

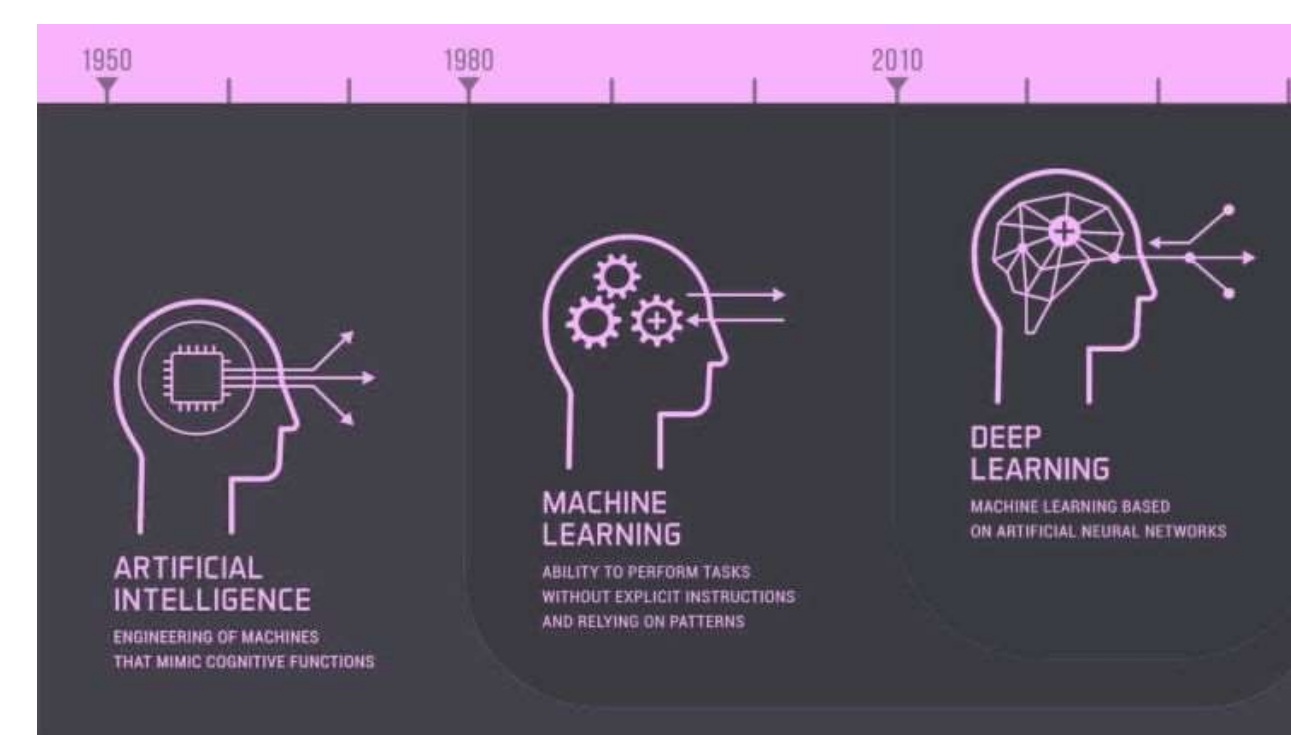
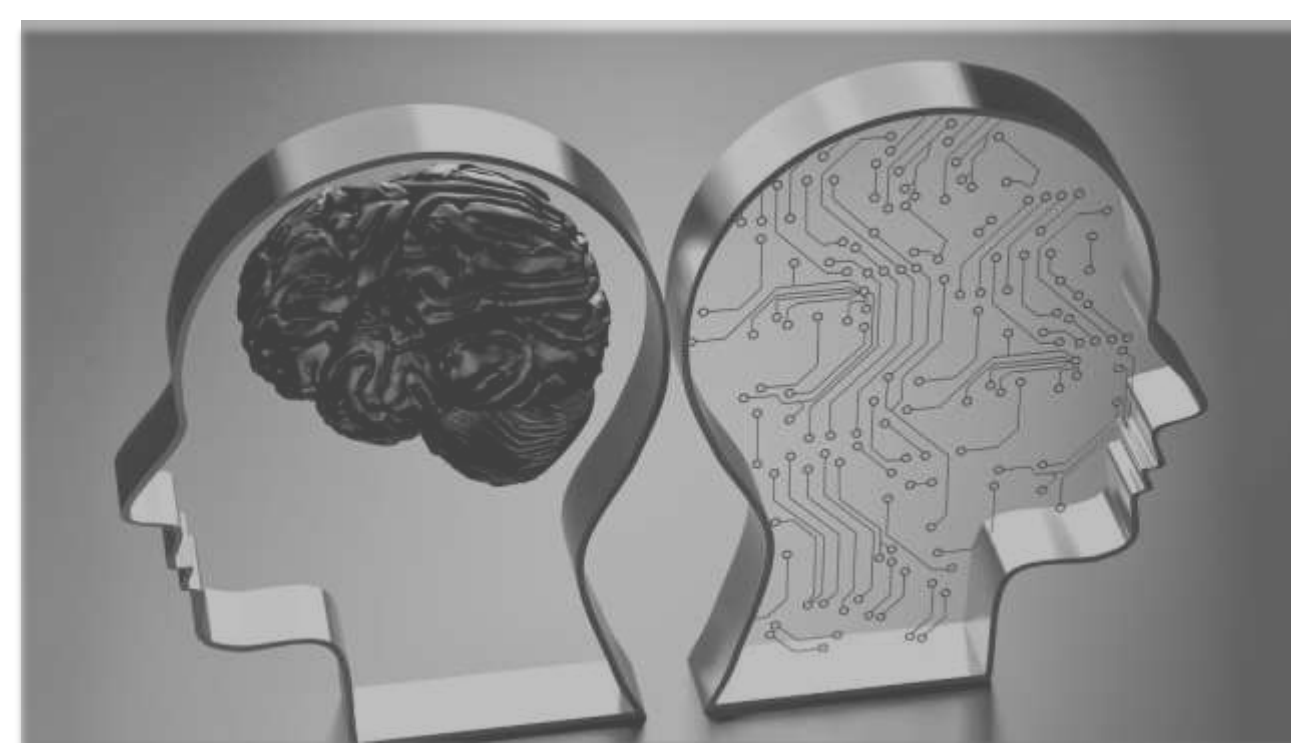
ChatGPT 3.5 versus divers niveaux d'experts médicaux dans le diagnostic des maladies ORL: Une étude comparative transversale

Mikhael Makhoul¹ · Antoine E. Melkane¹ · Patrick El Khoury¹ · Christopher El Hadi¹ · Nayla Matar¹

¹ Département d'Oto-rhino-laryngologie et de Chirurgie Cervico-Faciale, Hôtel-Dieu de France, Université Saint-Joseph, Boulevard Alfred Naccache, BP : 166830 Achrafieh, Beyrouth, Liban.

Introduction

- Les avancées récentes en intelligence artificielle (IA) ont permis la création de modèles linguistiques sophistiqués tels que le Chat-Based Generative Pre-Trained Transformer (ChatGPT). Lancé en novembre 2022, ChatGPT est l'application à la croissance la plus rapide de l'histoire, atteignant 100 millions d'utilisateurs actifs dès février 2023.
- Dans le secteur médical, ChatGPT pourrait transformer radicalement le diagnostic grâce à sa capacité à analyser en profondeur de vastes ensembles de données médicales et à apprendre de divers cas cliniques^{1,2,3,4}.
- Cependant, l'intégration de l'IA dans les soins de santé soulève des questions d'éthique et de confidentialité des données. Dans ce contexte, il est crucial de procéder à des tests rigoureux pour déterminer la fiabilité de ChatGPT en tant qu'outil d'aide aux soins de santé.



Objectifs

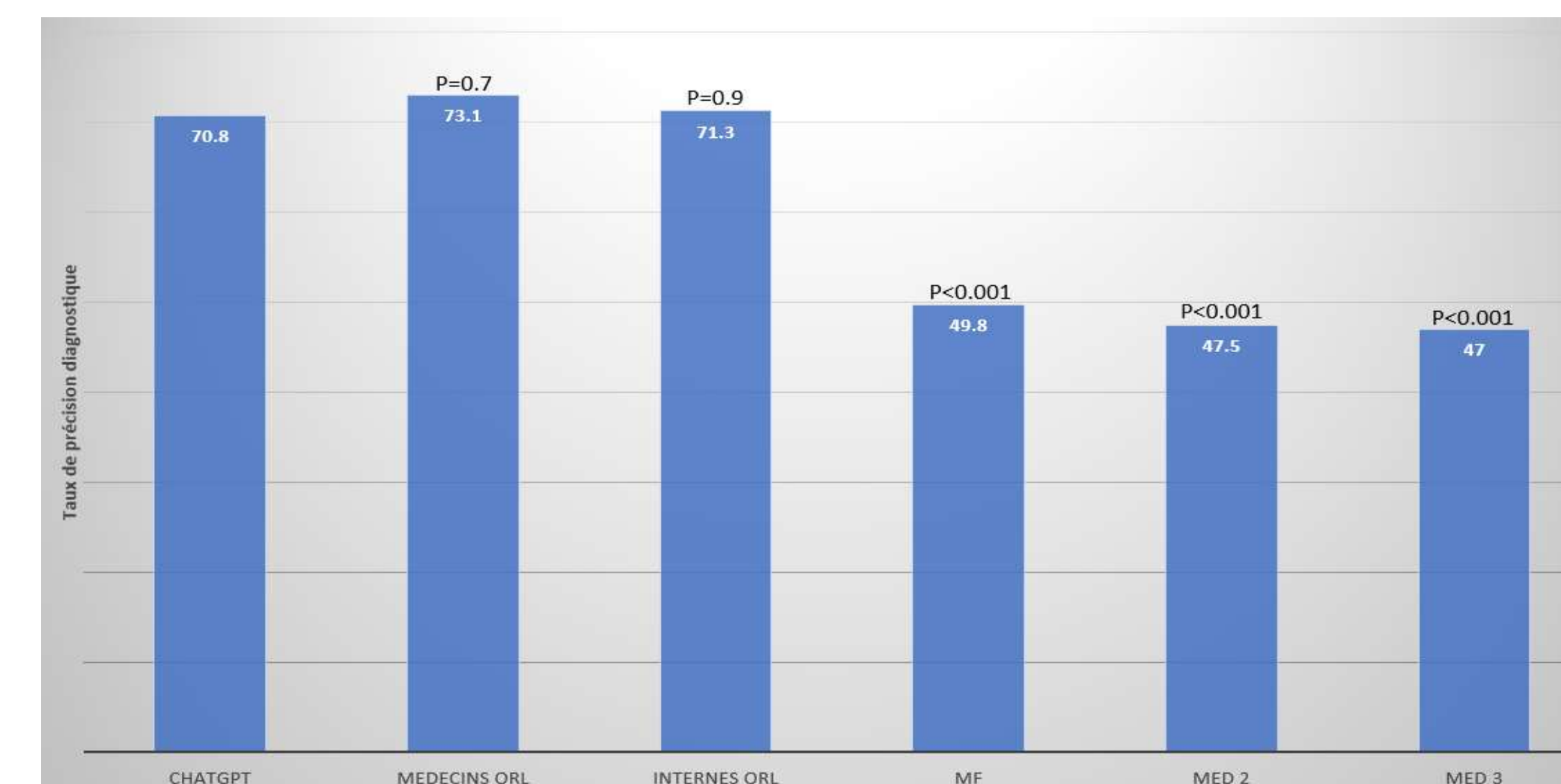
- Evaluer la précision diagnostique de ChatGPT pour diagnostiquer les pathologies ORL.
- Evaluer la précision diagnostique de ChatGPT par rapport à différents niveaux d'experts médicaux.
- Evaluer la capacité de ChatGPT à prioriser les diagnostics les plus critiques dans le domaine des pathologies ORL.

Méthodes et Matériels

- Trente-deux cas ont été présentés à ChatGPT 3.5, ainsi qu'à des médecins ORL, internes ORL, spécialistes en médecine de famille (MF), étudiants en deuxième année de médecine (Med2) et étudiants en troisième année de médecine (Med3).
- Huit cas réels ont été utilisés pour chacun des quatre principaux symptômes ORL : dysphonie, odynophagie, congestion nasale et otalgie
- Chaque participant a fourni trois diagnostics différentiels pour chaque cas.
- L'étude a analysé les taux de précision diagnostique ainsi que le taux de diagnostic le plus critique.
- Pour déterminer le diagnostic le plus critique, toutes les réponses de ChatGPT pour chaque cas ont été évaluées par deux enquêteurs, notant les diagnostics les plus critiques mentionnés.

Résultats

Le taux de précision de ChatGPT est de 70,8 %, sans différence significative par rapport aux médecins ORL ou aux internes ORL. Une différence significative existe entre ChatGPT et MF (49,8 %, $p < 0,001$), ainsi qu'entre ChatGPT et les étudiants en médecine (Med2 47,5 %, $p < 0,001$; Med3 47 %, $p < 0,001$).



Dans 68,75 % des cas, ChatGPT n'a pas mentionné le diagnostic le plus critique.

Conclusion

ChatGPT a démontré une précision comparable à celle des médecins et internes ORL pour diagnostiquer les pathologies ORL, surpassant les MF, ainsi que les Med2 et Med3. Cependant, il a montré des limitations dans l'identification du diagnostic le plus critique.

Références

- 1- Barat et al. (2023) "Appropriateness of recommendations provided by ChatGPT to interventional radiologists". *Can Assoc Radiol J*.
- 2- Rao et al. (2023) "Assessing the utility of ChatGPT throughout the entire clinical workflow". *Health Inform*.
- 3- Strong et al. (2023) "Performance of ChatGPT on free-response, clinical reasoning exams". *Med Educ*.
- 4- Cao et al. (2023) "Accuracy of information provided by ChatGPT regarding liver cancer surveillance and diagnosis". *Am J Roentgenol*.
- 5- Makhoul et al. (2024) "A cross-sectional comparative study: ChatGPT 3.5 versus diverse levels of medical experts in the diagnosis of ENT diseases". *Eur Arch Otorhinolaryngol*