

Quel rôle pour l'IA générative en médecine générale ?

par M. Sébastien DELETAILLE*

* CEO et fondateur de Rosa
Cantersteen 12
1000 Bruxelles
sebastien.deletaille@rosa.be

L'IA générative est appelée à jouer un rôle grandissant en médecine générale. De manière légitime, de nombreux soignant-es se demandent s'il s'agit d'une véritable avancée pour la santé ?

Si cette technologie est adoptée, ne faut-il pas d'abord surmonter certaines limitations techniques, poser certaines balises éthiques et surtout délimiter son rôle par rapport à l'humain ?

L'auteur déclare ne pas présenter de liens d'intérêts avec l'industrie pharmaceutique ou de dispositifs médicaux en ce qui concerne cet article.

| Prétest | Vrai | Faux |
|--|--------------------------|--------------------------|
| 1. ChatGPT, une application d'IA générative, est devenue l'application la plus rapidement adoptée de l'histoire de l'humanité. | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Il existe déjà des IA médecins généralistes autonomes. | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. En Belgique, le cadre réglementaire actuel interdit l'utilisation d'un logiciel de synthèse de notes de consultation. | <input type="checkbox"/> | <input type="checkbox"/> |

Réponses en page [35](#).

ABSTRACT

This article explores the growing role of generative AI in general medicine, presenting its concepts, challenges, and ethical considerations, while emphasizing the need for a better regulatory framework to maintain the importance of human interaction.

Keywords: generative artificial intelligence, digital health.

RÉSUMÉ

L'article explore le rôle croissant de l'IA générative en médecine générale, en présentant ses concepts, défis et considérations éthiques, tout en soulignant la nécessité d'un meilleur cadre réglementaire pour maintenir l'importance de l'interaction humaine.

Mots-clés: intelligence artificielle générative, santé digitale.

Introduction

L'utilisation de l'intelligence artificielle (IA) générative en médecine générale, et en particulier des systèmes tels que ChatGPT, suscite des réactions divergentes parmi les professionnel·les de santé. Alors que certain·es médecins généralistes voient dans ces technologies un vecteur d'innovation et de progrès, d'autres expriment des réserves et une certaine anxiété quant à leur impact sur le métier de soignant·e, la relation patient·e-médecin et le risque de remplacement ou d'automatisation.

De manière légitime, de nombreuses et nombreux soignant·es s'interrogent quant à l'utilisation de ces technologies au regard de la déontologie médicale, du principe de précaution ou simplement des vagues précédentes de digitalisation qui ont eu tendance à alourdir la charge des soignant·es au lieu de la simplifier. Intégrer des outils comme ChatGPT est-il donc une réelle avancée pour la santé ?

L'objectif de cet article est double. Premièrement, il vise à informer en définissant les concepts fondamentaux et en nommant les défis et considérations éthiques associés à l'adoption de l'IA générative en santé. En examinant quatre applications concrètes de ces technologies en médecine générale, nous illustrerons l'état des lieux en avril 2024. Deuxièmement, cet article propose une réflexion sur les perspectives futures, en mettant en lumière le besoin d'un meilleur cadre réglementaire tant au niveau européen qu'au niveau belge.

Au travers de cette analyse, nous cherchons à fournir aux professionnel·les de santé un aperçu nuancé, invitant les lecteurs et lectrices à réfléchir au rôle



qu'ils ou elles souhaitent accorder à l'IA générative en médecine générale.

Concepts fondamentaux en IA générative

IA vs IA générative ?

L'IA désigne un domaine de l'informatique consacré à la création de systèmes capables d'effectuer des tâches qui nécessitent habituellement l'intelligence humaine. Ces tâches incluent le raisonnement, l'apprentissage, la perception sensorielle, la communication et la prise de décision.

L'IA générative constitue une branche de l'IA caractérisée par sa capacité à générer de nouvelles données, comme des textes (p. ex. ChatGPT) ou des images (p. ex. DALL-E). Pour être « nouvelle », la donnée générée ne peut avoir de correspondant direct avec les données avec lesquelles le modèle a été entraîné.

L'ascension fulgurante de ChatGPT

En novembre 2022, OpenAI, une entreprise américaine, lance ChatGPT¹, un chatbot destiné à répondre à toute requête d'un-e utilisateur-ice, et dont la version 3.5 est en libre accès. Pour répondre aux questions posées, ChatGPT utilise une forme d'IA générative appelée « Grands Modèles de Langage » (« LLM », dérivé de l'anglais « Large Language Model ») qui excelle à simuler le langage humain.

Entraînés sur d'importants corpus de texte, ces modèles acquièrent une compréhension approfondie des structures linguistiques, peuvent générer du texte, traduire des langues, résumer des documents, répondre à des questions et accomplir d'autres tâches linguistiques avec une précision et une fluidité qui se rapprochent de celles d'un humain.

En seulement 2 mois, ChatGPT a atteint plus de 100 millions d'utilisateurs actifs², devenant l'application

adoptée la plus rapidement de l'histoire (détrônant le précédent record détenu par TikTok qui avait atteint le même seuil en 9 mois) (cf. tableau 1).

Dans la course à devenir le leader global de l'IA générative, il est important de noter l'ampleur des montants investis, essentiellement aux États-Unis, par des entreprises qui sont soutenues par les géants de la tech. Pour ne vous en citer que deux, OpenAI est soutenue par Microsoft et Anthropic l'est par Amazon.

Méthodes d'évaluation

Comme tout modèle, l'IA générative peut être évaluée. D'ailleurs, des plateformes publient des comparatifs des modèles (p. ex. Huggingface.co³) selon une série de dimensions :

- **précision linguistique** : mesure la capacité du modèle à générer du texte grammaticalement correct et sémantiquement précis. Cela inclut aussi sa performance sur des tâches spécifiques de traitement du langage comme la traduction, résumer un texte ou la génération de contenu ;
- **compréhension contextuelle** : évalue la capacité du modèle à comprendre le contexte d'une conversation ou d'un texte et à y répondre de manière pertinente, souvent testé à travers des tâches complexes de dialogue ou de réponses à des questions ;
- **cohérence et cohésion** : juge la capacité du modèle à produire des textes logiques et bien structurés sur des segments de textes longs, maintenant la pertinence par rapport au sujet initial ;
- **diversité du texte généré** : analyse si le modèle peut offrir une variété de réponses ou générer du contenu sur un large éventail de sujets sans tomber dans des réponses répétitives ou génériques ;
- **rapidité et performance en temps réel** : mesure le temps de réponse du modèle, crucial pour des applications en temps réel ;
- **efficacité en ressources de calcul** : évalue la quantité de ressources de calcul nécessaires pour exécuter le modèle, ce qui influence directement les coûts d'exploitation et la faisabilité de son déploiement à grande échelle ;
- **robustesse et fiabilité** : détecte la capacité du modèle à gérer des entrées inattendues ou

Tableau 1. Cartographie des principaux acteurs dans le domaine de l'IA.

| Entreprise | Produits | Fonds levés | Pays d'origine | Open source |
|------------|-----------------------|--------------------|----------------|-------------|
| OpenAI | ChatGPT, DALL-E, Sora | > 10 milliards USD | USA | Non |
| Anthropic | Claude | > 6 milliards USD | USA | Non |
| Alphabet | PaLM, Gemini | Non communiqué | USA | Non |
| Mistral.AI | Mistral | ~0.5 milliards USD | FR | Oui |
| Meta | LLaMA | Non-communiqué | USA | Oui |

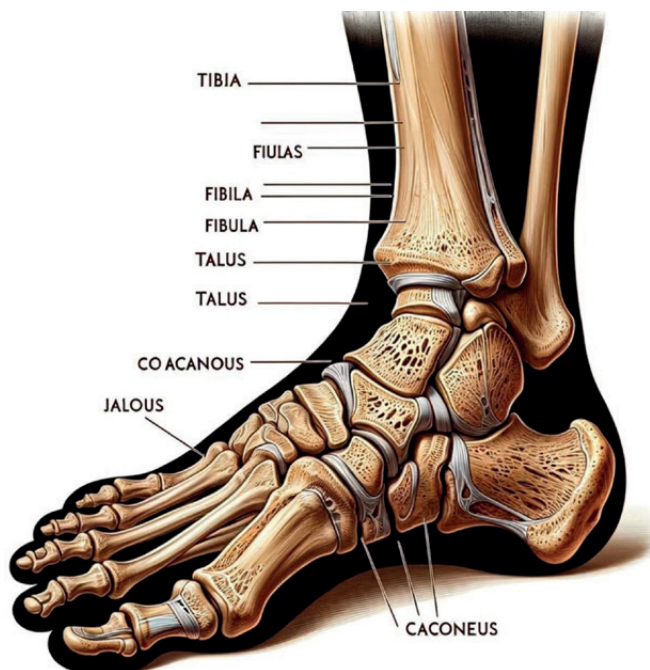
erronées et à produire des réponses de qualité constante, même sous des conditions de stress ou des scénarios d'usage intensif.

Les méthodes d'évaluation reposent sur un large catalogue d'indicateurs standardisés, de bases de données de test, voire de questions d'examens. Les IAs sont aussi évaluées en situation réelle, tant par les utilisateurs que par les chercheurs. En particulier, dans le monde médical, de plus en plus d'études montrent que ChatGPT peut réussir les examens de médecine⁴.

Défis et considérations éthiques

Dans le secteur de la santé, l'utilisation de l'IA générative pose des défis significatifs et soulève d'importantes considérations éthiques :

- **biais et équité** : pour entraîner une IA, il existe une très grande dépendance à la qualité des données d'entraînement. Par exemple, une IA entraînée sur l'ensemble des forums de Doctissimo performera moins bien qu'une IA entraînée sur l'ensemble des ouvrages de référence des études de médecine. Lorsque les données d'entraînement contiennent des biais (p. ex. une surreprésentation de mammographies de femmes eurasiennes), cela peut entraîner une performance moindre pour les groupes sous-représentés⁵ (p. ex. femmes d'origine subsaharienne) ;
- **risques d'hallucination** : les IA génératives peuvent parfois générer des informations fausses ou trompeuses, un phénomène appelé « hallucination ». Par exemple, dans l'anatomie du pied ci-dessous, DALL-E a halluciné de nombreux détails : les noms des os sont parfois incorrects (p. ex. « fiulas » au



lieu de « fibula »), certaines lignes « flottent » sans être liées à une annotation, les os sont poreux... Une hallucination pourrait avoir de graves conséquences dans un diagnostic ou une recommandation de traitement ;

- **responsabilité** : la définition claire de la responsabilité en cas d'erreurs médicales impliquant l'IA est nécessaire. Il faut des cadres juridiques adaptés pour clarifier les obligations tant des développeurs, des fournisseurs de soins que des utilisateurs finaux.

Applications en médecine générale

Dans cette section, nous explorons quatre applications spécifiques de l'IA générative en médecine générale, sans rentrer dans des applications plus générales (p. ex. chatbots...).

Deux points de mise en garde : d'une part, les exemples cités (datant d'avril 2024) peuvent rapidement devenir obsolètes en raison de l'évolution rapide de la technologie ; d'autre part, certaines applications mentionnées sont encore en phase d'évaluation. Ainsi, le principe de précaution prévaut avant tout déploiement à grande échelle.

Synthèse de notes de consultation

Nabla.com⁶ a développé une solution d'IA générative qui synthétise automatiquement les notes de consultation. Lors d'une consultation, Nabla utilise la reconnaissance vocale pour capter en temps réel la conversation entre le-la médecin et le-la patient-e. Cette technologie est spécifiquement entraînée pour reconnaître et orthographier correctement les termes médicaux (p. ex. termes médicaux, symptômes, conditions, noms de médicaments...).

Au cours de la consultation, l'IA analyse et filtre les informations cliniquement pertinentes en écartant les digressions et discussions hors sujet.

Une fois la consultation terminée, le système génère une note médicale détaillée et concise qui peut alors être intégrée au dossier médical informatisé. Les notes générées sont structurées et incluent les codes ICD-10 pertinents.

Cette application a-t-elle été évaluée ? À ce stade, non, aucune équipe de chercheurs n'a publié sur le sujet. Par contre, l'application est librement disponible et Nabla décrit comment l'IA est améliorée de jour en jour pour minimiser les hallucinations⁷ et améliorer la reconnaissance vocale⁸.



Vérification des symptômes

Ada Health⁹ a développé une application de vérification de symptômes qui aide les patient·es à comprendre leurs symptômes et à orienter leurs décisions de soins de santé (N.B.: cette app n'est pas disponible en Belgique à ce jour).

Le processus commence par une série de questions personnalisées que l'application pose au·à la patient·e pour clarifier et détailler les symptômes présentés, affinant ainsi son diagnostic préliminaire. Cette interaction est conçue pour imiter une consultation médicale, aidant à éliminer les causes moins probables et à identifier les symptômes qui nécessitent une attention médicale urgente. À la fin de cette évaluation, Ada fournit un rapport détaillé des conditions médicales possibles, accompagné de recommandations sur les étapes à suivre, qu'il s'agisse de consulter un·e médecin, de surveiller les symptômes ou de se rendre immédiatement aux urgences.

Cette application a-t-elle été évaluée? Continuellement. De nombreuses études montrent qu'Ada est plus efficace qu'un médecin pour toute une série de diagnostics¹⁰.

Aide à la décision d'urgence

Corti.AI¹¹ a développé un « co-pilote » destiné à assister les opératrices des centrales d'appels pour urgences médicales.

Dès qu'un·e patient·e compose un appel d'urgence, Corti commence à évaluer la situation en analysant les mots utilisés, le ton de voix, et d'autres indicateurs acoustiques pour identifier rapidement la gravité potentielle de la situation.

L'IA propose ensuite des recommandations instantanées aux opératrices, les aidant à poser les bonnes questions pour clarifier le diagnostic préliminaire et à donner des instructions de premiers secours, tout en lançant l'envoi d'une ambulance.

Cette application a-t-elle été évaluée? Oui, le co-pilote de Corti augmente la précision des décisions prises et réduit le temps nécessaire pour agir face à un·e patient·e en urgence cardiaque¹².

Réduire les tâches administratives dans le milieu hospitalier

Hippocratic.AI¹³ a été l'une des entreprises pionnières à entraîner un grand modèle de langage en médecine. Leurs équipes viennent de lancer des agents qui reprennent les tâches administratives du personnel hospitalier de manière autonome¹⁴.

Dans sa démonstration, l'agent est capable d'avoir une conversation téléphonique avec un patient, de lui demander de confirmer son identité, de vérifier l'adhésion thérapeutique et de résumer l'appel. D'autres agents existent pour assurer l'admission d'un patient, faire la décharge d'un patient, faire le suivi de factures...

Alors qu'Hippocratic.AI multiplie la création d'agents pour les tâches administratives, ils annoncent que la technologie n'est pas encore assez mature pour poser un diagnostic médical. Il n'existe donc pas encore d'IA médecin autonome.

Cette technologie a-t-elle été évaluée? Oui, l'outil a déjà passé deux phases d'évaluation (sur trois) et montre que l'agent est évalué comme étant plus sûr que du personnel hospitalier dans le cadre des tâches administratives.

Quel rôle pour l'humain ?

L'une des principales objections exprimées par les soignant·es face à l'intégration croissante de l'IA en médecine générale concerne le rôle incontournable de « l'humain » au côté du·de la patient·e.

Avec raison, les médecins soulignent l'importance cruciale des interactions humaines non seulement pour le diagnostic et le traitement, mais aussi pour la communication empathique et le soutien émotionnel qui sont essentiels dans la pratique de la médecine. Une IA ne peut pas regarder le·la patient·e dans les yeux ou lui tenir la main.

Malgré sa précision et sa rapidité, les médecins craignent que l'IA ne puisse pleinement comprendre ni répondre aux nuances subtiles des états émotionnels et psychologiques des patient·es, qui sont souvent révélés à travers des indices non verbaux et contextuels.

Alors que ces craintes sont légitimes, il est essentiel de noter que les sociétés qui développent des agents autonomes entraînent leurs IA à faire preuve d'empathie¹⁵. Dans leurs évaluations internes, ils comparent la perception d'empathie de l'IA à celle d'un humain¹⁶.

Une étude vient même de montrer que lorsqu'un·e patient·e doit évaluer le niveau d'empathie d'un échange écrit, le·la patient·e évalue l'IA comme étant plus empathique qu'un·e soignant·e¹⁷.



Cadre réglementaire manquant

Partout sur terre, la vitesse de l'innovation technologique met en exergue les lacunes du cadre réglementaire. Ce décalage entre innovation et réglementation est particulièrement prononcé en Belgique où l'absence de communication, d'initiatives et de propositions proactives de la part des autorités, des administrations et des corps représentatifs est préoccupante.

Dans certains cas, le cadre réglementaire est même amendé à contre-courant. Par exemple, l'utilisation d'outils de synthèse de notes de consultation intégrés à des outils de téléconsultation est indirectement interdite parce que la législation interdit l'enregistrement d'une téléconsultation¹⁸.

Au niveau européen, l'introduction de l'AI Act¹⁹ par l'Union européenne représente une tentative pour combler ces vides réglementaires en proposant un cadre législatif détaillé pour l'application de l'IA. Toutefois, l'expérience avec des législations antérieures comme le Règlement Général sur la Protection des Données (RGPD) suggère que le risque que ces textes deviennent obsolètes entre leur conception et leur mise en œuvre est réel. Ce retard dans l'adaptation des textes législatifs aux réalités technologiques courantes souligne un problème fondamental : le manque de flexibilité et d'agilité dans les processus réglementaires.

Conclusion

Nous dirigeons-nous vers un avenir où la médecine générale sera entièrement délivrée par ChatGPT? Non. Par contre, nous nous dirigeons vers un monde où l'IA générative va prendre une place de plus en plus prépondérante.

Dans un premier temps, cet article a cherché à expliquer comment des outils tels que ChatGPT arrivent à simuler les compétences clés du/de la médecin : l'écoute attentive, l'analyse, la communication claire et l'empathie. Ces outils sont entraînés pour performer aussi bien qu'un-e soignant-e. Alors que de nombreux défis existent encore, la technologie les adresse et fait d'énormes progrès en très peu de temps.

Cet article a également présenté quatre applications actuelles de l'IA générative en médecine, avec trois objectifs spécifiques : démontrer la variété des applications possibles, souligner que ces technologies sont déjà utilisées dans le monde réel et montrer qu'elles font l'objet d'évaluations scientifiques rigoureuses. Ces développements suggèrent que

nous nous approchons d'un point où l'IA pourrait surpasser la performance d'un-e médecin généraliste médian en termes de connaissances, voire d'empathie.

Le cadre réglementaire représente le principal défi de cette évolution. Comme c'est souvent le cas avec le progrès technologique, c'est à la société de décider la manière dont elle souhaite intégrer ces nouveaux outils : une technologie en soutien aux « ressources humaines » de la santé ou une technologie en remplacement de certaines tâches (ou certains rôles).

À l'avenir, alors que de plus en plus de patient-es pourraient consulter des IA génératives telles que ChatGPT avant de voir leur médecin, la question n'est pas tant de savoir si l'IA jouera un rôle en médecine générale, mais plutôt quel rôle elle devrait jouer. Il est donc crucial que les autorités réglementaires agissent rapidement pour définir quelles tâches doivent impérativement rester sous le contrôle direct des médecins, afin de garantir une intégration sûre et éthique de l'IA dans les pratiques médicales.

Bibliographie

1. ChatGPT [Internet]. [cité 27 avr 2024]. Disponible sur : <https://chat.openai.com>
2. Hu, Krystal. « ChatGPT sets record for fastest-growing user base – analyst note. » Reuters, 2 February 2023, <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01>
3. The Big Benchmarks Collection – a open-llm-leaderboard Collection [Internet]. 2024 [cité 27 avr 2024]. Disponible sur : <https://huggingface.co/collections/open-llm-leaderboard/the-big-benchmarks-collection-64faca6335a7fc7d4ffe974a>
4. Levin G, Horesh N, Brezinov Y, Meyer R. Performance of ChatGPT in medical examinations : A systematic review and a meta-analysis. *BJOG*. 2024 ; 131 (3) : 378-380. <https://doi.org/10.1111/1471-0528.17641>
5. Seyyed-Kalantari, L., Zhang, H., McDermott, M.B.A. et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med* 27, 2176-2182 (2021). <https://doi.org/10.1038/s41591-021-01595-0>
6. Nabla · L'assistant médical qui vous libère du temps [Internet]. [cité 27 avr 2024]. Disponible sur : <https://www.nabla.com/fr/>
7. Evaluating medical note generation · Nabla [Internet]. [cité 27 avr 2024]. Disponible sur : <https://www.nabla.com/blog/evaluating-note-generation/>
8. How we built the best speech-to-text engine for medical encounters · Nabla [Internet]. [cité 27 avr 2024]. Disponible sur : <https://www.nabla.com/blog/speech-to-text/>
9. Ada [Internet]. [cité 27 avr 2024]. Health. Powered by Ada. Disponible sur : <https://ada.com/>
10. Ada [Internet]. [cité 27 avr 2024]. Research and Publications. Disponible sur : <https://ada.com/studies/>
11. Corti – AI for Patient Consultations [Internet]. [cité 27 avr 2024]. Disponible sur : <https://www.corti.ai/>



12. Wenstrup, J., Havtorn, J.D., Borgholt, L. et al. A retrospective study on machine learning-assisted stroke recognition for medical helpline calls. *npj Digit. Med.* 6, 235 (2023). <https://doi.org/10.1038/s41746-023-00980-y>
13. Hippocratic AI [Internet]. [cité 27 avr 2024]. Hippocratic AI. Disponible sur: <https://www.hippocraticai.com>
14. Hippocratic AI [Internet]. [cité 27 avr 2024]. Video. Disponible sur: <https://www.hippocraticai.com>
15. Mukherjee, S., Gamble, P., Ausin, M.S., Kant, N., Aggarwal, K., Manjunath, N., Datta, D., Liu, Z., Ding, J., Busacca, S. and Bianco, C., 2024. Polaris: A Safety-focused LLM Constellation Architecture for Healthcare. <https://arxiv.org/abs/2403.13313>
16. Hippocratic AI [Internet]. [cité 27 avr 2024]. Foundation Model. Disponible sur: <https://www.hippocraticai.com/foundationmodel>
17. Mukherjee, S., Gamble, P., Ausin, M.S., Kant, N., Aggarwal, K., Manjunath, N., Datta, D., Liu, Z., Ding, J., Busacca, S. and Bianco, C., 2024. Polaris: A Safety-focused LLM Constellation Architecture for Healthcare. <https://arxiv.org/abs/2403.13313>
18. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med.* 2023 Jun 1 ; 183 (6): 589-596. doi: 10.1001/jamainternmed.2023.1838. PMID: 37115527; PMCID: PMC10148230. <https://pubmed.ncbi.nlm.nih.gov/37115527/>
19. Remboursement des consultations médicales à distance | INAMI [Internet]. [cité 27 avr 2024]. Disponible sur: <https://www.inami.fgov.be/fr/themes/soins-de-sante-cout-et-remboursement/les-prestations-de-sante-que-vous-rembourse-votre-mutualite/prestations-a-distance-par-des-medecins>
20. AI Act | Shaping Europe's digital future [Internet]. 2024 [cité 27 avr 2024]. Disponible sur: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

EN PRATIQUE, NOUS RETIENDRONS

1. L'IA se développe de manière fulgurante et ses applications en médecine générale vont s'accroître rapidement.
2. Les défis à relever sont les biais et l'équité, les risques d'hallucination et la responsabilité en cas d'erreur médicale.
3. Actuellement, les applications en médecine générale vont de la synthèse de notes de consultation à l'aide à la décision d'urgence.
4. Il n'est pas question de remplacer le médecin mais il est important et urgent de définir et de régler le rôle de l'IA en médecine générale en tant que soutien à la pratique, pour une utilisation éthique et sûre.

La Rédaction