifie qu'une organisation respecte des processus de gouvernance ; il opère sur des artefacts statiques (documents, configurations, logs ex post). Un socle vérifie qu'une trajectoire d'exécution préserve l'imputabilité ; il opère sur la dynamique en cours. Un système peut passer un audit ISO/IEC 42001 et échouer le socle. C'est précisément ce que démontre Doctronic : une organisation peut être conforme aux standards de gouvernance et déployer une chaîne dont les comportements émergents ne sont pas attribuables à l'exécution.

Le socle n'est pas la solution au problème. Il en est la condition nécessaire de solubilité : ce sans quoi aucune doctrine régulatoire fondée sur l'imputabilité ne peut s'appliquer à une chaîne agentique persistante. Cette précision protège la primitive contre la confusion entre falsifiabilité comportementale et conformité organisationnelle, qui sont des objets distincts.

12. Articulation avec les travaux antérieurs

La présente analyse prolonge trois lignes : la critique du paradigme LLM-centré au profit d'une architecture composite, la formalisation des Clinically-Informed Neural Networks comme classe de modèles internalisant des contraintes par construction, et le cadre RAISE comme doctrine de responsabilité architecturale en environnement régulé.

RAISE supposait que la responsabilité ne peut être assignée que là où elle est techniquement portable. La présente thèse complète cette supposition par sa contre-partie : en environnement agentique, la responsabilité ne peut être assignée que là où la

trajectoire est falsifiable. Le socle est l'extension de RAISE à un objet, la chaîne d'orchestration, que RAISE ne thématisait pas explicitement.

13. Conditions de validité de la thèse

Trois conditions, et seulement trois, sous lesquelles la thèse défendue ici cesserait de tenir.

- Premièrement, si une refonte régulatoire majeure (extension explicite du SaMD à la trajectoire d'exécution, ou formalisation doctrinale de la composition dynamique sous AI Act) intervenait dans un horizon court, la prédiction de domination du régime trois serait invalidée. Cette refonte n'est étayée par aucun signal régulatoire actuel ; elle reste théoriquement possible.
- Deuxièmement, si une normalisation technique publique (par exemple un standard CEN-CENELEC ou ISO sur l'auditabilité d'orchestration adopté à l'échelle européenne ou internationale, opposable juridiquement par voie d'harmonisation) intervenait avant la stabilisation jurisprudentielle, elle rendrait l'auto-régulation industrielle secondaire et non plus structurante. Cette éventualité est compatible avec les dynamiques en cours autour des standards d'auditabilité IA, mais le calendrier prévisible la situe en aval, pas en amont, des décisions juridictionnelles attendues.
- Troisièmement, si la non-localisabilité causale s'avérait empiriquement contournable par des mécanismes d'audit dynamique généralisables (par exemple un standard de typage régulatoire des transitions adopté par l'industrie), le socle deviendrait observable comme propriété par défaut, et la divergence cumulative cesserait de croître. Cette éventualité est précisément ce que le présent texte cherche à provoquer en formalisant le socle comme primitive.

14. Conclusion : l'instabilité structurelle

Le SaMD reste l'ancrage juridique central des dispositifs d'IA en santé. L'AI Act, par ses Articles 14, 15, 25, 56 et 95, prolonge cet ancrage par capture indirecte solide. Mais ni l'un ni l'autre ne formalise la trajectoire génératrice comme objet régulatoire autonome, c'est-à-dire la propriété qui caractérise la classe de systèmes désormais déployée.

Cette inadéquation n'est pas un état. C'est une trajectoire :

Les systèmes sont gouvernés là où ils ne décident pas, et décident là où ils sont peu gouvernés. L'écart se creuse à chaque cycle d'innovation. Ce n'est pas une divergence : c'est une instabilité structurelle.

Tant que l'instabilité demeure, la régulation effective ne sera pas produite par la doctrine. Elle sera produite par le triangle assureurs-tribunaux-industriels : les trois forces qui n'ont jamais fait que rattraper la doctrine quand celle-ci a manqué le réel. Trois prédictions, structurelles plutôt que datées, en découlent.

Le régime d'auto-régulation industrielle deviendra dominant sur l'orchestration agentique en santé avant toute évolution doctrinale formelle, codifié par les Codes of Practice de l'AI Act et par les terms of service des plateformes. La doctrine régulatoire publique le rejoindra par voie d'implémentation, non par voie législative.

La jurisprudence sur la PLD 2024/2853 produira, dans les années qui viennent, quelques décisions structurantes sur l'imputabilité des comportements émergents. Ces décisions définiront la responsabilité civile effective des déployeurs avant que l'AI Act ne soit révisé pour en codifier les principes.

Le SaMD survivra, mais sa centralité doctrinale sera relativisée. Le paysage régulatoire à venir est celui d'une coexistence : SaMD sur les dispositifs, AI Act sur les modèles, codes ex post sur les chaînes. Aucun de ces régimes ne saisira la totalité ; leur articulation sera le travail régulatoire de la décennie qui s'ouvre.

L'orchestration agentique n'est pas, en avril 2026, un objet régulatoire formel. Elle est déjà le principal vecteur de comportement dans les architectures qu'elle structure, et elle sera régulée (partiellement, indirectement, et hors doctrine) avant que la doctrine n'ait reconnu qu'elle l'était. C'est cela, l'instabilité structurelle.

La doctrine publique ne régulera pas l'agentique en premier. Elle codifiera, plus tard, les compromis imposés par ceux qui auront dû assurer, contractualiser et juger ses accidents.