

**Thiago Gonçalves dos Santos Martins<sup>1</sup>, Diogo Gonçalves dos Santos Martins<sup>2</sup>, Paulo Schor<sup>3</sup>**

<sup>1</sup>Universidade Federal do Rio de Janeiro, Campus Macaé

<sup>2</sup>Hospital Central da Aeronáutica

<sup>3</sup> Universidade Federal de São Paulo, Campus São Paulo

---

## Can we help GPT chat get smarter?

---

ChatGPT is a chatbot that uses deep learning algorithms to analyze and recognize patterns in text data, allowing it to predict and generate new word sequences in a contextually relevant way. It can be used for educational purposes, helping medical students and professionals access information and gain a better understanding of medical concepts. In addition, it has applications as aiding in diagnoses, clarifying medical doubts and supporting health professionals in their decision-making processes.<sup>1-3</sup>

The performance of this artificial intelligence has already been tested in areas of knowledge of medicine<sup>2,3</sup>. We decided to evaluate their ability to retain knowledge from the interaction with the user. We applied multiple-choice questions from the 2022 ophthalmology Brazilian Board test to the Gpt Chat algorithm. Questions with images were excluded. The responses applied were recorded and compared with the official answer template. When the algorithm got the answer wrong, we provided the correct answer. The algorithm flagged the error and immediately recognized the given answer as correct. The same test was applied one week apart, and the performance of the algorithm was compared.

In the first test of 50 questions, the algorithm had a correct answer of 26 questions in the first attempt and a correct answer of 22 questions in the second attempt. This test had 4 questions canceled by the bank and a question with a figure that was not used in the test. The performance of the algorithm then varies from 57.7% in the first attempt to 48.8% in the second attempt.

The second test consisted of 125 questions and none of them were canceled. The performance of Chat Gpt was 48.8% correct with a percentage of 46.4% on the second attempt.

In both attempts, the performance was lower than desired to be approved. Other chatbots tested such as Gabriel also had security locks for inputs of user-supplied information<sup>4,5</sup>.

New versions of the application are already being used, performing better in multiple-choice exams, capable of generating text in more languages, understanding images, and processing a greater number of words, ensuring a more fluid conversation with the user.<sup>4,5</sup>

While language models can significantly benefit from user feedback through fine-tuning or adaptive learning, there's a critical need to establish safeguards against potential risks inherent in this process. While user input can undoubtedly enhance the model's accuracy and predictive capabilities, there's a fine balance between improvement and the potential introduction of biases, misinformation, or privacy breaches. Implementing protective measures such as data sanitization, bias detection, limited exposure to feedback, and transparent feedback mechanisms is crucial. By ensuring careful oversight, transparency, and ethical considerations, language models can harness the power of user feedback responsibly, enhancing their performance while maintaining integrity and trustworthiness. While learning from users can lead to better performance, it also comes with potential risks, such as:

**Biases:** If the model learns from biased or inaccurate information, it can perpetuate and magnify those biases, leading to incorrect or unfair results.

**Malicious Inputs:** Allowing user feedback can open the system up to manipulation as malicious actors can try to influence the model with false or harmful information.

**Quality Control:** Ensuring the accuracy and quality of user-provided feedback is

crucial. Mechanisms should be in place to verify and moderate information, preventing misinformation from being incorporated into the model.

To address these concerns, a robust and transparent system with adequate quality control mechanisms would be essential. Implementing safeguards to avoid potential problems is vital when incorporating user feedback into language models.

Therefore, we can conclude from the conducted experiment that the performance of both evaluated algorithms did not improve after the correct responses were provided, indicating some mechanism of protection against user learning.

In conclusion, finding a balance between openness and security is paramount for fostering the development of increasingly potent and beneficial AI tools. As AI technology progresses, we anticipate that advancements in this domain will lead to the creation of even more efficient and dependable AI systems. Continuous evaluation and refinement are indispensable when utilizing language models like GPT Chat. AI models stand to gain from regular updates and retraining on pertinent data to accommodate new information and enhance their accuracy. By prioritizing this iterative process, we can ensure that AI technologies continue to evolve in a manner that maximizes their positive impact while safeguarding against potential risks.

## References

- 1- BISWAS, S. S. Role of Chat GPT in Public Health. **Annals of Biomedical Engineering**. v. 51, n. 5, p. 868-869, 2023.
- 2- GOBIRA, M. C.; NAKAYAMA, L. F.; MOREIRA, R. C.; REGATIERI, S. V.; ANDRADE, E. & BELFORT Jr, R. Performance of chatGPT 3.5 answering questions from the Brazilian Council of Ophthalmology Board Examination. **The Pan-American Journal of Ophthalmology**, v.5, n.1, p.17, 2023.
- 3- HOCH, C. C., WOLLENBERG, B., LÜERS, J. C., KNOEDLER, S., KNOEDLER, L., FRANK, K., ... & ALFERTSHOFER, M. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single- choice and multiple-choice board certification preparation questions. **European Archives of Oto-Rhino-Laryngology**, v.280, n.9, p. 4271-4278, 2023
- 4- ZHANG, C., ZHANG, C., ZHENG, S., QIAO, Y., LI, C., ZHANG, M., ... & HONG, C. S. A. Complete Survey on Generative AI (AIGC): Is ChatGPT from GPT-4 to GPT-5 All You Need? Available at: <https://doi.org/10.48550/arXiv.2303.11717>.

5- GABRIEL ARANHA. Available at: <https://gabrielaranha.com/>. Accessed on: 27 July. 2023.

<sup>1</sup>Thiago Gonçalves dos Santos Martins, Filiação acadêmico-profissional: Universidade Federal do Rio de Janeiro, Campus Macaé. e-mail: thiagogsmartins@yahoo.com.br

<sup>2</sup>Diogo Gonçalves dos Santos Martins, Filiação acadêmico-profissional: Hospital Central da Aeronáutica

<sup>3</sup>Paulo Schor, Filiação acadêmico-profissional: Universidade Federal de São Paulo, Campus São Paulo

Este artigo:

Recebido em: 30/10/2023

Aceito em: 10/04/2024

Como citar este artigo:

MARTINS, T.G.S.; MARTINS, D.G.S.; SCHOR, P. Can we help GPT chat get smarter? *Scientia Vitae*, v.17, n.45, ano 11, p. 13-17, abr./mai./jun. 2024.