Université Paris Cité

Diplôme Universitaire et Formation Qualifiante :

Intelligence artificielle (IA) appliquée à la santé,

Année Universitaire 2025

Mémoire

Le contrôle humain de l'IA : Perte de chance, ou régulation positive efficace ? L'Europe à l'épreuve de son approche réglementaire.

Présenté par :

Sophie Bouju, Directrice Médicale.
Manon Cattenoz, Rédactrice spécialisée - Droit de la santé, Lamy Liaisons.
Pierre-Frederic Degon, Abbott France.
Stephanie Kervestin, Déléguée générale alliance pour la recherche & l'innovation des industries de santé ariis – Paris.
Romain Leger, Médecin.
Sandrine Kueny, Médecin, Directrice de la délégation territoriale de la Marne ARS Grand-Est.

Résumé :

Ce mémoire explore la mise en œuvre du contrôle humain dans les systèmes d'intelligence artificielle (IA) appliqués à la santé, à la lumière du Règlement (UE) 2024/1689 (RIA), premier cadre juridique mondial en la matière.
L'article 14 du RIA impose aux systèmes d'IA à haut risque une supervision humaine afin de garantir leur utilisation éthique et sécurisée, notamment dans les domaines où les décisions algorithmiques peuvent entraîner des conséquences irréversibles sur la vie des patients. Cette exigence soulève des enjeux techniques, organisationnels et éthiques complexes, tant pour les développeurs que pour les professionnels de santé.

Le mémoire analyse les modalités concrètes de ce contrôle humain (*a priori* et *a posteriori*), les efforts français pour adapter le cadre européen, ainsi que les défis liés à l'explicabilité des systèmes et aux biais cognitifs des superviseurs. Une comparaison internationale met en lumière les divergences entre l'approche européenne (préventive et fondée sur les droits fondamentaux), américaine (plus permissive et centrée sur l'innovation) et chinoise (centralisée et orientée vers la stabilité sociale). Ces différences illustrent les tensions entre souveraineté technologique, compétitivité économique et protection des droits.

Enfin, l'étude souligne les obstacles à l'harmonisation réglementaire, notamment l'absence de standards techniques finalisés et le risque de disparités d'interprétation entre États membres. Pourtant, dans un contexte de compétition mondiale, le contrôle humain pourrait devenir un levier stratégique pour positionner l'Europe comme référence éthique en matière d'IA. Ce mémoire appelle à une mise en œuvre rigoureuse mais pragmatique de la supervision humaine, conciliant innovation technologique et protection des patients, et insiste sur la nécessité de former les acteurs de terrain pour rendre ce contrôle effectif et légitime.

I) Introduction

L'essor rapide de l'intelligence artificielle (IA) transforme profondément le domaine de la santé, en promettant des diagnostics plus rapides, des traitements personnalisés et une optimisation des ressources médicales. Toutefois, ces progrès soulèvent des défis éthiques majeurs, notamment lorsque des systèmes automatisés influencent des décisions vitales, affectant directement l'intégrité, la dignité et la vie des patients. Pour y faire face, l'Union européenne (UE) a mis en place le Règlement (UE) 2024/1689 (RIA), premier cadre réglementaire mondial. Adopté en mars 2024 et publié le 12 juillet, il prévoit un cadre unifié pour le développement et l'usage de l'IA dans l'UE, tous secteurs confondus. La question des droits fondamentaux des citoyens ainsi que la gestion des risques sont au cœur de ce Règlement. Ce n'est donc pas un texte sanitaire, mais il emporte de nombreuses conséquences dans le domaine de la santé. Le contrôle humain, prévu à l'article 14 du RIA, constitue l'une des exigences essentielles auxquelles doivent répondre les systèmes d'IA à haut risque. Ce contrôle humain dans le champ de la santé est une spécificité qu'il convient de comprendre dans son périmètre et moyens à mettre en œuvre (II), alors que d'autres réglementations dans le monde de même nature diffèrent dans leur approche, posant la question de l'applicabilité d'un tel contrôle à l'européenne, pour des produits de santé qui cherchent à être pensés et conçus pour le monde entier (III), appelant à des recommandations préconisées pour atteindre l'objectif initial fixé par le législateur européen (IV).

Ce mémoire propose d'analyser les règlementations des systèmes d'IA et plus spécifiquement la mise en œuvre du contrôle humain dans le champ spécifique de la santé, et dans un moment historique et en constante évolution de compétition mondiale pour la définition des standards de réglementation et par conséquent pour le leadership l'IA. Finalement, l'analyse des défis pratiques liés à la mise en œuvre de ce principe de garantie humaine permettra d'évaluer dans quelle mesure le contrôle humain peut réellement protéger les patients sans freiner l'innovation technologique indispensable aux progrès médicaux.

II) Le principe d'un contrôle humain de l'IA appliqué à la santé

A. L'article 14 du RIA

a. *Principe*

La notion de contrôle humain, ou de supervision humaine faisait déjà partie des travaux préparatoires du RIA, la Commission européenne ayant déjà abordé cette question depuis plusieurs années à travers son Groupe d'Expert de Hauts niveau sur l'IA. Dès 2019 ce groupe a élaboré **sept principes éthiques non contraignants pour l'IA**, destinés à contribuer à faire en sorte que l'IA soit digne de confiance et saine sur le plan éthique. Parmi ces sept principes, figure le **contrôle humain** comme le mentionne le **considérant 2**7 du RIA. L'enjeu de cette notion réside ainsi dans la mise à disposition d'outils d'IA développés « *au service des personnes, qui respecte la dignité humaine et l'autonomie de l'individu, et qui fonctionne de manière à pouvoir être contrôlé et supervisé par des êtres humains ».*

L'article 14 définit ainsi le **principe du contrôle humain et ses modalités**, de façon générale et transversale, sans s'appliquer spécifiquement au champ de la santé. D'une part, il concerne les **systèmes d'IA à haut risque** tant dans leur conception que leur développement « *de manière à ce que les personnes physiques puissent superviser leur fonctionnement et veiller à ce qu'ils soient utilisés comme prévu et à ce que leurs incidences soient prises en compte tout au long de leur cycle de vie* » (Considérant 73). Cette approche par le risque (1) doit permettre soit de prévenir ou de réduire les risques, lorsque cela est « **prévisible** », malgré l'application d'exigences complémentaires comme la documentation technique (traçabilité), la transparence, ou encore la robustesse etc… comme précisé notamment dans les articles 15 et 16. Sont concernés par cet article 14 les **fournisseurs**, dont la grande majorité sera constituée par des industriels de la santé (DM, DMN, éditeurs de logiciels), ainsi que les **déployeurs**, entendus ici comme étant notamment les établissements de santé, ou les professionnels de santé, voire les patients-utilisateurs.

b. *Moyens à mettre en œuvre : un contrôle a priori et/ou a posteriori*

Les moyens du contrôle humain doivent garantir que le système est « *soumis à des contraintes opérationnelles **intégrées** (*contrôle *a priori) qui ne peuvent pas être ignorées par le système lui-même, que le système répond aux ordres de l'opérateur humain et que les personnes physiques auxquelles le contrôle humain a été confié ont les compétences, la formation et l'autorité nécessaires pour s'acquitter de ce rôle* ». Il y a donc une approche laissée à la libre appréciation du fournisseur dans la conception de son produit : d'une part un contrôle intégré au système d'IA (SIA) « *lorsque cela est techniquement possible* », et **le cas échéant, des mesures complémentaires à destination du déployeur** (contrôle *a posteriori*).
Ce dernier se voit ainsi confier la responsabilité afin d'intervenir le cas échéant pour :

- Traiter des anomalies (ex. un traitement de l'image non-conforme)

- Identifier les biais de traitement (ex. logiciels d'aides à la décision clinique)

- Interpréter correctement les sorties du SIA

- Décider de ne pas utiliser le SIA, ou de modifier la sortie

- Intervenir pour arrêter le système, pour éviter des conséquences négatives sur un patient (ex. un laser pour une chirurgie)

c. *L'approche française*

Le Gouvernement français a fait de l'IA et de l'approche réglementaire une force, pour convaincre la communauté scientifique, de l'intérêt d'investir et de développer des SIA en Europe, sur la base d'une approche concertée et éthique. Dans le domaine de la santé, les administrations sanitaires sont toutes engagées sur la préparation d'un cadre adapté pour le développement des IA en santé. Le mois d'août 2025 devrait d'ailleurs être propice à l'annonce de l'autorité ou l'administration qui sera compétente pour la mise en œuvre du RIA en France pour la santé, avec a priori un rôle dévolu à l'ANSM ou à la CNIL. La HAS (2) travaille également depuis cinq ans sur un cadre d'évaluation des IA et prépare ses propres outils pour améliorer la qualité des soins et des accompagnements sociaux et médico-sociaux. Dans le cadre de la Feuille de route du numérique en santé, des premières actions sur l'IA ont été initiées, mais la Délégation au numérique en santé (DNS) doit annoncer d'ici l'été 2025 sa stratégie française IA en santé, dans laquelle les travaux éthiques menés par la DNS devraient tenir une bonne place, tout comme des propositions en faveur de la mise en œuvre du contrôle humain « à la française », qui pourrait être liée au remboursement des produits de santé. La notion de collège de garantie humaine fait partie des solutions proposées pour la mise en œuvre du contrôle a posteriori, tandis que l'AFNOR travaille par ailleurs sur la définition d'un cadre commun d'une norme, en miroir des travaux menés au niveau européen par le CEN-CENELEC.

B. Enjeux du contrôle humain

La mise en œuvre du contrôle humain pose plusieurs défis pour les acteurs concernés, tant la proposition de l'article 14 vise une approche transversale à tous les secteurs, et qui doit être appréhendée de façon particulière dans le champ de la santé, au regard des nombreuses exigences déjà existantes en matière de sécurité des produits de santé. D'une part, le défi technique posé par le contrôle humain imposera aux fournisseurs une montée en compétence majeure sur le plan technique, le contrôle *a priori* devant répondre à **une parfaite maîtrise dès la conception**, ce qui n'est pas sans interroger sur la nature de l'IA et sa conception *by design*. D'autre part, **le contrôle du respect de l'article 14 pour les IA à haut risque**, nécessite de la part de la Commission européenne la définition de normes harmonisées, dont les organismes notifiés chargés du marquage CE devront mettre en œuvre. Or, en l'état, les travaux de normalisation du CEN-CENELEC qui concernent 10 chantiers (3), n'ont toujours pas été finalisés et ce corpus jouera un rôle majeur dans l'exécution du RIA, en raison d'une présomption de conformité des normes harmonisées (Article 40) (4).

Enfin, le défi de mise en œuvre du contrôle humain résidera dans l'équilibre entre une mise en œuvre qui garantisse la protection des droits fondamentaux, la mise en place d'un cadre suffisamment prévisible et transparent pour favoriser l'accès à l'innovation pour les systèmes de santé et les patients. Cet équilibre à trouver sera dans la main des États membres qui seront en première ligne pour la mise en œuvre opérationnelle du RIA, selon un calendrier ambitieux sous 24 mois avec une entrée en vigueur du contrôle humain pour les IA à haut risque en août 2027. L'harmonisation de la « doctrine » du contrôle humain sera en effet un véritable défi et soulève plusieurs questions :

- Le contrôle *a priori*, qui sera opéré par les Organismes notifiés, sous la responsabilité des fournisseurs, en tant que partie intégrante d'un marquage CE, pourra-t-il être remis en cause par un contrôle *a posteriori* ?
- Une exigence locale pourra-t-elle être ajoutée, notamment en lien avec une prise en charge pour les assurés ?

III) Approches internationales de la supervision humaine de l'IA

La supervision humaine de l'intelligence artificielle est devenue un principe central dans l'élaboration des cadres de gouvernance à travers le monde, bien que les modalités d'application varient selon les contextes nationaux. À l'heure de la mondialisation, la régulation européenne doit être évaluée en comparaison avec les mesures prises par les autres pays, notamment les grandes puissances telles que les États-Unis et la Chine. En outre, le début de l'année 2025 a marqué un tournant historique, avec une actualité particulièrement riche en développements concernant les régulations de l'IA.

A. Les États-Unis, entre régulation et innovation

La **régulation générale de l'IA aux États-Unis** est actuellement **au cœur des décisions et des déclarations du nouveau Gouvernement américain** dessinant pas à pas au gré de l'actualité internationale de nouvelles orientations. Dès sa prise de fonction, le président Donald Trump a fait de l'IA un levier majeur de la guerre économique qu'il mène sur le plan international (5). En janvier 2025, pour promouvoir une approche plus permissive et axée sur l'innovation, la nouvelle administration a révoqué les exigences et obligations strictes imposées aux développeurs d'IA, telles que la réalisation de tests de sécurité et le partage des résultats avec le gouvernement avant toute mise sur le marché (6). À ce même moment, la Chine a lancé le LLM **DEEPSEEK-MoE**, beaucoup moins coûteux et de niveau comparable à d'autres LLM, renforçant ses ambitions de devenir leader mondial de l'IA en 2030 (7). Cette annonce a pour conséquence de renforcer la guerre commerciale entre la Chine et les États-Unis pour être leader mondial en IA. Les États-Unis se positionnent comme garants du développement d'IA protégeant les droits des citoyens. Ainsi, le Gouvernement américain, par la voix du vice-président J.D. Vance lors de IA Summit à Paris (11 février 2025), a affirmé son souhait de maintenir leur position de leader mondial en favorisant l'innovation et en mettant l'accent sur l'établissement de règles internationales tout en prenant le contre-pied de la réglementation européenne jugée trop contraignante (8). La politique américaine pour le développement de l'IA s'appuiera aussi sur un plan d'action qui devrait être publié en juillet 2025 et qui s'alimente d'une consultation publique (9). Les géants de l'IA US comme OpenAI ou Anthropic ont ainsi publié leurs contributions, autour du **principe de « liberté d'innover »** pour **accélérer le progrès tout en évitant des réglementations trop restrictives** qui pourraient avantager des concurrents étrangers, comme notamment la Chine (10,11). Ils plaident aussi pour des règles communes à tous, faisant référence à un cadre fédéral. En effet, en révoquant le décret présidentiel sur l'IA, Donald Trump a supprimé le **principal cadre fédéral** encadrant l'intelligence artificielle aux États-Unis, laissant place à une **régulation fragmentée au niveau des États**. Dans ce contexte de décentralisation, certains États, tels que la Californie, jouent le rôle de *leader*. Les lois récemment adoptées par la Californie convergent sur de nombreux points avec le RIA européen (12). Ainsi, pour l'utilisation de l'IA dans les soins de santé, de nouvelles lois garantissent que les médecins supervisent l'utilisation des outils d'IA dans certains contextes (13,14).

En résumé, la réglementation générale de l'IA aux États-Unis reste **conflictuelle et mouvante, avec une volonté de dérégulation pour favoriser l'innovation tout en évitant une fragmentation législative, et ce dans un contexte de compétition avec la Chine** (15). Toutefois sans obligations claires en matière d'éthique et de sécurité, les solutions d'IA américaines pourraient être perçues comme moins fiables ou plus risquées par les entreprises et consommateurs européens.

En matière de santé, c'est la Food and Drug Administration (FDA) qui supervise l'utilisation de l'IA dans les dispositifs médicaux, quand d'autres agences comme le National Institute of Standards and Technology (NIST) définissent les standards du cadre réglementaire dans le

domaine de l'IA. Les fournisseurs sont encouragés à prévoir des "boucles humaines" (*human-in-the-loop* 16) lors du déploiement de systèmes IA à impact élevé, tout en laissant une marge d'interprétation aux entreprises quant au niveau de supervision requis selon les contextes d'usage.

AdvaMed (17), l'association Medtech américaine, a publié le **22 Avril 2025 sa « Feuille de route pour une politique de l'IA »** destinée au Congrès et aux agences fédérales compétentes en matière de medtech afin de les **guider dans l'établissement de recommandations politiques** visant à garantir la protection de la vie privée et des données des patients tout en permettant à l'innovation de progresser efficacement. Ils prônent le **renforcement du rôle de la FDA en tant que principal régulateur** des technologies de santé basées sur l'IA.

Leurs principales recommandations sur la supervision humaine des systèmes d'IA dans les dispositifs médicaux portent sur : (i) une **surveillance réglementaire par la FDA,** qui évalue la **sécurité et l'efficacité** des algorithmes d'IA **avant et après** leur mise sur le marché, (ii) la nécessité de mettre à jour la loi HIPAA (Health Insurance Portability and Accountability Act) pour faciliter le partage sécurisé des grands ensembles de données nécessaires au développement de l'IA, (iii) une **approche harmonisée au niveau mondial** pour la réglementation des dispositifs d'IA.

Enfin, AdvaMed souligne le besoin d'une réflexion sur la création des **"*third-party quality assurance labs*"**. En effet, la FDA mais aussi d'autres parties prenantes comme the Coalition for Health AI (CHAI) (18) explorent la création d'un réseau de **"*health AI assurance labs*"** pour soutenir la gestion du cycle de vie de l'IA dans les dispositifs médicaux**, avec un accent sur la surveillance continue des performances.** Ils incitent à réfléchir en profondeur à l'utilisation des tiers pour gérer le cycle de vie de l'IA, en soulevant des préoccupations sur leur pertinence, la confidentialité, la propriété intellectuelle et la sécurité. Ils craignent aussi que leur mise en place ne crée des doublons avec la surveillance de la FDA, augmente les coûts et ralentisse la mise sur le marché. Pour l'instant, plutôt que de compter sur ces tiers, AdvaMed recommande que la FDA participe à l'élaboration et à la reconnaissance rapide de normes communes et crédibles pour assurer la qualité.

Face aux dernières annonces de réduction des effectifs de la FDA par l'administration Trump, on peut déjà s'interroger sur ses capacités pour surveiller tous les dispositifs IA.

B. Le Canada : un modèle à l'européenne

Depuis le début des années 2000, le Canada a posé les premières bases de sa régulation numérique avec la **Loi sur la protection des renseignements personnels et les documents électroniques (PIPEDA)** (19), axée sur la protection des données dans le secteur privé. Toutefois, cette loi n'aborde pas spécifiquement les enjeux liés à l'IA.

Face à l'accélération technologique, le Gouvernement canadien a proposé en 2022 le **projet de loi C-27** (20), comprenant notamment la **Loi sur l'intelligence artificielle et les données (LIAD)**. Inspirée d'une approche fondée sur les risques, la LIAD impose aux développeurs de systèmes d'IA à impact élevé (notamment en santé) des obligations de transparence, de réduction des risques de biais et de préservation de la sécurité humaine (21). Un **Commissaire à l'IA et aux données** devait être chargé de la surveillance et de l'application.

Cependant, en janvier 2025, le projet de loi a été suspendu en raison de la prorogation du Parlement, laissant le pays sans cadre fédéral spécifique pour l'IA. En parallèle, la **Stratégie**

pancanadienne en IA (22), lancée en 2017 et renouvelée en 2022, promeut le développement responsable de l'IA, mais sans pouvoir réglementaire contraignant.

À ce jour, la **supervision humaine** des systèmes d'IA au Canada, notamment dans le domaine sensible de la santé, repose sur des lignes directrices non obligatoires et sur les politiques internes des institutions, renforçant la nécessité d'une future législation claire et opérationnelle.

C. La Chine : une régulation centralisée, proactive et itérative

La Chine a été parmi les premiers pays à réglementer l'IA à travers la mise en place d'un cadre réglementaire complet pour encadrer le développement et l'utilisation de l'IA, avec pour objectif de devenir un **leader mondial en 2030**. Alors que la loi chinoise sur la protection des données s'inspire largement des principes et concepts de l'UE, la Chine **prend sa propre direction concernant le droit de l'IA et avance rapidement**. Elle construit un ensemble de règles ciblant différents problèmes liés à l'IA **d'une manière très itérative, à l'opposé du RIA** (23). Sans détailler la chronologie, l'encadrement des différentes IA est issu de réglementations publiées chaque année depuis 2021. Dès cette date, le principe de la **Crédibilité contrôlable est mis en place** : les systèmes d'IA doivent être fiables et sous contrôle humain. Notablement, en 2023, pour les IA génératives, il est imposé des obligations strictes aux développeurs, notamment en matière de supervision humaine et de conformité aux "valeurs socialistes fondamentales".

La régulation chinoise aspire à protéger en premier lieu les utilisateurs notamment face aux violations de leurs données personnelles par les entreprises. Le **fournisseur** est l'acteur majeur qui doit se conformer aux lois chinoises en matière d'IA. La dernière réglementation leur impose de procéder à une évaluation de la sécurité avant que leurs services soient disponibles en ligne. Ils deviennent responsables de toute violation de propriété intellectuelle ou encore de toute fuite d'informations personnelles. Le **développeur**, quant à lui, doit se conformer à un niveau élevé de contrôle technique, très difficile à atteindre pour les entreprises. De plus, les sociétés doivent subir une évaluation préalable pour rendre leur système d'IA disponible et une inspection de sécurité sera imposée aux logiciels chinois (24).

Par ailleurs, pour la Chine, le développement de **l'IA en santé** est un enjeu particulièrement important car ce pays, qui compte 1,4 milliard d'habitants, ne compte que 12 millions de médecins, essentiellement concentrés dans les métropoles de l'Est du pays. L'IA apparaît indispensable pour désenclaver des territoires entiers, véritables déserts médicaux. Le pays se veut pionnier avec le tout **premier hôpital virtuel entièrement piloté par l'IA** (25).

La Chine a donc aussi un cadre réglementaire de l'IA incluant des briques de contrôle humain et son approche fondamentale en matière de gouvernance de l'IA reste axée sur l'atténuation des dommages aux individus et le maintien de la stabilité sociale et du contrôle de l'État, tout en visant un leadership mondial en matière d'IA et en influençant le débat sur sa réglementation. Cela est évident dans leurs efforts pour développer des normes internationales qui pourraient leur offrir un avantage concurrentiel.

D. Ailleurs dans le monde

Le Gouvernement britannique a récemment annoncé son intention d'introduire une législation sur l'IA, qui devrait être adoptée dans le courant de l'année.

La Corée du Sud a récemment adopté la Basic Act on the Development of Artificial Intelligence and the Establishment of Trust, qui entrera en vigueur en janvier 2026.

Taïwan, le Brésil et le Chili ont également présenté des projets de loi sur l'IA. Comme le RIA, ces lois prévoient des obligations différentes en fonction du niveau de risque présumé d'un système d'IA.

Enfin, le 11 Février 2025, les autorités de protection des données de l'Australie, de la Corée du Sud, de la France, de l'Irlande, et du Royaume-Uni ont signé une déclaration commune sur la mise en place de cadres de gouvernance des données fiables afin d'encourager le développement d'une intelligence artificielle innovante et protectrice de la vie privée (26).

Ces exemples internationaux de réglementations de l'IA illustrent la diversité des approches, du contrôle étatique centralisé au modèle collaboratif, mais tous reconnaissent le rôle irremplaçable de la vigilance humaine pour garantir une IA fiable et éthique. Par ailleurs, le moment actuel est historique avec un "affrontement" entre les réglementations européennes, chinoises et américaines pour **définir les standards de la réglementation de l'IA et ainsi devenir le leader en IA.**

IV) Défis de mise en œuvre et perspectives

L'article 14 du RIA consacre l'obligation d'un contrôle humain pour les systèmes d'IA à haut risque, dans une logique de transparence, de sécurité et de responsabilité. Toutefois, sa mise en œuvre soulève des difficultés concrètes et universelles, tant sur le plan technique qu'humain, et interroge la capacité des cadres juridiques à suivre le rythme de l'innovation technologique.

A. Défis communs à toutes les juridictions

Un premier obstacle majeur réside dans le manque de **"*explainability*"** — ou explicabilité — de nombreux systèmes d'IA, en particulier ceux fondés sur des réseaux de neurones profonds. Ces modèles, souvent perçus comme des "boîtes noires", produisent des résultats qui ne peuvent pas toujours être interprétés ou justifiés de manière intelligible, même par leurs concepteurs. Or, le contrôle humain selon le principe d'agentivité, pour être effectif, suppose *a minima* une capacité à comprendre les mécanismes de décision de l'IA supervisée. L'absence d'explicabilité rend donc difficile, voire illusoire, le contrôle humain en temps réel, notamment dans les environnements critiques.

En parallèle, la **qualité de la supervision ou du contrôle** dépend directement des compétences des personnes chargées de ce rôle. Au-delà des aspects techniques, il faut souligner les **biais cognitifs** qui peuvent affecter les superviseurs eux-mêmes, notamment les biais **d'automatisation** :

- L'erreur **d'omission**, où l'opérateur tend à ne pas intervenir face à une erreur de la machine, par excès de confiance ;
- L'erreur **de commission**, où il suit une recommandation erronée de l'IA malgré des signes évidents de dysfonctionnement.

  Ces biais illustrent combien le contrôle humain ne peut se résumer à une simple présence passive, mais requiert une **formation approfondie**, une posture critique et des protocoles clairs d'intervention. Faute de cela, l'humain risque de devenir un simple figurant dans une boucle de décision dominée par la machine.

Enfin, le **rythme d'exécution** de certains systèmes d'IA, notamment dans le domaine médical, rend difficile toute intervention humaine rapide et éclairée. Il devient alors essentiel de repenser les modalités de supervision ainsi que l'éthique de l'outil, peut-être en amont (conception) voire en aval (audit), plutôt qu'en cours d'usage.

B. Spécificités du modèle européen

Le cadre européen, avec le RIA, se distingue par une volonté forte de **protection des droits fondamentaux**. La supervision humaine y est conçue comme une garantie contre les risques sociétaux de l'IA (discrimination, surveillance, double usage). Mais cette exigence, si elle n'est pas accompagnée de **ressources humaines et financières suffisantes**, risque de créer un décalage entre les obligations réglementaires et les capacités réelles des acteurs à les appliquer.

De plus, le texte laisse une certaine latitude aux États membres pour définir les modalités concrètes de supervision, ce qui pourrait entraîner une **hétérogénéité d'interprétation** au sein de l'UE, affaiblissant la cohérence du dispositif.

À cela s'ajoute un **risque d'échec dans la normalisation européenne**, dû à un **retard dans l'exécution du calendrier d'adoption secondaire** (actes délégués, standards harmonisés), mais aussi à la **forte influence d'acteurs industriels internationaux**, créant de **potentiels conflits de rôle, d'intérêts ou de missions**, notamment entre innovation, autorégulation et éthique.

Toutefois, la récente **déclaration conjointe** de plusieurs autorités de protection des données (dont la CNIL) sur la gouvernance des données rappelle que d'autres régions du monde partagent des **valeurs communes de transparence et de protection de la vie privée** *(Privacy by design)*, ce qui pourrait faciliter à terme une convergence éthique au-delà du seul cadre européen.

C. Perspectives internationales

Dans un contexte où l'IA est globalisée, l'absence d'un socle commun international sur la supervision humaine risque d'entraver la coopération et d'engendrer des distorsions de concurrence. L'UE, en promouvant une approche fondée sur la transparence et les droits humains, pourrait néanmoins influencer les standards globaux, à condition de dialoguer avec les autres grandes puissances technologiques.

À l'avenir, des solutions hybrides pourraient émerger, combinant supervision humaine et **assistance algorithmique**. Des outils d'aide à la décision, de visualisation ou de détection d'anomalies pourraient renforcer les capacités humaines sans les évincer. Cette approche pourrait représenter un compromis réaliste entre exigence éthique et faisabilité technique.

V) Conclusion

L'article 14 du RIA consacre le contrôle humain comme une exigence incontournable pour les systèmes d'IA à haut risque, en particulier dans le domaine de la santé, où les décisions prises par des algorithmes peuvent avoir des conséquences irréversibles sur la vie humaine. Cette orientation souligne la **volonté européenne d'encadrer l'innovation technologique** par une approche fondée sur le **respect de la dignité humaine et de la protection des patients**.

La comparaison avec les approches nord-américaines et chinoises montre que si la nécessité d'un contrôle humain est reconnue, ses modalités, son intensité et **ses finalités varient fortement selon les contextes culturels, économiques et politiques**. L'Union européenne fait de la supervision un pilier normatif, au service des droits fondamentaux, les États-Unis l'envisagent de manière modulée, à travers des recommandations non contraignantes et la Chine l'utilise comme levier de contrôle politique, intégré à une vision centralisée de l'innovation. **Ces différences posent la question de l'interopérabilité réglementaire** et du rôle que pourra jouer l'Europe dans la construction d'un référentiel global. En santé, ces écarts peuvent se traduire par **des risques de disparités** dans la qualité des soins, dans la transparence des décisions cliniques et dans la capacité des patients à faire valoir leurs droits.

La mise en œuvre effective du contrôle humain en santé posera **des défis importants** : **explicabilité** des modèles cliniques, **formation** des professionnels de santé au jugement critique face aux outils d'IA, et **prévention des biais** d'automatisation. Il faudra également porter une attention particulière à la **coordination** ainsi qu'au contrôle et à la supervision des outils IA lorsque leur utilisation est partagée, interdépendante ou interfacée, notamment pour des questions de **responsabilité** (exemple des moniteurs cardiaques implantables).
La dynamique actuelle de coopération internationale, notamment autour de la gouvernance éthique des données, ouvre cependant une voie vers une **convergence des principes de transparence**, de responsabilité et de protection des individus et favorise le développement **d'initiatives innovantes** en termes de régulation positive (Ethik-IA).

Ainsi, loin d'entraver le développement de l'intelligence artificielle en santé, une supervision humaine forte et bien conçue pourrait au contraire en renforcer la légitimité, en plaçant **la technologie au service du patient et non l'inverse**.

Nous remercions chaleureusement Mme Hélène Marin et M. David Gruson d'Ethik-IA, ainsi que M. Arnaud Augris du SNITEM, pour leur disponibilité et la qualité de leurs réponses, qui ont grandement enrichi notre réflexion.

Bibliographie :

1. S'agissant des IA interdites, voire la publication de la CE : https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act?utm_source=substack&utm_medium=email
2. https://www.has-sante.fr/jcms/p_3599637/fr/l-ia-en-sante-un-enjeu-majeur-pour-la-has-et-l-ensemble-du-systeme-de-sante
3. 10 chantiers de normalisation : gestion du risque, qualité/gouvernance des données, jounralisation (logs), transparence, contrôle humain, performance, robustesse, cybersécurité, gestion de la qualité et évaluation de conformité.
4. « *Les systèmes d'IA à haut risque ou les modèles d'IA à usage général conformes à des normes harmonisées ou à des parties de normes harmonisées dont les références ont été publiées au Journal officiel de l'Union européenne conformément au règlement (UE) n 1025/2012 sont présumés conformes aux exigences visées à la section 2 du présent chapitre ou, le cas échéant, aux obligations énoncées au chapitre V, sections 2 et 3, du présent règlement, dans la mesure où ces exigences ou obligations sont couvertes par ces normes* ».
5. https://www.lemonde.fr/economie/article/2025/01/22/le-mandat-trump-commence-par-une-pluie-de-centaines-de-milliards-dollars-destines-l-intelligence-artificielle_6509500_3234.html
6. https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.
7. https://www.lemonde.fr/economie/article/2025/01/25/la-chine-talonne-les-americains-dans-la-course-a-l-intelligence-artificielle_6514703_3234.html
8. https://www.presidency.ucsb.edu/documents/remarks-the-vice-president-the-artificial-intelligence-action-summit-paris-france
9. https://www.whitehouse.gov/briefings-statements/2025/02/public-comment-invited-on-artificial-intelligence-action-plan/#:~:text=The%20AI%20Action%20Plan%20will,from%20hindering%20private%20sector%20innovation
10. https://openai.com/global-affairs/openai-proposals-for-the-us-ai-action-plan/
11. https://www.anthropic.com/news/anthropic-s-recommendations-ostp-u-s-ai-action-plan?utm_source=openai[sB3]
12. la-regulation-de-lintelligence-artificielle-aux-etats-unis (1).pdf https://shs.cairn.info/revue-action-publique-recherche-et-pratiques-2024-4-page-32?lang=fr
13. AB 3030 Bill Text - AB-3030 Health care services: artificial intelligence. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB3030
14. SB 1120 - Bill Text - SB-1120 Health care coverage: utilization review. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1120
15. Two Futures of AI Regulation under the Trump Administration by Claudio Novelli, Akriti Gaur, Luciano Floridi :: SSRN - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5198926
16. https://iapp.org/news/a/-human-in-the-loop-in-ai-risk-management-not-a-cure-all-approach
17. AI Policy Roadmap
18. A Nationwide Network of Health AI Assurance Laboratories | Health Policy | JAMA | JAMA Network
19. Gouvernement du Canada. *Loi sur la protection des renseignements personnels et les documents électroniques (LPRPDE/PIPEDA)*, S.C. 2000, c. 5.
https://laws-lois.justice.gc.ca/fra/lois/P-8.6/
20. Gouvernement du Canada. *Projet de loi C-27 : Loi de mise en œuvre de la Charte du numérique de 2022*. Première lecture le 16 juin 2022.
https://www.parl.ca/DocumentViewer/fr/44-1/projet-loi/C-27/premiere-lecture
21. Innovation, Sciences et Développement économique Canada. *Charte du numérique : pour un avenir numérique canadien, fondé sur la confiance*, 2020.
https://ised-isde.canada.ca/site/charte-numerique/fr

22. CIFAR (Institut canadien de recherches avancées). *Stratégie pancanadienne en matière d'intelligence artificielle*, 2017 et mise à jour 2022.
https://cifar.ca/fr/ia-strategie/
23. https://pernot-leplay.com/fr/reglementation-ia-chine-europe-us/
24. https://naaia.ai/chine-ia-reglementation/
25. La Chine inaugure son premier « hôpital » reposant entièrement sur l'IA – Artificiële Intelligentie / Inteligencia Artificial https://hlrnet.com/sites/ai/?p=5813
26. Joint statement on building trustworthy data governance frameworks to encourage development of innovative and privacy-protective AI
27 Parlement européen & Conseil de l'Union européenne. *Proposition de règlement établissant des règles harmonisées concernant l'intelligence artificielle (AI Act)*. COM(2021) 206 final, 21 avril 2021.
https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A52021PC0206
28 Union européenne. *Charte des droits fondamentaux de l'Union européenne*, 2000/C 364/01. https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:12012P/TXT
29 High-Level Expert Group on AI (AI HLEG). *Ethics Guidelines for Trustworthy AI*. Commission européenne, avril 2019.
https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
30 OCDE. *Principes de l'OCDE sur l'intelligence artificielle*. Mai 2019.
https://oecd.ai/en/dashboards/ai-principles
31 Carvalho, P., Dos Santos Soares, D., & Ferreira, A. "*Automation bias in decision support systems*." *International Journal of Medical Informatics*, vol. 78, 2009, pp. 115–126.
https://doi.org/10.1016/j.ijmedinf.2008.06.003
32 CNIL, ICO, OAIC, PIPC, DPC. *Déclaration conjointe sur la construction de cadres de gouvernance des données fiables pour encourager le développement d'une IA innovante et protectrice de la vie privée*. 11 février 2025.