Bull Acad Natl Med 208 (2024) 536—547

Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

RAPPORT ET RECOMMANDATIONS DE L'ANM

Rapport 24-03. Systèmes d'IA générative en santé : enjeux et perspectives☆

Generative AI systems in healthcare: Challenges and prospects

Bernard Nordlinger[a,*], Claude Kirchner[b], Olivier de Fresnoye[c], au nom d'un groupe de travail de l'Académie nationale de médecine[1]

[a] *Académie nationale de médecine, 16, rue Bonaparte, 75006 Paris, France*
[b] *Comité national pilote d'éthique du numérique, Comité consultatif national d'éthique, Inria, Paris, France*
[c] *Projet Echopen, Paris, France*

Disponible sur Internet le 28 mars 2024

Résumé La santé est un des domaines majeurs d'application des technologies dites d'intelligence artificielle. Tous les domaines de la santé et toutes les spécialités sont concernés. Les systèmes d'intelligence artificielle générative (SIAgen) impressionnent par leur capacité à produire en quelques secondes des textes souvent pertinents, mais aussi parfois erronés. Leurs champs d'applications dans le domaine de la santé sont vastes et peuvent aller de l'aide à la rédaction de notes d'information à la rédaction de thèses ou de projets de programme de recherche. Pour les utiliser à bon escient il est important d'en connaître les principes de fonctionnement. Les SIAgen fonctionnent à partir d'auto-apprentissage basé sur un nombre extrêmement élevé d'exemples, ce qui est très différent de l'approche humaine, qui s'appuie sur l'expérience, le contexte et un système de valeurs. Ils génèrent des textes avec une grande

☆ Un rapport exprime une prise de position officielle de l'Académie nationale de médecine. L'Académie dans sa séance du mardi 5 mars 2024, a adopté le texte de ce rapport par 70 voix pour, 3 voix contre et 8 abstentions.

* Auteur correspondant.
Adresses e-mail : bernard.nordlinger@gmail.com (B. Nordlinger), claude.kirchner@inria.fr (C. Kirchner), olivierdefresnoye@gmail.com (O. de Fresnoye).

[1] Ce rapport est issu des travaux du groupe de travail sur les systèmes d'IA générative mis en place par l'Académie nationale de médecine de mai à octobre 2023. Il a réuni les personnes suivantes que nous remercions pour leur participation active, pour les documents qu'elles ont pu partager et pour leur relecture de la première version de ce rapport : Catherine Adamsbaum, Gilles Adda, Jean-François Allilaire, Raymond Ardaillou, Guillaume Assié, Emmanuel Bacry, Raja Chatila, André Chays, Bruno Clément, Pierre Corvol, Arthur Dauphin, Laurence Devillers, Ferdinand Dhombres, Emmanuel Didier, Nicolas Do Huu, Valérie Edel, Jean-Gabriel Ganascia, Isabelle Gentil, Henri Julien, Alexei Grinbaum, Thierry Hauet, Pierre Jannin, Dominique Lecomte, Arnold Migus, Patrick Netter, Olivier Palombi, Nadir Saoudi, Guy Vallancien, Eric Vivier, Laurence Watier.

https://doi.org/10.1016/j.banm.2024.03.005

rapidité mais ne sont pas entraînés à rechercher ou à dire la vérité. Une validation humaine est donc toujours nécessaire. Par ce rapport, l'Académie nationale de médecine explicite plusieurs de ces avancées pour la santé, décrit les enjeux d'éthique associés et recommande des points d'actions à mettre en œuvre sans délai.

Summary Healthcare is one of the major application fields of Artificial Intelligence technologies. All areas of healthcare and all specialties are concerned. Generative Artificial Intelligence systems are impressive in their ability to produce texts in a matter of seconds, often relevant, but sometimes erroneous. They can be used in a wide range of healthcare applications, from helping to write briefing notes to drafting theses and research programs. To use them properly, it is important to understand how they work. Large Language Models use neural networks trained on massive amounts of text data, which is very different from the human, experience-based approach. They generate language but are not trained to tell or search for the truth. Human validation is therefore always necessary. Through this report, the Académie nationale de médecine explains the resulting progress and discoveries for health, describes associated ethical issues and recommends action points to be implemented without delay.

Introduction

Dans l'histoire de la médecine il est peu de périodes où ont convergé des avancées majeures de la biologie, telle la mise au point en quelques mois de nouveaux vaccins, et celles apportées par d'autres sciences et technologies telles qu'actuellement celles du numérique et sa composante IA, avec la capacité de gérer en quelques secondes des milliards de données. La santé est un des domaines majeurs d'application des technologies dites d'intelligence artificielle et les potentiels sont tels que ni le public, ni les patients, ni les professionnels de santé ne peuvent rester à l'écart des enjeux, des bénéfices et des limites de ces nouvelles avancées.

Tous les domaines de la santé et toutes les spécialités sont ou seront concernés par les progrès des technologies numériques, de l'analyse des images diagnostiques obtenues par radiographie, scanner ou imagerie par résonance magnétique, l'aide au diagnostic et au choix des traitements, à la mesure de l'efficacité des soins et de l'organisation du système de santé. Tous ont en commun qu'une validation humaine est indispensable avant leur mise en œuvre, nous allons en décrire les enjeux et les conditions permettant cette supervision humaine.

Depuis quelques mois les systèmes d'intelligence artificielle générative (SIAgen) ont franchi une nouvelle étape. Leur capacité à produire des textes en quelques secondes impressionne. Leurs champs d'applications sont très vastes et peuvent aller de l'aide à la rédaction de notes d'information à l'aide à la rédaction de thèses ou de projets de programme de recherche. Elles font l'objet de campagnes médiatiques intenses qui mélangent information et fiction et qui suscitent fantasmes et craintes.

Les SIAgen fonctionnent à partir d'auto-apprentissage basé sur un nombre extrêmement élevé d'exemples, ce qui est très différent de l'approche humaine, basée sur l'expérience, la signifiance et la recherche de vérité. Cela leur permet de trouver un mot à partir du précédent et du contexte et de générer des textes avec une grande rapidité mais sans qu'actuellement les sources des textes générés soient connues. Même si association de mots ne signifie pas automatiquement causalité, les résultats impressionnent par la fréquence élevée de réponses pertinentes, au risque de méconnaître les erreurs souvent appelées « hallucinations ».

Le but de ce rapport est de faire le point sur l'intérêt et les risques de l'utilisation des SIAgen dans le domaine de la santé et de proposer des points d'actions immédiats.

L'évolution du sujet est très rapide et des mises à jour ultérieures seront nécessaires.

Les systèmes d'IA générative et les modèles de fondation

Les systèmes d'intelligence artificielle générative (SIAgen) sont des systèmes numériques capables de produire de multiples résultats, des textes comme des images ou des vidéos, à des fins diverses telles que la production de compte-rendus, la traduction, la production de code informatique, l'aide au diagnostic ou à la décision, la synthèse de structures comme l'impression 3D, ... Ce rapport se focalisera essentiellement sur les aspects textuels avec une ouverture vers les images dans le contexte général de la médecine. Cette description de ce que sont ces SIAgen est largement inspirée de l'avis 7 [1] du Comité National Pilote d'Éthique du Numérique (CNPEN).

Les premiers exemples de modèles[2] de génération de texte, comme GPT-2 (GPT signifiant *Generative Pretrained Transformers*), ou de génération d'images, comme DALL-E ou *Stable Diffusion*, ont montré un potentiel pour de multiples applications. Les systèmes d'IA générative pour la langue sont souvent utilisés pour des interfaces d'agents

[2] Ces modèles sont des algorithmes permettant de produire des textes ou des images similaires aux données d'apprentissage qui ont servi à les construire.

conversationnels (*chatbots*) : ChatGPT construit par OpenAI (et sa variante *Microsoft Copilot* [anciennement Bing Chat]) est fondé sur le grand modèle de langue GPT-4, et le chatbot *Gemini* (anciennement *Bard*) construit par Google est basé sur le modèle PaLM (*Pathways Language Model)*.

Les SIAgen répondent à des demandes ou requêtes

Les SIAgen répondent à des demandes ou requêtes (souvent appelées prompts) en produisant de nouvelles données, par exemple la séquence de mots la plus probable après le prompt, à partir de caractéristiques communes apprises sur un corpus de données de très grande taille. Ces systèmes se servent donc de modèles de fondation, selon l'appellation *Foundation Model* introduite à l'université de Stanford, qui permettent de produire un résultat présentant un certain degré de similarité avec les données d'apprentissage qui ont servi à le construire. Un modèle de fondation est un réseau de neurones (numériques ou aussi dit artificiels [2]), profond[3], entraîné sur une grande quantité de données non annotées[4], généralement par apprentissage auto-supervisé[5]. Les grands modèles de langue (LLM pour *Large Language Model*) sont des modèles de fondation entraînés sur un corpus de textes. Ils ouvrent de nouvelles perspectives et introduisent un nouveau paradigme dans le traitement de la langue, mais aussi dans le traitement des signaux multimodaux (son, image, vidéo, etc.). Ces modèles pré-entraînés sur de grands corpus peuvent être optimisés pour réaliser une nouvelle application, appelée génériquement SIAgen, en utilisant peu de données supplémentaires spécifiques à cette tâche.

Les techniques d'apprentissage machine

Les techniques d'apprentissage machine qui sont à la base des systèmes d'intelligence artificielle dont il est question ici, produisent des modèles exprimant des corrélations statistiques entre les éléments des données (segments de mots, parties d'images) utilisées pour leur entraînement. Les systèmes d'IA générative combinent, dans diverses phases, trois techniques de l'apprentissage statistique :

- l'apprentissage non-supervisé (ou auto-supervisé) qui produit des modèles corrélatifs de données sans annotation a priori ;
- l'apprentissage supervisé qui permet d'affiner ces modèles en les entraînant sur des données spécifiques et en filtrant certains résultats, enfin ;
- l'apprentissage par renforcement qui permet d'optimiser les performances du système au travers de la sélection des meilleurs résultats.

Dans la méthode RLHF (*Reinforcement Learning with Human Feedback*), l'apprentissage par renforcement permet d'accorder les résultats avec les préférences d'annotateurs humains exprimées pendant le stade supervisé, dans le but de rendre ces réponses conformes aux valeurs humaines, sans que la signification de ces valeurs soit appréhendée par les systèmes. L'arrivée en masse des systèmes d'IA générative est récente mais les architectures et les techniques d'apprentissage automatique qui les sous-tendent existent depuis plusieurs décennies. Toutefois, elles ont beaucoup évolué ces dix dernières années. L'approche actuelle est celle des réseaux de neurones pour apprendre la distribution des données[6] et produire des résultats similaires, mais rarement identiques, à ces données d'apprentissage. Les modèles les plus connus sont les réseaux antagonistes génératifs (GANs [3]) et plus récemment les « transformers »[7].

Pour entraîner un transformer et créer un modèle de fondation de type LLM, les textes sont décomposés par un algorithme en suites de caractères, appelés *tokens*, qui ne forment pas nécessairement des mots signifiant (ayant une signification). Le transformer, qui est un réseau de neurones, est entraîné par auto-apprentissage sur les données du corpus divisées en *tokens* représentés sous forme de vecteurs de « plongement lexical » (*word embedding*). La taille des vecteurs est par exemple de 512 dans GPT-3.5. Les transformers s'appuient sur l'hypothèse distributionnelle selon laquelle des mots qui se trouvent dans des contextes d'apparition similaires tendent à avoir des sens similaires[8]. L'hypothèse distributionnelle et les modèles vectoriels de représentation des *tokens* permettent de calculer une distance entre ceux-ci. Quand cette distance est petite, la proximité des vecteurs dans l'espace vectoriel correspond à une certaine parenté. Les vecteurs des *tokens* se retrouvant dans des contextes similaires dans le corpus d'apprentissage ont tendance à devenir proches les uns des autres. De plus, un transformer met en œuvre un mécanisme de calcul appelé « mécanisme d'attention », qui permet d'ajuster le poids de chaque *token* en fonction de tous les autres. Un transformer apprend ainsi les régularités (relations) les plus saillantes entre les *tokens*, sans être influencé par l'ordre de ceux-ci.

Il existe deux grandes familles de transformers :

- les modèles de type GPT (OpenAI) qui sont entraînés à prédire le *token* suivant dans une séquence. Le contexte considéré est donc réduit aux *tokens* qui précèdent ;
- les modèles de type BERT (*Bidirectional Encoder Representations from Transformers*) de Google sont entraînés sur ce qui précède et ce qui suit le *token*. Quand on leur

[3] Les réseaux de neurones dits profonds sont des réseaux de neurones numériques dont le nombre de couches (ou la profondeur d'analyse) est grand.

[4] Les données non annotées sont des données qui n'ont pas fait l'objet d'annotation décrivant ce qu'elles représentent (par exemple, la photographie d'un cheval).

[5] L'apprentissage est dit auto-supervisé lorsqu'il est sous le contrôle direct du processus algorithmique considéré. Il est supervisé lorsqu'il peut utiliser les annotations portées (en général) par des humains.

[6] C'est-à-dire la manière dont les données d'apprentissage sont réparties dans l'ensemble des données possibles.

[7] Vaswani et al. ''Attention Is All You Need'', 31st Conference on Neural Information Processing Systems (Neurips 2017), Long Beach, CA, USA.

[8] Firth J. R. ''You shall know a word by the company it keeps'' — ''Tu connaîtras un mot par ses fréquentations'' (1957).

propose une phrase avec un *token* manquant, ils sont capables de produire le *token* le plus probable dans ce contexte.

Notons que deux modèles de type BERT ont été entrainés spécifiquement sur des données en langue française, FlauBERT[9] et CamenBERT[10].

Les hyperparamètres des modèles de fondation

Les hyperparamètres des modèles de fondation (le nombre de couches dans un réseau de neurones, la dimension des vecteurs des *tokens*, la taille du dictionnaire de *tokens*, etc.) sont déterminants pour la structure du modèle et pour l'entraînement du modèle. Pour un chatbot qui utilise un modèle de fondation, la taille de l'historique est déterminante pour les performances du modèle (OpenAI GPT3.5 : 8000 *tokens* — OpenAI GPT4 : 32000 *tokens* — Anthropic Claude : 100 000 *tokens*). Souvent, ces hyperparamètres ne sont pas dévoilés pour des raisons de cybersécurité ou de confidentialité. Un paramètre clé est celui de la « température » qui exprime le degré d'aléa dans le choix des *tokens*. À une température élevée, le modèle est plus « créatif » car il peut générer des sorties plus diversifiées, tandis qu'à une température basse, le modèle tend à choisir les sorties les plus probables, ce qui rend le texte généré plus prévisible. L'ajustement des paramètres est important dans la conception d'un modèle et peut avoir un impact significatif sur sa performance. En général, le réglage des hyperparamètres est un processus long, procédant par essai et erreur, bien qu'il existe des recherches sur l'automatisation de ces choix.

La taille des modèles de fondation

La taille des modèles de fondation peut être impressionnante. C'est en mars 2020 qu'OpenAI annonce GPT-3, modèle doté de 175 milliards de paramètres. La course au plus gros modèle est en cours, le nombre de paramètres de GPT-4 n'étant pas dévoilé officiellement. Bard, construit par Google, utilise le modèle de fondation PaLM entraîné avec 540 milliards de paramètres. Le modèle Chinois WuDao 2.0 de BAAI, utilise 1 750 milliards de paramètres. Il n'est pas certain, et c'est le sujet de recherches actives, que des modèles encore plus grands apporteraient des performances plus élevées. Ainsi, Google a également publié PaLM-2 avec moins de paramètres que son prédécesseur PaLM [4]. Ces modèles de langue gigantesques posent aujourd'hui la question de leur impact environnemental et de la dépense d'énergie nécessaire à leur élaboration.

L'intégration des valeurs sociales et des filtres dans les systèmes d'intelligence artificielle générative

Les LLM peuvent produire des résultats qui n'ont pas de sens et peuvent être dangereux. Ces résultats peuvent prendre de nombreuses formes, y compris du contenu nuisible tel qu'un discours de haine, un diagnostic erroné ou du contenu pédopornographique. Dans une quête de neutralité, les systèmes d'IA générative sont optimisés avec des filtres construits par les concepteurs. De plus, dans le RLHF, l'annotateur reçoit des instructions pour guider ses choix. Les valeurs sociales traduites dans les filtres, comme l'évitement des biais, sont donc liées aux êtres humains qui testent les systèmes ainsi qu'aux choix des concepteurs. Aujourd'hui, ce processus n'est ni transparent ni vérifié. La méthode d'évaluation adversaire par les équipes humaines, appelée *red teaming*, a été étendue au-delà de son domaine d'origine en cybersécurité et appliquée aux LLM. Elle désigne l'utilisation de nombreux types de sondages, de tests et d'attaques des systèmes d'IA (par exemple, par injection de prompts [5]) afin de mettre à jour les biais ou les comportements émergents de ces modèles.

Les modèles d'IA générative sont souvent multilingues

Les modèles d'IA générative sont souvent multilingues [6], c'est-à-dire qu'ils sont construits à partir de corpus dans plusieurs langues, avec le plus souvent pour langue actuellement dominante l'anglais ou le chinois. La génération de textes dans certaines langues peu dotées de corpus peut être rendue plus performante grâce à ces systèmes multilingues. Comme nous l'avons noté plus haut, il existe cependant des modèles de fondation en français (c'est-à-dire pré-entraînés sur des corpus francophones), par exemple FlauBERT, CamemBERT. En entraînant le même algorithme sur des corpus de textes asiatiques ou français, on obtiendrait certainement des représentations numériques différentes. Les modèles produiraient alors des textes ayant des nuances différentes. Le langage possède des ambiguïtés complexes et il est imprégné des représentations spécifiques à la culture considérée.

Pour conclure cette partie visant à fixer le vocabulaire et donner une intuition du mode de fonctionnement de ces systèmes numériques, il convient d'insister sur les points suivants :

- *la recherche de vérité n'est pas intégrée dans l'état actuel du développement des SIAgen :* les textes générés le sont par des moyens stochastiques, ils n'intègrent pas de représentation logique ni de critères de vérité ;
- *la qualité des données utilisées pour l'entraînement des réseaux de neurones est fondamentale*. Mais la quantité des données nécessaire à cet entraînement, données recueillies en particulier sur internet, ne permet pas actuellement de vérifier leur qualité ni leur diversité. Les résultats produits par ces systèmes, malgré les phases de test RLHF, peuvent être clairement erronés. De plus, on assiste maintenant via ces systèmes génératifs à la mise en ligne massive de textes générés et non vérifiés ou bien construits à des fins de manipulation ou de désinformation et dont les sources et les auteurs ne sont usuellement ni connus ni publiés.

[9] FlauBERT: Unsupervised Language Model Pre-Training for French, 2020.
[10] CamemBERT: a Tasty French Language Model, 2020.

Applications remarquables des SIAgen

Aide au diagnostic et au choix du traitement

La mise au point de systèmes d'aide à la décision clinique est depuis longtemps un objectif des spécialistes de l'IA. Mais les résultats ne sont pas à la hauteur des espoirs malgré des investissements financiers très importants dans ce secteur.

Si on demande à ChatGPT s'il peut aider à proposer un diagnostic médical, la réponse est prudente : « Non, ChatGPT ne peut pas proposer un diagnostic médical. Je ne suis pas un médecin ni un professionnel de la santé, et il est important de ne pas se fier à des informations médicales provenant d'une source non qualifiée pour diagnostiquer des problèmes de santé. Les diagnostics médicaux nécessitent une évaluation approfondie par un professionnel de la santé, comme un médecin, un infirmier ou un spécialiste médical, qui prend en compte les antécédents médicaux, les symptômes, les tests et les examens appropriés pour établir un diagnostic précis. Si vous avez des préoccupations concernant votre santé, je vous encourage vivement à consulter un professionnel de la santé. Les informations médicales en ligne peuvent être utiles pour obtenir des informations générales, mais elles ne remplacent pas une évaluation médicale appropriée. »[11].

On ne peut qu'approuver, et pourtant... Les outils d'intelligence artificielle appliquée à la santé se sont, jusqu'à présent, essentiellement intéressés à des secteurs spécifiques comme l'analyse d'images, radiographiques, scanographiques, par IRM ou anatomopathologiques. En 2023 près de 700 dispositifs utilisant des technologies d'IA dans le domaine de la santé ont été approuvés par la FDA[12]. Mais, pour l'instant les outils d'aide au diagnostic proposant une synthèse des différentes modalités diagnostiques, telle que peut la faire un médecin, n'ont pas fait la preuve de leur efficacité, à l'image d'IBM Watson [7].

De nombreux algorithmes ont été proposés pour une variété de diagnostics, sans qu'ils trouvent beaucoup d'échos auprès des cliniciens qui se montrent réticents à adopter des outils d'IA diagnostique dont ils ne comprennent ni le fonctionnement ni comment ils peuvent s'intégrer dans leur pratique médicale.

Ces difficultés sont aussi dues à des raisons structurelles. Certaines sont liées au manque de généralisation ou de reproductibilité de certains algorithmes. D'autres sont plus spécifiques au numérique en santé, comme l'entraînement sur des données non représentatives de la population générale par manque de diversité, d'âges, de genre, de race ou d'ethnie ou une évaluation insuffisante des algorithmes en clinique, ne prenant pas en compte la diversité des pratiques. Les systèmes d'IA diagnostique apprennent en imitant les experts, en exploitant des exemples et en utilisant le feed-back des utilisateurs, tout progrès impliquant une collaboration des cliniciens et des développeurs d'algorithmes.

Les systèmes d'IA génératives ouvrent de nouvelles opportunités, mais celles-ci dépendent des données avec lesquelles elles ont été entraînées : entraînées avec des données générales comme ChatGPT elles ne peuvent fournir que des réponses générales ; de nouveaux systèmes comme MedPalm 2 entraînés avec des données médicales sont susceptibles d'apporter des réponses plus spécifiques. Mais il est important que les utilisateurs gardent à l'esprit que ces systèmes ne sont pas entraînés à fournir un diagnostic exact mais à produire un texte. Et il y a un pas majeur entre proposer un diagnostic à partir d'un résumé de cas clinique publié dans un journal où les SIAgen peuvent exceller [8,9], et la synthèse, faite par un médecin, des données cliniques et paracliniques (images, biologie) d'un patient réel. Se pose bien sûr la question des responsabilités : les systèmes logiciels en tant que tels n'ont pas de responsabilité attribuable, contrairement aux personnes physiques ou morales qui les utilisent ou celles qui les conçoivent. Dans ce contexte la responsabilité des personnels de santé peut être engagée, soit pour ne pas avoir pris le recul nécessaire relativement à une proposition de diagnostic trop rapidement validée ou bien à contrario pour avoir pris une direction différence de celle suggérée par un SIA. Ces questions de responsabilité seront en particulier abordées dans la mise en œuvre du règlement européen sur l'IA qui va également concerner les concepteurs des SIA, en particulier génératifs.

Avant la généralisation de l'application clinique de ces outils en vie réelle le chemin sera encore long, et le premier objectif des développeurs d'algorithmes est de fournir une aide aux professionnels de santé, voire au grand public, mais pas de remplacer le médecin..., en tout cas pour l'instant. Les utilisateurs potentiels devront par ailleurs être conscients des risques d'introduire dans ces systèmes des données personnelles, car elles ne sont pas gérées localement mais entreposées à distance avec le risque qu'elles soient mal utilisées, vendues, etc.

Apports à l'imagerie médicale

L'analyse d'image par les techniques d'apprentissage machine a été l'une des premières applications de ces techniques dès les années 1960. Mais c'est grâce aux techniques d'analyse par des réseaux neuronaux profonds développées dans les années 2010 que l'imagerie médicale a très fortement progressé en précision et que ces technologies sont maintenant très largement utilisées : il est significatif que parmi les 691 autorisations attribuées par la FDA à des dispositifs médicaux incluant des techniques d'IA, 77 % concernent la radiologie [10]. Comme le précise l'avis commun du CCNE et du CNPEN [11], ces techniques se prêtent particulièrement à la classification et à la segmentation d'images issues d'examens radiologiques tels que les scanners, CT scans, IRM, radiographies conventionnelles, échographies, examens de rétine, de peau, et à l'analyse d'images histologiques. Dans ce contexte déjà très avancé, les apports des SIAgen sont au moins de deux ordres : une amélioration potentielle de la qualité et de la rapidité d'analyse, et la capacité à générer de nouvelles images ou des textes d'analyse ou de synthèse.

Les applications médicales vont bénéficier de l'amélioration importante des capacités des modèles de fondation et permettre d'obtenir des images avec une meilleure résolution, moins de bruit de fond, donc plus

[11] Citation de ChatGPT 3.5 en novembre 2023.
[12] U.S. Food and Drug Administration, https://fda.gov.

précises et comportant moins d'erreurs ou d'imprécisions. Cela bénéficiera à l'a*natomo-cyto-pathologie (anapath) comme aux examens radiologiques*.

Les capacités générationnelles des SIAgen permettent de créer des images médicales synthétiques de haute qualité qui pourront être utilisées pour la formation des professionnels de la santé en créant des jeux de données d'entraînement diversifiés. La simulation de pathologies spécifiques dans des images médicales pourra être utile à la formation des médecins et des chercheurs, ou pour tester la robustesse des algorithmes de diagnostic. En *anatomie pathologique*, les SIAgen pourront aider à la segmentation automatique des différentes structures tissulaires dans les images histologiques, aux analyses multiparamétriques, à l'analyse quantitative, facilitant la compréhension des caractéristiques spécifiques des tissus pathologiques.

Les SIAgen pourront servir à assister ou même à automatiser la rédaction de comptes rendus radiologiques en proposant des analyses préliminaires, permettant d'alléger la charge de travail des radiologues, à la condition qu'ils soient in fine validés par un humain.

L'intégration de ces technologies dans l'imagerie médicale devra prendre en compte les questions de confidentialité, de transparence des algorithmes, la nécessité d'une supervision humaine, et d'une formation des professionnels, condition à leur adoption.

Médicaments

La découverte conventionnelle de médicaments est un processus long, coûteux et aux résultats aléatoires. Les chimistes recherchent une nouvelle molécule qui répond à un cahier des charges multiparamétrique : elle doit être nouvelle, brevetable, active sur la cible biologique sélectionnée, sélective et dépourvue de toxicité rédhibitoire. Cette identification de candidats-médicaments potentiels se fait par le biais d'expériences laborieuses. La durée moyenne de développement d'un médicament dépasse souvent 12 ans, pour un coût allant d'un à trois milliards de dollars.

Les systèmes d'intelligence artificielle générative (SIAgen) ont le potentiel de révolutionner ce processus en le rendant plus rapide et moins coûteux et en permettant d'examiner beaucoup plus d'hypothèses de nouvelles molécules qui répondent au cahier des charges qu'il ne serait possible de le faire en laboratoire. Le marché de la découverte de médicaments par l'IA générative de 150M$ en 2023 pourrait être multiplié par 10 dans les années à venir. Actuellement les grandes entreprises pharmaceutiques ont tendance à laisser ce secteur à risque à des « startups » spécialisées.

Ces systèmes d'IA génératives ne génèrent pas des textes, comme ChatGPT, mais des molécules nouvelles grâce à un générateur de molécules[13]. Ils exploitent des algorithmes d'apprentissage automatique tels que les réseaux antagonistes génératifs (GAN), les réseaux de neurones récurrents (RNN) et les réseaux de neurones graphiques (GNN), pour générer de nouvelles structures moléculaires. Le système génère d'abord des molécules qui répondent en partie au cahier des charges puis le modèle itère pour affiner la sélection jusqu'à ce qu'il trouve des molécules ayant de très bonnes probabilités de répondre aux critères requis. Les molécules présélectionnées sont testées virtuellement à l'aide de modèles prédictifs (« fine tuning », renforcement), puis la molécule doit être chimiquement validée. Pour ce faire, les molécules doivent être ensuite synthétisées en laboratoire, une phase qui fait aussi appel à des technologie numériques, puis testées. Le gain de productivité par rapport au temps nécessaire jusqu'à maintenant à un chimiste est estimé à 50 % pour la phase de conception et un peu moins pour les phases ultérieures du développement.

Les systèmes d'IA générative peuvent également aider à optimiser des molécules existantes, en générant des modifications de leurs structures moléculaires qui améliorent l'efficacité, la sécurité ou la spécificité des médicaments. Les systèmes d'IA générative peuvent aussi être appliqués à l'identification de nouvelles cibles thérapeutiques grâce à l'analyse de grandes quantités de données, et à l'aide à la conception de médicaments entièrement nouveaux pour cibler des maladies spécifiques. Enfin les SIAgen peuvent aussi aider à prédire les interactions potentiellement dangereuses entre médicaments [12].

Organisation des établissements hospitaliers et relations avec les professionnels de santé

Les systèmes de santé et l'organisation des hôpitaux font face à des difficultés chroniques en raison notamment des manques de personnel et de moyens qui peuvent compromettre leurs missions. Les SIAgen et les modèles de langage de grande dimension peuvent aider à améliorer de différentes manières l'organisation hospitalière pour la rendre plus efficace et ce faisant améliorer la qualité des soins. En voici des exemples.

L'optimisation de la gestion des ressources en moyens humains et en équipements dans les hôpitaux

L'analyse des tendances épidémiologiques, climatiques, historiques peuvent permettre de prévoir les besoins en personnels et en lits d'hospitalisation et les répartir dans différents secteurs en fonction des besoins. Les systèmes de planification automatisée, en s'appuyant sur ces prédictions, devraient permettre d'assigner les ressources de manière plus efficiente.

L'amélioration de la coordination des soins au sein des équipes multidisciplinaires

L'amélioration de la coordination des soins au sein des équipes multidisciplinaires, grâce à des outils de communication intelligents qui pourront mettre à la disposition des professionnels les informations pertinentes pour permettre une prise de décision multidisciplinaire éclairée et rapide.

Les aides à la décision

Les aides à la décision sont traitées par ailleurs dans ce rapport. Elles peuvent être particulièrement utiles à l'hôpital, notamment à l'usage des praticiens juniors, par exemple

[13] Communication de Nicolas Do Huu, Iktos à l'Académie nationale de médecine, 2023.

en signalant les interactions médicamenteuses grâce à des bases de données constamment mise à jour.

Le recours aux SIAgen, comme à un moindre degré au numérique à l'hôpital aura des *conséquences sur les métiers de l'hôpital* : des nouvelles compétences pourront être créées, d'autres disparaîtront, la plupart devront s'adapter aux nouvelles technologies. Dans ce contexte d'évolution forte il sera important que les métiers en contact direct avec les patients soient préservés, voire renforcés.

Les modèles de langage peuvent aider à générer quasiment en temps réel des comptes rendus d'hospitalisation, des comptes-rendus opératoires, des notes cliniques, des comptes-rendus de réunion de concertation pluridisciplinaires (RCP), des lettres de liaison entre l'hôpital et les médecins traitants. Les SIAgen paraissent aussi très bien adaptés à la rédaction des lettres d'information personnalisées adressées aux patients. Dans tous les cas, ces communications écrites nécessiteront une validation humaine attentive et la préservation de la confidentialité nécessaire.

L'automatisation de certaines tâches administratives hospitalières, notamment le suivi des parcours patients, peut être facilitée grâce aux SIAgen par exemple pour permettre d'adapter les plannings en temps quasi réel. L'intervention d'acteurs publics ou privés pour la gestion des rendez-vous médicaux pourra bénéficier des outils de SIAgen et devra être soumise à des règles strictes pour protéger les données personnelles des patients et des professionnels de santé.

La *logistique hospitalière*, comme celle d'une entreprise, a bénéficié de l'apport du numérique pour optimiser les commandes de fournitures médicales, de produits pharmaceutiques, de dispositifs médicaux ou de consommables quotidiens, en fonction des besoins, afin d'aider à réduire les coûts, éviter le gaspillage tout comme éviter les pénuries. Les SIAgen pourront aussi permettre de nouvelles optimisations en temps réel.

Les agents conversationnels sont déjà opérationnels dans certains établissements, pour *guider et informer les patients*. L'utilisation des SIAgen pourra améliorer la qualité et la rapidité des réponses afin de guider les patients à travers les procédures hospitalières tout en réduisant la charge sur le personnel.

Dans toutes ces applications, il est important de permettre un recours à une aide humaine en cas de difficulté avec des outils numériques qui peuvent être deshumanisants et difficiles à maîtriser pour des patients ou des personnes en situation de stress ou en difficulté.

Recherche en biologie et en santé

La recherche et l'innovation en biologie-santé connaissent une transformation significative grâce à l'intégration des SIAgen et des LLM introduits dans la section 2. Cette intégration présage une augmentation de l'efficacité, de la précision et de la vitesse à laquelle les données peuvent être générées, analysées et partagées mais elle n'est pas exempte d'inconvénients importants.

La recherche se heurte aujourd'hui aux limites de nos capacités intellectuelles à colliger, organiser et donner du sens à des données massives et hétérogènes d'origines variées :

- techniques expérimentales à haut débit, notamment en génomique, imagerie ou chimie médicinale ;
- entrepôts de données de santé, épidémiologiques, biologiques et cliniques ;
- production très rapide de connaissances publiées dans la littérature scientifique[14].

Si des progrès importants ont été réalisés dans les capacités de stockage et de vitesse de traitement de ces données, les SIAgen fournissent un apport important pour améliorer les performances de leur extraction et de leur analyse.

Pour décrire les processus biologiques et physiopathologiques complexes, la recherche progresse en nourrissant des analyses systémiques susceptibles de générer des hypothèses originales et interdisciplinaires. Elles devraient bénéficier de la puissance des SIAgen capables d'organiser, de structurer et d'intégrer des données massives et hétérogènes issues de disciplines variées qui travaillent souvent en silo [13]. Si les modèles actuels n'ont pas encore atteint un niveau d'analyse suffisant, des LLM spécialisés à des domaines spécifiques devraient rapidement voir le jour avec la capacité de prédire des phénotypes en intégrant des données allant des analyses génomiques et protéiques jusqu'aux données phylogéniques et environnementales des organismes [14]. De plus, ces modèles peuvent permettent une augmentation du nombre des données considérées. Cette approche interdisciplinaire assistée par les SIAgen ouvre la voie aux études systémiques nécessaires à notre compréhension des maladies multigéniques et multifactorielles qui sont la conséquence d'interactions entre génotype, phénotype et environnement.

De même la compréhension de la survenue des maladies infectieuses et la mise en œuvre de stratégies thérapeutiques adaptées se heurtent à la complexité des facteurs impliqués. Les techniques actuelles de prédiction ont montré leurs limites au cours de la pandémie de la COVID-19, que ce soit pour sa progression ou son amplitude [15]. Les SIAgen pourront être intégrés aux modèles classiques, mathématiques et statistiques pour aider à déterminer des paramètres critiques en temps réel à partir de cas cliniquement documentés. Cependant, les SIAgen n'ont pas encore atteint le niveau de maturité nécessaire et par exemple ont pu produire des résultats erronés sur le COVID-19 du fait de biais méthodologiques bien identifiés depuis [16].

L'impact de l'IA générative intervient à tous les niveaux de la démarche scientifique. Ainsi, la formulation d'une question scientifique s'appuie sur une revue générale de la littérature scientifique qui peut être structurée par l'utilisation de SIAgen. Cependant, l'utilisation de tels outils ne permet pas de sourcer avec acuité le texte produit, ce qui est un manque majeur pour établir une analyse bibliographique documentée. De plus, cette analyse rétrospective, si elle contribue à construire une base de travail ne peut

[14] Par exemple, PubMed comprend plus de 36 millions de citations de littérature biomédicale provenant de MEDLINE, de journaux scientifiques en sciences de la vie et de la santé, et de livres en ligne. https://pubmed.ncbi.nlm.nih.gov/.

suffire à générer des hypothèses originales, par essence prospectives et issues de la créativité et de l'inventivité du chercheur. De même les choix des modèles expérimentaux, des technologies à mettre en œuvre et du plan expérimental gagneront à une extraction structurée de l'existant par l'utilisation de SIAgen, mais avec la limite du non-référencement déjà mentionné et à la condition que l'expérimentateur reste maître de sa démarche scientifique. Un exemple récent concerne l'analyse par un SIAgen approprié des effets secondaires générés par la radiothérapie [17].

Outre son impact à venir dans la pratique médicale [18] évoqué par ailleurs dans ce rapport, les SIAgen présentent un fort potentiel en recherche translationnelle et clinique [19,20]. Ces systèmes devraient contribuer à apporter des informations aujourd'hui difficilement accessibles, et faciliter le déroulement des essais cliniques, notamment :

- l'identification de groupes de patients pertinents au sein de la population, susceptibles d'être inclus dans une étude clinique, avec la possibilité d'appariement aux bases publiques de données de santé (SNDS, PMSI) ;
- la gestion des consentements dynamiques [21] ;
- la définition des « baselines » et le suivi à distance des patients inclus dans les essais ;
- l'intégration de données multiparamétriques dans le corpus de données des essais cliniques, notamment le style de vie, les paramètres environnementaux et les données de génomique ;
- la création de groupes contrôles[15] synthétiques, à partir de patients virtuels basés sur les données de vie réelle.

L'utilisation des SIAgen devrait donc accélérer la conduite des essais cliniques, tout en les enrichissant et en réduisant leurs coûts [22]. Cependant, la puissance d'extraction des données et leur mise en commun dans un réseau mondial dont le contrôle est loin d'être unifié soulèvent des questions éthiques majeures, reprises section 4, sur la protection des personnes, leur vie privée ainsi que sur la pertinence des résultats ainsi obtenus. De plus, les droits de protection intellectuelle doivent être pris en compte dans toute utilisation des SIAgen, ce qui peut soulever des questions sur la pertinence des modèles économiques de l'innovation, de la production de médicaments innovants et de la mise au point de nouvelles stratégies thérapeutiques.

L'utilisation abusive de SIAgen en recherche peut également avoir des conséquences préjudiciables, comme récemment décrit par Taloni et al. [23]. Les auteurs ont utilisé GPT-4 associé à une ADA (*Advanced Data Analysis*), un modèle qui peut effectuer des analyses statistiques et créer des visualisations de données. Le système a permis la comparaison des résultats de deux procédures chirurgicales et a conclu à tort qu'un traitement était meilleur que l'autre, les auteurs ayant fourni volontairement des jeux de données falsifiés. Cette expérience montre qu'il est possible de produire des résultats semblant consistants à partir de fausses mesures sur des patients inexistants, ou de répondre à des questionnaires imaginaires, ou encore de générer un ensemble de données expérimentales sans aucun support. Outre ces conduites condamnables, il ne peut être exclus que, sans intention de falsification, des informations de piètre qualité ou fausses [24] soient incorporées dans un jeu de données et exploitées sans distinction par un SIAgen. L'impact serait double, d'une part sur l'étude désignée et d'autre part dans les études à venir puisque ces résultats erronés peuvent être à leur tour utilisés pour générer de nouveaux modèles de fondation.

La communication des résultats scientifiques vers les pairs est une activité à part entière du processus de recherche. Les SIAgen peuvent avoir un impact important sur les différents modes de transmission des résultats de recherches via les conférences et les présentations visuelles, les publications scientifiques [25] et leurs résumés [26], les thèses et mémoires de stage, les brevets d'invention, les rapports scientifiques et les demandes de contrats. La standardisation des publications en biologie-santé bouleverse les modes de communication des résultats avec une séparation entre la disponibilité en accès ouvert des données, y compris agrégées dans des LLM, et la publication dans des journaux à comité de lecture des hypothèses et des travaux originaux fruits de la créativité humaine qui ne peut être remplacée par des SIAgen. Cette démarche est déjà largement engagée, comme c'est le cas dans d'autres disciplines telles que l'informatique, les mathématiques, la physique des particules, l'astrophysique, la climatologie etc. Pour permettre ce saut qualitatif de la production scientifique, il est essentiel que l'utilisation des SIAgen en recherche soit parfaitement transparente et déclarée dans les publications, et qu'elle soit supervisable par les pairs. De même l'examen de la qualité des publications et des thèses, du recrutement et de l'activité des chercheurs et des demandes de contrats, ainsi que la formation [27] bénéficieront des SIAgen en première intention. La conséquence la plus immédiate est la possibilité de contrôler rapidement les éventuels plagiats et violation de droits d'auteur, et d'évaluer l'originalité des travaux par rapport à l'existant. Mais ces étapes essentielles devront être contrôlées par des experts scientifiques pour garantir l'éthique et l'intégrité scientifique. In fine, si les SIAgen peuvent contribuer aux processus d'évaluation de la recherche, seul un avis collégial par les pairs reste le garant de la validation des données scientifiques, de la justesse des décisions et des orientations stratégiques pour le progrès des connaissances au bénéfice de tous.

Enseignement

La santé est un des principaux domaines d'application des systèmes numériques intégrant des techniques d'intelligence artificielle et l'utilisation de systèmes d'IA génératives (SIAgen) dans le cadre du soin est inéluctable. Il faut distinguer, les SIAgen au service du professionnel de santé dans la pratique de la médecine et donc comment former les futurs médecins à savoir l'utiliser, et d'autre part l'utilisation des SIAgen au service de la formation des médecins dont la valeur ajoutée est d'assister les apprenants dans l'acquisition de connaissances théoriques et du raisonnement clinique.

Former à l'utilisation des systèmes d'IA générative. Le recours à ces outils rend indispensable que tous les futurs

[15] Stéphanie Allassonnière à l'Académie de médecine, 2023.

soignants aient une connaissance minimale des principes de fonctionnement, des principes éthiques et des règlements.

Une formation initiale aux SIAgen intégrant leurs principes de fonctionnement, leurs potentialités et limites médicales et éthiques est indispensable aujourd'hui à tous les professionnels de santé et doit être rendue obligatoire.

Les étudiants en médecine devront continuer à apprendre et maîtriser les connaissances de base. Mais les outils d'IA et d'IAgen pourront efficacement effectuer les tâches répétitives, résumer, et même analyser et proposer des décisions basées sur leur gestion statistique des données.

Les étudiants disposeront ainsi de plus de temps pour le raisonnement, les analyses cognitives, une meilleure compréhension des spécificités de chaque patient, et la communication. Les outils numériques ne remplaceront pas les médecins, mais au contraire des craintes que le numérique et en particulier les SIAgen détournent les étudiants et futurs médecins de la médecine clinique, on peut espérer que l'aide qu'elle apportera en libérant du temps passé à remplir des dossiers, permettra aux étudiants d'apprendre leur métier en passant plus de temps au chevet du patient.

Il est assez vraisemblable que l'essentiel de la formation à ces outils d'IA générative se fera « sur le tas », et l'adaptation des contenus d'enseignement en santé seront liés aux cas d'usage.

Une nouveauté vient d'être imposée par la délégation du numérique en Santé pour toutes les filières d'enseignement en Santé : l'obligation de 28 h de cours de numérique en Santé. Au programme doivent être abordées les données, la cybersécurité, la communication, les outils et la télésanté. Elle se rajoute à différents enseignements indirects des premiers cycles, notamment sur des aspects éthiques et réglementaires dans les modules « santé société humanité », ou sur certains aspects méthodologiques du traitement des données massives dans les modules de statistiques dans les filières médecine/odonto/pharma/maïeutique. Avec une échéance de mise en place à la rentrée 2024. C'est un progrès, mais qui pour le moment ne porte pas de mention spécifique des outils d'IA ou des SIAgen.

L'interdisciplinarité est indispensable au développement et à l'utilisation des systèmes d'IA en santé, combinant trois axes : l'expertise médicale, l'expertise technique informatique et mathématique, et enfin l'expertise éthique, réglementaire et entrepreneuriale. L'interdisciplinarité effective passe par l'effacement des frontières entre ces trois axes, et par l'invention d'un nouveau métier interdisciplinaire l' « AI-designer » en santé.

Les systèmes d'IA générative au service de la formation. Les études de Santé en France sont impactées par deux phénomènes concomitants qui sont à l'origine d'une réflexion générale autour de la formation.

Le premier concerne l'usage intensif du numérique tout au long de la formation des médecins. Par exemple, l'Université Numérique en Santé et en Sport [28] propose un environnement d'apprentissage national, unique en son genre, qui permet un usage innovant du numérique à toutes les étapes du cursus : « e-learning », examens, entraînement, validation de stage, portfolio, saisie des actes, corpus... Cet espace numérique dédié, partagé et utilisé par tous, est un vecteur de diffusion et d'appropriation de l'innovation sans pareil.

Le deuxième phénomène est d'avantage conceptuel puisqu'il concerne la bascule souhaitée vers une approche par compétences de toutes les formations de santé afin que les diplômes dispensés correspondent à des savoir-faire et des savoir-être, en accord avec les attentes des patients et de leurs proches. Ce profond changement pédagogique impose de repenser tout l'enseignement, de la première année de sélection à la dernière année d'internat, avec le développement de centres de simulation et la mise en place de parcours et de suivi personnalisé pour chaque apprenant. Cette transition ouvre le champ des possibles et facilite l'intégration de nouveaux outils comme les SIAgen.

Le principe de mettre les SIAgen au service des apprenants et des enseignants, mais de ne pas chercher à les remplacer, est aujourd'hui largement partagé. Comme ces nouveaux outils impactent les apprenants mais aussi les enseignants, ils offrent l'occasion d'en co-construire les usages. Ainsi, le doyen de la faculté de médecine d'Harvard, Bernard S. Chang récemment interrogé sur cette question, invite ses étudiants et ses enseignants à utiliser largement les systèmes d'IA générative [29]. Les SIAgen redonnent une place majeure au texte et à l'écriture : formuler des questions (dont nous avons vu qu'elles étaient souvent appelées des prompts) à destination de l'outil nécessite une réflexion et un entraînement à la programmation de ces prompts si l'on veut obtenir des réponses pertinentes. Cette initiation est d'autant plus importante qu'elle recouvre des thématiques générales[16] ou bien spécifique à la médecine[17] ou encore des techniques d'attaques des prompts initialement donnés par les concepteurs d'un SIAgen[18].

Les applications des SIAgen dans la formation des futurs médecins sont nombreuses, en particulier : créer automatiquement des épreuves de contrôle dématérialisées avec des réponses commentées ; simuler des cas cliniques réalistes ; produire du contenu pédagogique personnalisé ; enrichir les manuels de formation en proposant des exemples, en illustrant ou en en reformulant le contenu enseigné ou encore en produisant des fiches de synthèse.

En résumé, on peut retenir les messages suivants :

- l'utilisation de systèmes d'IA génératives dans le cadre de la santé et du soin est inéluctable ;
- Une formation initiale aux connaissances générales en numérique et en particulier aux techniques d'IA générative est indispensable à tout soignant et doit être rendue obligatoire ;
- le temps libéré grâce à l'utilisation de ces outils numériques pourra être consacré à l'apprentissage pratique au lit du malade et en centre de simulation ;
- l'usage intensif du numérique pour la formation des médecins proposé par l'UNESS (Université Numérique en Santé et en Sport) devrait être soutenu et étendu ;
- l'approche par compétences et les mises en situation vont réguler les parcours de formation et les SIA vont

[16] Prompt programming for large language models: beyond the few-shot paradigm.

[17] Prompt engineering for healthcare: methodologies and applications.

[18] Ignore previous prompt: attack techniques for language models.

accompagner le suivi individuel nécessaire à chaque étudiant pour mener un parcours de formation personnalisé.

Enjeux d'éthique

Le numérique et en particulier les applications comme les systèmes d'intelligence artificielle générative transforment profondément et durablement l'ensemble de notre société et tout particulièrement l'ensemble de nos systèmes de santé nationaux. Identifié en particulier par Hans Jonas, le décalage entre la rapidité de l'innovation technologique et notre capacité à en comprendre et maîtriser les conséquences est au cœur des enjeux d'éthique. Comme le souligne le CNPEN dans son avis sur les enjeux d'éthique des SIAgen, « Ce décalage est susceptible de générer pendant plusieurs années des tensions anthropologiques, psychologiques, économiques, sociales, politiques et culturelles ». Nous soulignons ici les enjeux d'éthique soulevés par l'utilisation à grande échelle des SIAgen dans le cadre de notre système de santé national, en suivant en particulier les réflexions développées dans les avis 3 et 7 du CNPEN et les avis communs du CCNE et du CNPEN sur le diagnostic médical et l'IA ou encore sur les plateformes de données de santé.

Il ne serait pas éthique de ne pas pouvoir bénéficier de l'aide que peuvent apporter dans le cadre de la santé les systèmes d'IA et en particulier les SIAgen. Mais ces systèmes présentent des risques avérés de leur conception à leur utilisation et à leurs évolutions. Ils doivent donc faire l'objet de considérations précises et claires en considérant les modèles de fondation mis sur le marché et les SIAgen comme des systèmes à haut risque dans le cadre de la finalisation du règlement européen sur l'intelligence artificielle (IA Act).

Le rapport à la vérité et à la signification. Comme nous l'avons explicité dans la section 2, la manière dont les SIAgen sont conçus repose d'abord sur des techniques d'apprentissage machine complétées par différentes approches permettant de renforcer la crédibilité des assertions faites par le système. Mais à aucun moment, dans l'état actuel des techniques mises en œuvre, ni la recherche de vérité, ni la mise en contexte de la signification des textes générés ne sont mises en œuvre. Il peut en résulter des textes (ou des images et vidéos) fondamentalement incorrects dont l'expression peut sembler pertinente mais dont les conséquences doivent être maîtrisées par une supervision humaine comprenant bien les principes de fonctionnement de ces systèmes, à même de prendre le recul approprié sur les résultats produits par un SIAgen et capable de détecter les erreurs et « hallucinations » éventuelles.

Maintenir la distinction entre un texte écrit par un être humain et un texte généré par un SIAgen. Cette distinction est essentielle d'une part pour mettre en œuvre de manière appropriée la supervision humaine dont nous venons de voir l'importance, en particulier dans le contexte médical. Elle l'est d'autre part car les nouveaux contenus générés par un SIAgen peuvent être utilisés pour l'apprentissage de nouveaux modèles de fondation et donc propager le cas échéant des biais ou des informations incorrectes dans la création de nouveaux SIAgen. Mais il faut souligner la difficulté à savoir identifier ou marquer de manière pérenne les contenus textuels générés par une machine alors que ce tatouage (*watermarking*) peut-être de bonne qualité pour ce qui concerne des images générées automatiquement.

Le rôle fondamental de la langue d'expression. Les SIAgen sont construits à partir de corpus d'apprentissage exprimés dans diverses langues humaines. Actuellement l'anglais et le chinois représentent chacun de l'ordre de 30 % de la taille des corpus utilisés alors que le français représente typiquement de l'ordre de 13 %. Or des éléments fondamentaux de la culture, de l'histoire de la pensée, des pratiques quotidiennes sont propres à chaque pays ou région. La prise en compte des enjeux d'éthique sous-jacents à l'utilisation de tel ou tel langage, ou à la spécialisation d'une application à un langage particulier constituent une priorité tant de souveraineté culturelle qu'une prise en compte de spécificités, par exemple médicales, de populations particulières.

Les souverainetés scientifique, technologique et numérique doivent être prises en compte pour nous permettre de maîtriser en France et en Europe les SIAgen de leur conception à leur réalisation et usages. La maîtrise des données, en particulier médicales, doit faire l'objet d'une attention particulière tant dans le respect de leur confidentialité que dans la représentativité de la très grande diversité de la population vivant sur l'ensemble du territoire français. Les pouvoirs publics doivent veiller à ce que les concepteurs de modèles de fondation conçus en France et en Europe puissent prendre en compte nos spécificités culturelles, sociétales et linguistiques, en particulier pour les rendre disponibles dans les SIAgen spécifiques à la médecine et la santé. Cela suppose aussi une réflexion sur les modèles économiques du développement des SIAgen qui, tout en prenant en compte les souverainetés scientifique, technologique et numérique, permette d'une part de respecter la propriété intellectuelle des sources de données utilisées (voir par exemple la position récente du New-York Times [30]) et d'autre part mette en œuvre des modèles économiques équitables et respectueux du bien commun, permettant la mise en œuvre des SIAgen et de leur maintenance et évolutions.

L'éducation au numérique et en particulier aux SIAgen doit être intégrée à la formation de tous les personnels soignants. Ceci permettra d'une part de former à l'utilisation raisonnée des SIAgen en clinique et pour la recherche. Cela doit permettre aussi aux personnels soignants d'expliciter aux patients l'intérêt des outils numériques et des SIAgen, tout en sachant être transparents sur la maîtrise des outils utilisés et en étant à même d'expliquer les résultats fournis par ces outils en termes médicaux.

Maîtriser l'impact environnemental des Systèmes d'IA générative. Nous l'avons vu au travers de différents exemples, les coûts environnementaux et en énergie des modèles de fondation puis leur mise en œuvre dans une grande variété de SIAgen sont particulièrement importants. En particulier pour ce qui concerne les applications en santé, il est nécessaire de développer des métriques des empreintes environnementales de ces différents systèmes aux différentes étapes de leur développement et de leur mise en œuvre et évolutions.

L'académie de médecine recommande

Dans le contexte qui vient d'être présenté et dans une dynamique d'évolution internationale extrêmement rapide, l'Académie nationale de médecine recommande à tous les acteurs du système de santé, publics ou privés, aux citoyens et aux pouvoirs publics, la prise en compte des recommandations suivantes.

- Tous les professionnels de santé doivent être formés à l'usage des Systèmes d'intelligence artificielle générative (SIAGen).
- L'usage des systèmes d'intelligence artificielle générative par les professionnels de santé doit se généraliser ; il serait contraire à l'éthique de se passer de l'aide de ces outils.
- Il faut absolument éviter de communiquer les données personnelles des patients et des professionnels de santé à des SIAgen dont la maîtrise en France ou en Europe n'est pas clairement établie : il n'y a pas de « petites » données.
- Toute utilisation des SIAGen doit faire l'objet d'une supervision humaine et le temps consacré au colloque singulier patient/soignant doit être préservé voire renforcé quels que soient les systèmes numériques utilisés.
- La responsabilité des différents acteurs, concepteurs, utilisateurs, doit être clairement déterminée par une réglementation prenant rapidement en compte le règlement européen sur l'IA pour les systèmes numériques en santé.
- La cyber sécurité des établissements de santé et des plateformes de données de santé doit être traitée comme une priorité première et absolue.
- La propriété intellectuelle des données servant à l'apprentissage des modèles doit être garantie et notamment celle des publications scientifiques.
- Les recherches sur les applications en santé des SIAGen doivent être fortement soutenues au niveau national et européen afin de réduire la dépendance aux entreprises étrangères et de progresser vers une souveraineté numérique européenne.
- L'utilisation des SIAgen doit se faire dans le respect des principes éthiques, notamment développés dans les avis 3 et 7 du Comité National Pilote d'Ethique du Numérique.
- Les impacts environnementaux et énergétique de la conception, la mise en œuvre et l'utilisation des SIAgen en santé doivent être précisément mesurés et pris en compte dans l'évaluation de l'impact environnemental du système de santé.

Déclaration de liens d'intérêts

Les auteurs déclarent ne pas avoir de liens d'intérêts.

Remerciements

Nous remercions Bruno Clément pour sa contribution au chapitre recherche, Maï Kouyaté pour son aide et les membres de la commission 5 pour leurs retours sur la première version du rapport.

Références

[1] Comité national pilote d'éthique du numérique. Systèmes d'intelligence artificielle générative : enjeux d'éthique. Avis 7 du CNPEN, 30 juin 2023. [En ligne] Disponible sur : https://www.ccne-ethique.fr/fr/publications/avis-7-du-cnpen-systemes-dintelligence-artificielle-generative-enjeux-dethique (consulté le 26/03/2024).
[2] Touzet C. Les réseaux de neurones artificiels, introduction au connexionnisme : cours, exercices et travaux pratiques. EC2. Collection de l'EERIE; 1992 https://hal.science/hal-01338010.
[3] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. Proc 27th Int Conf Neural Info Proc Sys 2014;2:2672—80.
[4] Google. Ai across Google: Palm2. [En ligne] Disponible sur https://ai.google/discover/palm2/ (consulté le 26/03/2024).
[5] Liu Y, Jia Y, Geng R, Jia J, Gong NZ. Prompt injection attacks and defenses in LLM-integrated applications. arXiv 2023, 2310.12815(cs) [Submitted on 19 Oct 2023].
[6] BigScience Blog. Introducing the world's largest open multilingual language model: BLOOM. [En ligne] Disponible sur : https://bigscience.huggingface.co/blog/bloom (consulté le 23/03/2024).
[7] Guillaud H. Watson: l'Intelligence artificielle en ses limites. In « Le Monde.fr » [En ligne] Disponible sur : https://www.lemonde.fr/blog/internetactu/2017/10/07/watson-lintelligence-artificielle-en-ses-limites (consulté le 23/03/2024).
[8] Eriksen AV, Möller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. NEJM AI 1; 2023, http://dx.doi.org/10.1056/Alp2300031.
[9] McDuff D, et al. Towards accurate differential diagnosis with large language models; 2023, http://dx.doi.org/10.48550/arXiv.2312.00164.
[10] Marolleau A, Baumard C. IA dans les dispositifs médicaux : 16 sociétés françaises ont obtenu une autorisation de mise sur le marché auprès de la FDA. Publié le 21 novembre 2023. [En ligne] Disponible sur : https://www.mind.eu.com/health/industrie/ia-dans-les-dispositifs-medicaux-15-societes-francaises-ont-obtenu-une-autorisation-de-mise-sur-le-marche-aupres-de-la-fda (consulté le 23/03/2024).
[11] Comité consultatif nationale d'éthique. Comité national pilote d'éthique du numérique. Diagnostic Médical et Intelligence Artificielle: Enjeux Éthiques. Avis commun du CCNE et du CNPEN, Avis 141 du CCNE, Avis 4 du CNPEN; 2022 [En ligne] Disponible sur : https://www.ccne-ethique.fr/fr/publications/avis-141-du-ccne-et-4-du-cnpen-diagnostic-medical-et-intelligence-artificielle-enjeux (consulté le 23/03/2024).
[12] Djeddi, et al. Advancing drug—target interaction prediction: a comprehensive graph-based approach integrating knowledge graph embedding and ProtBERT pretraining. BMC Bioinform 2023;24:488, http://dx.doi.org/10.1186/s12859-023-05593-6.
[13] Hassoun S, Jefferson F, Shi X, Stucky B, Wang J, Rosa E. Artificial intelligence for biology. Integr Comp Biol 2022;61:2267—75, http://dx.doi.org/10.1093/icb/icab188.
[14] Burnett KG, Durica DS, Mykles DL, Stillman JH, Schmidt C. Recommendations for advancing genome to phenome research in non-model organisms. Integr Comp Biol 2020;60:397—401, http://dx.doi.org/10.1093/icb/icaa059.
[15] Kuhl E. Data-driven modeling of COVID-19. Lessons learned. Ext Mech Lett 2020;40:100921, http://dx.doi.org/10.1016/j.eml.2020.100921.
[16] Roberts K, Alam T, Bedrick S, Demner-Fushman D, Lo K, Soboroff I, et al. Searching for scientific evidence in a pandemic: an overview of TREC-COVID. J Biomed Inform 2021;121:103865, http://dx.doi.org/10.1016/j.jbi.2021.103865.

[17] Wu DJ, Bibault JE. Pilot applications of GPT-4 in radiation oncology: summarizing patient symptom intake and targeted chatbot applications. Radiother Oncol 2024;190:109978, http://dx.doi.org/10.1016/j.radonc.2023.109978.
[18] Lee P, Bubeck S, Petro J. Benefits, limits and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388:1233—9, http://dx.doi.org/10.1056/NEJMsr2214184.
[19] Malik S. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11:887, http://dx.doi.org/10.3390/healthcare11060887.
[20] Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023;47:33, http://dx.doi.org/10.1007/s10916-023-01925-4.
[21] Comité consultatif national d'éthique. Avis 136 : l'évolution des enjeux éthiques relatifs au consentement dans le soin. 15 avril 2021 [En ligne] Disponible sur : https://www.ccne-ethique.fr/fr/publications/avis-136-levolution-des-enjeux-ethiques-relatifs-au-consentement-dans-le-soin (consulté le 23/03/2024).
[22] Ghim JL, Ahn S. Transforming clinical trials: the emerging roles of large language models. Transl Clin Pharmacol 2023;31:131—8, http://dx.doi.org/10.12793/tcp.2023.31.e16.
[23] Taloni A, Scorcia V, Giannaccare G. Large Language model advanced data analysis abuse to create a fake data set in medical research. JAMA Ophthalmol 2023;141:1174—5, http://dx.doi.org/10.1001/jamaophthalmol.2023.5162.
[24] Bradley D, Menz et Al. Health disinformation use case highlighting the urgent need for artificial intelligence vigilance, weapons of mass disinformation. JAMA Intern Med 2024;184:92—6, http://dx.doi.org/10.1001/jamainternmed.2023.5947.
[25] Conroy G. Scientists used ChatGPT to generate an entire paper from scratch, but is it any good? Nature 2023;619:443—4.
[26] Else H. Astracts written by CHATGPT fool scientists. Nature 2023;613:423.
[27] Editorial. Why teachers should explore ChatGPT's potential, despite the risks. Nature 2023;623:457—8, http://dx.doi.org/10.1038/d41586-023-03505-5.
[28] Université numérique en santé et sport. Site internet. [En ligne] Disponible sur : https://www.uness.fr (consulté le 23/03/20244).
[29] Hswen Y, Abbasi J. AI will — and should — change medical school, says Harvard's Dean for medical education. JAMA 2023;330:1820—3, http://dx.doi.org/10.1001/jama.2023.19295.
[30] Grynbaum MM, Mac R. The Times Sues OpenAI and Microsoft Over A.I. Use of copyrighted work. The New York times 27 décembre 2023. [En ligne] Disponible sur : https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html (consulté le 23/03/2024).