

HUG: Hôpital cantonal de Genève

mardi 5 novembre 2023

Chat GP-T, Les chatbots et la pratique de la médecine

Dre Meyssan Nehme, Dr Frédéric Schneider

Quelques bases sur la technologie

Lorsque l'on donne: "Une hernie inguinale" à ChatGPT 3.5, ça se tourne vers 175mia de paramètres, 300 mia de "training tokens" et 50 mille mots de vocabulaire, pour compléter la phrase en donnant physiopathologie, symptômes, causes...

Il est également possible de nourrir une base de donnée à chatGPT et de l'entraîner sur celle-ci à travers un renforcement PPO, ce qui donne un modèle spécifique.

Le renforcement PPO "proximal Policy Optimization" est proche du modèle pavlovien: L'IA teste des solutions et fait des erreurs. A chaque pas réussi, elle reçoit une récompense virtuelle, ce qui la pousse vers les solutions justes.

Le RAG: Reality Augmented Generation

ChatGPT a appris énormément de données disponibles sur internet mais cela reste parfois limité, l'IA pose alors la question au modèle RAG, qui va chercher la réponse sur internet, et sur une éventuelle base de données privée fournie par l'utilisateur.

Contrairement au renforcement PPO, cette méthode demande peu de temps et de ressources. Il permet à chatGPT de fournir des réponses plus précises et complètes.

Le projet confiance

Soutenu par la fondation privée des HUG, ce projet a pour but de déterminer l'utilité d'un chatbot destiné aux patients en médecine de premier recours.

Les stratégies du SMPR sont nourries à un chatbot pré-entraîné, en utilisant un modèle de langage large (LLM). Ces informations scriptées pour des professionnels vont donc devenir accessibles, en termes de langage notamment, pour le tout-venant.

Après des cycles de test et de couches de contrôle, le projet est mis sur pied pour étude.

Plus sur le projet confiance en fin de colloque...

Les chatbots sont étudiés dans de nombreux domaines en médecine: maladies chroniques (diabète, BPCO, IC...), oncologie (diagnostic radiologique, prise de décision, soutien émotionnel...), comportements (tabac, hygiène...), populations vulnérables (personnes âgées, proches aidants...)

C'est particulièrement intéressant pour la catégorie comportements, où le chatbot peut être présent juste au bon moment dans le combat contre une addiction, par exemple.

Toutefois, les études restent petites et ciblées sur une seule problématique. Il manque un chatbot avec une prise en charge plus transversale, ce qui est plus complexe.

Evolution des chatbots

- 1966 ELIZA: premier chatbot capable de simuler une conversation
- 1995 ALICE, développement de l'AIML (Artificial Intelligence Mark-Up language)
- 2011 SIRI (apple), premier assistant vocal
- 2014 Xiaoice (Microsoft Asie), 660 mio d'utilisateurs en 2022
- 2022 ChatGPT 3.5 (OpenAI), 100 mio d'utilisateurs en 2 mois

Acceptabilité

En 2019, [une étude](#) montre que c'est plutôt accepté et utilisé pour: rechercher des services de santé locaux, des informations générales sur la santé, et pour prendre un RDV médical.

Les utilisateurs sont également ouverts à utiliser les chatbot pour faire des recherches plus générales sur la santé, bien que la fiabilité perçue soit un obstacle, tout comme l'attitude individuelle à l'égard des robots et des conversations avec un ordinateur.

Le chatbot "HUG contact" répond aux questions administratives concernant les HUG depuis septembre 2021, avec une supervision humaine à chaque étape. Sur 26'600 utilisateurs uniques, le bot a pu répondre à 76% (match) des questions, ce qui est supérieur à la moyenne.

Avant le chatbot, il y avait un formulaire. Dans la première semaine de mise en service, il y a 4x plus d'interactions avec le chatbot qu'avec le formulaire. La clef c'est la supervision humaine, qui permet de répondre aux questions trop complexes pour le bot, ce qui aurait un effet décevant autrement.

Mots clefs les plus recherchés: covid, dépistage, rdv puis → contact service, rdv, facturation

Le chatbot "Rafael post-covid"

Inspiré de "HUG contact", il arrive en 2021 pour soutenir la population avec des questions sur le post-covid. Il sait dire qu'il ne sait pas, plutôt que d'halluciner une réponse fausse, ce qui est essentiel.

Il est basé sur un processus de langage naturel (NLP, natural language processing). Cela demande au bot d'analyser les intentions derrière la question pour pouvoir sélectionner la réponse qui lui correspond, ce qui est limitant.

Son taux de réussite est encore meilleur, à 82%. Il permet une économie de 500k-800k USD. Les mots clefs les plus recherchés sont: vertige, fatigue, maux de tête.

La nouvelle génération de chatbot LLM, Largeur Modèle de Langage, et le IA conversationnelles représentent un ensemble de fonctionnalités:

- Reconnaissance vocale
- Traduction automatique
- Génération de langage naturel
- Reconnaissance de caractères
- Reconnaissance d'écriture manuscrite
- Recherche d'informations

Les chatbots LLM sont limités de façon importante par:

- Le coût environnemental de l'entraînement et de l'utilisation: très énergivores
- Biais issus des jeux d'entraînements tirés d'Internet
- Difficulté à comprendre le contexte, en particulier s'il est implicite: même si le chatbot à l'air de comprendre, une grande partie du contexte lui échappe.
- Risque d'hallucinations: Erreur factuelle, "mensonge", fréquent

Le [Fondation Model Transparency Index](#) permet de classer les nouveaux chatbots. Il prend en compte la transparence sur les données d'entraînement, la protection des utilisateurs et la fiabilité de l'outil.

Une majorité des outils disponibles sur le marché sont extrêmement peu transparents. Seuls deux (Llama 2 de Meta et BLOOMZ de Big Science) ont un score > 50%.

Les biais des IA conversationnelles

- Priorisation des données d'entraînement en anglais
- Renforcement des stéréotypes de genre, race ou religions
→ Il y a plus de probabilités que chatgpt donne aux pédiatres un pronom masculin, alors que c'est une spécialité majoritairement féminine (en anglais du moins).
- Pas de prise en compte des évolutions sociales: La connaissance de la situation sociale s'arrête à la fin de l'entraînement.
- Biais de confirmation, en lien avec la façon dont la question est posée. Un biais qui n'était pas un problème avec les moteurs de recherche.

Enjeux éthiques:

- Confidentialité des données, fiabilité, transparence
- Comment les chatbots vont modifier la relation patient-médecin

Les chatbots représentent un changement majeur dans la recherche d'information sur la santé. Microsoft et d'autres moteurs de recherche ont déjà intégré un chatbot.

[Une étude](#) compare chatGPT et le moteur de recherche de Google pour des questions d'éducation aux patient.es sur la dépression du post-partum.

Les réponses de chatGPT sont bien meilleures que Google, plus complètes. Néanmoins, et en particulier chez les personnes qui n'ont pas l'habitude, cela comporte encore des risques.

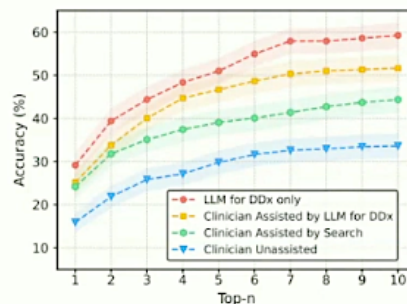
L'IA et les chatbots doivent renforcer la connexion humaine, la relation médecin-patient et non la remplacer. C'est à garder en tête à chaque étape de création de ces outils.

Cas d'usages de chatGPT: Assistance à la décision médicale, Gestion de l'information (rapports), Education des médecins (vignettes cliniques), parcours de soin.

Il y a quelques jours, une [nouvelle étude](#) compare la précision d'une évaluation médicale par des cliniciens à celle de cliniciens soutenus par un moteur de recherche ou un LLM.

La précision diagnostique est fort supérieure chez les LLM, même comparé aux cliniciens + LLM.

Cependant, l'utilisation des IA et des chatbots est formellement proscrite pour l'assistance à la décision médicale, car la transmission des informations du patient ne peut être correctement protégée.



Le projet confiance va offrir une alternative locale de confiance, basée sur les stratégies SMPR, avec une évaluation éthique, juridique, et de la sécurité d'utilisation et des données.

