

Questionnements (éthiques) sur l'intelligence artificielle en santé

DOSSIER THÉMATIQUE

L'intelligence artificielle (IA) est présente dans notre vie quotidienne sous de multiples formes : internet et vente en ligne, robots* professionnels ou domestiques, assistants vocaux (Alexa, Siri...), aide à la circulation des véhicules, traduction instantanée, suivi de notre santé (sommeil, activité physique), agents conversationnels génératifs (ChatGPT, Le Chat - Mistral IA...).

Nos réactions sont variées entre utilisation enthousiaste et résistance craintive.

Dans le domaine de la santé, **les conséquences sociétales induites par les systèmes d'intelligence artificielle** ont conduit le Comité consultatif national d'éthique (CCNE) à inscrire ce thème dans les « États généraux de la bioéthique » (EGBE) organisés avec les Espaces de réflexion éthique régionaux (ERER) pour recueillir l'expression de la société avant la prochaine révision de la loi de bioéthique.

Le thème du « numérique* en santé », dont les usages de systèmes d'intelligence artificielle en constituent la forme la plus actuelle, a déjà été abordé dans ce même cadre.

L'ERE Occitanie vous propose donc un atelier-débat sur le thème de « Intelligence artificielle en santé » dont l'objectif est double :

- **sensibiliser les citoyens aux questions éthiques soulevées par l'IA en santé ;**
- **en débattre avec vous pour recueillir vos avis et les rapporter au niveau national.**

Nous vous adressons ce document pour préparer vos réflexions et nos discussions.

N.B. : les applications de l'intelligence artificielle dans les neurotechnologies, que ce soit dans le domaine médical ou hors domaine médical, ne sont pas abordées ici mais dans les ateliers « Neurotechnologies ».

1- L'intelligence artificielle¹

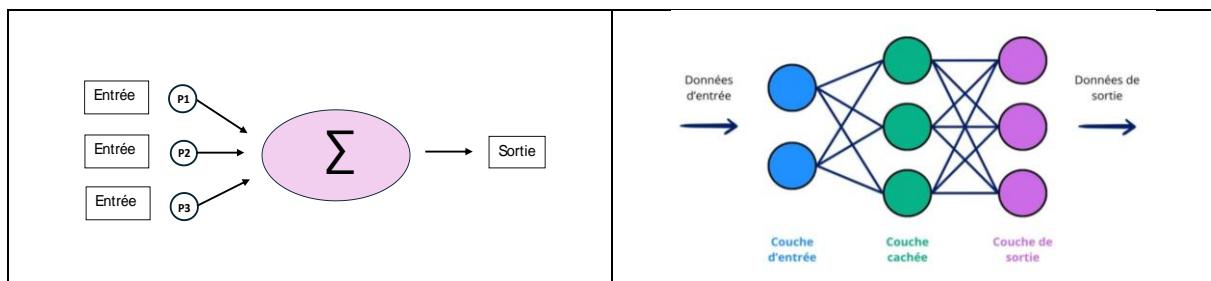
L'expression « intelligence artificielle » est utilisée dans plusieurs contextes avec des significations différentes. Il s'agit avant tout d'un **domaine de recherche** fondé en 1955-1956 dont l'objectif est de concevoir des programmes informatiques qui simulent l'intelligence ou les capacités cognitives humaines à certains degrés. Les termes désignant les fonctions réalisées sont ainsi souvent inspirés de ceux utilisés pour les capacités humaines : la perception de l'environnement et sa modélisation, la représentation de connaissances extraites de données, l'apprentissage automatique de modèles représentant des données, le traitement du langage naturel, la résolution de problèmes et le raisonnement, la planification d'actions ou de mouvements (en robotique), ou l'interaction avec d'autres systèmes ou avec des êtres humains. Ces programmes peuvent traiter des images, du texte, du son ou tout autre type de signal, ou en produire, et agir sur leur environnement. Les algorithmes issus de ce domaine de recherche sont mis en œuvre dans **des systèmes d'intelligence artificielle** utilisés en

¹ D'après : Raja Chatila, Claude Kirchner, Sylvain Petitjean, Catherine Tessier - Chapitre 5 « L'intelligence artificielle : quels enjeux éthiques pour la recherche ? » In : *L'éthique de la recherche*, Quae, à paraître 2026.

vue de telle ou telle application comme la reconnaissance faciale, le diagnostic médical ou la production de textes.

On constate aussi que cette expression est également utilisée, depuis quelques années, comme synonyme de « système à base d'apprentissage machine » ou encore de « système d'intelligence artificielle générative », voire pour qualifier tous les systèmes numériques.

Les méthodes d'apprentissage automatique (ou *apprentissage machine*, *apprentissage statistique*)* construisent des modèles statistiques en corrélant des éléments extraits de très grandes quantités de données. Outre des avancées méthodologiques et l'accroissement de la puissance de calcul, c'est justement la disponibilité de masses de données, indispensable pour élaborer une statistique, qui a favorisé leur émergence. Très souvent, les méthodes d'apprentissage automatique utilisent des *réseaux de neurones formels* (voir figure) de très grande taille (*apprentissage profond*) parce que ceux-ci peuvent approximer les fonctions mathématiques implicites complexes qui représentent la distribution statistique des données.



Neurone formel et réseau de neurones formels

D'après *Tout comprendre (ou presque) sur l'intelligence artificielle* (O. Cappé, C. Marc - CNRS éditions, 2025)

Le processus d'entraînement ou d'apprentissage est fondé sur une itération des valeurs des poids synaptiques du réseau, l'objectif étant de classer les éléments contenus dans les données en fonction de leurs similarités. Le résultat de ce processus est un **modèle statistique des données** qui est ensuite exploité pour identifier à quelles classes de similarité appartiennent de nouvelles données (on parle de **modèles prédictifs**) ou pour produire, à partir d'une requête, des réponses probables selon la corrélation de celle-ci avec les distributions statistiques du modèle (**modèles génératifs**). Il est important de noter qu'il n'y a pas de raisonnement causal dans l'ensemble de ce processus.

Les systèmes génératifs fondés sur des grands modèles de langage (**Large Language Models – LLM**) sont largement connus depuis fin 2022 avec la mise à disposition du grand public de ChatGPT.

2- Usages des systèmes d'intelligence artificielle en santé

La facilité d'utilisation et la disponibilité des systèmes d'intelligence artificielle, la possibilité de « dialogue » en langage naturel et la rapidité des réponses concourent à multiplier leurs usages dans le domaine de la santé :

- pour les professionnels de santé : assistance au diagnostic, lecture d'images, surveillance à distance, génétique, prothèses et exosquelettes, assistance au

secrétariat et gestion, rédaction de comptes rendus, recherche (nouveaux traitements), etc.

- pour les patients : suivi des activités (activités physiques, sommeil...), information sur les maladies et les traitements, téléconsultations et amélioration de l'accès aux « soins » dans les déserts médicaux, etc.

En revanche, les risques liés à l'usage de ces systèmes ne sont pas toujours bien identifiés par les utilisateurs, comme la diffusion incontrôlée de **données de santé*** vers des tiers, souvent de grandes entreprises étrangères.

Par exemple, une enquête réalisée en 2025 au CHU de Nancy a révélé qu'un professionnel sur cinq utilisait des systèmes d'intelligence artificielle, 23% tous les jours et 81% une fois par semaine – soit ChatGPT (60%), soit Le Chat Mistral AI (10%) – souvent sans encadrement, formation, ni sécurisation, pourtant prévus par la loi de bioéthique de 2021.

En outre les systèmes d'intelligence artificielle fondés sur l'apprentissage machine comportent des défauts intrinsèques (données d'entraînement insuffisantes ou inadaptées résultats erronés, faux positifs ou faux négatifs...). Ils peuvent être utilisés pour produire facilement de fausses informations (textes, schémas, images, vidéos), y compris dans les publications scientifiques.

3- La protection des données à caractère personnel en France et en Europe

La protection des données à caractère personnel peut être faible dans les applications « grand public » proposées par des entreprises extra-européennes, mais plus robuste dans les applications médicales grâce à la réglementation.

- **La loi relative à l'informatique, aux fichiers et aux libertés de 1978** a placé la France parmi les premiers pays dotés d'une protection des données numériques. Elle a posé le principe que *l'informatique doit être au service de chaque citoyen* et créé une autorité de contrôle pour le respect de la loi, la Commission nationale de l'informatique et des libertés (CNIL). Elle a défini les droits des personnes vis-à-vis de leurs données à caractère personnel et a interdit que des décisions entraînant des conséquences pour les personnes (décision judiciaire, octroi de crédit bancaire, etc.) soient prises sur le seul fondement d'un traitement automatisé de données à caractère personnel.
- **D'autres législations** interdisent les discriminations (droits de l'homme), protègent la vie privée (code civil), interdisent l'utilisation des tests génétiques par les compagnies d'assurances (loi de mars 2002).
- Au niveau européen, le **Règlement général sur la protection des données (RGPD)**, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel est appliqué depuis 2018 dans l'Union européenne mais aussi en dehors du territoire de l'Union européenne pour les

citoyens de l'Union européenne ce qui les protège face aux pays qui n'ont pas institué le même niveau de protection comme les États-Unis

- **Le Règlement Européen sur l'intelligence artificielle (IA Act, 2024)** définit quatre niveaux de risque pour les systèmes d'intelligence artificielle : risque inacceptable (systèmes interdits), haut risque, risque limité, risque minimal. Les dispositifs médicaux qui incluent des techniques d'intelligence artificielle sont classés « à haut risque », ce qui impliquera un certain nombre d'obligations (contrôles, transparence, traçabilité, robustesse).

Les données de santé bénéficient d'une protection particulière dans toutes ces législations parce qu'elles se rattachent à la vie privée et sont considérées comme des données à caractère personnelles sensibles*.

4- Repères éthiques

Les usages de systèmes d'intelligence artificielle en santé interrogent les principes de bioéthique et d'éthique du numérique (autonomie de la personne, bienfaisance et non-malfaisance, justice – ou équité auxquels L Floridi a proposé de rajouter *l'explicabilité*).

- Autonomie : chaque personne doit conserver son pouvoir de décision. Les professionnels de santé doivent donner aux patients « une information claire, loyale et adaptée » sur leur utilisation de systèmes d'intelligence artificielle d'aide à la décision avant de recueillir leur consentement.
- Bienfaisance – non malfaisance : le premier objectif de l'usage de systèmes d'intelligence en santé devrait être l'amélioration du service médical rendu au patient. Ces systèmes ne devraient pas détériorer la relation humaine dans le soin (par exemple, l'envoi automatique de résultats d'examens sans accompagnement humain).
- Justice – équité : le développement des usages de systèmes d'intelligence artificielle en santé devrait éviter toute discrimination et toute exploitation des vulnérabilités aux niveaux individuel et collectif. Cependant, certains patients peuvent avoir des difficultés d'accès aux systèmes numériques ou des difficultés de compréhension des enjeux.
- Explicabilité : le professionnel de santé doit être capable d'expliquer au patient les résultats issus d'un système d'intelligence artificielle. Cependant, les cheminements menant aux résultats produits par certains systèmes (en particulier fondés sur des techniques d'apprentissage machine) peuvent être difficiles à expliquer, en particulier au public non informé.

D'autres questions d'éthique plus générales sont liées à l'exploitation des personnes qui annotent les données pour l'apprentissage machine, à l'impact environnemental de ces systèmes et à la dépendance envers de grandes sociétés technologiques étrangères.

En outre, l'évolution des métiers du soin et des formations, ainsi que la transformation des savoirs et savoir-faire des professionnels doivent être considérées.

Glossaire complémentaire

Numérique² : ensemble des applications des technologies de l'information. Le numérique a pour périmètre scientifique et technique l'informatique, une partie des mathématiques, de la physique et des sciences cognitives ; il inclut la robotique et l'intelligence artificielle.

Algorithm³ : description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée.

Apprentissage⁴

- **supervisé** : pendant l'apprentissage, la réponse correcte est fournie au système pour organiser les données en classes prédéfinies. L'apprentissage supervisé part d'un flux d'exemples étiquetés, c'est-à-dire associant à chacun une catégorie. Ces exemples servent à « entraîner » le système d'apprentissage. Ensuite, une fonction est construite pour retrouver l'étiquette de chaque exemple. Il existe une multitude de techniques d'apprentissage supervisé : construction d'arbres de décisions, programmation logique inductive, machines à noyaux, apprentissage profond, etc.
- **non-supervisé** : l'algorithme cherche des régularités dans les données afin de construire différentes classes, qui ne sont pas établies préalablement.
- **par renforcement** : l'algorithme recherche l'action la plus prometteuse par optimisation d'une fonction dite « objectif » ou récompense, fournie préalablement.

Robot⁵ : machine, commandée par un ordinateur, qui se meut dans l'espace physique et y effectue des actions.

Donnée de santé⁶ : les données de santé sont des données à caractère personnel relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique [...] qui révèlent des informations sur l'état de santé de cette personne.

Données à caractère personnel sensibles⁷ : les données sensibles forment une catégorie particulière des données à caractère personnel. Elles comprennent les données qui révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale ; les données génétiques, les données biométriques aux fins d'identifier une personne physique de manière unique, les données concernant la santé ou les données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

² D'après Comité national pilote d'éthique du numérique - *Pour une éthique du numérique*. Coordonné par Éric Germain, Claude Kirchner, Catherine Tessier, PUF, février 2022

³ CNIL - [Algorithme](#)

⁴ Idem note 2

⁵ Jean-Paul Laumond - *La robotique : une récidive d'Héphaïstos*. Leçon inaugurale prononcée au Collège de France, 2012

⁶ CNIL - [Qu'est-ce ce qu'une donnée de santé ?](#)

⁷ [Règlement général pour la protection des données, article 9](#)

Bibliographie

- **Ouvrages**

- O. Cappé, C. Marc - *Tout comprendre (ou presque) sur l'intelligence artificielle*. CNRS éditions 2025
- S. Diamond, L. Masseron, J.P. Mueller - *L'intelligence artificielle pour les nuls*. First Interactive, 3^e édition 2025
- G. Bronner - *Apocalypse cognitive*. Le Livre de poche 2021
- L. Devillers - *Des robots et des hommes – Mythes, fantasmes et réalités*, Plon 2017
- L. Floridi - *L'Éthique de l'intelligence artificielle – Principes, défis et opportunités*. Éditions Mimésis 2023
- L. Julia - *L'Intelligence artificielle n'existe pas*. First 2019
- L. Julia - *IA génératives pas créatives*. Le Cherche-midi 2025
- É. Sadin - *Le désert de nous-mêmes*. L'échappée 2025

- **Avis et recommandations**

- Avis n°141 du CCNE et n°4 du CNPEN - [Diagnostic Médical et Intelligence Artificielle : Enjeux Éthiques](#), 2022
- Avis n°143 du CCNE et n°5 du CNPEN - [Plateformes de données de santé : enjeux d'éthique](#), 2023
- Haute autorité de santé - [Premières clefs d'usage de l'IA générative en santé dans les secteurs sanitaire, social et médico-social](#), décembre 2025
- UNESCO - [Recommandation sur l'éthique de l'intelligence artificielle](#), 2021
- Lignes directrices en matière d'éthique pour une IA digne de confiance. Groupe d'experts de haut niveau sur l'intelligence artificielle. Commission Européenne - <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, 2018

Ce document a été élaboré avec l'aide de Catherine Tessier, directrice de recherche et référente intégrité scientifique et éthique de la recherche à l'ONERA, membre du Comité consultatif national d'éthique du numérique.