

Consulter “Docteur ChatGPT”, est-ce une bonne idée ?

Santé Open AI, Anthropic ou Perplexity sont en train de développer des versions de leur chatbot dédiées uniquement à la santé. Quels sont les avantages et les risques pour le patient ?

C'est une nouvelle tendance dans laquelle s'engouffrent les grandes entreprises de l'intelligence artificielle. Open AI, Anthropic ou Perplexity sont en train de développer des versions de leur chatbot IA (ChatGPT, Claude...) dédiées uniquement à la santé. Des versions tests sont disponibles depuis ces dernières semaines et sont accessibles dans plusieurs pays, avant un lancement plus global. Le potentiel est énorme: les thèmes en lien avec la santé figurent en effet parmi les plus nombreuses questions posées aux modèles massifs de langage (LLM). *“Les acteurs qui poussent à ces interfaces dédiées à la santé ont développé ces sites car ils ont parfaitement compris l'enjeu économique majeur, cadre le D^r Jean-Emmanuel Bibault, spécialiste français de l'usage de l'IA en médecine. Chaque semaine, plus de 230 millions de personnes posent des questions médicales et de santé à ChatGPT. Soit 10-15 % des requêtes totales des téléphones. C'est énorme.”*

L'ambition de ces nouvelles interfaces est d'aller au-delà des versions classiques de ChatGPT et consorts: proposer aux usagers des réponses plus utiles et plus personnalisées, en s'appuyant sur leurs données réunies par l'application: documents médicaux téléversés, données d'applications de bien-être, dossiers médicaux parfois, selon les pays et les conditions d'accès. Chat GPT Health doit ainsi *“combiner l'intelligence de ChatGPT et vos informations de santé, de manière sécurisée”*, vante Open AI. Les versions d'Anthropic et de Perplexity sont comparables.

Une vieille histoire

Depuis l'aube d'Internet, les patients ont toujours recherché des informations en ligne concernant leur état de santé. Mais désormais, les patients s'appuient de plus en plus sur des réponses générées par l'IA, observe Catherine Lietaer, médecin généraliste à Woluwe-Saint-Lambert. *“Dans les requêtes faites à ces chatbots, il y a de tout. Cela peut viser à comprendre le résultat d'un examen pour lequel ils ne verront pas le spécialiste avant un mois. Ou des demandes du type: Mon voisin a un cancer du côlon, que dois-je faire pour éviter cette maladie? Mon médecin m'a prescrit ce médicament, est-ce vraiment le meilleur? J'ai mal au ventre, à la tête et de la fièvre, qu'est-ce que ça peut être?”*, énumère la médecin.

Pour la D^r Lietaer, ce recours à l'IA est à valoriser, car *“un patient qui s'intéresse à sa santé et cherche à comprendre est quelqu'un qui a envie de se prendre en main. Que ChatGPT conseille à l'utilisateur de poser telle ou telle question au médecin peut en outre être constructif. Son diagnostic peut également donner des idées au médecin. Vu l'usage grandissant par les patients, on va en tout cas devoir composer avec. Ce serait une erreur de dire: ‘N'allez pas voir sur Internet’. Mais comme en plomberie, il faut savoir reconnaître quand passer la main à un spécialiste. Je vois que les patients tentent d'utiliser l'IA au lieu d'aller chez le médecin, pour ne pas devoir attendre pour un rendez-vous, par exemple. Mais une vraie réponse médicale s'accompagne de l'examen de la personne. Par ailleurs, pour pouvoir donner un traitement vraiment personnalisé à un patient, il y a énormément de paramètres et l'IA n'aura jamais accès à tous.”*

L'intelligence artificielle diffère entre outre des “simples” recherches en ligne effectuées jusqu'ici sur différents sites web, dont les patients se méfient a priori. *“Ici,*

ils ont peut-être une confiance un peu trop aveugle en une IA censée déjà avoir trié l'information pour eux, regrette celle qui est aussi administratrice de la Société scientifique de médecine générale (SSMG). Et comme toujours avec l'IA, si on est capable d'évaluer la réponse, on peut voir si elle part dans le bon sens ou pas. Si on n'est pas du tout compétent, on peut très vite se faire leurrer par une IA qui aura halluciné.”

Il hallucine toujours

Car malgré ses améliorations certaines au cours de ses diverses itérations, ChatGPT souffre en effet toujours d'hallucinations en matière de santé. C'est-à-dire qu'il peut produire un résultat absurde tout en affichant une grande confiance en lui. *“Le risque d'hallucinations, lorsque le système utilise vos propres données, est certes minimisé, mais il existe bel et bien, car le modèle exploite toujours une grande quantité d'informations issues de ses données d'entraînement (données publiques provenant d'études et autres), prévient le D^r Henry Bair, spécialiste américain de l'usage du numérique en médecine. Par exemple, si vous êtes un patient qui a du cholestérol et que vous lui demandez ce que les médecins recommandent avec un taux précis, il y a une grande possibilité d'hallucinations. Parce qu'il ne comprend pas vraiment la question que vous posez... Il se contente d'analyser les données d'entrée: le nombre, le mot ‘triglycéride’, le mot ‘cholestérol’, et consulte sa base de données pour déterminer, statistiquement, ce qui est le plus susceptible d'apparaître. Il ne sait pas ce qu'est le cholestérol!”*

“ChatGPT ne comprend pas la question que vous lui posez. Le cholestérol, il ne sait pas ce que c'est!”

D^r Henry Bair

Ophthalmologue et spécialiste de l'usage numérique en santé

De récentes recherches mettent aussi déjà en évidence les manquements lorsque ChatGPT Health réalise du “triage médical” (quelle action choisir en cas de souci de santé). Dans une étude visant à

évaluer les risques du chatbot, une équipe de l'hôpital Mont Sinai (New York) a créé des dizaines de cas sanitaires. Ces scénarios allaient d'affections bénignes, comme un rhume, pouvant être gérées à domicile, à des cas nécessitant une prise en charge immédiate, comme une crise d'asthme sévère. *“ChatGPT Health a obtenu les meilleurs résultats en matière de triage des cas modérément urgents, nécessitant par exemple une consultation médicale dans les deux jours, détaille Alvira Tyagi, une des auteurs. Malheureusement, nous avons constaté des défaillances importantes aux deux extrêmes: les cas tout à fait non urgents et les véritables urgences. Pour les cas non urgents, plus de 50 % des situations étaient surtriées. Quant aux véritables urgences, nous avons constaté un sous-triage dans 52 % des cas. Donc, dans plus de la moitié des cas, le système recommandait des soins moins urgents que ceux que les médecins jugeraient appropriés dans la situation. C'est très préoccupant, car ce sont précisément ces types de situations qui risquent d'empêcher un patient d'avoir les soins qu'il mérite.”*

Un patient chinois n'est pas un patient européen

La qualité des données intégrées dans ces modèles de langage santé (applications de bien-être pas toujours fiables, erreurs fréquentes dans les dossiers...) tout comme leur origine géographique posent également question. *“Pour l'instant, ces modèles sont entraînés sur des manuels d'écoles de médecine, etc., mais avec leur sophistication potentielle, dans le futur, ils le seront de plus en plus sur des données patients, note le D^r Bibault. Si ces données proviennent de patients américains ou chinois et sont utilisées en*