

Université de Technologie de Compiègne

# IA Générative en santé

PR A24



Adèle Couvez | Paul Coulombel  
10/01/2025

# Table des matières

|   |    |
|---|----|
| 1. Introduction.....                                      | 1  |
| 2. Cadre d'étude.....                                     | 3  |
| a. Définition et exemples.....                            | 3  |
| b. Méthodologie.....                                      | 8  |
| 3. Contexte de déploiement de l'IA Gen en santé.....      | 9  |
| a. Les entreprises privées, les acteurs majoritaires..... | 9  |
| b. Formations et savoir-faire du personnel soignant.....  | 13 |
| c. Disparité de déploiement.....                          | 16 |
| 4. Défis éthiques et sociaux.....                         | 18 |
| a. Biais et discriminations.....                          | 18 |
| b. Gestion des données et du consentement.....            | 22 |
| c. Le manque d'explicabilité et ses conséquences.....     | 24 |
| d. Responsabilité des différents acteurs.....             | 31 |
| e. Gouvernance des données et des modèles.....            | 34 |
| f. Impacts environnementaux.....                          | 36 |
| 5. Possibles reconfigurations du système médical.....     | 38 |
| 6. Conclusion.....  | 45 |
| Bibliographie.....  | 47 |
| Sitographie.....  | 51 |

## 1. Introduction

Depuis quelques années, le nombre d'intelligences artificielles (IA) offrant la possibilité de générer du texte, des images mais aussi du son et des vidéos s'est multiplié. Le déploiement de ces IA génératives (IA Gen) prend des proportions considérables. Le nombre d'utilisateurs réguliers s'élève aujourd'hui à plusieurs centaines de millions de personnes, notamment avec la mise à disposition au grand public de ChatGPT, développé par Open AI, en version gratuite. La quantité et la variété des modèles créés explosent et les possibilités d'usages sont pratiquement infinies. Ces modèles rédigent des comptes-rendus, traduisent, corrigent ou résumant des textes, effectuent des recherches internet, créent ou modifient des images, produisent du contenu sonore (voix, musique), interagissent avec l'utilisateur, etc. La popularité de ces IA Gen est favorisée par le développement de *Large Language Model* (LLM) traitant et générant du texte comme le ferait un humain. Elles sont disponibles en ligne, faciles d'utilisation, employables pour des tâches très variées et surtout, elles répondent aux demandes (prompt) en un temps record. Même si le premier résultat donné n'est pas celui escompté, il est simple de modifier la demande pour spécifier les changements à opérer et de le régénérer comme souhaité, seulement quelques secondes après.

Le domaine médical n'échappe pas à l'arrivée de ces nouveaux outils. L'IA Gen est perçue comme révolutionnaire : elle serait la clé permettant au secteur de la santé d'être plus efficace, plus équitable et plus personnalisé. Son déploiement auprès du personnel médical et des patients est présenté comme indispensable et inévitable. De plus en plus d'entreprises proposent des solutions incluant de l'IA Gen pour répondre aux questions des médecins sur un traitement ou des dosages, rédiger des comptes-rendus, aider au diagnostic, répondre aux inquiétudes éventuelles des patients, etc. Ces outils amélioreraient la qualité des soins, la prise en charge des patients et les conditions de travail du personnel soignant, notamment en gagnant du temps sur les tâches administratives. Doctolib propose par exemple un « assistant de consultation » écoutant la consultation médicale, prenant des notes en temps réel et retranscrivant l'échange entre le patient et le soignant sous la forme d'une synthèse organisée<sup>1</sup>. Le temps de rédaction par le médecin est réduit et son attention peut se tourner exclusivement vers le patient. Mais, savoir que la consultation est écoutée ne modifierait-il pas l'interaction entre le médecin et le patient ? Les avantages vantés des IA Gen se retrouvent-ils dans la pratique ? Comment les informations personnelles inévitablement échangées lors de la consultation sont-elles traitées ? Sachant que les LLM se basent sur des calculs statistiques et non une recherche de la vérité, peuvent-ils interpréter et retranscrire la discussion avec sens et fidélité ? Comment intégrer les informations non verbales (timbre de voix, gestuelle, expression faciale, etc.) dans la synthèse ?

---

<sup>1</sup> « Intelligence artificielle : Doctolib lance l'Assistant de consultation ». 2024. About Doctolib - France. 15 octobre 2024. <https://about.doctolib.fr/news/intelligence-artificielle-doctolib-lance-lassistant-de-consultation/>.

Certes, l'intégration d'IA Gen en santé peut apporter des avantages, mais elle ne peut se faire sans prendre en compte les limites et les enjeux éthiques. La protection des données, l'implication de la responsabilité individuelle des soignants, le manque de fiabilité ou de formation à l'utilisation de ces outils peuvent freiner leur déploiement dans le domaine médical. Aujourd'hui, aucun modèle génératif n'est parfaitement intégré dans une pratique, la majorité d'entre eux sont encore en phase de développement, de test ou de certification (pour obtenir le marquage CE ou le statut de dispositif médical). De nombreux professionnels se montrent réticents à l'idée d'employer ce type d'outils et l'utilisation de ces derniers reste encore très marginale. Comment expliquer ce décalage entre l'engouement général pour les IA Gen et la réalité du terrain ? À quelles questions font face les potentiels futurs utilisateurs du secteur médical lorsqu'ils envisagent d'adopter un système d'IA Gen ? Quelles problématiques sont identifiées lors des tests sur le terrain ? Quels sont les risques et les limites liés à l'utilisation de ces modèles, en particulier dans le domaine de la santé ?

Pour répondre à ces questions, nous nous sommes intéressés dans un premier temps, au cadre actuel de déploiement des IA Gen tant du point de vue économique, infrastructurel et pédagogique. Dans un second temps, nous avons abordé les problématiques éthiques liées à l'utilisation de l'IA Gen dans le secteur de la santé. Enfin, nous abordons les potentielles reconfigurations que l'intégration de l'IA Gen pourrait apporter au domaine médical.

## 2. Cadre d'étude

### a. Définition et exemples

Dans ce document, nous identifions l'IA Gen comme les modèles informatiques basées sur des réseaux de neurones profonds pouvant générer du contenu à partir d'une base de données. Ce contenu peut être du texte, des images, du son, des vidéos, etc. et peut mélanger ces différentes formes. Les modèles sont développés à partir d'une quantité massive de données (une centaine de milliards de mots dans le cas des LLM) qui empêche la validation humaine exhaustive. Les IA Gen retrouvées en santé peuvent être des IA générales utilisées dans d'autres domaines ou des IA développées spécifiquement pour un usage médical. Nous avons répertorié dans le tableau suivant (Tableau 1) l'ensemble des cas d'utilisation d'IA Gen en santé que nous avons identifiés. Dans les exemples de modèle que nous donnons, certains sont déjà commercialisés, d'autres sont encore en phase de recherche ou ne sont qu'au stage d'idée.

|  | Utilisateurs                   | But recherché                                    | Mise en œuvre   | Comment elle est/veut être intégrée ? | Exemples                |
|--|--------------------------------|--|---|---------------------------------------|-------------------------|
| Diagnostic                                   | Patients                       | Pré-diagnostic patient                           | Établir un diagnostic à domicile et déterminer la durée avant laquelle il faudrait consulter un médecin | Chatbot conversationnel               | Ada Health <sup>2</sup> |
| Synthèse de dataset                          | Chercheurs                     | Anonymiser les données                           | Utiliser des données personnelles tout en respectant les RGPD   | Préparation des bases d'apprentissage | Octopize <sup>3</sup>   |
| Information pour les professionnels médicaux | Médecins, personnels soignants | Gain de temps, faciliter l'accès à l'information | Regrouper les informations médicales et les restituer   | Chatbot conversationnel               | PulseLife <sup>4</sup>  |
| Information pour le patient                  | Patients                       | Informier le patient                             | Poser des questions à Chat GPT  | Chatbot conversationnel               | Chat GPT <sup>5</sup>   |
|  | Patients                       | Faciliter l'accès à l'information                | Résumer les analyses médicales en les contextualisant   | Textes + schémas                      | Kiro <sup>6</sup>       |

2 Morse, Keith E., Nicolai P. Ostberg, Veena G. Jones, et Albert S. Chan. 2020. « Use Characteristics and Triage Acuity of a Digital Symptom Checker in a Large Integrated Health System: Population-Based Descriptive Study ». *Journal of Medical Internet Research Research* 22 (11). <https://doi.org/10.2196/20549>.

3 « Accueil – Octopize ». 2025. 5 janvier 2025. <https://www.octopize.io/>.

4 « PulseLife, La Médecine Fondée Sur Les Preuves ». 2025. PulseLife. 5 janvier 2025. <https://pulselife.com/fr-fr>.

5 Lee, Tsung-Chun, Kyle Staller, Vlaicu Botoman, Mythili P. Pathipati, Sanskriti Varma, et Braden Kuo. 2023. « ChatGPT Answers Common Patient Questions About Colonoscopy ». *Gastroenterology* 165 (2): 509-511. <https://doi.org/10.1053/j.gastro.2023.04.033>.

6 « Kiro : Construire la biologie de demain ». 2025. Kiro. 5 janvier 2025. <https://www.kiro.bio>.

| Suivi patient / surveillance    | Patients                | Gain de temps, meilleure prise en charge patients | Informar, suivre et répondre aux questions du patient  | Chatbot conversationnel       | Vik <sup>7</sup>                               |
|---------------------------------|-------------------------|---|--|-------------------------------|--|
| Préparation aux consultations   | Oncologues              | Gain de temps                                     | Résumer les symptômes du patient avant une consultation  | QCM + chatbot conversationnel | GPT 4 <sup>8</sup>                             |
|                                 | Urgences                | Gain de temps, meilleure prise en charge patients | Trier les patients en fonction de la gravité de leur état et de leurs symptômes à l'arrivée aux urgences | Chatbot conversationnel       | Chat GPT <sup>9</sup>                          |
| Comptes rendus et consultations | Médecins généralistes   | Gain de temps                                     | Résumer les symptômes du patient avant une consultation  | QCM + compte rendu            | Jaide <sup>10</sup>                            |
|                                 | Professionnels médicaux | Gain de temps                                     | Rédiger un compte rendu de la consultation   | Micro + synthèse de textes    | Doctolib <sup>11</sup> , Jaide <sup>12</sup> , |

7 Bibault, Jean-Emmanuel, Benjamin Chaix, Arthur Guillemassé, Sophie Cousin, Alexandre Escande, Morgane Perrin, Arthur Pienkowski, Guillaume Delamon, Pierre Nectoux, et Benoît Brouard. 2019. « A Chatbot Versus Physicians to Provide Information for Patients With Breast Cancer: Blind, Randomized Controlled Noninferiority Trial ». *Journal of Medical Internet Research Research* 21 (11). <https://doi.org/10.2196/15787>. Et Chaix, Benjamin, Jean-Emmanuel Bibault, Arthur Pienkowski, Guillaume Delamon, Arthur Guillemassé, Pierre Nectoux, et Benoît Brouard 2019. « When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot ». *JMIR Cancer* 5, n° 1. <https://doi.org/10.2196/12856>.

8 Wu, David JH, et Jean-Emmanuel Bibault. 2024. « Pilot applications of GPT-4 in radiation oncology: Summarizing patient symptom intake and targeted chatbot applications ». *Radiotherapy and Oncology* 190:109-978. <https://doi.org/10.1016/j.radonc.2023.109978>.

9 Zhang, Yanhui, Haolong Pei, Shihan Zhen, Qian Li, et Fengchao Liang. 2023. « Chat Generative Pre-Trained Transformer (ChatGPT) usage in healthcare ». *Gastroenterology & Endoscopy* 1 (3): 139-143. <https://doi.org/10.1016/j.gande.2023.07.002>

10 « Jaide | Generative AI for patients and physicians ». 2025. 5 janvier 2025. <https://www.jaide.care/>.

|                   |                                       |             |  |   |                              |  |
|-------------------|---------------------------------------|-------------|--|---|------------------------------|--|
| résumés           | Imagerie                              | Radiologies | Gain de temps                          | Rédiger un compte rendu à partir d'un diagnostic donné par une IA d'analyse d'image | Chatbot conversationnel      | Nabla <sup>13</sup>  |
|                   |                                       | Chirurgiens | Meilleure prise en charge patients     | Améliorer la qualité visuelle des vidéos chirurgicales                              | Vidéo sur écrans d'opération | Milvue <sup>14</sup> , Incepto <sup>15</sup> , Gleamer <sup>16</sup> |
| Imagerie médicale | Formation des professionnels de santé | Étudiants   | Former les soignants                   | Simuler un patient  | Chatbot conversationnel      | Chat GPT <sup>18</sup>   |
|                   |                                       | Oncologues  | Former et tester les soignants         | Répondre aux questions des soignants et leur en poser en retour pour les tester     | Chatbot conversationnel      | Watson assistant IBM <sup>19</sup>                                   |
| Chirurgie         |                                       | Chirurgiens | Gain de temps + diminution des risques | Répondre aux questions durant l'opération   | Chatbot conversationnel      | Chat GPT <sup>20</sup>   |

11 « Intelligence artificielle : Doctolib lance l'Assistant de consultation ». 2024. About Doctolib - France. 15 octobre 2024. <https://about.doctolib.fr/news/intelligence-artificielle-doctolib-lance-assistant-de-consultation/>.

12 « Jaide | Generative AI for patients and physicians ». 2025. 5 janvier 2025. <https://www.jaide.care/>.

13 « Nabla · L'assistant médical qui vous libère du temps ». 2025. 5 janvier 2025. <https://www.nabla.com/fr/>.

14 « Milvue | New Tech, New Care ». 2025. Milvue. 5 janvier 2025. <https://www.milvue.com/>.

15 « Incepto ». 2025. Incepto Medical. 5 janvier 2025. <https://incepto-medical.com/>.

16 « Gleamer Homepage AI for Radiology – Gleamer ». 2025. 5 janvier 2025. <https://www.gleamer.ai/>.

17 « Incepto ». 2025. Incepto Medical. 5 janvier 2025. <https://incepto-medical.com/>.

18 <https://link.springer.com/article/10.1007/s10439-023-03232-y>. Li, Wenbo, Yinxi Zhang, et Fengmin Chen. 2023. « ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad ? » *Annals of Biomedical Engineering* 51 (9): 1892-1897. <https://doi.org/10.1007/s10439-023-03232-y>.

19 Kovacek, David, et James C. L. Chow. 2021. « An AI-Assisted Chatbot for Radiation Safety Education in Radiotherapy ». *IOP SciNotes* 2 (3): 1-9. <https://doi.org/10.1088/2633-1357/ac1f88>.

20 Li, Wenbo, Yinxi Zhang, et Fengmin Chen. 2023. « ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad ? » *Annals of Biomedical Engineering* 51 (9): 1892-1897. <https://doi.org/10.1007/s10439-023-03232-y>.

|                            |             |  |  |                                      |                        |
|----------------------------|-------------|--|--|--------------------------------------|------------------------|
| Application à la recherche | Chirurgiens | Gain de temps + diminution des risques | Identifier les différents plans opératoires                        | Chatbot conversationnel              | Chat GPT <sup>21</sup> |
|                            | Chirurgiens | Gain de temps + diminution des risques | Expliquer au patient les dangers spécifiques de son opération      | Chatbot conversationnel              | Chat GPT <sup>22</sup> |
|                            | Chercheurs  | Idées pour la recherche                | Donner des pistes d'amélioration et de continuation des recherches | Chatbot conversationnel              | Chat GPT <sup>23</sup> |
|                            | Chercheurs  | Gain de temps                          | Créer des nouvelles molécules et tester leur efficacité, etc.      | Itérations successives, exhaustivité | Pfizer <sup>24</sup>   |

Tableau 1 : Cas d'usages réels et envisagés d'IA Gen en santé et exemples.

<sup>21</sup> *Ibid.*

<sup>22</sup> *Ibid.*

<sup>23</sup> Lahat, Adi, Eyal Shachar, Benjamin Avidan, Zina Shatz, Benjamin S. Glicksberg, et Eyal Klang. 2023. « Evaluating the Use of Large Language Model in Identifying Top Research Questions in Gastroenterology ». *Scientific Reports* 13 (1): 1-6. <https://doi.org/10.1038/s41598-023-31412-2>.

<sup>24</sup> L'Oeil de la E-santé, réal. 2024. #ESANTE24 | « L'IA et l'IA générative, quel impact pour le secteur de la santé ? » [https://www.youtube.com/watch?v=wPzWf8\\_ZIY](https://www.youtube.com/watch?v=wPzWf8_ZIY).

## b. Méthodologie

Dans cette étude, nous avons combiné une analyse théorique, à travers une revue de littérature scientifique et une analyse empirique via des entretiens qualitatifs avec des personnes en lien avec le secteur médical ou l'IA Gen. L'objectif de ces derniers était de confronter l'étude bibliographique avec les utilisations réelles de l'IA Gen sur le terrain. Nous avons ainsi illustré ou nuancé les concepts théoriques évoqués avec des perspectives concrètes et mis en évidence, le cas échéant, de potentiels écarts.

Pour cela, nous avons tenté d'entrer en contact directement avec des professionnels hospitaliers ou cliniciens, des entreprises développant des solutions d'IA Gen, des chercheurs, des instituts chargés de l'innovation ou de l'assistance au développement et à la mise en conformité des projets des entreprises.

Malgré plusieurs tentatives, nos démarches pour entrer en contact avec des professionnels de santé n'ont pas abouti. Ce constat peut s'expliquer par le manque de temps que le personnel médical peut accorder, illustrer un désintérêt pour les sujets concernant l'innovation et/ou suggérer que l'adoption de l'IA Gen en santé demeure aujourd'hui peu commune ou non revendiquée. Les personnes avec qui nous avons pu nous entretenir sont toutes, du fait de leur métier, plus ou moins concernées par l'innovation technologique.

Dans la suite, nous nous appuierons notamment sur six entretiens semi-directifs (menés entre octobre et décembre 2024) dont trois avec des concepteurs de solutions numériques intégrant de l'IA Gen, un avec une personne gérant un Entrepôt de Données de Santé (EDS) et deux avec des personnes responsables de l'innovation dans deux institutions différentes.

### *Entretien 1*

Avec un dirigeant d'une entreprise française spécialisée dans la génération de données synthétiques, avec moins de 10 employés et un capital social supérieur à 300k euros en 2022.

### *Entretien 2*

Avec un commercial d'une entreprise française qui développe des logiciels d'analyse d'images médicales par IA, employant moins de 100 personnes et ayant un capital social supérieur à 3M euros en 2022.

### *Entretien 3*

Avec un ingénieur informatique dans une entreprise canadienne spécialisée dans le traitement d'images médicales par IA Gen, qui prévoit avoir moins de 10 employés en 2025.

#### *Entretien 4*

Avec une responsable de projet pour un Entrepôt de Données de Santé (EDS) au sein d'un Centre Hospitalier Universitaire (CHU) français. Ce projet a pour objectif de collecter, structurer et gérer les données hospitalières.

#### *Entretien 5*

Avec une membre du pôle de direction de l'innovation dans un CHU français.

#### *Entretien 6*

Avec une directrice des projets de soins et de la santé numérique dans une structure française spécialisée dans les soins à domicile, qui prend en charge environ 45000 patients par an en 2023.

Durant nos recherches, nous avons participé à distance à la conférence du Moment de l'Intelligence Artificielle (MIAM) sur l'IA Gen en santé. Cette conférence s'est déroulée à Montréal le 24/10/2024, en collaboration avec l'École de l'Intelligence Artificielle en Santé (ÉIAS). Nous avons également assisté en présentiel aux tables rondes du Grand Remues-Méninges le 28/11/2024. Le Club santé & digital avait organisé cet événement sur le thème : l'IA en santé, une fausse promesse ?

### 3. Contexte de déploiement de l'IA Gen en santé

#### a. Les entreprises privées, les acteurs majoritaires

Aujourd'hui, la majorité des IA sont issues du secteur privé. En effet, en 2021, l'Europe a investi un milliard de dollars, les États-Unis 1,5 milliard et le secteur privé 340 milliards dans le développement d'IA<sup>25</sup>. Le développement d'une IA demande un gros investissement initial qui impacte la R&D des startups devant le rembourser. Par exemple, lors d'un entretien, un employé d'une entreprise développant une IA analysant des images médicales, nous a expliqué que son entreprise privilégiait le développement d'IA Gen « support » pour la rédaction de comptes-rendus, au développement d'IA Gen synthétisant des images médicales, qu'il pense être le « futur » de leur spécialité. Leur priorité est l'augmentation de leur chiffre d'affaires. Pour vendre plus d'abonnements et rembourser leurs investissements, ils développent des outils complétant leurs offres auprès des clients qui utilisent déjà leurs services. Ces gros investissements impactent également le produit en lui-même. Afin de toucher le plus de clients possibles, les entreprises développent des IA pouvant être utilisées dans des environnements variés, avec des protocoles et des façons de travailler différents<sup>26</sup>. Quelle est la réelle utilité de ces outils et quels sont leurs impacts sur le métier des utilisateurs ? Ces IA ne sont pas forcément adaptées aux besoins réellement éprouvés sur le terrain par leurs utilisateurs et ne s'insèrent pas efficacement dans tous les univers médicaux. Le remboursement de ces investissements peut impliquer un prix du produit final très élevé<sup>27</sup>. Deux questions se posent alors : qui sera contraint de payer ce surplus ? Et l'apport de l'outil vaut-il concrètement le prix auquel il est proposé ?

Lorsqu'une entreprise veut commercialiser un modèle d'IA, elle doit posséder une autorisation de mise sur le marché comme le marquage CE en Europe. Cette autorisation dépend des lois en vigueur dans la région où est utilisé l'outil. Les évaluateurs s'appuient sur les normes mises en place pour accorder ou non la commercialisation d'une IA. Dans le cadre plus restreint de la santé, peu de normes existent. Anne-Laure Bailly, consultante en affaires médicales et réglementaires, explique comment se déroule aujourd'hui une évaluation pour une autorisation de commercialisation d'une IA utilisée dans le domaine médical. Plusieurs points sont inspectés<sup>28</sup>:

- Le bénéfice pour le patient, l'utilisateur et la santé publique lors de l'emploi de l'IA qui est proposée.

25 OMS. 2024. « *Ethics and Governance of Artificial Intelligence for Health: Large Multi-Modal Models* ». *WHO Guidance*. 1st ed. Geneva: World Health Organization.

26 Grand Remue-Méninges, Fondation Université Paris Cité, 12 rue de L'École de Médecine, 28/11/2024.

27 Bubien, Yann, et Vincent Vuiblet. 2023. « La révolution numérique de la santé ». *Esprit*, n° 6, 77-87.  
<https://doi.org/10.3917/espri.2306.0077>.

28 Bpifrance, réal. 2024. *Comment l'IA révolutionne l'imagerie médicale et la médecine proactive*.  
<https://www.youtube.com/watch?v=oLogkWBvyr8>.

- Les risques liés à l'utilisation de l'IA, les conséquences d'une erreur de celle-ci et son impact sur le corps médical dans son ensemble. Avec l'inexplicabilité des IA génératives, il est difficile de vérifier les algorithmes. Ce sont donc les bases de données avec lesquelles le modèle a été entraîné et testé qui sont évaluées. Ces bases de données doivent représenter la population qui utilisera cet outil et contenir une grande diversité.
- L'évolution des modèles au cours de leur emploi est aussi un point clé de l'autorisation. Cela doit être prévu par l'entreprise qui développe l'IA générative afin que les modèles puissent s'adapter au mieux à la population qui les utilise.
- La présence et le contrôle d'un expert sur le résultat donné par l'outil informatique. La garantie humaine est pour le moment encore très présente dans les systèmes de comptes-rendus, mais devient très compliquée à mettre en place dans le développement de chatbots ayant pour but de parler directement avec un patient.

Ce temps de vérification peut durer plusieurs mois, ralentit encore la commercialisation des modèles d'IA et ajoute des frais supplémentaires pour les startups développant des IA. C'est le cas du fondateur d'une startup que nous avons rencontré.

Une fois l'autorisation de commercialisation obtenue, les entreprises doivent démarcher leurs clients et vendre leurs prestations. L'accessibilité au marché dépend beaucoup du pays dans lequel l'IA générative est déployée. Par exemple, aux États-Unis, un grand nombre de professionnels de santé sont regroupés sous un même organisme privé comme Kaiser. Lorsqu'une entreprise vend l'utilisation d'un de ses modèles d'IA, le contrat est directement signé par l'organisme privé. C'est l'ensemble des cliniques sous l'organisme privé qui bénéficie de l'IA. Le nombre de contrats signés par les prestataires est très faible mais leur prix équivaut à plusieurs dizaines de fois le prix proposé pour une seule clinique. Ce modèle économique est plus risqué, un rapport de force se crée entre l'organisme privé et le prestataire. Cependant, le ratio gain/temps pour les entreprises proposant des IA est bien meilleur qu'avec les autres modèles économiques et permet de diminuer les coûts de la branche commerciale de l'entreprise<sup>29</sup>. En France, l'effort commercial que fournissent les entreprises pour atteindre un chiffre d'affaires équivalent est bien plus important. En effet, le lien entre les différents hôpitaux et cliniques n'est pas présent et un contrat pour chaque clinique voulant leurs produits doit être signé<sup>30</sup>. De plus, toujours avec l'exemple américain, les organismes privés jouent le rôle des assureurs et possèdent les cliniques et les hôpitaux. Lors d'une table ronde<sup>31</sup>, Alexandre Lebrun, CEO de Nabla, explique que les organismes privés ont un intérêt financier à utiliser des outils préventifs et curatifs. Si l'efficacité préventive d'une IA est prouvée et peut leur faire gagner de l'argent, ils la mettront très rapidement en place. En France, la mutuelle qui contribue à notre système de santé et qui s'occupe de tous les frais liés aux « petites maladies » n'a pas la charge de payer les frais des maladies graves. Or les IA préventives, qui ont pour objectif de réduire le nombre de patients atteints de maladies graves, ne rentrent pas dans le cadre des consultations liées aux maladies graves, elles sont donc à la charge des mutuelles. Cela réduit la volonté des acteurs français à introduire ces nouvelles technologies et complique les ventes des startups évoluant sur ce marché. Le développement des startups françaises à l'international dépend fortement de leur croissance à l'échelle nationale. Or la volonté générale, prônant la souveraineté des modèles d'IA, veut fluidifier la création et la croissance des startups françaises<sup>32</sup>. Améliorer l'accessibilité au marché pour ces entreprises est-il envisageable ? Si oui, qui doit fournir des efforts pour l'améliorer et quelle forme doivent-ils prendre ? Les professionnels de santé doivent-ils se regrouper en corps de métier, le gouvernement pourrait-il mettre en place des outils reliant les utilisateurs aux développeurs ? Les entreprises peuvent-elles collaborer pour se développer mutuellement ?

29 L'Oeil de la E-santé, réal. 2024. #ESANTE24 | « L'IA et l'IA générative, quel impact pour le secteur de la santé ? » [https://www.youtube.com/watch?v=wPXzWf8\\_ZIY](https://www.youtube.com/watch?v=wPXzWf8_ZIY).

30 Carpentier, Manu, réal. 2024. *Le JT de la Radiologie*. Healthtech Innovation Paris. <https://www.jtradiologie.com/Video/view/<https://www.jtradiologie.com/Video/view/xs3mzvu>.

31 L'Oeil de la E-santé, réal. 2024. #ESANTE24 | « L'IA et l'IA générative, quel impact pour le secteur de la santé ? » [https://www.youtube.com/watch?v=wPXzWf8\\_ZIY](https://www.youtube.com/watch?v=wPXzWf8_ZIY).

32 Carpentier, Manu, réal. 2024. *Le JT de la Radiologie*. Healthtech Innovation Paris. <https://www.jtradiologie.com/Video/view/<https://www.jtradiologie.com/Video/view/xs3mzvu>.

Une autre considération pour les entreprises de ce secteur est la différence d'équipements entre le milieu industriel et les hôpitaux. En effet, Kendrys Legenty, directeur général de la Fabrique des Soignants, a expliqué durant le Grand Remue-Méninges que les infrastructures informatiques des hôpitaux datent des années 1990. La vitesse de génération de texte par des LLM par exemple, peut passer d'une seconde sur des serveurs récents à quelques secondes dans les hôpitaux et ainsi devenir un frein dans l'utilisation de l'outil par les soignants. Toujours d'après Kendrys Legenty, la modernisation du parc informatique français se fait de manière progressive, en passant d'infrastructures des années 90 aux années 2000, puis 2010, etc. Ce choix permet une meilleure adaptabilité des utilisateurs et permet aussi de garder un lien entre tous les hôpitaux qui possèdent des technologies similaires. Une autre solution serait de passer des infrastructures des années 90 à des infrastructures des années 2020, comme cela est fait pour certains parcs informatiques africains. Dans le cas des IA d'analyses d'images médicales, pour contourner ce problème, les entreprises mettent à disposition des serveurs pour faire tourner leurs modèles d'IA. Les images sont donc exportées de l'hôpital pour les traiter. Cependant, cette méthode ne peut être faite pour des données sensibles et non anonymes que des IA Gen peuvent traiter au sein d'un hôpital. Un nouveau mode de fonctionnement doit donc être mis en place entre le secteur privé et les hôpitaux. Si les hôpitaux veulent utiliser des modèles d'IA Gen traitant les données personnelles des patients, ils devront investir dans des serveurs informatiques et dans l'écosystème les entourant (experts en cybersécurité, en matériel informatique, etc.). Malgré la diminution des prix des abonnements avec le retrait de la location des serveurs, le prix global pour une présence de l'IA dans les services hospitaliers devrait augmenter en comparaison à une IA utilisée sur des serveurs externes. La division des tâches permettrait une baisse globale des prix. En effet, la gestion des serveurs et de leur sécurité était réalisée par des entreprises expertes dans ce domaine qui, avec plusieurs clients, pouvaient réduire leurs coûts de prestation. Pour les hôpitaux, le déploiement de ces outils demande un investissement pour les infrastructures puis un coût de maintenance et de gestion élevé pour la taille de leur parc.

Pour conclure, le secteur de l'IA médicale est confronté à de nombreux obstacles. Les investissements importants, le long délai d'obtention des autorisations, les problèmes d'accessibilité aux marchés français ou encore le manque d'infrastructures informatiques performantes au sein des hôpitaux, rendent difficile le développement de petites startups. Le manque de connaissance sur l'IA de la part des soignants peut être considéré comme un frein au nombre de ventes qu'ils peuvent conclure.

## b. Formations et savoir-faire du personnel soignant

L'arrivée des IA Gen est généralement perçue comme une perte de savoir-faire et une délégation de certaines tâches à l'outil informatique. Cependant, leur application requiert des savoir-faire et des connaissances, générales et/ou spécifiques à l'outil. Les soignants devront sélectionner les informations à transmettre à l'IA Gen, formuler des requêtes, interpréter les résultats, etc. Leur formation doit donc aborder l'utilisation et le fonctionnement des IA. Cependant, la formation initiale de certains corps de métiers est déjà saturée<sup>33</sup>. Avec le déplacement des concours nationaux au début de la sixième année, les étudiants en médecine n'ont que très peu de temps pour aborder et intégrer les connaissances demandées. Une formation superficielle à l'IA pourrait probablement être intégrée facilement dans leur formation, néanmoins, les experts présents lors du Grand Remue-Méninges défendaient l'importance d'une étude approfondie de l'IA par les étudiants pour connaître le fonctionnement interne de ces outils et leurs risques. Jean-Emmanuel Bibault suggérait malgré tout d'intégrer un bloc informatique/IA lors de la deuxième année de formation, lorsque la charge de travail est la moins élevée, en même temps que le bloc sur les biostatistiques. Est-il utile d'étudier seulement ponctuellement et en début de formation un sujet qui évolue aussi rapidement et qui sera différent lorsque les étudiants seront diplômés ? Une autre solution serait d'intégrer dans la formation continue des médecins les connaissances nécessaires à l'utilisation de l'IA. D'après Guillaume Assié, directeur de la Chaire IA en santé de l'Unité de Formation de Recherche (UFR) de Médecine de l'Université Paris, les prestataires d'IA en santé seraient les plus aptes à former les soignants à tous les aspects liés aux produits qu'ils vendent. La coordination et la supervision de ces prestataires restent tout de même un problème à régler. La question de l'utilité des formations sur le fonctionnement des IA a été soulevée par plusieurs experts comme Nicolas Castoldi, Co-directeur de l'Espace Ethique Île-de-France et co-Président du Club Santé & Digital, prônant une utilité temporelle. Ces connaissances sur le fonctionnement des IA seraient nécessaires lors de l'adoption de l'IA par la communauté et par les patients, puis lorsque ces outils seront particulièrement bien implantés, les connaissances sur leur application seraient suffisantes. L'exemple avancé est l'utilisation des scanners dans le domaine médical. Lors de leur adoption, les médecins étaient formés et connaissaient le fonctionnement et les principes sur lesquels reposaient les scanners, alors qu'aujourd'hui, ces connaissances ne font plus partie de la formation d'un médecin et peu d'entre eux pourraient expliquer leur fonctionnement précis.

33 Grand Remue-Méninges, Fondation Université Paris Cité, 12 rue de L'École de Médecine, 28/11/2024.

La grande quantité de données que peut traiter une IA permet l'emploi d'une diversité importante d'informations. Lors du développement des outils informatiques, des recherches peuvent mener à l'utilisation de données que les soignants ne maîtrisent pas nécessairement. Avec l'arrivée d'une nouvelle IA, le professionnel de santé devra dans certains cas être formé à la récupération de ces nouvelles données et à l'interprétation de celles-ci. Par ailleurs, les connaissances médicales au sein des équipes développant des outils informatiques sont rares et recherchées. Les experts du Grand Remue-Méninges aimeraient développer des formations complémentaires comme des masters pouvant être suivis par les étudiants de médecine afin d'augmenter le nombre de médecins pouvant travailler avec les entreprises développant ces outils et ainsi réduire l'écart de connaissance entre le monde médical et le monde industriel.

Les connaissances spécifiques à une IA concernent les performances de celle-ci dans les différents cas où elle est utilisée. Ces savoirs permettraient d'augmenter l'utilité et l'efficacité des modèles sur le terrain<sup>34</sup>. Cette connaissance empirique est acquise par la pratique et l'utilisation répétée d'un même outil. Elle ne peut pas être acquise lors de la formation des soignants au regard de la diversité des IA présentes sur le marché et des différences de mises en place au sein des services. Une étude<sup>35</sup> menée par Ethan Goh *et al.* ayant pour objectif de mettre en évidence l'influence d'un LLM (GPT4) sur le diagnostic réalisé par des médecins soulève différentes questions. Le score médian de diagnostic établi dans l'étude, pour GPT4 seul, est de 90 % et pour des médecins disposant seulement des ressources conventionnelles de 74 %. Cependant, le score des médecins disposant de GPT4 est de 76 %, soit uniquement deux pourcents de plus que les médecins sans GPT4. La connaissance théorique des hallucinations est assimilée par les utilisateurs de ces outils. Cependant, l'absence d'expérience de diagnostics d'un médecin avec IA réduit la confiance que le soignant peut lui donner et amène à soupçonner une hallucination ou une erreur qui n'en est pas obligatoirement une. L'explicabilité, que nous évoquerons plus tard, est un point important de la connaissance de l'outil. En effet, avec une IA explicable, la connaissance empirique des IA peut devenir une connaissance théorique et universelle qui faciliterait l'appropriation de l'outil par les nouveaux utilisateurs. Cette connaissance spécifique peut aussi être remise en question avec l'évolution des bases de données que nous avons vue précédemment. Cette évolution impacte directement les résultats de l'IA Gen et ainsi, il est difficile d'obtenir une expérience solide sur l'utilisation de celle-ci. De la même façon qu'avec l'étude précédente, l'évolution des bases de données rendant en théorie l'IA plus performante ne serait-elle pas compensée par la perte de performance liée à l'impossibilité de connaître son outil ?

34 OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

35 Goh, Ethan, Robert Gallo, Jason Hom, Eric Strong, Yingjie Weng, Hannah Kerman, Joséphine A. Cool, et al. 2024. « Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial ». *JAMA Network Open* 7 (10). <https://doi.org/10.1001/jamanetworkopen.2024.40969>.

La perte de connaissances et d'expériences qu'engendre la délégation de certaines tâches à l'IA Gen est un point très sensible en médecine. En effet, comment doit réagir le soignant, le service ou l'hôpital lorsqu'une panne, un bug ou une cyberattaque advient sur les IA ? Des réponses doivent être préparées pour subvenir aux besoins des patients malgré l'absence de ces outils<sup>36</sup>. À quel point le gouvernement est-il prêt à investir pour que les soignants continuent à être formés à des activités qui sont déléguées aux IA ? Les protocoles et les infrastructures peuvent-ils permettre la cohabitation entre l'IA et l'humain afin de garder un savoir-faire estimé indispensable ? Si la décision de garder un certain savoir-faire est approuvée, une réévaluation de l'utilité des IA au sein des différents services doit être réalisée. Par ailleurs, les prestataires de ce secteur devraient, eux aussi, participer à ces réflexions et proposer des solutions en cas de problème (bugs, sécurité des données, panne des serveurs, etc.) lors de l'utilisation de leurs modèles. Cependant, ce n'est pas dans l'intérêt de ces entreprises de proposer ce genre de service. Cela dégraderait l'image de leurs produits en évoquant les possibles problèmes auxquels ils pourraient faire face et augmenterait considérablement leurs frais de R&D pour peu de valeur ajoutée sur leurs produits<sup>37</sup>. Un autre point soulignant l'importance de l'expérience pour un médecin concerne les tests d'utilisation de l'IA aujourd'hui. En effet, naturellement, l'idée pré reçue serait de penser que ce sont les jeunes médecins et les internes qui seraient les plus enclins à utiliser ces nouveaux outils informatiques, alors que ce sont majoritairement les seniors qui veulent les tester et les utiliser<sup>38</sup>. Nicolas Castoldi explique cela par la peur des jeunes médecins de rencontrer un problème technique avec l'outil. Prenons l'exemple de l'IA Gen rédigeant le compte rendu d'une consultation, si un problème survient et que l'IA Gen n'a pas noté les informations explicitées lors de la consultation, un médecin expérimenté peut gérer la situation en reliant les informations qu'il a retenues avec les cas qu'il connaît, etc. Un jeune médecin, quant à lui, n'a pas l'aisance d'un médecin expérimenté pour rédiger un compte rendu à partir des informations superficielles qu'il a retenues.

Pour conclure, l'application de l'IA en santé doit être accompagnée d'une formation solide pour tous les soignants. Cette formation théorique et pratique est indispensable pour une implantation contrôlée et sereine de l'IA dans les différents métiers. La formation initiale des soignants ne doit pas être négligée, l'humain ne peut pas se reposer entièrement sur les IA pour les soins et les problèmes techniques doivent être prévenus. Il faut aussi souligner que les formations, en général, ne sont pas parfaites et n'assurent pas une assimilation totale des informations qui y sont présentes.

36 Grand Remue-Méninges, Fondation Université Paris Cité, 12 rue de L'École de Médecine, 28/11/2024.

37 *Ibid.*

38 *Ibid.*

### c. Disparité de déploiement

Afin de comprendre comment l'IA Gen se déploie ou pourrait se déployer en santé, nous nous sommes intéressés aux facteurs facilitant l'implantation de systèmes d'IA dans un service médical. La présence de l'IA en imagerie médicale est très importante par rapport aux autres domaines. Environ 58 % des logiciels comportant de l'IA et validés par la FDA (Food and Drug Administration) sont à destination du secteur de la radiologie<sup>39</sup>. Comment expliquer cette disparité ? Pourquoi les entreprises se sont-elles concentrées principalement sur ce secteur ? L'IA Gen pourrait-elle aussi s'implanter fortement dans un service en particulier ?

Tout d'abord, les réseaux de neurones convolutifs sont très bien adaptés aux traitements d'images grâce à leur capacité d'accepter un très grand nombre de paramètres en entrée (pour chaque pixel de l'image). Les bonnes performances et le potentiel de ces approches influencent le choix des nouvelles startups et les incitent à aller dans le domaine de l'imagerie médicale. Un deuxième point expliquant ces chiffres est la présence importante de la technologie au sein de cette profession. Les radiologues sont régulièrement amenés à travailler avec des dispositifs comprenant des algorithmes augmentant leur résolution et diminuant leur temps d'acquisition. Ainsi, les radiologues reconnaissent volontiers, avant l'IA, l'apport des avancées technologiques dans leur métier et y étaient plus familiers. Cette numérisation précoce a permis de produire et de conserver une quantité massive de données dans les *Pictures Archiving and Communicating System* (PACS). Le stockage et le regroupement de ces données ont facilité l'entraînement et donc le développement de modèles d'IA assez performants pour que leur utilisation ne soit ni une gêne ni une perte de temps mais un réel bénéfice pour les utilisateurs. Par ailleurs, les radiologues interagissent moins avec leurs patients que d'autres corps de métiers en santé. Les discussions entre le patient et le médecin pour établir un diagnostic, n'étant pas, à ce jour, remplacées par l'IA, favorise encore son implantation au sein des services de radiologie<sup>40</sup>.

39 Mignot, Léo, et Émilien Schultz. 2022. « Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé ». *Réseaux : communication, technologie, société*, Contrôler l'intelligence artificielle ?, 232-233 : 65-97. <https://doi.org/10.3917/res.232.0065>.

40 *Ibid.*

Nous pourrions ajouter à cela la facilité de l'anonymisation des images médicales. En effet, le format DICOM permet facilement d'identifier et de supprimer les informations sensibles présentes et de brouiller les pixels contenant des annotations personnelles. Cette anonymisation permet un export des images médicales pour l'entraînement des IA et la production de diagnostics sur des serveurs externes aux hôpitaux. Le deuxième point que nous pourrions apporter est la similitude des paramètres d'entrée d'une image médicale par rapport à un prompt. Comme nous le verrons plus tard, la sensibilité des IA Gen face aux changements légers des tournures de phrases et autres, pose des problèmes lors de son application. Les images médicales, quant à elles, ne changent que très peu entre elles, ce qui permet une utilisation plus aisée de l'outil. L'ensemble de ces points peuvent expliquer en partie l'implantation de l'IA dans le domaine de l'imagerie médicale. Cependant, il reste difficile à partir de ces informations de prévoir ou d'identifier un domaine médical propice au développement de l'IA Gen. En effet, les données médicales par nature confidentielles et ne possédant qu'une faible diversité, ne permettent pas un apprentissage optimal des IA Gen. Aucun service n'est prédisposé à l'emploi de l'IA Gen, sauf la radiologie qui utilise déjà l'IA, mais qui, comme nous l'avons vu, la déploie comme support aux autres IA déjà implantées. La quantité de données pour l'apprentissage des IA Gen est importante et la formation des Entrepôts de Données de Santé (EDS) est trop récente pour avoir accumulé assez. De plus, la complexité et la diversité de ces données rendent difficile leur anonymisation et leur utilisation simple.

## 4. Défis éthiques et sociaux

### a. Biais et discriminations

S'assurer que les modèles d'IA Gen garantissent l'inclusion et l'équité est essentiel<sup>41</sup>. Leur introduction dans le domaine de la santé doit favoriser l'égalité du traitement et de la prise en charge des patients et non pas engendrer, perpétuer ou aggraver une quelconque forme de discrimination. Elle ne doit pas être excluante envers une personne ou un groupe de personnes en raison de leur sexe, genre, race, âge, classe sociale, handicap, origine ethnique, orientation sexuelle, etc. Pour que les systèmes d'IA Gen et leur utilisation ne soient pas inéquitables, les biais quelle que soit leur origine, doivent être considérés et réduits au maximum.

L'une des premières sources de biais est la base de données sur laquelle les modèles sont entraînés. En effet, même si le programme est très performant, si les données d'entraînement sont inexactes ou biaisées, les résultats qui en découleront le seront aussi<sup>42</sup>. Les modèles pourraient même exacerber ou amplifier ces biais<sup>43</sup>. Ces derniers ont plusieurs origines, notamment les préjugés sociaux, des données déséquilibrées ou mal étiquetées<sup>44</sup>.

Tout d'abord, les données sur lesquelles l'algorithme se base peuvent être imprégnées de biais résultants de stéréotypes (d'ethnie, d'orientation sexuelle, etc.). En effet, toute donnée est produite à un moment par des êtres humains et reflète donc les préjugés présents dans notre société. L'exemple phare illustrant cette idée est l'IA de recrutement utilisée par Amazon à partir de 2015<sup>45</sup>. Cette IA associait à chaque candidat une note sur cinq points pour faciliter le processus de recrutement. Il s'est avéré qu'elle n'évaluait pas les candidats de manière neutre : elle associait fréquemment de plus mauvaises notes aux femmes, surtout pour les postes à responsabilité. Ce modèle a été entraîné sur les centaines de milliers de CV soumis à l'entreprise sur dix années. Or, la grande majorité des cadres recrutés durant cette période étaient des hommes. Le modèle a donc appris à sélectionner, pour deux CV similaires, les candidats masculins. Ainsi, les inégalités de sexe dans le monde professionnel, dans le secteur de la technologie et plus précisément pour les postes à responsabilité étaient inscrites dans les données. L'IA étant entraînée sur ces dernières, elle a perpétué, si ce n'est renforcé, le sexisme qui était déjà présent.

41 OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

42 Bertail, Patrice, David Bounie, Stéphan Cléménçon, et Patrick Waelbroeck. 2019. « Algorithmes: Biais, Discrimination et Équité ». *HR Today*, n° 58.

43 Besse, Philippe, Aurèle Besse-Patin, et Céline Castets-Renard. 2020. « Legal and Ethical Implications of Artificial Intelligence Algorithms in the Health Field ». *Questions de société* 8 (3).

44 Ministère de la Cybersécurité et du Numérique du Québec. 2024. « Guide des bonnes pratiques d'utilisation de l'intelligence artificielle générative », 1-25.

45 Dastin, Jeffrey. 2018. « Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women ». *Reuters*, sect. World. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>.

Une autre source de biais dans les données peut être liée à un manque de représentativité des données<sup>46</sup>. Le manque de données sur un sous-groupe introduit un déséquilibre lors de l'apprentissage de l'IA. Cette sous-représentation d'un sous-groupe engendre alors une diminution des performances de l'IA lorsqu'elle traite un cas qui le concerne. Veiller à ce que les données soient représentatives de ce qui est étudié est nécessaire pour garantir l'égalité entre les utilisateurs. Toutefois, dans certains cas, les données disponibles ne le permettent pas. Par exemple, une étude a montré que 96 % des bases génomiques en 2009 étaient tirées de la population européenne et 81 % en 2016<sup>47</sup>. Ainsi, seules des populations d'origine ethnique très spécifique sont représentées. De même, dans le cas de certaines maladies, telles que les maladies cardiovasculaires, la majorité des études cliniques sont réalisées sur des hommes<sup>48</sup>. Lorsque ces pathologies atteignent des femmes, elles sont moins bien dépistées et traitées, par manque d'information et d'études approfondies dissociant les deux sexes. Ainsi, les données disponibles aujourd'hui ne sont pas équitables entre les populations et au sein même des populations. Elles varient en fonction de l'origine ethnique, du sexe, etc. Les modèles d'IA peuvent alors se montrer plus performants pour certains types de personnes et être discriminants vis-à-vis d'autres.

L'Organisation Mondiale de la Santé (OMS) souligne d'ailleurs l'importance de permettre l'accès et l'utilisation de modèles d'IA performants par tous les pays<sup>49</sup>. Les caractéristiques et les besoins des populations pouvant grandement varier en fonction des régions, il est important que les groupes de personnes auprès desquelles les données sont récoltées, soient les mêmes que ceux auxquels les modèles sont dédiés<sup>50</sup>. Il peut s'agir des données en elles-mêmes ou de la forme qu'elles prennent et en particulier de la langue dans laquelle elles sont exprimées. En effet, le langage est imprégné de représentations propres aux cultures considérées et possède des ambiguïtés<sup>51</sup>. Si la majorité des textes utilisés comme source sont dans une certaine langue, le modèle sera moins efficient pour d'autres (d'autant plus si elles sont très différentes), ce qui peut notamment poser problèmes avec les chatbots.

46 Besse, Philippe, Aurèle Besse-Patin, et Céline Castets-Renard. 2020. « Legal and Ethical Implications of Artificial Intelligence Algorithms in the Health Field ». *Questions de société* 8 (3).

47 Popejoy, Alice B., et Stephanie M. Fullerton. 2016. « Genomics Is Failing on Diversity ». *Nature* 538 (7624): 161-164. <https://doi.org/10.1038/538161a>.

48 Bastian-Pétrél, Kathleen, Jessica L. Rohmann, Sabine Oertelt-Prigione, Marco Piccininni, Katja Gayraud, Michelle Kelly-Irving, et Nathalie Bajos. 2024. « Sex and Gender Bias in Chronic Coronary Syndromes Research: Analysis of Studies Used to Inform the 2019 European Society of Cardiology Guidelines ». *The Lancet Regional Health – Europe* 45. <https://doi.org/10.1016/j.lanepe.2024.101041>. Et Tobb, Kardie, Madison Kocher, et Renée P. Bullock-Palmer. 2022. « Underrepresentation of women in cardiovascular trials- it is time to shatter this glass ceiling ». *American Heart Journal Plus: Cardiology Research and Practice* 13:100-109. <https://doi.org/10.1016/j.ahjo.2022.100109>.

49 OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

50 Bertail, Patrice, David Bounie, Stéphan Cléménçon, et Patrick Waelbroeck. 2019. « Algorithmes : Biais, Discrimination et Équité ». *HR Today*, n° 58.

51 Nordlinger, Bernard, Olivier De Fresnoye, et Claude Kirchner. 2024. « Systèmes d'IA générative en santé : enjeux et perspectives ». *Académie Nationale de Médecine* 208 (5): 536-547.

Le manque de données, le déséquilibre en termes de représentativité et la présence de stéréotypes dans les données servant à l'apprentissage d'un modèle d'IA introduisent des biais importants dans les résultats qu'elle fournit, amenant à la perpétuation ou l'accentuation de discriminations d'ores et déjà présentes dans notre société. La création de bases de données les plus représentatives et avec le moins de biais possible ainsi que leur contrôle sont donc essentiels pour garantir l'équité et l'inclusivité des modèles. Cependant, certaines IA, dont les IA Gen, sont entraînées sur des quantités massives de données, or plus leur nombre augmente, plus la régulation devient complexe. Pour limiter l'apparition de biais dans les réponses fournies par l'IA ou même pour restreindre la propagation de fausses informations, le programme en lui-même peut être modifié.

Une alternative complémentaire pour réguler les réponses de l'IA est d'agir directement sur le programme. La quantité massive de données d'entraînement ne pouvant pas être exempte de biais, une certaine forme de neutralité pourrait être atteinte grâce à une intervention dans le code même des IA. Cette quête de neutralité passerait par la construction de filtres restreignant la possibilité de propager des discours nuisibles, haineux ou des réponses biaisées<sup>52</sup>. Les concepteurs et développeurs seraient chargés de valider et de reformuler les résultats si nécessaire pour obtenir une IA équitable et inclusive<sup>53</sup>. Mais comment tester tous les résultats alors qu'ils sont par définition infiniment variés ?

L'application de tels filtres requiert de déterminer au préalable les valeurs sociales que l'on souhaite protéger ou défendre et la manière dont nous voulons le faire. Déterminer ce que peut générer ou non une IA que ce soit en termes de texte, d'image ou autres, demande d'identifier les thèmes sensibles et de fixer une limite entre ce qui est toléré et interdit. Bannir certains mots ou expressions considérés comme insultants ou établir les conditions sous lesquelles certains thèmes doivent être abordés demande de définir ce qui est discriminant ou non, les thèmes sensibles, etc. Par exemple, lorsqu'une IA aborde des théories démenties par des preuves scientifiques (telles que « la Terre est plate »), elle pourrait être obligée de confronter les arguments de ces théories avec les études scientifiques démontrant l'inverse (comme le fait ChatGPT actuellement). Quels mots sont discriminants ? Le sont-ils dans tous les contextes ? Quelles affirmations l'IA ne peut-elle pas soutenir ? Comment définir les thèmes qui doivent être nuancés ? Contrôler ce qui est possible et impossible implique inévitablement l'intégration de valeur au sein même des IA. Dans le cas présent, l'influence humaine sur le programme est souhaitée, mais, il est aussi possible que des biais s'insèrent dans le programme sans que les développeurs en aient conscience.

52 *Ibid.*

53 Ministère de la Cybersécurité et du Numérique du Québec. 2024. « Guide des bonnes pratiques d'utilisation de l'intelligence artificielle générative », 1-25.

Le programme étant écrit par des êtres humains, des biais peuvent s'introduire sans que les développeurs y prêtent attention. Les biais cognitifs influencent la manière de programmer et orientent donc les algorithmes<sup>54</sup>. Concevoir un modèle suppose de faire des choix, de privilégier une orientation par rapport à d'autres et traduit à terme une certaine perception des phénomènes, pouvant biaiser les réponses de l'IA<sup>55</sup>. Outre, les biais cognitifs identifiés (biais de « du mouton de Panurge », d'anticipation, de confirmation, de « corrélation illusoire », etc.), la culture du programmeur, ses préjugés et sa perception du monde conditionnent sa façon de coder. Par exemple, la majorité des professionnels du domaine informatique étant des hommes, s'ils ne sont pas sensibilisés aux stéréotypes, le risque d'introduction de biais de genre est très important<sup>56</sup>. L'écueil est double : des biais sont glissés dans le modèle et souvent de manière inconsciente.

Le langage de programmation autorise lui aussi certaines actions et en interdit d'autres. Chaque langage possède ses avantages et ses inconvénients et favorise ainsi certaines orientations. Il n'est pas neutre, il influence la manière de coder et donc le fonctionnement du modèle d'IA.

Pour conclure, les IA ne sont pas des outils neutres. Des biais volontaires ou involontaires, de nature et d'origine variées s'introduisent tout au long de leur conception : de la programmation à l'apprentissage, et même, tout au long de leur vie. L'évaluation des biais présents et leur prise en compte doit donc être continue. Ces considérations sont d'autant plus importantes que les IA peuvent sembler neutres : la grande quantité de données sur laquelle elle se base leur permettrait d'être impartiales et la programmation serait une activité objective. Les utilisateurs doivent donc être informés et capables d'identifier les potentiels biais présents dans les réponses afin d'en tenir compte dans leur interprétation. La compréhension du modèle et des résultats fournis est donc essentielle pour prendre en considération les biais et ainsi ne perpétuer ou n'aggraver aucune forme de discrimination.

54 Bertail, Patrice, David Bounie, Stéphan Cléménçon, et Patrick Waelbroeck. 2019. « Algorithmes : Biais, Discrimination et Équité ». *HR Today*, n° 58.

55 *Ibid.*

56 Bonifay Emmanuelle, et Sereno, Sophie. 2022. « Genre, Algorithme et Droit ».

## b. Gestion des données et du consentement

Le développement et la performance des IA dépendent fortement des bases de données sur lesquelles elles sont entraînées. Les développeurs font varier la taille, la qualité et la forme de celles-ci pour perfectionner leurs IA. Cependant, plusieurs problèmes se posent lorsque ces processus sont appliqués au domaine médical. Les données de santé des patients sont, par nature, personnelles et sensibles. Le consentement du patient est donc indispensable pour la récupération, le stockage et l'utilisation de ses informations<sup>57</sup>. La quantité de données nécessaire pour développer des modèles d'IA Gen est conséquente et rend ce travail très laborieux. Les hôpitaux ont mis en place des EDS permettant de stocker l'ensemble des données patient récupérées lors de consultations au sein des hôpitaux. Le consentement pour le stockage des données dans ces entrepôts est donné implicitement, en l'absence de contre-indication de la part du patient. Les données sont donc récupérées, stockées et peuvent être utilisées dans le cadre de projets de recherche. Une autorisation délivrée par la Commission Nationale de l'Informatique et des Libertés (CNIL) aux hôpitaux est tout de même nécessaire pour l'ouverture et la gestion d'un EDS. Il est important de souligner que cette autorisation est fondée sur une liste de critères à respecter qui évolue au cours du temps, mais que les autorisations, elles, ne sont pas remises en cause. Un EDS créé en 2019 peut ne pas respecter la liste de critères de 2024. L'autorisation définitive délivrée par la CNIL et le consentement implicite des patients lors de leurs venues à l'hôpital remettent en cause ce système de récupération de données. Toutefois, les EDS examinent chaque demande d'utilisation des données avec un jury interne à l'hôpital. Ce jury analyse la démarche scientifique, la nécessité des données demandées pour le projet de recherche et l'application des futurs modèles produits avec les données de l'EDS. Les hôpitaux mettent ensuite à disposition des serveurs informatiques locaux pour produire les modèles des différents projets de recherche sans exporter les données personnelles récoltées. Une fois les modèles créés, le jury doit donner son accord pour les exporter de l'EDS. Lorsqu'il n'est pas possible d'utiliser les serveurs informatiques locaux et malgré l'interdiction d'exporter les données, certains EDS autorisent cette exportation en faisant signer un document d'utilisation<sup>58</sup>. Cependant, lors d'un entretien avec un chercheur étudiant une maladie rare, ce dernier nous a avoué que ses confrères et lui partageaient directement par mail les « données patient » de leur hôpital respectif pour accélérer les transmissions et réduire les coûts de recherche. En effet, les demandes d'exports de données sont étudiées individuellement et peuvent être monnayées. Certains chercheurs aimeraient créer un règlement permettant de définir des cas généraux pouvant couvrir 80 % des cas et ainsi fluidifier les demandes et éviter la transmission non sécurisée de données sensibles. De plus, après avoir discuté avec certains experts des maladies rares (chercheurs, présidents d'association, etc.), une réelle divergence entre la volonté de partage des patients ayant des maladies rares et les autres est apparue. L'implication de ces patients dans leur maladie est plus importante et une vraie volonté de partager leurs données est présente. Dans ce genre de cas, c'est la réglementation qui restreint la transmission et l'utilisation de certaines « données patient ».

57 Entretien n°4.

58 *Ibid.*

Un moyen utilisé actuellement pour s'affranchir du consentement du patient est l'anonymisation. Après la récupération des données médicales d'une personne, il est possible de créer un avatar de ce patient possédant des caractéristiques proches<sup>59</sup>. Toutefois, cet avatar doit avoir des caractéristiques assez éloignées pour que l'identification du patient *a posteriori* ne soit pas faisable ou que les efforts demandés pour l'identifier soient considérés comme trop importants<sup>60</sup>. Avec cette méthode, le consentement du patient est nécessaire seulement pour la création de l'avatar. Ces nouvelles données peuvent ensuite être utilisées et stockées librement<sup>61</sup>. Des projets de recherche hospitaliers dans ce domaine sont conduits et certaines entreprises proposent déjà ce service. Basée sur une approche symbolique, ils utilisent des algorithmes permettant de :

- Si les données sont dans des textes, récupérer les informations importantes s'y trouvant afin de structurer la base de données (sous forme de tableau, une ligne équivaut à un point de la base de données).
- Anonymiser la base de données structurée.
- Réduire le bruit et augmenter les performances des modèles d'IA qui utiliseront cette base de données.

Des LLM peuvent être utilisés par la suite s'il est indispensable que la forme de la base de données soit textuelle. Ils ont donc pour but de générer des textes à partir des données synthétiques. Cette réécriture peut modifier la mise en forme des textes de manière aléatoire ou d'une façon à retrouver la même mise en forme qu'originellement. L'approche symbolique est indispensable pour garantir une anonymisation. La génération de données à partir d'IA Gen ne peut pas prouver une anonymisation complète, car des données de sa base d'apprentissage peuvent être retrouvées dans les modèles créés<sup>62</sup>. La réduction du bruit passe par une modification de la base de données. Les points très éloignés des autres, correspondant à des erreurs de mesures ou des personnes avec des caractéristiques hors normes, sont exclus de la base de données ou rapprochés de la norme afin de diminuer « le bruit des marginaux ». Cette course au pourcentage entrave alors les principes d'équité : les personnes aux caractéristiques hors normes ne sont pas représentées dans la base de données. Lorsque ces bases de données sont ensuite utilisées pour entraîner d'autres modèles d'IA, ce manque d'équité se transmet et peut s'accroître. À terme, les modèles d'IA pourraient être efficaces seulement pour les personnes ayant des caractéristiques les plus représentées dans la population. Les modifications de la base de données ou des modèles d'IA afin d'obtenir de meilleures performances introduisent donc des biais importants.

59 Entretien n°1 et n°4.

60 Julie Follet.

61 Entretien n°1.

62 « Solutions – Octopize ». 2025. 5 janvier 2025. <https://www.octopize.io/solutions#section-FAQ>.

De la même façon, des recherches sur le développement d'IA Gen synthétisant des bases de données à partir de données existantes sont en cours<sup>63</sup>. Le but de ces IA serait par exemple de constituer à partir de deux images médicales une nouvelle avec des caractéristiques de l'une et de l'autre. Si, sur une image médicale, un patient avec une morphologie A possède une tumeur et sur une deuxième image un patient avec une morphologie B n'en possède pas, l'IA Gen pourrait permettre de générer une image médicale d'un patient possédant une tumeur et la morphologie B. Ces IA soulèvent des questions d'utilités sur l'ajout réel d'information dans la base de données et des questions éthiques sur la reproduction de biais, la nécessité de recourir à ces techniques alors que le système médical possède déjà les données nécessaires, l'exclusion de l'humain, etc. En effet, la génération de données est nécessaire lorsque la quantité de données disponibles est trop faible pour le développement de modèles performants. Cela augmentera-t-il la taille des bases de données d'apprentissage ? Cette quantité de données et les biais qui peuvent y être amplifiés<sup>64</sup> sont-ils contrôlables et vérifiables ? L'IA explicable sur ce type de données synthétiques maintient-elle toujours cette explicabilité ? Une boucle sans fin de générations de données et d'entraînement de modèles générant de nouvelles données apparaîtra-t-elle ? La complexité et la taille des algorithmes empêcheront-elles la réparation et la supervision humaine ? Des IA ayant pour but de réparer des IA et de superviser les bases de données seront-elles nécessaires ?

Pour conclure, les performances des IA sont très liées à leurs bases d'apprentissage. En santé, la sensibilité des données traitées complexifie l'apprentissage et la création de modèles d'IA. Le recours à l'anonymisation peut être un moyen de sécuriser les données personnelles des patients tout en les utilisant pour développer de nouveaux outils informatiques. En allant plus loin, il serait possible d'augmenter la taille des bases d'apprentissage grâce à des données synthétiques, ce qui peut potentiellement augmenter les performances des modèles, mais soulève de nouveau de nombreuses questions éthiques.

### c. Le manque d'explicabilité et ses conséquences

L'explicabilité est considérée par de nombreux experts comme l'un des piliers éthiques des intelligences artificielles<sup>65</sup>. Selon « Les lignes directrices en matière d'éthique pour une IA digne de confiance<sup>66</sup> » définies par la Commission européenne, elle fait partie des quatre principes définissant une « IA digne de confiance », aux côtés du respect de

63 Carpentier, Manu, réal. 2024. *Le JT de la Radiologie*. Healthtech Innovation Paris. <https://www.jtradiologie.com/Video/view/<https://www.jtradiologie.com/Video/view/xs3mzvu>. Et Entretien n°2.

64 Bertail, Patrice, David Bounie, Stéphan Cléménçon, et Patrick Waelbroeck. 2019. « Algorithmes: Biais, Discrimination et Équité ». *HR Today*, no 58.

65 Coeckelbergh, Mark. 2019. « Artificial Intelligence: Some Ethical Issues and Regulatory Challenges ». *Technology and Regulation* 1:31-34. <https://doi.org/10.26116/techreg.2019.003>. Et OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

66 Veerappa, Manjunatha, et Salvo Rinzivillo. 2023. « Special Theme : eXplainable AI ». *ERCIM*, n° 134, 8-37.

l'autonomie humaine, de la prévention de toute atteinte et de l'équité. Ces quatre piliers sont considérés comme des « impératifs éthiques<sup>67</sup> » permettant d'assurer les droits fondamentaux.

Cependant, bien que cette notion ait une place centrale dans les questionnements éthiques autour des IA, il n'existe aujourd'hui aucune définition consensuelle<sup>68</sup>. Malgré tout, plusieurs définitions se rejoignent : dans le guide présentant les caractéristiques des IA dignes de confiance, l'explicabilité suppose que « les décisions prises par un système d'IA puissent être comprises et retracées par des êtres humains<sup>69</sup> ». La CNIL la définit aussi comme étant « la capacité de mettre en relation et de rendre compréhensible les éléments pris en compte par le système d'IA pour la production d'un résultat<sup>70</sup> ». Dans les deux cas, la compréhension du résultat et des processus qui y ont mené à partir des données d'origine sont évoqués.

De manière générale, la notion d'explicabilité renvoie à la capacité de rendre certains aspects d'une IA compréhensibles par des humains. Il peut s'agir d'explicitier le déroulement de l'apprentissage de l'IA à partir des données d'origine ou bien le processus qui a mené une IA à prendre telle ou telle décision à partir d'un certain prompt. Deux types d'explicabilité sont d'ailleurs souvent distingués : l'explicabilité locale et globale<sup>71</sup>.

La première, l'explicabilité locale, correspond à la prédiction unique de la réponse d'un système d'IA à partir d'une demande spécifique. L'objectif est de mettre en évidence les relations existantes entre les informations présentes dans la demande (ou prompt) avec la réponse qui est donnée par l'IA ainsi que de retracer les raisonnements qui ont mené à ce résultat précis. Tandis que la seconde, l'explicabilité globale, concerne le modèle dans son ensemble. Elle évalue la capacité à rendre compréhensible la totalité des résultats possibles à partir de toutes les valeurs des variables d'entrées. Elle permet une vérification de tous les cas et la certitude de ne pas obtenir d'hallucinations lors de son utilisation.

67 *Ibid.*

68 Vilone, Giulia, et Luca Longo. 2021. « Notions of explainability and evaluation approaches for explainable artificial intelligence ». *Information Fusion* 76 (décembre):89-106. <https://doi.org/10.1016/j.inffus.2021.05.009>.

Et Longo, Luca, Mario Brcic, Federico Cabitza, Jaesik Choi, Roberto Confalonieri, Javier Del Ser, Riccardo Guidotti, et al. 2024. « Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions ». *Information Fusion* 106 (juin):1-22. <https://doi.org/10.1016/j.inffus.2024.102301>.

69 Veerappa, Manjunatha, et Salvo Rinzivillo. 2023. « Special Theme : eXplainable AI ». *ERCIM*, n° 134, 8-37.

70 « Explicabilité (IA) ». 2025. CNIL. 5 janvier 2025. <https://www.cnil.fr/fr/definition/explicabilite-ia>.

71 Longo, Luca, Mario Brcic, Federico Cabitza, Jaesik Choi, Roberto Confalonieri, Javier Del Ser, Riccardo Guidotti, et al. 2024. « Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions ». *Information Fusion* 106 (juin):1-22. <https://doi.org/10.1016/j.inffus.2024.102301>.

Bien que l'explicabilité soit considérée comme un des fondements éthiques des IA, la complexité technique des modèles connexionnistes compromet son application dans la pratique<sup>72</sup>. En effet, avec l'essor des IA se basant sur des réseaux de neurones et le *deep learning*, il est de moins en moins aisé de comprendre le raisonnement suivi par l'IA<sup>73</sup>. La multiplication des IA connexionnistes par rapport aux IA symboliques opacifie techniquement nos outils informatiques<sup>74</sup>. Ces IA, dont nous ne connaissons et ne contrôlons pas le fonctionnement précis, sont aussi appelées « boîtes noires ». C'est notamment le cas des IA Gen, fondées sur une architecture de réseaux de neurones profonds et entraînées avec une quantité massive de données.

Face au développement d'IA de moins en moins explicables, y compris par les développeurs et les concepteurs, l'explicabilité et son évaluation en termes de fiabilité et de validité, deviennent des problématiques majeures<sup>75</sup>. C'est dans ce contexte qu'a émergé l'IA explicable, un domaine de recherche visant à développer des méthodes permettant à l'utilisateur de comprendre facilement, de rendre intelligibles les raisonnements qui ont mené une IA à telle ou telle décision ou prédiction<sup>76</sup>. La compréhension des mécanismes sous-jacents des IA et de leurs résultats est d'autant plus importante lorsque ces IA « boîte noire » sont utilisées dans des domaines considérés à hauts risques<sup>77</sup>, tels que la santé ou la finance.

En dépit des nombreuses recherches menées pour tenter de rendre les IA Gen plus explicables, certaines problématiques ne sont toujours pas résolues et requièrent une analyse plus approfondie<sup>78</sup>. Expliquer les raisons qui ont poussé un modèle à générer un résultat ou à prendre une décision en particulier, n'est pas toujours possible. Il est d'ailleurs tout autant difficile de déterminer quelles combinaisons de facteurs d'entrée y ont contribué<sup>79</sup>. Aujourd'hui, les IA fondées sur des réseaux de neurones, telles que les IA Gen, ne sont pas explicables. Il est impossible de comprendre ou d'exercer une forme de contrôle sur leur fonctionnement, même pour les concepteurs.

72 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

73 Veerappa, Manjunatha, et Salvo Rinzivillo. 2023. « Special Theme : eXplainable AI ». *ERCIM*, n° 134, 8-37. <https://ercim-news.ercim.eu/images/stories/EN134/EN134-web.pdf>

74 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

75 Vilone, Giulia, et Luca Longo. 2021. « Notions of explainability and evaluation approaches for explainable artificial intelligence ». *Information Fusion* 76 (décembre):89-106. <https://doi.org/10.1016/j.inffus.2021.05.009>.

76 *Ibid.*

77 *Ibid.*

78 *Ibid.*

79 Veerappa, Manjunatha, et Salvo Rinzivillo. 2023. « Special Theme : eXplainable AI ». *ERCIM*, n° 134, 8-37.

Le manque de moyens techniques ou de connaissances ne permet pas à lui seul d'expliquer ces lacunes en termes d'explicabilité. En effet, il est dans la nature même de certaines IA, notamment celles se fondant sur des réseaux de neurones, entraînées sur une très grande quantité de données et se référant à des calculs statistiques complexes, de rendre difficile l'établissement du lien entre la demande avec le résultat. Ces caractéristiques intrinsèques aux IA basées sur des réseaux de neurones rendent impossible la démonstration du processus ayant mené au résultat et compliquent l'établissement d'un lien entre les données d'entrée et celles de sortie. Ainsi, il est compliqué, voire impossible d'obtenir une suite de justifications sous forme de décisions logiques qui permette d'expliquer le résultat d'une IA connexionniste, contrairement aux IA symboliques. De plus, de légères modifications dans les données d'entrée peuvent grandement modifier la réponse de l'IA<sup>80</sup>. Les utilisateurs doivent donc apprendre à rédiger au mieux leurs demandes pour obtenir la réponse avec les informations, le niveau de précision et la forme souhaités. Cette sensibilité des IA Gen, c'est-à-dire, cette grande variabilité des résultats donnés pour des demandes similaires, les rend également moins fiables aux yeux des utilisateurs.

La question de l'explicabilité des IA est d'autant plus importante dans les domaines à hauts risques puisqu'aujourd'hui, il n'existe aucune réglementation obligeant les concepteurs à une quelconque exigence d'explicabilité, ni obligeant les utilisateurs du secteur médical à utiliser exclusivement des IA explicables. Ainsi, de nombreux modèles génératifs ont été développés ou sont en développement pour des applications dans le monde médical.

Les décisions prises par le personnel médical et de manière plus générale par tout professionnel travaillant dans le domaine de la santé (personnels administratifs, d'innovation, etc.) peuvent avoir des répercussions graves, notamment pour les patients. Si l'intelligence artificielle intervient dans les processus décisionnels mais aussi administratifs dans le domaine de la santé, son utilisation se doit d'être d'autant plus encadrée et transparente.

Par exemple, supposons qu'un médecin soit assisté par une IA Gen (et donc non explicable) pour établir un diagnostic. Lorsque ses résultats et ceux de l'IA divergent, il lui est impossible de comprendre pourquoi. Les seules possibilités s'offrant à lui sont de croire aveuglément en l'IA, de maintenir son diagnostic malgré les contradictions, de faire des analyses complémentaires afin de vérifier les résultats obtenus ou alors de solliciter l'avis de ses collègues. Comment l'IA peut-elle être un outil si les réponses qu'elle apporte ne sont pas justifiées ? Dans quelle mesure une IA dont nous ne pouvons pas retracer le raisonnement est une aide pour son utilisateur ? Une IA non explicable peut-elle réellement aider à la prise de décision ?

80 *Ibid.*

Ces questions s'étendent d'ailleurs à tout type d'utilisation d'IA qui serait non explicable, les réponses de l'IA peuvent mettre en doute celles des humains. Comprendre son outil, son mécanisme, ses failles et ses limites sont des prérequis pour pouvoir s'en servir à bon escient, pour en avoir une utilisation éclairée et surtout pour qu'il soit bénéfique à son utilisateur. L'explicabilité d'une IA permet aussi de mettre en évidence les biais ou les erreurs dans le raisonnement qu'elle a suivi. L'utilisateur peut alors comprendre la réponse donnée et donc potentiellement identifier l'origine d'une erreur afin de la corriger (en modifiant les données d'entrées par exemple) ou de la prendre en compte dans son interprétation des résultats. La possibilité de contester les résultats d'une IA de manière constructive ne peut se faire sans explication, sans connaissance du modèle et de son fonctionnement<sup>81</sup>. Sans explicabilité, il ne s'agit pas de comprendre, mais plutôt de croire en l'IA, ce qui soulève davantage une question lorsque les conséquences des décisions prises à partir de son utilisation peuvent s'avérer graves (mauvais diagnostic, mauvais choix de traitement pour le patient, mauvaises informations ou conclusions dans les comptes rendus, mauvaises informations transmises aux patients, etc.)

En outre, le manque d'explicabilité peut être à l'origine d'une baisse de performance des modèles. Au-delà de la possibilité de contester les résultats fournis par une IA, l'explicabilité a pour objectif de soutenir et d'améliorer la confiance humaine accordée aux réponses du modèle<sup>82</sup>. Ainsi, face à la méfiance des utilisateurs, les performances d'un modèle peuvent diminuer. L'expérience menée par Ethan Goh *et al.* montre en particulier qu'une IA Gen seule (GPT4) possède de meilleures performances pour établir un diagnostic que les médecins accompagnés par cette même IA (score de 90 % contre 76 %, et 74 % pour les médecins seuls)<sup>83</sup>. Ces résultats peuvent s'expliquer par un manque de confiance des médecins dans les conclusions de GPT4. Ils privilégient le diagnostic qu'ils ont fondé sur leur propre logique à celui établi par l'IA dont le raisonnement n'est pas explicité. L'impossibilité de comprendre les résultats d'une IA et de les contester peuvent donc être, entre autres, des facteurs empêchant la création d'une relation constructive entre l'IA et son utilisateur. De plus, prendre en compte les performances des modèles lors de leur utilisation par des médecins se révèle être nécessaire. En effet, les performances des modèles fondés sur des réseaux de neurones sont, à ce jour, supérieures à celles de modèles plus explicables. Mais dans la pratique, ils peuvent s'avérer moins performants, notamment à cause de la méfiance des utilisateurs résultant du manque d'information sur le raisonnement de l'IA et donc de l'impossibilité de vérifier si une hallucination est présente.

81 *Ibid.*

82 Longo, Luca, Mario Brcic, Federico Cabitza, Jaesik Choi, Roberto Confalonieri, Javier Del Ser, Riccardo Guidotti, et al. 2024. « Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions ». *Information Fusion* 106 (juin):1-22.  
<https://doi.org/10.1016/j.inffus.2024.102301>.

83 Goh, Ethan, Robert Gallo, Jason Hom, Eric Strong, Yingjie Weng, Hannah Kerman, Joséphine A. Cool, et al. 2024. « Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial ». *JAMA Network Open* 7 (10). <https://doi.org/10.1001/jamanetworkopen.2024.40969>.

Dans la pratique, les médecins semblent privilégier la performance des systèmes<sup>84</sup>. Le souci de compréhension et d'explication du fonctionnement de l'algorithme passerait après la performance prédictive. Si le modèle utilisé présente de très bons résultats dans les tâches qu'il doit réaliser, le besoin d'explicabilité serait moindre, y compris pour les applications les plus critiques<sup>85</sup>. La priorité est de savoir à quel point les prédictions ou les réponses sont justes. Ce qui importe c'est que l'IA réalise les tâches prescrites avec le moins d'erreurs possibles. Considérer des IA comme des « boîtes noires » ne serait donc pas un problème tant que les performances du modèle sont les meilleures.

Cette supériorité de l'exigence de performance par rapport à celle d'explicabilité se creuserait davantage une fois que les IA sont déployées dans les organisations. Les indicateurs de performances deviennent prioritaires face aux efforts d'explicabilité au point parfois d'entraver les travaux visant à gagner en compréhension<sup>86</sup>. Dans certains cas, un arbitrage peut devenir nécessaire entre l'amélioration des performances (au détriment de la compréhension) et le renforcement de l'explicabilité du système (pouvant réduire sa précision)<sup>87</sup>.

Cette substitution, dans la pratique, de l'exigence d'explicabilité à celle de performance introduit de nouvelles problématiques éthiques. Si un modèle n'est pas explicable, l'unique moyen de l'évaluer et de le comparer aux autres, repose sur les critères de performance. Les recherches se focalisent alors sur l'optimisation de problèmes ou de critères de réussite déjà identifiés, excluant ainsi l'investigation de nouveaux types de raisonnement<sup>88</sup>. De plus, comme évoqué précédemment avec l'entreprise Octopize, si l'objectif fixé n'est pas de rendre les IA plus explicables et que l'exigence qui prime est celle de la performance, des modèles peuvent être créés et modifiés dans le seul but d'augmenter la validité globale des résultats (au détriment de problématiques éthiques et notamment d'équité).

Enfin, un dernier élément concernant l'explicabilité doit être pris en compte : le destinataire des explications. En effet, même si à l'avenir les modèles de réseaux de neurones s'avéraient plus explicables, ils ne le seraient pas forcément pour n'importe qui. Une IA peut être rendue compréhensible et intelligible pour un développeur, un concepteur ou un expert du domaine mais ne pas l'être pour l'utilisateur final (professionnel de santé, patient, etc.) ou un régulateur. La forme et le niveau de détail attendus en termes d'explicabilité varient donc également en fonction de la personne à qui sont destinées les explications. En fonction du destinataire, un compromis entre une explication précise, potentiellement très complexe et une description plus appréhendable mais simplifiée et omettant des informations significatives, devra sans doute s'effectuer<sup>89</sup>.

84 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

85 *Ibid.*

86 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

87 Veerappa, Manjunatha, et Salvo Rinzivillo. 2023. « Special Theme : eXplainable AI ». *ERCIM*, n° 134, 8-37.

88 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

Donc, malgré l'absence d'une définition qui fasse l'unanimité, certains aspects de l'explicabilité font consensus, notamment son rôle majeur dans la garantie du caractère éthique des IA, d'autant plus dans les domaines à hauts risques tels que la santé. Cependant, aujourd'hui, nous sommes incapables techniquement d'expliquer le fonctionnement des IA connexionnistes et dans la pratique, le critère de performance semble être prioritaire. Au-delà même de rendre compréhensible les modèles d'IA, une transparence concernant les choix opérés lors de leur conception et la manière dont ils influencent et façonnent les processus de prise de décisions sont indispensables<sup>90</sup>. Finalement, est-il acceptable d'utiliser des IA non explicables en santé ou dans quelconque domaine à haut risque ? Dans le cas contraire, étant donné qu'il est aujourd'hui impossible d'expliquer les modèles fondés sur des réseaux de neurones, doit-on s'en priver malgré leurs performances très élevées ? Est-il possible de se soustraire à l'exigence d'explicabilité et si oui, dans quelles conditions ? Être informé des performances, des limites et des biais est-il suffisant pour en avoir un usage éthique ?

89 Grote, Thomas, et Philipp Berens. 2025. « On the Ethics of Algorithmic Decision-Making in Healthcare | Journal of Medical Ethics ». *Journal of Medical Ethics*, janvier. <https://jme.bmj.com/content/46/3/205.long>.

90 Veerappa, Manjunatha, et Salvo Rinzivillo. 2023. « Special Theme : eXplainable AI ». *ERCIM*, n° 134, 8-37.

#### d. Responsabilité des différents acteurs

L'arrivée des IA génératives s'accompagne également de questionnements concernant la responsabilité des différentes parties prenantes. Lorsqu'un problème est rencontré (mauvais diagnostic, fausses informations insérées dans un compte rendu médical, etc.), qui est responsable ? Quand le résultat de l'IA est erroné et que l'utilisateur n'a pas su le détecter, est-il le seul responsable ? Qui doit s'assurer que les informations sensibles ou personnelles ne sont accessibles qu'aux personnes autorisées ? Établir les responsabilités qui incombent à chaque acteur serait primordial pour intégrer des IA, en particulier dans le domaine de la santé<sup>91</sup>. Or, ces questions se complexifient dans le cas des IA Gen. En effet, lorsqu'une IA est explicabile, il est plus facile de prendre en compte de manière éclairée sa réponse. Dans le cas contraire, il faut faire confiance au modèle, croire en la réponse donnée, sans pouvoir la comprendre. Revient-il aux concepteurs d'assurer des performances très élevées permettant aux utilisateurs d'avoir confiance ? Ou revient-il aux utilisateurs de vérifier par leurs propres moyens chaque réponse donnée par l'IA ?

Le faible niveau d'explicabilité, la différence (parfois très significative) des réponses alors que la demande n'a que légèrement varié et l'apprentissage dynamique, font des IA Gen, des modèles peu prévisibles et peu contrôlables. Cette complexité technologique complique l'attribution des responsabilités<sup>92</sup>. La tendance actuelle des concepteurs d'IA Gen tout domaines confondus est de se dédouaner de toute responsabilité concernant les résultats donnés par l'IA. L'impossibilité d'expliquer leur modèle et la complexité technique justifieraient qu'ils ne soient pas responsables des réponses données par l'IA et de ce que les utilisateurs en font. Il reviendrait donc à l'utilisateur de savoir bien formuler sa demande et surtout de savoir interpréter et de prendre le temps de vérifier les résultats qu'il obtient. Il est donc jugé responsable de son utilisation de l'IA. Si les résultats entraînent des erreurs, des décisions discutables ou quelconque impact négatif, c'est lui qui devrait y répondre et non pas l'IA en elle-même ou ses concepteurs.

Pour remédier à ces problèmes, la responsabilité (ou tout du moins une partie) pourrait être remise entre les mains des personnes et des organisations qui développent ces technologies. Rendre les concepteurs responsables en cas de conséquence négative permettrait d'aboutir à des IA plus justes<sup>93</sup>. Cependant, cela pourrait inciter les médecins à ne prendre aucune responsabilité et à suivre les décisions de l'IA même s'ils avaient diagnostiqué une autre maladie. Un point sous-jacent serait lié aux jeunes médecins qui, par facilité ou par manque de confiance, utiliseraient excessivement l'IA et n'obtiendraient pas l'expérience du métier. De plus, la prise de responsabilité des développeurs d'IA serait une forme d'exclusion des médecins des soins des patients et nous pourrions envisager un début d'autonomie de l'IA en santé.

91 Nordlinger, Bernard, Olivier De Fresnoye, et Claude Kirchner. 2024. « Systèmes d'IA générative en santé : enjeux et perspectives ». *Académie Nationale de Médecine* 208 (5): 536-547.

92 Yeung, Karen. 2019. « Étude sur les incidences des technologies numériques avancées (dont l'intelligence artificielle) sur la notion de responsabilité, sous l'angle des droits humains ». *Responsibility and AI*, 1-100.

93 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

La protection des données personnelles est également souvent transférée aux utilisateurs. Dans un guide de bonne pratique non spécifique à la santé sur l'intelligence artificielle générative publié par le gouvernement québécois, il est stipulé que « le personnel demeure imputable de l'utilisation qu'il fait des données de son organisation<sup>94</sup> » et la première bonne pratique individuelle serait de « ne jamais divulguer d'informations personnelles, confidentielles ou sensibles<sup>95</sup> ». Est-il envisageable de suivre les mêmes directives dans le domaine de la santé ?

Tout d'abord, il faut définir ce que l'on entend par information « sensible » ou « personnelle ». Selon la CNIL, une donnée personnelle correspond à « toute information se rapportant à une personne physique identifiée ou identifiable<sup>96</sup> ». Il est précisé qu'une personne peut être identifiable à partir d'une seule donnée, mais aussi de la conjonction de plusieurs données. De même, selon la CNIL, les données sensibles « forment une catégorie particulière des données personnelles<sup>97</sup> » et comprennent entre autres « le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé<sup>98</sup> ».

Interdire la divulgation de données personnelles telles que l'âge, le sexe, les antécédents ou des particularités significatives (sportif de haut niveau, etc.) d'un patient peut nuire au bon fonctionnement d'une IA dédiée à la santé (pour établir un diagnostic par exemple). Les systèmes d'IA peuvent donc requérir le traitement, voire la conservation de données personnelles, bien qu'elles puissent permettre l'identification du patient ou de l'utilisateur<sup>99</sup>. Établir les modalités de récupération, de traitement et de conservation des données est donc essentiel pour respecter le Règlement Général de Protection des Données (RGPD)<sup>100</sup> et en déduire les responsabilités de chacune des parties.

La notion de responsabilité, notamment du personnel soignant pouvant être amené à se servir de ce type d'outil, doit également être pensée selon plusieurs points de vue dans la pratique. Aujourd'hui, la responsabilité de l'utilisation de l'IA est généralement à la charge des utilisateurs. Or, si une personne est jugée garante de l'interprétation des résultats et de l'influence de ces derniers sur les décisions qu'elle prend par la suite, elle peut être tenue responsable d'avoir suivi les réponses de l'IA, de ne pas avoir pris de recul. Mais, il peut aussi lui être reproché d'avoir pris une direction trop différente de celle suggérée<sup>101</sup>. Donc, elle est responsable de s'être fiée à l'IA, mais aussi de ne pas s'y être fiée, de ne pas avoir suivi les indications fournies par le modèle.

94 Ministère de la Cybersécurité et du Numérique du Québec. 2024. « Guide des bonnes pratiques d'utilisation de l'intelligence artificielle générative », 1-25. p.17.

95 *Ibid.* p.23.

96 « Donnée personnelle ». 2025. CNIL. 5 janvier 2025. <https://www.cnil.fr/fr/definition/donnee-personnelle>.

97 « Donnée sensible ». 2025. CNIL. 5 janvier 2025. <https://www.cnil.fr/fr/definition/donnee-sensible>.

98 *Ibid.*

99 « IA : comment être en conformité avec le RGPD ? » 2025. CNIL. 5 janvier 2025.

<https://www.cnil.fr/fr/intelligence-artificielle/ia-comment-etre-en-conformite-avec-le-rgpd>.

100 *Ibid.*

101 Nordlinger, Bernard, Olivier De Fresnoye, et Claude Kirchner. 2024. « Systèmes d'IA générative en santé : enjeux et perspectives ». *Académie Nationale de Médecine* 208 (5): 536-547.

De plus, les utilisateurs doivent savoir dans quels contextes il est possible d'avoir recours à une IA, quelles sont les conditions pour l'employer de manière bénéfique et connaître les limites du modèle<sup>102</sup>. Ils peuvent par exemple être considérés responsables de ne pas l'avoir employé pour les applications prévues ou de ne pas lui avoir fourni toutes les informations nécessaires pour obtenir des réponses justes.

Cet engagement de la responsabilité individuelle des membres du personnel médical peut être à l'origine du rejet de l'utilisation des IA Gen. Il peut dissuader les praticiens d'utiliser ce type d'outils, que ce soit par la charge morale trop importante, par manque de temps pour vérifier les réponses données, par la difficulté de respecter toutes les règles et recommandations, etc. Lors d'un entretien, une personne chargée de développer des parcours innovants pour un établissement permettant le soin à domicile des patients, nous a confié être réticente à l'idée d'implanter ce type d'outils dans leur système. L'établissement propose aujourd'hui une plateforme grâce à laquelle les patients peuvent converser avec un infirmier afin de demander des renseignements sur le traitement, des symptômes, etc. Or, l'infirmier peut mettre jusqu'à sept heures pour apporter une réponse. Mais, malgré la possibilité que le patient obtienne une réponse instantanée avec un chatbot, l'implantation de ce type d'outil n'est pas envisagée. La gravité des conséquences pour les patients si l'IA ne donne pas une réponse juste est trop grande. Ainsi, le manque de confiance en la fiabilité des résultats introduit une responsabilité morale trop importante pour imaginer en utiliser.

L'absence d'expert lors d'utilisations de chatbots par les patients ouvre également de nouveaux questionnements quant aux responsabilités qui incombent aux utilisateurs. Lorsque l'IA fournit une réponse erronée, qui est responsable ? L'entreprise n'ayant pas développé un système assez performant ? Le patient pour avoir suivi les indications de l'IA sans avoir consulté un professionnel ? Le cas échéant, l'organisme ayant proposé ce type d'outil ?

Pour terminer, la question de la responsabilité ne doit pas nous détourner de nos objectifs. Définir un responsable est généralement considéré comme une manière d'assurer les principes éthiques, d'inciter à respecter les normes et les valeurs établies. En effet, assigner les charges qui incombent à chacun permet d'établir la personne devant être à l'origine des indemnités en cas de problème<sup>103</sup>. La possibilité que « le responsable » soit puni devrait le dissuader d'agir en dehors des règles établies. Cependant, mettre en place des sanctions permet de légitimer certains comportements ou certaines situations et n'incite pas forcément à changer la situation. Par exemple, dans le cas d'une fuite de données personnelles, demander le dédommagement des personnes concernées, sans qu'aucune autre mesure visant à éviter que cela se reproduise soit prise, invisibilise le problème de fond. Ce qui importe c'est que les données ne fuient pas, pas de savoir qui devra verser la compensation financière.

102 OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

Petitprez, Eugénie. 2024. *Les enjeux liés à l'utilisation de l'intelligence artificielle en matière de santé : regards croisés*. Ceprisca. Colloques.

103 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

Pour conclure, définir les responsabilités de chacun permettrait de prendre en compte les conséquences potentielles sur les personnes affectées par l'utilisation d'IA. Cependant, une grande partie des responsabilités en cas de problème repose aujourd'hui sur les utilisateurs. La notion de responsabilité est souvent liée à celle de causalité, l'auteur des responsables<sup>104</sup>. Or, la diffusion de modèles non explicables et en particulier génératif obscurcit davantage le lien entre une situation donnée et la personne qui est à son origine. Enfin, déterminer les responsabilités qui incombent à chacun et les sanctions ne doit pas freiner la gestion des problèmes en amont.

### e. Gouvernance des données et des modèles

Il est important de s'intéresser plus largement à la question de la gestion des données, personnelles ou non, entrées lors de l'utilisation de l'IA. Sont-elles stockées ? Si oui, par qui et où ? Sous quelles modalités ? Quelles sont les personnes qui y ont accès ? Pour quelles raisons sont-elles stockées ? Ces questions précèdent d'ailleurs la question de la responsabilité et de la confidentialité des informations personnelles. En effet, si les données ne sont pas sauvegardées, c'est-à-dire directement supprimées après l'utilisation, les modalités de protection des données sont toutes autres.

Cependant, les informations entrées sont généralement conservées par le modèle, afin par exemple de permettre un suivi d'un utilisateur sur le long terme ou de participer à l'amélioration continue de l'IA. Se questionner sur l'hébergement des données est d'autant plus important que ce marché est dominé par les GAFAM. Si les données proviennent d'un pays en Europe mais qu'elles sont stockées dans des serveurs aux États-Unis, sont-elles tout de même protégées par le droit européen ? La réponse n'est pas évidente, puisque ce sera l'extraterritorialité des droits américains et européens qui sera mise en jeu. Il s'agit d'une question de rapport de force, il est donc toujours possible que ces données ne soient pas protégées par le droit européen.

La gestion des données et plus généralement l'établissement des règles s'appliquant aux concepteurs d'IA Gen sont d'autant plus mis à l'épreuve par l'arrivée de modèles d'IA Gen open source. Ces modèles tels que LLaMA (de Meta) ou ceux développés par Mistral AI, introduisent de nouvelles problématiques de gouvernance. Ils sont accessibles et téléchargeables par quiconque, ce qui peut limiter les possibilités de régulations. Cette mise à disposition favorise la généralisation de l'utilisation des modèles d'IA développés par les entreprises les plus puissantes. Ils sont employés tels quels ou servent de base pour développer des IA spécialisées dans certains domaines ou pour des usages précis. La possibilité que la majorité des modèles d'IA Gen se basent seulement sur quelques modèles phares accentue les problématiques de gouvernance et d'éthique. Comment contrôler et réguler les modèles d'IA en open source ? Quelles règles s'appliquent, celles du pays développeur ou du pays dans lequel l'IA est utilisée ? Quelle place les grandes entreprises développant ses modèles accordent-elles à l'éthique ?

104 *Ibid.*

Par exemple, certaines entreprises américaines de ce domaine se sont engagées auprès des autorités à identifier les meilleures pratiques pour un développement et un déploiement responsable de gros modèles d'IA. Elles doivent aussi collaborer avec les autorités réglementaires pour partager les connaissances concernant la confiance et les risques associés à l'utilisation de l'IA. Cependant, l'OMS note un écartement ou une expulsion des équipes chargées de l'éthique des IA de ces entreprises pour éviter un ralentissement du développement de certains types d'IA. L'OMS explique dans un rapport : « *Ethics teams in one company, for example, that had recommended halting the release of a new LMM changed their documents and downplayed previously documented risks*<sup>105</sup>. »

Dans son guide sur les *Large Multi-modal Model* (LMM), l'OMS souligne plusieurs problématiques liées aux modèles open source<sup>106</sup>. L'intégration des principes éthiques et l'amélioration de la gestion des risques qui serait permise par les modèles open source reste incertaine. Ces derniers permettraient en théorie de favoriser la transparence et la performance grâce à la contribution, aux retours des utilisateurs ainsi qu'à l'accès au code permettant une évaluation critique. Mais, la surveillance des données d'entraînement, des performances et des résultats reste difficile. L'entreprise Meta propose par exemple des licences pour accéder à un modèle open source mais avec des conditions restrictives. Elle ne met donc pas entièrement son modèle à disposition, non conformément aux principes open source<sup>107</sup>. Ces modèles sont aussi plus vulnérables et peuvent faciliter le contournement des mesures de sécurité. L'OMS encourage donc les gouvernements à investir davantage dans des LMM afin de pouvoir exiger que ces derniers soient largement accessibles et de permettre une supervision publique car aujourd'hui, la majorité des investissements se font dans le secteur privé<sup>108</sup>. Pour certains, les modèles open source devraient même être interdits afin de contrôler finement le développement des IA génératives et de limiter leur développement à des acteurs réglementés<sup>109</sup>.

L'OMS préconise également la discussion et le développement de règles concernant les IA au niveau mondial<sup>110</sup>. Une gouvernance internationale permettrait de garantir le respect de normes de sécurité et d'efficacité minimales, mais aussi de mettre en place des réglementations assurant le respect des principes éthiques. Cette gouvernance ne doit pas seulement être mise en place par les pays les plus puissants, elle doit inclure tous les pays.

105 OMS. 2024. « *Ethics and Governance of Artificial Intelligence for Health: Large Multi-Modal Models* ». *WHO Guidance*. 1st ed. Geneva: World Health Organization.

106 OMS. 2024. « *Ethics and Governance of Artificial Intelligence for Health: Large Multi-Modal Models*. » *WHO Guidance*. 1st ed. Geneva: World Health Organization.

107 Maffulli, Stefano. 2023. « Meta's LLaMa 2 License Is Not Open Source ». Open Source Initiative. 20 juillet 2023. <https://opensource.org/blog/metals-llama-2-license-is-not-open-source>.

108 OMS. 2024. « *Ethics and Governance of Artificial Intelligence for Health: Large Multi-Modal Models* ». *WHO Guidance*. 1st ed. Geneva: World Health Organization.

109 Cusumano, Michael A. « Generative AI as a New Innovation Platform ». *Communications of the ACM* 66, n° 10: 18-21. <https://doi.org/10.1145/3615859>.

110 OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

Par ailleurs, l'insertion (ou non) des IA Gen dans le domaine médical ne doit pas se faire sans assurer la souveraineté du personnel médical et de la population en général. Garantir que les décisions prises quant à l'introduction d'outils d'IA Gen en santé et aux modalités d'implantation sont discutées avec les personnes concernées, est primordial. Or, la situation actuelle n'offre pas les conditions nécessaires (en particulier du temps disponible) pour que le personnel soignant aborde ces questions.

## f. Impacts environnementaux

Un dernier enjeu majeur lié au développement et au déploiement de l'IA Gen, dans quelque domaine que ce soit, est l'impact environnemental. En effet, garantir l'éthique d'un modèle d'IA, c'est-à-dire, s'assurer qu'il n'est pas nuisible pour les êtres humains, mais aussi, pour tout être vivant, passe par la prise en compte des enjeux écologiques tout au long de sa vie<sup>111</sup>.

Tout d'abord, les modèles génératifs demandent une énorme puissance de calcul qui requière elle-même une très grande quantité d'énergie. L'apprentissage mais aussi l'utilisation des modèles d'IA Gen sont très énergivores et rejettent des gaz à effet de serre dans l'atmosphère. Les émissions de ces derniers sont proportionnelles aux besoins en termes de capacité de calcul des modèles et donc du nombre de paramètres composant ces algorithmes<sup>112</sup>. Bien qu'en fonction de la taille et du type de modèle d'IA Gen, l'empreinte carbone varie, ces IA accentuent significativement la demande en énergie dédiée aux data centers et donc les émissions de CO<sub>2</sub><sup>113</sup>. Par exemple, une requête sur ChatGPT génère en moyenne 0,27 kilogramme de dioxyde de carbone dans l'atmosphère<sup>114</sup>. Si une personne effectue plus de dix interactions par jour, passé un an, près d'une tonne de CO<sub>2</sub> aura été émise, soit la moitié de ce qu'une personne devrait dégager pour suivre les Accords de Paris.

111 Smuha, Nathalie. 2019. « Lignes directrices en matière d'éthique pour une IA digne de confiance ». [ethics\\_guidelines\\_for\\_trustworthy\\_ai-fr\\_87FE7A3C-D03D-9305-81A653DDA84B5A60\\_60427.pdf](#). et Nordlinger, Bernard, Olivier De Fresnoye, et Claude Kirchner. 2024. « Systèmes d'IA générative en santé : enjeux et perspectives ». *Académie Nationale de Médecine* 208 (5): 536-547.

112 Le Goff, Thomas. 2023. « Recommandations pour une action publique en faveur d'une IA générative respectueuse de l'environnement ». <https://hal.science/hal-04371031v2>.

113 *Ibid.*

114 « On explore les impacts environnementaux de ChatGPT. Françaises, français, mes chers compatriotes. » 2024. <https://www.radiofrance.fr/franceinter/podcasts/la-terre-au-carre/la-terre-au-carre-l-actu-du-mercredi-27-novembre-2024-5584893>.

Outre la forte consommation énergétique de ces modèles, leur création et leur implantation requièrent l'exploitation de nombreuses ressources naturelles. La phase d'entraînement implique une forte consommation d'eau pour le refroidissement des data centers, pouvant atteindre plusieurs centaines de milliers de litres (pour GPT3)<sup>115</sup>. De plus, le développement de ces outils augmente le besoin en infrastructures et en matériaux, notamment en matériaux rares tels que le silicium, le cobalt, le lithium ou encore le cuivre. Ces derniers sont difficiles à extraire et à raffiner, leur exploitation s'accompagne généralement de l'artificialisation des sols, de la pollution de l'eau et des terres ou encore de la déforestation. Tout le secteur du numérique est confronté à ces problématiques, mais l'arrivée de l'IA Gen modifie grandement le besoin et fait passer ces dernières à l'échelle supérieure<sup>116</sup>.

Ainsi, l'impact environnemental du développement et du déploiement des IA Gen ne peut pas être écarté des enjeux éthiques. L'entraînement, l'insertion et l'emploi de celles-ci augmentent considérablement les émissions de gaz à effet de serre et l'exploitation des ressources naturelles. Inciter ou favoriser l'utilisation systématique de ce type d'outil, en les intégrant par exemple dans les pratiques médicales ordinaires, accroîtrait fortement les conséquences des activités humaines sur la nature, au détriment de la préservation de conditions de vie favorables sur Terre. Même si les modèles génératifs s'avéraient très performants et offraient la possibilité d'améliorer significativement la prise en charge médicale, est-elle tout de même souhaitable ? Peut-elle être envisagée pour tous les pays ? Les potentielles retombées positives à court terme des IA Gen ne doivent pas faire oublier les conséquences à moyen ou long terme de la pollution ou de l'endommagement de la biodiversité. Qui bénéficiera de ces outils ? Qui sera impacté par les conséquences, en particulier de dégradation de l'environnement ? Sous quelles conditions pourra-t-on considérer que l'IA Gen est davantage bénéfique que délétère pour l'humanité ?

115 Le Goff, Thomas. 2023. « Recommandations pour une action publique en faveur d'une IA générative respectueuse de l'environnement ». <https://hal.science/hal-04371031v2>.

116 *Ibid.*

## 5. Possibles reconfigurations du système médical

L'une des premières peurs face à l'introduction de l'IA, y compris de l'IA Gen, dans un secteur donné, est liée à l'idée que ces outils remplaceraient les professionnels. Cependant, de nombreux exemples historiques tels que la machine à vapeur, les ordinateurs ou encore internet, démentent la disparition complète d'emplois<sup>117</sup>. L'insertion de nouveaux dispositifs techniques se traduit plutôt par une transformation, une reconfiguration plus ou moins profonde et brutale des métiers. L'environnement et les conditions de travail changent, les connaissances et les savoir-faire se transforment mais il ne s'agit pas d'une substitution ou d'une destruction de métiers. Quelles sont les reconfigurations que pourrait subir le secteur médical avec l'insertion d'IA Gen ?

L'IA change l'environnement dans lequel elle se déploie et peut entraîner une transformation des métiers qui étaient présents. Ces modifications sont liées aux fonctions de l'IA et aux objectifs visés lors de leur développement. L'argument temporel avancé par une grande partie des prestataires d'IA Gen est un point central dans son implantation en santé. Celui-ci est associé au manque de personnels qui entraîne une dégradation des relations entre les patients et les soignants et une baisse du bien-être au travail des employés médicaux. Peu d'études sont portées sur l'évaluation de la réduction du temps passé par les médecins sur la synthèse du rapport des consultations, mais si cela s'avère être le cas, ce temps gagné augmentera-t-il la durée moyenne des consultations ? Où sera réinvesti le temps gagné ? Permettra-t-il une meilleure qualité de prise en charge des patients ? Cette durée n'est pas fixe et peut varier selon la volonté du médecin ou de la clinique<sup>118</sup>. Effectivement, la quantité de patients dans le besoin d'un suivi ou d'une consultation favorise l'augmentation de la cadence de travail des médecins. Cependant, malgré l'augmentation de la cadence, les agendas des professionnels de santé restent surchargés. L'arrivée de l'IA et par conséquent d'un nouveau « temps libre » du médecin, peut devenir une opportunité pour les soignants de réduire le nombre de jours moyens pour obtenir une consultation avec eux<sup>119</sup>. Cette supposition de la modification de la quantité et non de la qualité des soins est soutenue par la gestion managériale des hôpitaux<sup>120</sup>. En effet, le déficit constant de ces organismes et leur manque de personnel, les obligent à fermer des lits qui pourraient être réouverts si chaque soignant pouvait s'occuper de plus de patients en même temps. La libéralisation du secteur de la santé alimente encore plus la disposition des organismes à utiliser des outils améliorant l'efficacité des employés<sup>121</sup>.

L'idée générale voyant le cœur du métier des médecins comme une suite de consultations et d'élaborations de diagnostic, présente l'IA Gen comme une opportunité de

117 Bettache, Mustapha, et Laurie Foisy. [Bettache, Mustapha, et Laurie Foisy](#). 2019. « Intelligence artificielle et transformation des emplois ». *Question(s) de management* ? 25 (3): 61-67.

<https://doi.org/10.3917/qdm.193.0061>

118 Grand Remue-Ménages, Fondation Université Paris Cité, 12 rue de L'École de Médecine, 28/11/2024.

119 *Ibid.*

120 Grimaldi, André. 2013. « Les hôpitaux en quête de rentabilité : à quel prix ? » *Revue du MAUSS* 41 (1): 25-33. <https://doi.org/10.3917/rdm.041.0025>.

121 Mignot, Léo, et Émilien Schultz. 2022. « Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé ». *Réseaux : communication, technologie, société*, Contrôler l'intelligence artificielle ?, 232-233 : 65-97. <https://doi.org/10.3917/res.232.0065>.

recentrer ce métier et d'écarter toutes les tâches administratives obligatoires. Pourquoi la partie administrative est perçue comme une tâche annexe pour les métiers médicaux ? Quel peut être l'impact de la suppression de ces tâches sur les activités du métier considérées comme les plus nobles ? Les compétences médicales sont rares et très estimées par la population, encore plus depuis la crise sanitaire de 2020. À l'inverse, la rédaction de documents et la partie administrative de toutes les activités est perçue comme une perte de temps et un travail ennuyeux. C'est pourquoi la réalisation d'un diagnostic est souvent considéré comme une activité à forte valeur ajoutée alors que celle de rédaction de comptes rendus serait d'une faible plus-value<sup>122</sup>. Cependant, cette partie rédactionnelle annexe ne fait-elle pas partie intégrante d'un diagnostic ? La rédaction des symptômes, la mise en forme et le lien pouvant être fait entre ceux-ci peuvent permettre au médecin d'identifier les possibles maladies dont peut souffrir un patient. Ce temps de rédaction n'est pas seulement composé du temps de frappe à l'ordinateur. La suppression ou la forte réduction de ce temps pourrait impacter la qualité du diagnostic fourni par le médecin. En effet, une relecture et un changement rapide au clavier ou à l'oral des informations présentes ne permet pas une aussi bonne intégration que la rédaction des informations.

De plus, certaines IA commencent à diagnostiquer des patients à partir de leurs symptômes, de leurs images médicales, etc. Elles commencent à pratiquer des activités considérées à forte valeur ajoutée. Tel est le cas des IA d'analyses d'imageries médicales qui couplées aux IA Gen produisent des comptes rendus sans aide humaine<sup>123</sup>. L'utilisation de ces technologies relève d'une contradiction forte. Les radiologues payent un service qui tend à remplacer leurs principales activités et les entreprises doivent se démarquer de leurs concurrentes tout en préservant et séduisant leurs clients. Cela entraîne une réappropriation de l'outil par les praticiens qui souhaitent conserver et améliorer leurs soins. Le corps des radiologues délimite les applications de l'IA en introduisant des distinctions dans leurs diagnostics<sup>124</sup>. L'IA serait donc utilisée pour la détection de fractures qui est considérée comme une tâche simple et peu intéressante par les radiologues. La supervision humaine reste tout de même en place, mais ces derniers peuvent se consacrer aux tâches qu'ils considèrent les plus captivantes de leur métier. L'implantation de l'IA Gen dans les différents métiers que nous avons pu évoquer pourra potentiellement se faire de cette façon. La régulation de l'IA peut se faire par les praticiens eux-mêmes qui acceptent ou non de payer un service proposé par des entreprises du secteur privé. Cependant, ce type de régulation peut se faire seulement si ce sont les praticiens qui décident de façon éclairée l'implantation ou non des outils informatiques dans leur service ou leur cabinet. En effet, reprenons le cas des États Unis, ce sont les organismes privés qui signent les contrats pour l'installation des différents outils dans leurs cliniques. Les praticiens auraient donc l'obligation d'utiliser des outils qu'ils ne souhaiteraient pas réellement utiliser.

122 Ministère de la Cybersécurité et du Numérique du Québec. 2024. « Guide des bonnes pratiques d'utilisation de l'intelligence artificielle générative », 1-25. et Douetteau, Florian. 2024. « Pourquoi l'IA générative va révolutionner le rapport au travail ? » *Administration* 281 (1): 70-72. <https://doi.org/10.3917/admi.281.0070>.

123 Mignot, Léo, et Émilien Schultz. 2022. « Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé ». *Réseaux : communication, technologie, société*, Contrôler l'intelligence artificielle ?, 232-233 : 65-97. <https://doi.org/10.3917/res.232.0065>.

124 *Ibid.*

Même dans le cas où l'utilisation de l'IA par les professionnels n'est pas rendue obligatoire, son insertion ne résulte pas forcément d'une volonté particulière ou d'un besoin auquel elle serait l'unique solution. En effet, lors d'un entretien, une employée de l'équipe d'innovation d'un CHU, nous a confié que son équipe envisageait d'implanter un système d'IA Gen suite à la proposition d'un patient. Ce dernier, travaillant dans le domaine, aurait soumis l'idée de gérer la rédaction de compte-rendu à l'aide d'une IA Gen. Ainsi, l'engouement général autour de l'IA Gen peut inciter de nouvelles personnes à les utiliser alors qu'elles n'avaient pas imaginé y avoir recours.

Les IA modifient l'environnement dans lequel elles se trouvent. Que ce soient les infrastructures nécessaires à leur implantation, les protocoles ou la connaissance de la présence de l'IA chez le patient ou le soignant, le secteur s'en trouve largement modifié<sup>125</sup>. L'emploi de l'IA passe obligatoirement par la numérisation de chaque étape du parcours du patient. Lors d'une consultation, le patient et le soignant ne sont pas seuls, la présence du matériel médical, d'un écran ou d'un enregistrement audio (comme avec l'IA de Nabla ou de Doctolib) influence leur relation. L'écran, déjà intégré dans de nombreuses consultations, met de la distance entre le patient et le professionnel de santé<sup>126</sup>. L'enregistrement audio, quant à lui, pourrait mettre mal à l'aise un patient. Savoir que la consultation est enregistrée, même si cela reste confidentiel, peut modifier la manière dont il s'exprime. Le patient qui se sait être écouté, ou le médecin qui veut augmenter l'efficacité de l'outil en déclarant oralement le plus d'informations possibles pour ne pas avoir à les écrire à la fin de la consultation, changent leurs interactions. Cela peut entraîner une confusion et une peur pour le patient qui ne comprend pas le vocabulaire médical employé. La personnification de l'IA par la population peut être perçue comme l'arrivée d'une troisième personne dans la salle de consultation qui serait l'outil informatique. Cependant, cet outil peut être perçu par le patient comme un inconnu et influencer sa manière de parler, comme s'il communiquait à une personne qu'il n'avait jamais vue. Il peut ne pas exprimer toutes les informations qu'il aurait dites au médecin seul ou les formuler différemment. Le temps gagné par le médecin avec l'IA sur la partie administrative pourrait être compensé par un temps plus long pour mettre à l'aise le patient pour qu'il puisse parler librement. La présence d'enregistrement audio pourrait tout de même permettre un recul de l'ordinateur du médecin qui est une barrière physique entre les deux parties<sup>127</sup>.

Le développement d'IA Gen à disposition du grand public rend l'accès à l'information encore plus simple. Un patient peut se renseigner sur ses symptômes ou sa maladie avant d'aller voir un médecin. L'IA fournit une explication claire, parfois plus compréhensible que celle d'un médecin, de toutes les maladies et du parcours des patients dans le cas d'opérations comme pour une coloscopie<sup>128</sup>, une pose de prothèse de genou, etc. Lorsque le patient se présente devant un médecin, il peut s'être renseigné ou non. Le médecin doit donc s'adapter à

125 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

126 Grand Remue-Ménages, Fondation Université Paris Cité, 12 rue de L'École de Médecine, 28/11/2024.  
127 *Ibid.*

128 Lee, Tsung-Chun, Kyle Staller, Vlaicu Botoman, Mythili P. Pathipati, Sanskriti Varma, et Braden Kuo. 2023. « ChatGPT Answers Common Patient Questions About Colonoscopy ». *Gastroenterology* 165 (2): 509-511. <https://doi.org/10.1053/j.gastro.2023.04.033>.

une plus grande diversité de profils sans savoir lors de ses échanges à quel point le patient est informé. De plus, le patient peut plus facilement remettre en cause la décision et le diagnostic de son médecin. Le soignant doit alors se justifier quant à ses décisions. Cette justification peut destinée au patient, mais peut aussi intervenir, comme nous l'avons vu précédemment, en cas de problème, si le médecin a suivi ou n'a pas suivi le diagnostic de l'IA. Elle change donc les rapports de force entre les différents acteurs du milieu.

Des entreprises, comme Kiro, développent des IA Gen dans le but de vulgariser la médecine et de la rendre accessible pour tous. Kiro propose une interface web pour les médecins, les laboratoires d'analyses médicales et les patients dans le but de centraliser les comptes rendus d'analyses. Le médecin a donc accès numériquement aux résultats et une IA Gen synthétise les résultats en fonction des caractéristiques du patient entrées par le laboratoire (femmes enceintes, pathologie particulière, etc.). De la même façon, le patient peut avoir accès à ses résultats immédiatement après leur publication. L'IA Gen permet d'expliquer et de rendre compréhensible tous les chiffres présents sur la feuille d'analyse. Elle prend en compte les caractéristiques de la personne et le contexte pour lui indiquer l'intervalle dans lequel elle devrait être. Mais où positionner la limite entre ce qui peut être annoncé directement au patient et ce qui ne peut pas l'être<sup>129</sup> ? Lorsqu'un patient passe une IRM ou un scanner, ce doit être le médecin généraliste qui communique les résultats. Si le résultat des images ou des analyses médicales indique une maladie grave, l'annonce au patient doit se faire de manière à ce que, tout de suite après, une solution, un traitement ou une explication de la suite des événements soit donné. En effet, chaque patient réagit différemment, la confirmation d'une pathologie sévère peut être source de stress, d'angoisse, etc. L'accompagnement par une personne formée est nécessaire afin de le soutenir et lui expliquer les prochaines étapes de son parcours.

Lors de l'entretien avec l'employée de l'équipe d'innovation du CHU, cette dernière nous a partagé leur projet d'implantation d'IA Gen dans leur travail. Leur métier consiste à faire le lien entre les entreprises de développement des dispositifs médicaux (DM) du secteur privé et les professionnels de santé présents au CHU. Les entreprises les consultent pour leur exposer un prototype ou une idée et son équipe le fait tester par les potentiels futurs utilisateurs. Par exemple, si une entreprise de développement de robots chirurgicaux les contacte, ils mettent en scène une opération pour tester l'outil dans des conditions les plus proches de la réalité. Un protocole d'évaluation est préparé à partir de protocoles types en fonction du DM et de l'avancée du projet, c'est-à-dire, si c'est un premier prototype ou la phase finale de R&D. Lors du test dans la salle d'opération, un membre du pôle d'innovation est présent pour guider les expériences et note tout ce qu'il observe et toutes les remarques du chirurgien. Ce protocole est réalisé plusieurs fois avec des équipes différentes (chirurgien/membre de l'équipe d'innovation). Ensuite, à partir des notes des différents membres, un compte rendu général est rédigé et transmis à l'entreprise.

129 Morse, Keith E., Nicolai P. Ostberg, Veena G. Jones, et Albert S. Chan. 2020. « Use Characteristics and Triage Acuity of a Digital Symptom Checker in a Large Integrated Health System: Population-Based Descriptive Study ». *Journal of Medical Internet Research* 22 (11). <https://doi.org/10.2196/20549>.

L'idée apportée par une personne extérieure à leur équipe (et extérieure au domaine médical) est de déléguer la rédaction du compte rendu général et du protocole à une IA Gen. Cette implantation soulève de nombreux problèmes. Tout d'abord, concernant la rédaction du protocole, n'entraînerait-elle pas une perte de savoir-faire ? Les résultats de ces tests découlent majoritairement du protocole, c'est un point très important du processus qui peut demander une grande expérience du métier, savoir sur quels points passer le plus de temps, quels aspects introduisent des défauts, etc. Même si la relecture humaine reste présente, l'implantation de l'IA modifiera l'approche du protocole et l'orientera de manière irréversible par rapport à la création d'un protocole entièrement à la main. De plus, le temps passé en moins sur cette partie du processus n'est pas forcément du temps gagné. En effet, ne pas connaître le sujet et le découvrir en lisant le protocole rend les tests en conditions réelles moins fluides et potentiellement moins pertinents.

De la même façon, la rédaction du compte rendu général par une IA Gen pose des problèmes. Pour former la base d'apprentissage, une homogénéisation et une standardisation des rapports doivent être réalisées. La qualité des rendus y est très liée. Pour le moment, chaque employé a une manière d'organiser et de rédiger qui lui est propre et qui dépend du DM, de l'avancement du prototype et des résultats. Trouver une organisation de compte rendu assez souple pour englober tous les cas est difficile, voire impossible. La qualité du compte rendu final peut en être dégradée et cette façon de rédiger peut ne pas convenir à certains employés.

Après cette phase d'entraînement, l'IA peut être utilisée. De nouveaux inconvénients doivent être pris en compte. La prise de note des employés lors des tests peut être faite sur papier ou sur tablette en fonction de la préférence de chacun. Afin que l'IA puisse écrire le compte rendu final, l'ensemble des notes doivent lui être transmises. Une réécriture informatique doit donc être réalisée après chaque test. L'utilisation de différentes couleurs, tailles de mot, symboles et abréviations, ne pouvant pas être comprise par l'IA, il faut retranscrire numériquement tous ces détails. Avec cet outil, les employés n'auraient plus besoin de rédiger les comptes rendus généraux, mais de nouvelles missions devront être réalisées et requièrent de modifier ce qu'ils faisaient déjà. Changer ses habitudes, s'adapter aux nouvelles tâches, etc. peut représenter une perte de temps, une charge mentale pour les employés et une potentielle perte de qualité du rapport final. Des problèmes de confidentialité peuvent aussi être soulevés dans ce domaine qui comporte des idées et des DM fraîchement brevetés et en cours d'amélioration.

Pour conclure, l'arrivée de l'IA Gen peut potentiellement reconfigurer le monde médical, les manières de travailler, les rapports de forces entre les différents intervenants, etc. Ces changements impliquent de redécouper individuellement et globalement les missions et les responsabilités de chacun. Les modifications apportées aboutiront probablement à une évolution des identités professionnelles des différents corps de métiers<sup>130</sup>. Les problèmes rencontrés possèdent généralement des bases communes, mais il est important de distinguer les cas d'utilisation pour déterminer les règles et lois à mettre en place. En effet, les risques rencontrés varient en fonction du domaine d'application et des données qui y sont traitées.

130 Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

En supposant que l'emploi de l'IA augmente de manière notable la qualité des soins, de nouvelles questions se poseront sur le droit d'accès à l'IA des patients et le droit d'exercer librement des médecins<sup>131</sup>. Est-il envisageable de rendre obligatoire l'utilisation de l'IA lors d'un diagnostic médical ? Si oui, comment le mettre en place ? Est-il judicieux de garder privé ce secteur ? L'assurance maladie peut-elle exiger l'emploi d'une IA qui n'augmente pas la qualité des soins, mais optimise le temps d'un professionnel (IA Gen pour compte rendu par exemple) ? Une solution possible serait d'appliquer le droit à l'IA l'IA peut être appliqué sans contraindre les médecins à l'utiliser. Cette obligation pourrait être portée par les hôpitaux, il serait envisageable de rendre indispensable la présence d'un médecin pouvant utiliser l'IA au sein de chaque hôpital. Ainsi, les médecins pourront choisir d'utiliser ou non une IA et le patient aura aussi le choix d'y avoir accès. En revanche, il reste indispensable d'informer le patient pour qu'il comprenne le rôle de l'IA dans son parcours de soin<sup>132</sup>.

131 Carpentier, Manu, réal. 2024. *Le JT de la Radiologie*. Healthtech Innovation Paris.

<https://www.jtradiologie.com/Video/view/<https://www.jtradiologie.com/Video/view/xs3mzvu>.

132 OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

## 6. Conclusion

L'IA Gen est un outil informatique très puissant pouvant avoir des applications dans de très nombreux domaines numérisés. La quantité importante de données produites par le milieu médical permet le développement et la création de ce type d'outils spécialisés dans la santé. De nombreux systèmes conversationnels dont le but est de faciliter l'accès à l'information des patients et des soignants, rédiger des comptes rendus médicaux ou encore fluidifier les consultations, sont en développement.

Ces IA sont majoritairement développées et proposées par des entreprises privées. Cette particularité influence le déploiement de ces outils et pose de nombreuses questions pratiques. Comment réglementer et contrôler ces IA ? Aujourd'hui ces vérifications sont chronophages et réalisées au cas par cas, est-ce envisageable à grande échelle ? Les infrastructures informatiques usagées des hôpitaux compliquent l'emploi des IA Gen qui demande de nombreux serveurs et des investissements conséquents.

La diversité et la complexité des IA Gen impliquent de posséder des connaissances et des savoir-faire pour les utiliser au mieux. Comment et quand les soignants doivent-ils être formés ? Ponctuellement, durant la formation initiale des soignants n'est peut-être pas suffisant au vu de l'évolution rapide des IA Gen. La formation continue par les prestataires d'IA Gen est envisagée mais elle doit être organisée et guidée par une organisation nationale ou internationale. Laisser un outil informatique automatiser des tâches entraîne une perte d'expérience voire de connaissance de la part des personnes qui les réalisaient. Or, le savoir médical est très important, est-il envisageable de le perdre ? Comment intégrer ces outils sans pour autant perdre des connaissances indispensables en retour ? L'expérience acquise avec l'arrivée des réseaux de neurones en imagerie médicale peut-elle aider à orienter le déploiement de l'IA Gen en santé ?

Les risques élevés intrinsèques à ce secteur rendent l'IA Gen et son utilisation controversée. En effet, un point clé des outils informatiques est l'explicabilité. Ce concept reprend la compréhension des décisions et résultats prisent par une IA. La quantité de paramètres des IA Gen la rend, à ce jour, inexplicable mais elle possède, dans de nombreux cas, les plus hautes performances du marché. Sa seule performance justifie-t-elle sa présence dans des domaines à hauts risques ? En pratique, un résultat inexplicable peut-il être utilisable et constructif par un médecin et dans un diagnostic ? En cas d'erreur, qui est responsable ? L'entreprise le développant ? Le soignant l'ayant utilisé ? Et dans le cas des dispositifs directement utilisés par les patients, est-ce à l'utilisateur que revient toutes les charges ? L'importance de ces questions ne doit pas entraver l'objectif premier qui est de réduire un maximum les risques et les erreurs de leurs utilisations.

Les financements nécessaires pour développer les IA Gen favorisent les multinationales. Pour ne pas ralentir ou entraver le développement de leurs produits, ces entreprises, pour la majorité américaine, n'écoutent pas ou écartent leurs équipes éthiques. Le déploiement de modèles open source peut répondre à certains problèmes éthiques mais augmente par la même occasion les risques de voir apparaître des utilisations d'IA à des fins répréhensibles pénalement. Les enjeux environnementaux, la consommation énergétique, en eau, etc. des serveurs doivent être pris en compte dans le déploiement à grande échelle de ces outils, ce qui est, à ce jour, très peu évoqué dans le domaine de la santé.

Les performances des IA sont très largement corrélées à la quantité et à la qualité des données utilisées. Ces données comportent des erreurs et des biais qui sont retranscrites dans le modèle. Lors du déploiement de celui-ci, il est nécessaire de les connaître et de les réduire afin que l'arrivée de l'IA ne creuse pas les inégalités et les disparités entre les personnes, mais à l'inverse les réduise. En santé, les données traitées sont personnelles et possiblement sensibles. Une attention particulière au stockage, à la divulgation, à l'utilisation, etc. est nécessaire et le consentement du patient est obligatoire. Certains algorithmes sont développés dans le but d'anonymiser les données récupérées, pour qu'à partir des informations synthétiques il soit impossible d'identifier la personne qui en est l'origine. Cette approche permet de n'avoir besoin de l'accord du patient que lors de la première étape d'anonymisation.

Enfin, la présence de l'IA Gen en santé pourrait reconfigurer un grand nombre de métiers qui y sont liés. De nouvelles tâches devront être accomplies, d'autres seront modifiées, le patient pourra plus facilement remettre en cause la parole du médecin avec un deuxième avis venant de l'IA Gen. Exceptés quelques métiers qui y sont déjà soumis, ces changements ne sont que très peu visibles à ce jour. Se questionner avant tout sur l'insertion ou non de l'IA en santé est donc essentiel. Puis, dans le cas où elle est souhaitée, sa place et les modalités de son utilisation devront être pensées pour garantir une médecine sûre et désirée.

## Bibliographie

Bastian-Pétrél, Kathleen, Jessica L. Rohmann, Sabine Oertelt-Prigione, Marco Piccininni, Katja Gayraud, Michelle Kelly-Irving, et Nathalie Bajos. 2024. « Sex and Gender Bias in Chronic Coronary Syndromes Research: Analysis of Studies Used to Inform the 2019 European Society of Cardiology Guidelines ». *The Lancet Regional Health – Europe* 45. <https://doi.org/10.1016/j.lanepe.2024.101041>.

Bertail, Patrice, David Bounie, Stéphan Cléménçon, et Patrick Waelbroeck. 2019. « Algorithmes: Biais, Discrimination et Équité ». *HR Today*, n° 58.

Besse, Philippe, Aurèle Besse-Patin, et Céline Castets-Renard. 2020. « Legal and Ethical Implications of Artificial Intelligence Algorithms in the Health Field ». *Questions de société* 8 (3).

Bettache, Mustapha, et Laurie Foisy. 2019. « Intelligence artificielle et transformation des emplois ». *Question(s) de management ?* 25 (3): 61-67. <https://doi.org/10.3917/qdm.193.0061>.

Bibault, Jean-Emmanuel, Benjamin Chaix, Arthur Guillemassé, Sophie Cousin, Alexandre Escande, Morgane Perrin, Arthur Pienkowski, Guillaume Delamon, Pierre Nectoux, et Benoît Brouard. 2019. « A Chatbot Versus Physicians to Provide Information for Patients With Breast Cancer: Blind, Randomized Controlled Noninferiority Trial ». *Journal of Medical Internet Research* 21 (11). <https://doi.org/10.2196/15787>.

Bonifay Emmanuelle, et Sereno, Sophie. 2022. « Genre, Algorithme et Droit », 2022.

Bubien, Yann, et Vincent Vuiblet. 2023. « La révolution numérique de la santé ». *Esprit*, n° 6, 77-87. <https://doi.org/10.3917/espri.2306.0077>.

Chaix Benjamin, Jean-Emmanuel Bibault, Arthur Pienkowski, Guillaume Delamon, Arthur Guillemassé, Pierre Nectoux, et Benoît Brouard 2019. « When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot ». *JMIR Cancer* 5, n° 1. <https://doi.org/10.2196/12856>.

Cusumano, Michael A. 2023. « Generative AI as a New Innovation Platform ». *Communications of the ACM* 66, n° 10: 18-21. <https://doi.org/10.1145/3615859>.

Coeckelbergh, Mark. 2019. « Artificial Intelligence: Some Ethical Issues and Regulatory Challenges ». *Technology and Regulation* 1:31-34. <https://doi.org/10.26116/techreg.2019.003>.

Dastin, Jeffrey. 2018. « Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women ». *Reuters*, sect. World. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>.

Douetteau, Florian. 2024. « Pourquoi l'IA générative va révolutionner le rapport au travail ? » *Administration* 281 (1): 70-72. <https://doi.org/10.3917/admi.281.0070>.

Goh, Ethan, Robert Gallo, Jason Hom, Eric Strong, Yingjie Weng, Hannah Kerman, Joséphine A. Cool, et al. 2024. « Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial ». *JAMA Network Open* 7 (10). <https://doi.org/10.1001/jamanetworkopen.2024.40969>.

Grimaldi, André. 2013. « Les hôpitaux en quête de rentabilité : à quel prix ? » *Revue du MAUSS* 41 (1): 25-33. <https://doi.org/10.3917/rdm.041.0025>.

Grote, Thomas, et Philipp Berens. 2025. « On the Ethics of Algorithmic Decision-Making in Healthcare | Journal of Medical Ethics ». *Journal of Medical Ethics*, janvier. <https://jme.bmj.com/content/46/3/205.long>.

Kovacek, David, et James C. L. Chow. 2021. « An AI-Assisted Chatbot for Radiation Safety Education in Radiotherapy ». *IOP SciNotes* 2 (3): 1-9. <https://doi.org/10.1088/2633-1357/ac1f88>.

Lahat, Adi, Eyal Shachar, Benjamin Avidan, Zina Shatz, Benjamin S. Glicksberg, et Eyal Klang. 2023. « Evaluating the Use of Large Language Model in Identifying Top Research Questions in Gastroenterology ». *Scientific Reports* 13 (1): 1-6. <https://doi.org/10.1038/s41598-023-31412-2>.

Le Goff, Thomas. 2023. « Recommandations pour une action publique en faveur d'une IA générative respectueuse de l'environnement ». <https://hal.science/hal-04371031v2>.

Lee, Tsung-Chun, Kyle Staller, Vlaicu Botoman, Mythili P. Pathipati, Sanskriti Varma, et Braden Kuo. 2023. « ChatGPT Answers Common Patient Questions About Colonoscopy ». *Gastroenterology* 165 (2): 509-511. <https://doi.org/10.1053/j.gastro.2023.04.033>.

Li, Wenbo, Yinxu Zhang, et Fengmin Chen. 2023. « ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad ? » *Annals of Biomedical Engineering* 51 (9): 1892-1897. <https://doi.org/10.1007/s10439-023-03232-y>.

Longo, Luca, Mario Brcic, Federico Cabitza, Jaesik Choi, Roberto Confalonieri, Javier Del Ser, Riccardo Guidotti, et al. 2024. « Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions ». *Information Fusion* 106 (juin):1-22. <https://doi.org/10.1016/j.inffus.2024.102301>.

Méadel, Cécile, et Madeleine Akrich. 2010. « Internet, tiers nébuleux de la relation patient-médecin ». *Les Tribunes de la santé* 29 (4): 41-48. <https://doi.org/10.3917/seve.029.0041>.

Mignot, Léo, et Émilien Schultz. 2022. « Les innovations d'intelligence artificielle en radiologie à l'épreuve des régulations du système de santé ». *Réseaux : communication, technologie, société*, Contrôler l'intelligence artificielle ?, 232-233 : 65-97. <https://doi.org/10.3917/res.232.0065>.

Ministère de la Cybersécurité et du Numérique du Québec. 2024. « Guide des bonnes pratiques d'utilisation de l'intelligence artificielle générative », 1-25.

Morse, Keith E., Nicolai P. Ostberg, Veena G. Jones, et Albert S. Chan. 2020. « Use Characteristics and Triage Acuity of a Digital Symptom Checker in a Large Integrated Health System: Population-Based Descriptive Study ». *Journal of Medical Internet Research* 22 (11). <https://doi.org/10.2196/20549>.

Nordlinger, Bernard, Olivier De Fresnoye, et Claude Kirchner. 2024. « Systèmes d'IA générative en santé : enjeux et perspectives ». *Académie Nationale de Médecine* 208 (5): 536-547.

OMS. 2021. « Éthique et Gouvernance de l'Intelligence Artificielle pour la santé : Orientations de l'OMS ». 2021. <https://www.jstor.org/stable/resrep40880?seq=5>.

OMS. 2024. « Ethics and Governance of Artificial Intelligence for Health: Large Multi-Modal Models ». WHO Guidance. 1st ed. Geneva: World Health Organization.

Petitprez, Eugénie. 2024. *Les enjeux liés à l'utilisation de l'intelligence artificielle en matière de santé : regards croisés*. Ceprisca. Colloques.

Popejoy, Alice B., et Stephanie M. Fullerton. 2016. « Genomics Is Failing on Diversity ». *Nature* 538 (7624): 161-164. <https://doi.org/10.1038/538161a>.

Smuha, Nathalie. 2019. « Lignes directrices en matière d'éthique pour une IA digne de confiance ». [ethics\\_guidelines\\_for\\_trustworthy\\_ai-fr\\_87FE7A3C-D03D-9305-81A653DDA84B5A60\\_60427.pdf](https://ethics_guidelines_for_trustworthy_ai-fr_87FE7A3C-D03D-9305-81A653DDA84B5A60_60427.pdf).

Tobb, Kardie, Madison Kocher, et Renée P. Bullock-Palmer. 2022. « Underrepresentation of women in cardiovascular trials- it is time to shatter this glass ceiling ». *American Heart Journal Plus: Cardiology Research and Practice* 13:100-109. <https://doi.org/10.1016/j.ahjo.2022.100109>.

Veerappa, Manjunatha, et Salvo Rinzivillo. 2023. « Special Theme : eXplainable AI ». *ERCIM*, n° 134, 8-37.

Vilone, Giulia, et Luca Longo. 2021. « Notions of explainability and evaluation approaches for explainable artificial intelligence ». *Information Fusion* 76 (décembre):89-106. <https://doi.org/10.1016/j.inffus.2021.05.009>.

Vuarin, Louis, et Véronique Steyer. 2023. « Le principe d'explicabilité de l'IA et son application dans les organisations ». *Réseaux* 240 (4): 179-210. <https://doi.org/10.3917/res.240.0179>.

Wu, David JH, et Jean-Emmanuel Bibault. 2024. « Pilot applications of GPT-4 in radiation oncology: Summarizing patient symptom intake and targeted chatbot applications ». *Radiotherapy and Oncology* 190:109-978. <https://doi.org/10.1016/j.radonc.2023.109978>.

Yeung, Karen. 2019. « Étude sur les incidences des technologies numériques avancées (dont l'intelligence artificielle) sur la notion de responsabilité, sous l'angle des droits humains ». *Responsibility and AI*, 1-100.

Zhang, Yanhui, Haolong Pei, Shihan Zhen, Qian Li, et Fengchao Liang. 2023. « Chat Generative Pre-Trained Transformer (ChatGPT) usage in healthcare ». *Gastroenterology & Endoscopy* 1 (3): 139-143. <https://doi.org/10.1016/j.gande.2023.07.002>

## Sitographie

« Accueil – Octopize ». 2025. 5 janvier 2025. <https://www.octopize.io/>.

Bpifrance, réal. 2024. *Comment l'IA révolutionne l'imagerie médicale et la médecine proactive*. <https://www.youtube.com/watch?v=oLogkWBvyr8>.

Carpentier, Manu, réal. 2024. *Le JT de la Radiologie*. Healthtech Innovation Paris. <https://www.jtradiologie.com/Video/view/<https://www.jtradiologie.com/Video/view/xs3mzvU>.

« Donnée personnelle ». 2025. CNIL. 5 janvier 2025. <https://www.cnil.fr/fr/definition/donnee-personnelle>.

« Donnée sensible ». 2025. CNIL. 5 janvier 2025. <https://www.cnil.fr/fr/definition/donnee-sensible>.

« Explicabilité (IA) ». 2025. CNIL. 5 janvier 2025. <https://www.cnil.fr/fr/definition/explicabilite-ia>.

« Gleamer Homepage AI for Radiology – Gleamer ». 2025. 5 janvier 2025. <https://www.gleamer.ai/>.

« IA : comment être en conformité avec le RGPD ? » 2025. CNIL. 5 janvier 2025. <https://www.cnil.fr/fr/intelligence-artificielle/ia-comment-etre-en-conformite-avec-le-rgpd>.

« Incepto ». 2025. Incepto Medical. 5 janvier 2025. <https://incepto-medical.com/>.

« Intelligence artificielle : Doctolib lance l'Assistant de consultation ». 2024. About Doctolib - France. 15 octobre 2024. <https://about.doctolib.fr/news/intelligence-artificielle-doctolib-lance-lassistant-de-consultation/>.

« Jaide | Generative AI for patients and physicians ». 2025. 5 janvier 2025. <https://www.jaide.care/>.

« Kiro : Construire la biologie de demain ». 2025. Kiro. 5 janvier 2025. <https://www.kiro.bio>.

L'Oeil de la E-santé, réal. 2024. #ESANTE24 | « L'IA et l'IA générative, quel impact pour le secteur de la santé ? » [https://www.youtube.com/watch?v=wPXzWf8\\_ZIY](https://www.youtube.com/watch?v=wPXzWf8_ZIY).

Maffulli, Stefano. 2023. « Meta's LLaMa 2 License Is Not Open Source ». Open Source Initiative. 20 juillet 2023. <https://opensource.org/blog/metas-llama-2-license-is-not-open-source>.

« Milvue | New Tech, New Care ». 2025. Milvue. 5 janvier 2025. <https://www.milvue.com/>.

« Nabla · L'assistant médical qui vous libère du temps ». 2025. 5 janvier 2025.  
<https://www.nabla.com/fr/>.

« On explore les impacts environnementaux de ChatGPT. Françaises, français, mes chers compatriotes. » 2024. <https://www.radiofrance.fr/franceinter/podcasts/la-terre-au-carre/la-terre-au-carre-l-actu-du-mercredi-27-novembre-2024-5584893>.

« PulseLife, La Médecine Fondée Sur Les Preuves ». 2025. PulseLife. 5 janvier 2025.  
<https://pulselife.com/fr-fr>.

« Solutions – Octopize ». 2025. 5 janvier 2025.  
<https://www.octopize.io/solutions#section-FAQ>.

« Vope ». 2025. Vope Medical. 5 janvier 2025. <https://www.vopecare.ai>.