

A Comparative Analysis of ChatGPT and Google's AI's "Bard" in Medicine

Ethan Waisberg, M.B, B.Ch, B.A.O^{1*}, Joshua Ong, M.D², Mouyad Masalkhi³, Nasif, Zaman⁴, Prithul, Sarker⁴, Andrew G. Lee, M.D⁵⁻¹², Alireza Tavakkoli, Ph.D⁴

¹ Department of Ophthalmology, University of Cambridge, Cambridge, United Kingdom

² Department of Ophthalmology and Visual Sciences, University of Michigan Kellogg Eye Center, Ann Arbor, MI, United States

³ University College Dublin School of Medicine, Belfield, Dublin, Ireland

⁴ Human-Machine Perception Laboratory, Department of Computer Science and Engineering, University of Nevada, Reno, Reno, Nevada, United States

⁵ Center for Space Medicine, Baylor College of Medicine, Houston, Texas, United States

⁶ Department of Ophthalmology, Blanton Eye Institute, Houston Methodist Hospital, Houston, Texas, United States

⁷ The Houston Methodist Research Institute, Houston Methodist Hospital, Houston, Texas, United States

⁸ Departments of Ophthalmology, Neurology, and Neurosurgery, Weill Cornell Medicine, New York, New York, United States

⁹ Department of Ophthalmology, University of Texas Medical Branch, Galveston, Texas, United States

¹⁰ University of Texas MD Anderson Cancer Center, Houston, Texas, United States

¹¹ Texas A&M College of Medicine, Texas, United States

¹² Department of Ophthalmology, The University of Iowa Hospitals and Clinics, Iowa City, Iowa, United States

* Corresponding author email address: ew690@cam.ac.uk

Received: 2023-12-05

Reviewed: 2023-12-20

Revised: 2023-12-23

Accepted: 2023-12-25

Published: 2023-12-29

Background: Bard AI, an AI chatbot developed by Google, emerged as a response to the success of OpenAI's ChatGPT. Bard utilizes natural language processing and machine learning techniques to emulate human-like dialogue.

Objectives: In this paper, we wanted to compare the Bard's performance to that of ChatGPT at various medical and surgical related tasks.

Methods: The responses generated were then examined by three doctors based at three different institutions to compare the performance of each AI chatbot for each specific prompt.

Results: Bard had the ability to generate a discharge summary, summarize medical literature, and recommend relevant medical guidelines. However, Bard's generated responses were not always clinically appropriate and contained both minor and major errors. Bard and ChatGPT will likely be followed by even more capable AI systems.

Conclusions: As these new tools are released it is important that they be viewed cautiously, ensuring that patient safety remains the fundamental priority.

Keywords: generative artificial intelligence; large language model; GPT-3.5; LLM; transformer network

How to cite this article:

Waisberg E, Ong J, Masalkhi M, Zaman N, Sarker P, Lee AG, Tavakkoli A. Comparative Analysis of ChatGPT and Google's AI's "Bard" in Medicine. *N Asian J Med.* 2023;1(2):37-43. [10.61838/kman.najm.1.2.5](https://doi.org/10.61838/kman.najm.1.2.5)

INTRODUCTION

Bard AI, an artificial intelligence (AI) chatbot developed by Google, emerged as a response to the success of OpenAI's ChatGPT (<http://chat.openai.com/>) in the field of AI-generated content and conversation [1]. Bard (<https://bard.google.com/chat>), similar to ChatGPT, is an AI-driven chatbot that utilizes natural language processing and machine learning techniques to emulate

human-like dialogues [2]. Nevertheless, Bard distinguishes itself by eliciting reactions from the online community through the utilization of its sophisticated language model, PaLM 2 (Pathways Language Model 2), specifically engineered to demonstrate exceptional proficiency in areas such as common-sense reasoning, logic, and mathematics [1]. The upgraded model significantly improves Bard's capacity to deliver



comprehensive and logically structured responses in comparison to its previous iterations [1].

The Bard system operates by leveraging a comprehensive dataset known as Infiniset, which encompasses dialog data sourced from public forums, C4 data, code documents, and other relevant sources. The utilization of this training data enhances Bard's capacity to comprehend language and context at a profound level, thereby enabling it to actively participate in substantial and meaningful dialogues. The primary objective of Bard is to retrieve information and deliver succinct responses, thereby differentiating itself from conventional search engine results through the provision of additional links for users to delve into for more information [3]. Moreover similar to ChatGPT, the involvement of Bard as a personal assistant could potentially facilitate healthcare providers in effectively writing operative notes [4], helping clinicians code AI algorithms [5], accessing pertinent medical data and making appointments swiftly. With access to up-to-date

information, Bard has the potential to revolutionize medicine, surgery and reduce workloads for healthcare professionals worldwide.

In this paper, we wanted to compare the Bard's performance to that of ChatGPT (GPT 3.5) at various medical and surgical related tasks. These prompts were all executed in August 2022, with Bard (update 2023.07.13) and ChatGPT version 3.5. The responses generated were then examined by three doctors at three different institutions, across two continents (E.W., J.O, A.G.L) to determine which AI chatbot had a higher level of performance for each specific prompt. Firstly, we wanted to examine Bard's ability to write a discharge summary (Figure 1A). Although discharge summaries are often time-consuming and relatively repetitive tasks, high levels of detail are required to ensure accurate and reliable medical documentation. Poorly written discharge summaries may also lead to other patient safety risks such as missed follow-up appointments and medication errors.

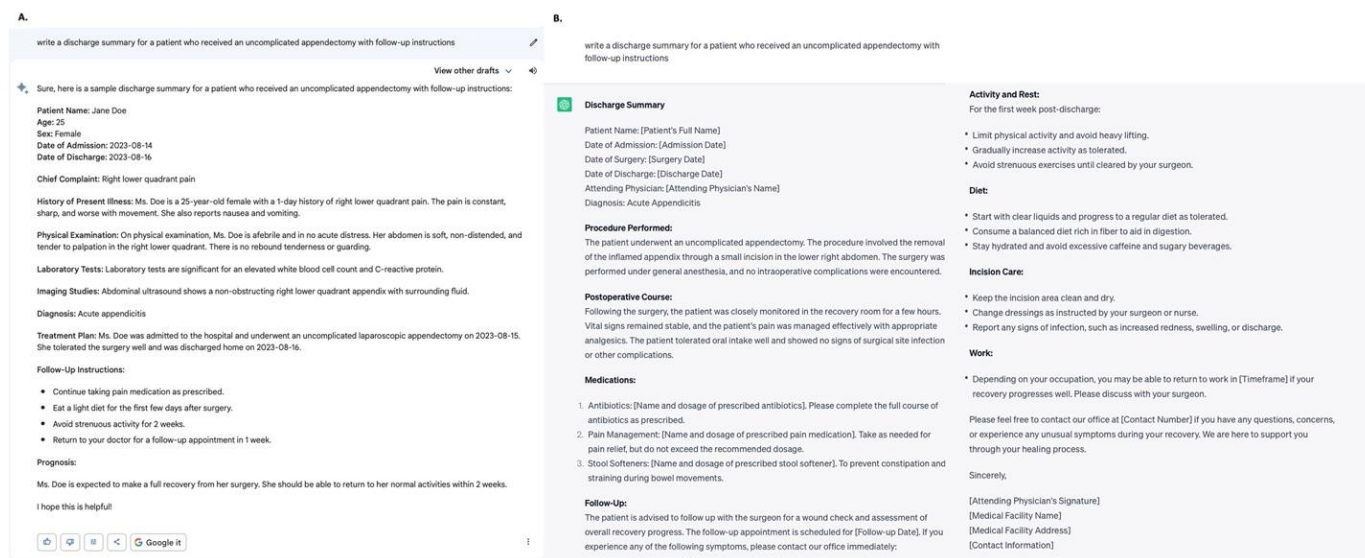


Figure 1. A: Generated by Bard from the prompt “write a discharge summary for a patient who received an uncomplicated appendectomy with follow-up instructions”. B: ChatGPT’s response to the same prompt.

The discharge summary provided by Bard included many details which did not need to be included in a discharge summary (e.g., initial physical exam finding prior to surgery and the discussion of pre-surgical lab values and imaging study results). Additionally Bard also fabricated a patient name (Jane Doe), patient age (25), date of admission (2023-08-14), discharge date (2023-08-16), symptoms and lab values. Therefore, if this note was used by a medical provider, much of the note would

have to be deleted, revised, or re-written correctly. The Bard-generated note also did not include any space for an attending physician’s name or contact information, which is a major error and an essential component of a discharge note. In contrast, the discharge note provided by ChatGPT was much more clearly written with separate headings indicating the procedure performed, post-operative course, medications, clear follow-up instructions and a section for the attending physicians

signature and contact information. Additionally, no patient information was fabricated by ChatGPT and sections were marked with open fields (e.g., [Patient's Full Name], [Surgery Date]).

Next, we wanted to examine Bard's ability to summarize clinical trial information. Staying up to date with the latest clinical trials is essential for clinicians to ensure

they are following the most up to date practices to optimize patient outcomes. Clinicians are also extremely busy, and reading all of the recently published clinical trials that are published every week is not feasible for most. We asked Bard to summarize the latest trials in treating chronic obstructive pulmonary disease (Figure 2).

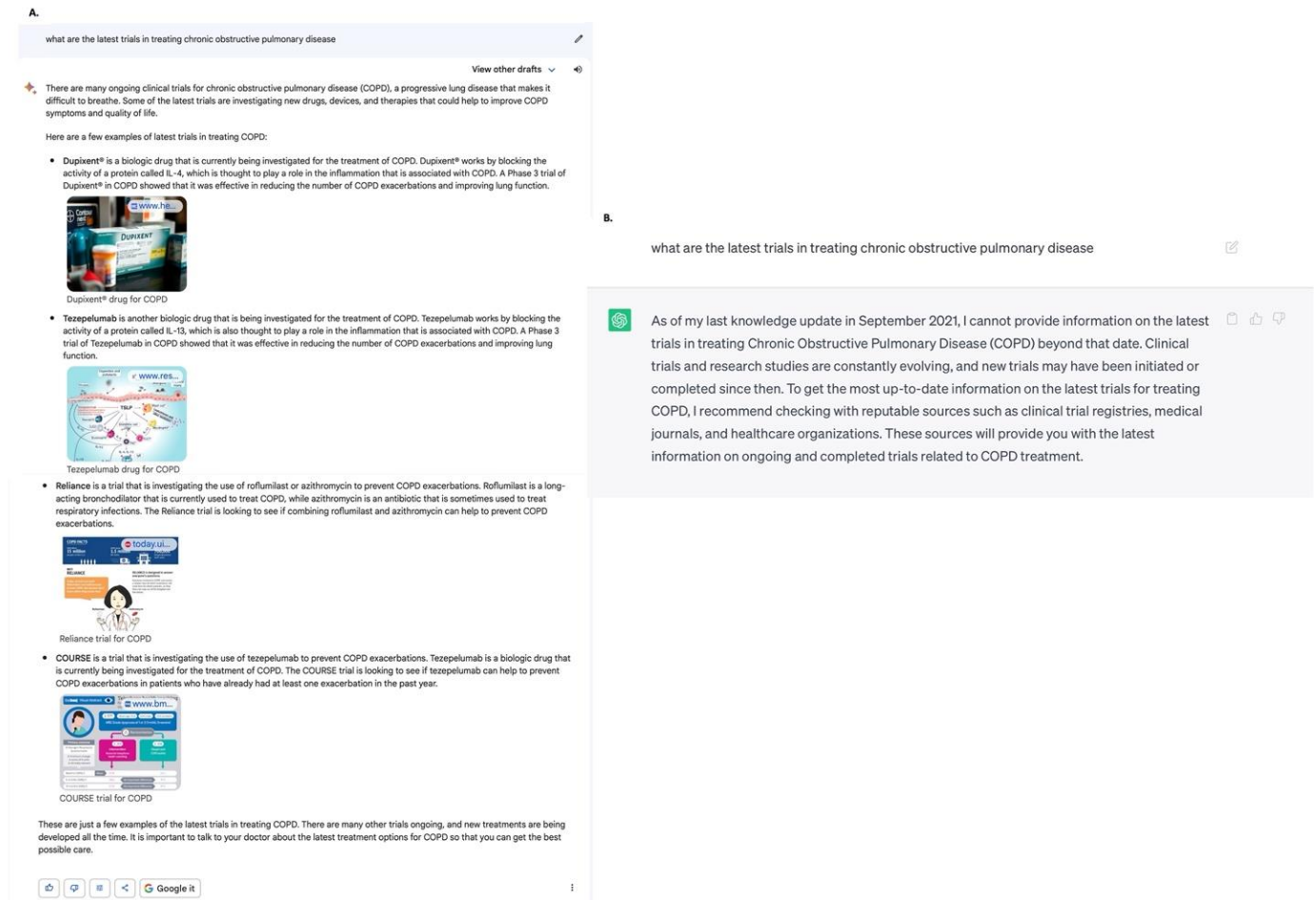


Figure 2. A: Generated from the prompt “what are the latest trials in treating chronic obstructive pulmonary disease”. B: ChatGPT’s response to the same prompt.

Bard described four recent clinical trials related to chronic obstructive pulmonary disease. These trials reported were published relatively recently such as the phase 3 trial results of Dupilimab (published July, 2022). However, upon closer inspection, the “Phase 3 Trial of Tezepelumab in COPD” was a non-existent trial and reference (i.e., an AI “hallucination”) that was generated by Bard. The latest stage clinical trial currently investigating the use of Tezepelumab in adults with COPD is the COURSE trial (currently at Phase 2a).[6] The most likely reason for this fabrication was because of an

abstract published on the use of tezepelumab in patients with severe asthma [7]. This abstract was presented at a conference session titled “Prediction of exacerbations in patients with COPD” which could potentially explain the error produced by the chatbot. However, errors in study reporting such as this can be extremely misleading for clinicians and scientists and study reporting from Bard should be used with extreme caution. Additionally, no referencing is provided by Bard, so verifying the credibility of its outputs can be time consuming, and impractical.

In contrast, ChatGPT was unable to provide any information and stated that the last knowledge update was in September 2021, and instead recommends clinical trial registries, healthcare organizations and medical journals for more up to date information.

We then examined Bard's ability to provide the latest information on the latest guidelines related to generative adversarial networks (GANs) in medicine. GANs are a revolutionary deep learning technique that can allow for image translation between various imaging modalities [8].

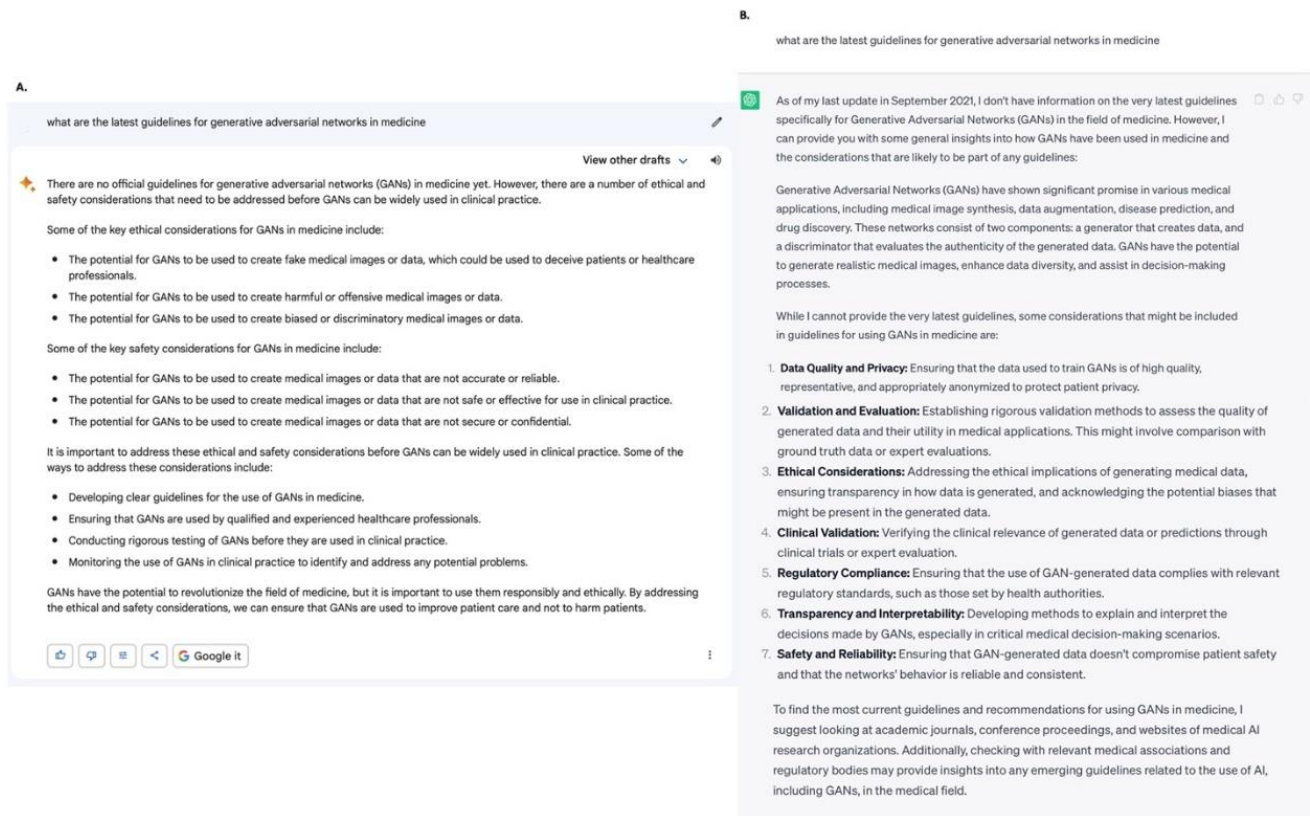


Figure 3. A: Generated by Bard from the prompt “what are the latest guidelines for generative adversarial networks in medicine”. B: ChatGPT’s response to the same prompt.

While both Bard and ChatGPT did not provide specific guidelines as these do not yet exist, both mention key considerations for the usage of GANs in medicine including acknowledging potential biases, ensuring GAN generated data does not compromise patient safety and establishing rigorous validation methods. These AI-generated responses were both specific and accurate.

Finally we wanted to examine Bard’s ability to interpret a medical image by providing it with an image of autosomal dominant optic atrophy (Figure 4).[9] The most recent update of Bard integrated Google Lens, allowing images to be added alongside prompts.

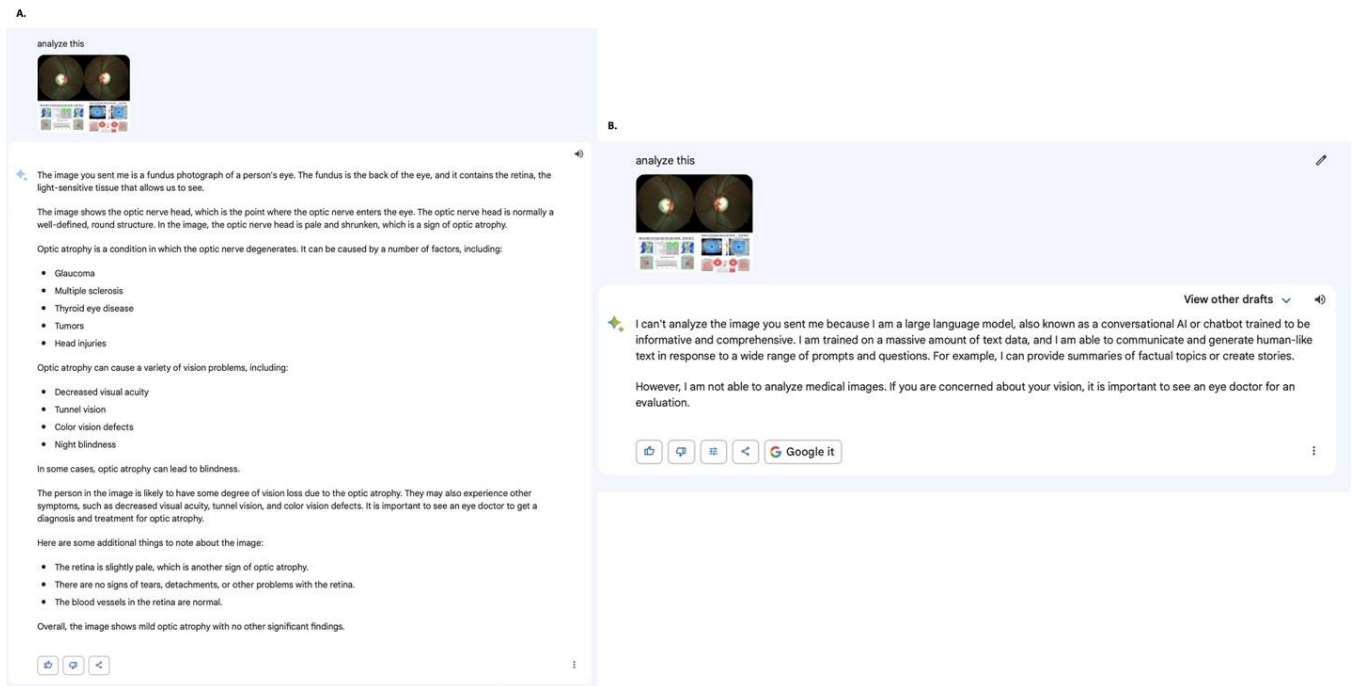


Figure 4. A: Providing Bard with a fundus photograph image of autosomal dominant optic atrophy. The image on the left had original title data (Figure 4A) while the image analyzed on the right had an anonymized title. Reprinted without changes from Waisberg E, Micieli JA. Neuro-Ophthalmological Optic Nerve Cupping: An Overview. *Eye Brain*. 2021 Dec 14;13:255-268. doi: 10.2147/EB.S272343 under Creative Commons Attribution - Non Commercial (unported, v3.0) License.

The initial analysis provided by Bard (Figure 4A) was extremely impressive, stating that the image was a fundus photograph and optic atrophy was present in this image. However, it is important to note that the file name of the figure was not renamed after being downloaded and was titled “An - 18 - year - old - man - with - dominant - optic - atrophy - DOA - The - fundus - photographs - top - panel.png”. We then anonymized this file name and re-tested the analysis (Figure 4B). Following this file name change, Bard was unable to analyze this image and stated “I am not able to analyze medical images”.

LIMITATIONS

While Bard is a highly promising tool, it is important to note that Bard is still actively under development, and errors can be made in the generated outputs. Concerns have previously been posed for ChatGPT producing incorrect or even non-existent information or references.[10] An AI chatbot cannot achieve a human-like understanding of text, which may lead to misinterpretations of subtleties in wording.[11] A false response by Bard or another AI chatbot is often referred to as a “hallucination”. In a medical scenario this is

particularly dangerous, as even subtle errors can put patient safety at risk, and a response generated by a chatbot may also appear very convincing. The hallucination above in Figure 2A shows how minor misinterpretations by Bard of a study can lead to inaccurate outputs that can appear correct. As training with more medical-specific data occurs, the performance of Bard and ChatGPT will likely continue to improve.[12, 13] However because of the reported errors and the risk to patient safety, Bard should not yet be used clinically, and further research is required to improve response accuracy.

Similar to other AI-chatbots, Bard functions as a black box model and it is often no clear how outputs are generated. This adds an additional layer of difficulty when interpreting the accuracy of the generated output. It must also be noted that Bard was trained on an immensely large dataset of text from various different sources which are not always reliable such as social media posts and news articles, rather than only relying on peer-reviewed sources. This can lead Bard to produce factual inaccuracies. In addition to this, Bard can only remain as unbiased as the data it is trained on. Caution

should be taken to not allow AI chatbots such as Bard to increase inequalities in medicine.

Protection of patient data and complying with local data storage regulations is also essential for Bard (such as General Data Protection Regulation (GDPR), HIPAA etc.). Google currently collects all Bard conversations, and they may be read and annotated by human reviewers for quality improvement purposes. For this reason, Bard should not currently be used to process any confidential patient medical information.

Regarding image analysis, in Figure 4B, Bard stated that it does not analyze medical images. It also appears that Bard is reliant on using text-based image metadata to aid in its image analysis. Therefore, medical image analysis using Bard should not be used clinically or interpreted with a high degree of caution until further improvements are made.

Another final concern regarding Bard is reproducibility. When asked the exact same question, with the same wording, Bard can produce multiple different responses. This lack of consistency in response generation can lead to further difficulties when evaluating the responses generated by Bard.

CONCLUSION

AI chatbots such as Bard, have the potential to revolutionize both medicine and surgery. While Bard is still labelled as an “Experiment” the possibilities of an effective AI chatbot go far beyond the capabilities described here. Bard and ChatGPT will likely be followed by even more capable AI systems. As these new tools are released it is important, they are approached cautiously, ensuring that patient safety remains the fundamental priority. Finetuning large language models will be the next step in improving the clinical performance of LLMs, with the incorporation of data from real medical records to improve the knowledge base of these models [14].

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

ACKNOWLEDGEMENTS

None.

COMPETING INTERESTS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

AUTHORS' CONTRIBUTIONS

All authors contributed equally.

DECLARATION

ChatGPT was employed to meticulously review language imperfections, enhancing the manuscript's fluency and overall quality [15, 16].

REFERENCES

1. An important next step on our AI journey 2023 [updated 2023/02/06]. Available from: <https://blog.google/technology/ai/bard-google-ai-search-updates/>.
2. Waisberg E, Ong J, Masalkhi M, Kamran SA, Zaman N, Sarker P, et al. GPT-4: a new era of artificial intelligence in medicine. *Ir J Med Sci*. 2023. [PMID: 37076707] [DOI]
3. Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng H-T, et al. LaMDA: Language Models for Dialog Applications. 2022.
4. Waisberg E, Ong J, Masalkhi M, Kamran SA, Zaman N, Sarker P, et al. GPT-4 and Ophthalmology Operative Notes. *Ann Biomed Eng*. 2023. [PMID: 37266720] [DOI]
5. Waisberg E, Ong J, Kamran SA, Masalkhi M, Zaman N, Sarker P, et al. Bridging artificial intelligence in medicine with generative pre-trained transformer (GPT) technology. *J Med Artif Intell*. 2023;6:13-. [DOI]
6. Fieldes M, Bourguignon C, Assou S, Nasri A, Fort A, Vachier I, et al. Targeted therapy in eosinophilic chronic obstructive pulmonary disease. *ERJ Open Res*. 2021;7(2):00437-2020. [PMID: 33855061] [PMCID: PMC8039900] [DOI]
7. Korn S, Bourdin A, Chupp G, Colice G, Ambrose CS, Kmita K, et al., editors. Effect of tezepelumab on the proportion of exacerbation-free patients with severe, uncontrolled asthma in NAVIGATOR. *Airway pharmacology and treatment*; 2021 2021/09/052021. [DOI]
8. Ong J, Waisberg E, Kamran SA, Paladugu P, Zaman N, Sarker P, et al. Deep learning synthetic angiograms for individuals unable to undergo contrast-guided laser treatment in aggressive retinopathy of prematurity. *Eye*. 2023. [PMID: 37500752] [DOI]
9. Waisberg E, Micieli JA. Neuro-ophthalmological optic nerve cupping: An overview. 2021. [PMID: 34934377] [PMCID: PMC8684388] [DOI]
10. Alser M, Waisberg E. Concerns with the usage of ChatGPT in Academia and Medicine: A viewpoint. *American Journal of Medicine Open*. 2023:100036. [DOI]
11. Waisberg E, Ong J, Masalkhi M, Zaman N, Sarker P, Lee AG, Tavakkoli A. Google's AI chatbot “Bard”: a side-by-side comparison with ChatGPT and its utilization in ophthalmology. *Eye*. 2023. [DOI]
12. Waisberg E, Ong J, Paladugu P, Kamran SA, Zaman N, Lee AG, Tavakkoli A. Challenges of Artificial Intelligence in

- Space Medicine. *Space: Science & Technology*. 2022;2022:1-7. [DOI]
13. Paladugu PS, Ong J, Nelson N, Kamran SA, Waisberg E, Zaman N, et al. Generative Adversarial Networks in Medicine: Important Considerations for this Emerging Innovation in Artificial Intelligence. *Ann Biomed Eng*. 2023. [PMID: 37488468] [DOI]
14. Wang G, Yang G, Du Z, Fan L, Li X. ClinicalGPT: Large Language Models Finetuned with Diverse Medical Data and Comprehensive Evaluation. 2023.
15. Dergaa I, Saad HB. Joining hands for better health: The inception of the New Asian Journal of Medicine. *New Asian Journal of Medicine*. 2023;1(1):1-3. [DOI]
16. Chtourou H, Guelmami N, Trabelsi K, Dergaa I. The Beginning of Our Journey: The Launch of the Tunisian Journal of Sports Science and Medicine. *Tunisian Journal of Sports Science and Medicine*. 2023;1(1):1-4. [DOI]