

COMMENT OPEN

Check for updates

How to use large language models in ophthalmology: from prompt engineering to protecting confidentiality

Oliver Kleinig 1,2^S, Christina Gao^{1,2}, Joshua G. Kovoor^{1,2}, Aashray K. Gupta^{2,3}, Stephen Bacchi^{1,2,4} and Weng Onn Chan^{1,2}

© The Author(s) 2023

Eye (2024) 38:649-653; https://doi.org/10.1038/s41433-023-02772-w

INTRODUCTION

Large language models (LLM) are complex predictive text algorithms that generate sentences and paragraphs from a prior sequence of words [1]. With appropriate training, they can act conversationally by generating text following a user-provided instruction or question. Commonly known modern LLMs include OpenAl's GPT-3 and GPT-4, Google's PaLM-2, and Meta's Llama-2. These models vary in design and performance. Information on accessing these models is included in the supplementary information.

Each implementation of the same LLM (e.g. GPT-4) may behave differently due to differences in 'fine-tuning' training, plugin access

[2], and initial instructions [3]. Early evidence suggests, when answering ophthalmology examination questions, GPT-4 via the ChatGPT or New Bing chatbot is the most performant [2, 4–6]. Understanding how to effectively employ LLMs may have significant utility for clinical education and, in future, clinical practice.

WHAT CAN LLMS DO?

LLMs may be effective in language tasks like writing, editing, and summarizing. For querying, LLMs excel at questions that are not easily web-searchable, or fact-finding without searchable keywords. LLMs can also be queried to write code in R and other

Box 1. The different outputs from ChatGPT (July 20 GPT-4 version) when asking "Briefly describe conjunctivitis" and "Briefly describe the epidemiology, pathophysiology, diagnosis, and treatment of conjunctivitis"

Query	Briefly describe conjunctivitis.	Briefly describe the epidemiology, pathophysiology, diagnosis, and treatment of conjunctivitis.
Response	 Conjunctivitis, often referred to as "pink eye", is an inflammation or infection of the conjunctiva, the transparent membrane that lines the inner part of the eyelids and covers the white part of the eyes. This condition can be caused by several factors: 1. Bacterial or viral infections: These are the most common causes, often resulting from a cold or other respiratory infection. These forms of conjunctivitis are contagious and can spread easily. 2. Allergies: Allergic conjunctivitis happens when the body reacts to an allergen, such as pollen, dust mites, or animal dander, and is not contagious. 3. Iritants: Exposure to irritants like chlorine in swimming pools, smog, smoke, or certain cosmetics can also lead to conjunctivitis. Common symptoms include redness of the eyes, increased tear production, a gritty feeling in the eyes, itchiness, and a discharge that can form a crust during sleep. Treatment varies based on the cause. Bacterial conjunctivitis rails will be treated with antibiotics, while viral conjunctivitis often resolves on its own. Allergic conjunctivitis might be treated with antihistamines, and conjunctivitis from irritants usually clears up by avoiding the triggering factor. 	 Epidemiology: Conjunctivitis, also known as pink eye, is a common condition worldwide. Its incidence varies based on the cause, whether it's viral, bacterial, allergic, or due to other causes like chemical exposure or certain diseases. Viral conjunctivitis is often associated with respiratory infections and can spread easily in close-knit communities, such as schools and nursing homes. Bacterial conjunctivitis is often seen in children and may occur with an ear infection. Allergic conjunctivitis is linked to seasonal allergies or exposure to specific allergens. Pathophysiology: The conjunctiva is a clear tissue covering the white part of the eye and the inside of the eyelids. Conjunctivitis is the inflammation of this tissue. Depending on the cause, the pathophysiology differs. In bacterial and viral conjunctivitis, the pathogens invade and irritate the conjunctiva, causing inflammation. In allergic conjunctivitis, exposure to allergens triggers an immune response, leading to inflammation. Chemical or irritant conjunctivitis occurs when the conjunctiv is is often diagnosed based on a clinical examination and patient history. Symptoms can include redness in the white of the eye, increased tear production, a feeling of grittiness in the eye, itching, burning, blurred vision, and an increased sensitivity to light. A patient's recent history of a cold, exposure to a person with pink eye, or a history of allergies can also provide clues. In some cases, a sample of the discharge from the eye may be taken for laboratory analysis to confirm the diagnosis and to differentiate the cause. Treatment: Treatment for conjunctivitis depends on the cause. Viral conjunctivitis typically resolves on its own without treatment, although cool compresses and artificial tears can be used to soothe the eye. In all activitis, the eye is often rinsed to remove the irritant, and then cool compresses and artificial tears can be eve used to rowne the irritant, and then cool consult and artificial tears

¹Royal Adelaide Hospital, Adelaide, SA, Australia. ²University of Adelaide, Adelaide, SA, Australia. ³Gold Coast University Hospital, Gold Coast, QLD, Australia. ⁴Flinders University, Adelaide, SA, Australia. ^{is}email: oliver.kleinig@student.adelaide.edu.au

Received: 1 September 2023 Revised: 19 September 2023 Accepted: 21 September 2023 Published online: 5 October 2023

Box 2. Using ChatGPT (GPT-4, August 3 Version) to create an iterate on a patient handout for cataracts surgery

Prompt	Create me a patient handout for cataracts surgery.
Response	Cataract Surgery: Patient Handout What is a Cataract? A cataract is a clouding of the eye's natural lens, which lies behind the iris and the pupil. Cataracts are the most common cause of vision loss in people over the age of 40 and are the advantage of this decrement of this decrement of the eye's natural lens, which lies behind the iris and the pupil. Cataracts are the most common cause of vision loss in people over the age of 40 and are the
	What is Cataract Surgery?
	Cataract surgery is a procedure to remove the lens of your eye and, in most cases, replace it with an artificial lens. When Is Surgery Recommended?
	- Vincuity in reading, driving, or performing daily tasks - Vision impairment that cannot be corrected with glasses
	Preparing for Surgery
	- Eye measurement tests
	- General health evaluation <u>Medication</u>
	- Inform your doctor about all medications you are taking - You may be advised to stop certain medications
	<u>Fasting</u> - Typically, you will be asked to fast for 6–12 h before surgery
	- Usually 15–20 min per eye <u>Steps</u>
	2. Small Incision: Cut near the cornea
	3. Lens Removal: Broken up and vacuumed out 4. Artificial Lens Insertion: A foldable lens is inserted
	Immediate Care
	- You'll be observed for complications for a short period - You may wear a protective eye shield or patch
	- Antibiotic eye drops
	- Anti-Inflammatory eye drops
	- Avoid strenuous activities for a few weeks - No swimming for up to two weeks
	- Avoid futbring your eyes Risks and Complications
	- Intection - Bleeding
	- Retinal detachment - Secondary cataract
	Follow-Up
	- One week after surgery
	- One month after surgery FAQs On the second s
	<u>Q: Is the procedure paintul?</u> A: Most patients report minimal discomfort, often describing a sensation of pressure.
	<u>Q: When will my vision improve?</u> A: Many people report clear vision within several hours, but each person heals differently.
	A: No, but a condition called "secondary cataract" might occur, which can be easily treated.
	Surgeon: Dr. [Name]
	Contact: [Contact Information] Please consult your healthcare provider for personalized medical advice.
	Disclaimer: This handout is for informational purposes only and should not be considered as medical advice."
Prompt	Give me specific and targeted feedback on this handout.
Response	Certainly! The handout is well-structured and provides a comprehensive overview of what a patient needs to know about cataract surgery. However, there are some areas where it could be improved for slaving doubt and approximately doubt approximately doubt and approximately doubt and approximately doubt approxi
	Strengths:
	 Use of Subheadings: Subheadings help to break down the text into manageable sections. EVG: The is a valuable addition as it participate particular parageable sections.
	Areas for Improvement: 1. Targinglage While the bandout user relatively simple language modical terms like "Petipal detectment" or "Secondary saturat" may need evaluations for the language
	 Visual Aids: The inclusion of diagrams or images could make the information easier to digest, particularly for visual learners. For instance, a diagram of the eye could be useful for available and a schorer time.
	 Personalization: Including a section where patients could jot down their own questions for the doctor might make the handout more interactive and personalized. Personalization: (ar visual acutity their own pre-operative tests it might be useful to list common tests like a 'slit-lamp examination' or 'visual acutity test' to patients how what to expect
	 Protocol Symptoms: A brief section on what normal and abnormal post-op symptoms look like could prepare patients for the recovery phase. Forst-Op Symptoms: A brief section on what normal and abnormal post-op symptoms look like could prepare patients for the recovery phase.
	7. Disclaimers and Citations: You include a general disclaimer, but it might be beneficial to cite sources or guidelines that the handout is based on.
	 Bisks and Complications: This section could be more detailed. It would be beneficial to include how often these complications cour and what steps are taken to mitigate them. Checklist: A pre-operative and post-operative checklist: could be a good addition. Patients could use it to track their preparations and recovery.
	11. Downloadable Resources: If this handout is also made available online, links to further reading or video explanations could be included. By addressing these points the bandout can offer more denth while still remaining accessible to patients. Overall it serves as a colid foundation and could be even more informative
	and engaging with these enhancements.

Prompt Implement this feedback on the above patient handout

sponse Cataract Surgery: Patient Handout What is a Cataract?	
A cataract is a clouding of the eye's natural lens, which lies behind the iris and the pupil. Cataracts are the most common cause of vision loss in people over the age of 40 and a principal cause of blindness worldwide. [Diagram of the Eve Showing Cataract Location]	are the
What is <u>Cataract Surgery</u> ? <u>What is Cataract Surgery</u> ? <u>Cataract surgery</u> is a procedure to remove the lens of your eye and, in most cases, replace it with an artificial lens.	
When Is Surgery Recommended? - Difficulty in reading, driving, or performing daily tasks - Vision impairment that contract be corrected with classes	
- Decreasing for Europein	
Pre-operative Tests	
- Slit-lamp Examination - Visual Acuity Test	
- cyc measurement tests Medication	
- Inform your doctor about all medications you are taking - You may be advised to stop certain medications	
- Typically, you will be asked to fast for 6–12 h before surgery	
The Procedure Duration	
- Usually 15–20 min per eye	
1. Local Anesthesia: Eye drops to numb the area	
2. Small Incision: Cut near the cornea 3. Lens Removal: Broken up and vacuumed out	
4. Artificial Lens Insertion: A foldable lens is inserted	
Arter Surgery Immediate Care	
- You'll be observed for complications for a short period	
Medication	
- Antibiotic eye drops - Anti-Inflammatory eve drops	
Activity	
- Avoid strenuous activities for a rew weeks	
- Avoid rubbing your eyes	
- Normal: Mild disconfort, cloudy vision	
- Abnormal: Severe pain, vision loss Risks and Complications	
- Infection: Rare but serious, Immediate medical attention needed.	
- <i>Bieeaing:</i> Rate and usually minor. - <i>Retinal Detachment</i> : Separation of retina from the back of the eye; requires immediate attention.	
- <i>Secondary Cataract</i> : Cloudiness that can occur months or years later; can be easily treated. - <i>Pressure increase inside the eye</i> : May require medication.	
Follow-Up - One day after surgery	
- One week after surgery	
- One month after surgery FAOs FAOs	
Q: Is the procedure painful?	
A: Most patients report minimum discontrate, orien describing a sensation of pressure. Q: When will my vision improve?	
A: Many people report clear vision within several hours, but each person heals differently.	
A: No, but a condition called "secondary cataract" might occur, which can be easily treated.	
Important Contacts Surgeon: Dr. [Name]	
Contact: [Contact Information]	
NUISE'S STATION: [L'ONTACT INTOTMATION] Emergency Contact: [Contact Information]	
Your Notes	
- [Your questions for the doctor]	
- [Your medication schedule]	
Pre-Op Checklist	
- [] Completed all tests - [] Confirmed medications	
- [] Started fasting	
Post-Up Checklist - [] Follow-up appointments scheduled	
- [] Medication taken as advised	
Disclaimer: This handout is for informational purposes only and should not be considered as medical advice. It is based on current medical guidelines as of [Year].	
	ve. For

*This process can be iterated multiple times to further refine the handout.

**Note the outputs have been altered slightly for formatting purposes, no changes to content were made.

The LLM is first asked to create the patient handout. Then the LLM is prompted to give targeted feedback on the above handout. Lastly, the LLM is tasked with implementing such feedback.

languages. LLMs perform mathematical operations inaccurately [7]. Plugins like embedded calculators (Supplementary Information) can address specific LLM limitations.

LLMs are engaged through text input. Slight variations in input

text can cause significant variations in output [8]. Resultingly,

high-quality prompting, or 'prompt engineering', is critical in optimizing LLM outputs.

Asking the right question

It is widely accepted that providing specific instructions to LLMs yields more useful outputs [9]. For example, requesting a summary of a condition's epidemiology, pathophysiology, diagnosis, and treatment is generally superior to asking for general

ENGAGING LLMS

information (Box 1). Additionally, using specifiers such as "Provide a brief/detailed summary of", "Explain at a high-school/collegegrade/consultant level", and "Present your findings in a table/as dot points" sets specific parameters to allow LLMs to generate more useful outputs.

Preparing the LLM

'Priming' LLMs for expected inputs and responses is an emerging practice [2]. For example: "I am going to provide X; I want you to output Y". This strategy increases the chance of a useful output. Example outputs can also be offered to LLMs to style-mimic. Online libraries of example 'guiding prompts' are emerging (for example https://www.learnprompt.org/act-as-chat-gpt-prompts/).

Notably, certain instructions could worsen output quality. ChatGPT's outputs are most accurate after explaining its reasoning [10]. Therefore, prompts that explicitly request LLMs omit reasoning could result in lower-quality outputs.

Additionally, during the same session, LLMs will consider previous inputs and outputs in response [7]. Given their textcompletion 'urge', LLMs will tend to mimic the format of prior outputs. This property may lower user workload as guiding prompts need only be typed once. Beginning a new session is prudent for unrelated conversations to reduce memory retention bias from prior outputs [5, 11].

It is also an emerging practice to tell LLMs, "If you are not sure about the answer, say you don't know". These uncertainty prompts may reduce, but not prevent, inaccurate outputs in certain tasks [12].

Completing tasks

As LLMs are effective information summarizers, they may streamline the non-creative parts of ophthalmological writing, such as abstract creation [13, 14] or medical letter generation [15]. However, ChatGPT may omit or hallucinate information [14, 15], so human oversight is required [15]. Additionally, the creative components of medical writing (for example, drawing conclusions from data) are most accurately performed by a credible human [13].

As these models are pre-trained to complete text, LLMs occasionally fail to summarize or edit incomplete documents. This is due to the 'urge' to complete the inputted text. In these situations, it is best to provide LLMs with clearly fragmented or completed text.

Powerful prompts

Powerful prompts guide LLMs in higher-order tasks, such as requesting more information for a conclusion. An LLM employed to diagnose ophthalmological conditions could be asked: "Patient X is [known information]. Keep asking me questions about patient X until you have sufficient information to arrive at a diagnosis." This method can minimize uncertainty and reduce the generality of responses provided by an LLM.

Another powerful prompt pattern is asking LLMs for feedback. For example, it is possible to ask LLMs for feedback on a patient handout and then request the LLM implement such feedback, as shown in Box 2. This process can be iterated multiple times.

Lastly, when unsure of how to prompt an LLM, it is possible to ask LLMs like ChatGPT to ask clarifying questions about a desired task until it has sufficient information to write its own prompt. The supplementary information includes an example promptgenerating prompt [16].

Online or local

Critically, data provided to online LLMs such as ChatGPT and Bard may be periodically reviewed by humans [17] or be used to train models [18]. Due to this insecurity, confidential data, even deidentified, should *never* be shared with an online LLM [19]. It is possible to run some LLMs on-device, for example GPT4ALL

(https://gpt4all.io) and Llama-2 (https://ai.meta.com/llama/). However, this may result in an unfavorable trade-off between model capability and speed on consumer-grade systems. Consequently, enterprise-hosted and institutionally and ethics committee approved LLMs may be the most appropriate LLM implementation for analysis of confidential information.

WHAT LIMITS LLMS?

While powerful tools, LLMs have well-documented limitations [20]. Most significantly being inaccuracies [20], potential lack of knowledge of specialist fields and, for GPT-3 and GPT-4, knowl-edge of events and advancements following September 2021 [7]. Furthermore, LLM's performance may vary; ChatGPT's capability (GPT-3 and GPT-4) declined in early 2023 [21], potentially due to increased safety training [10]. In the future, less safe models explicitly trained for professionals, such as Med-PaLM-2, may allow ophthalmologists to access more powerful LLMs.

CONCLUSION

The use of LLMs in ophthalmology is emerging and largely untested. Understanding the function of an LLM as a powerful but fallible predictive text algorithm may guide optimal use of this technology. When using LLMs, choosing deliberately powerful prompts such as the forms suggested above will result in the highest quality use of LLMs in ophthalmology.

REFERENCES

- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 6000–10. (NIPS'17).
- Kleinig O, Gao C, Bacchi S. This too shall pass: the performance of ChatGPT-3.5, ChatGPT-4 and New Bing in an Australian medical licensing examination. Med J Aust. [Internet]. https://doi.org/10.5694/mja2.52061.
- Edwards B. Ars Technica. 2023 [cited 2023 Jul 23]. Al-powered Bing Chat spills its secrets via prompt injection attack [Updated]. https://arstechnica.com/ information-technology/2023/02/ai-powered-bing-chat-spills-its-secrets-viaprompt-injection-attack/.
- Raimondi R, Tzoumas N, Salisbury T, Di Simplicio S, Romano MR. Comparative analysis of large language models in the Royal College of Ophthalmologists fellowship exams. Eye. 2023;9:1–4.
- Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci. 2023;3:100324.
- Cai LZ, Shaheen A, Jin A, Fukui R, Yi JS, Yannuzzi N, et al. Performance of generative large language models on ophthalmology board–style questions. Am J Ophthalmol. 2023;254:141–9.
- OpenAI. GPT-4 Technical Report [Internet]. arXiv [Preprint] 2023 [cited 2023 Jul 23]. Available from: http://arxiv.org/abs/2303.08774.
- Reeves B, Sarsa S, Prather J, Denny P, Becker BA, Hellas A, et al. Evaluating the performance of code generation models for solving Parsons problems with small prompt variations. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V 1 [Internet]. New York, NY, USA: Association for Computing Machinery; 2023. p. 299–305. https://doi.org/10.1145/ 3587102.3588805.
- Ekin S. Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices [Internet]. TechRxiv [Preprint] 2023 [cited 2023 Jul 23]. Available from: https://www.techrxiv.org/articles/preprint/Prompt_Engineering_For_ChatGPT_A_ Quick_Guide_To_Techniques_Tips_And_Best_Practices/22683919/2.
- Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. Sparks of artificial general intelligence: early experiments with GPT-4 [Internet]. arXiv [Preprint] 2023 [cited 2023 Jul 23]. Available from: http://arxiv.org/abs/ 2303.12712.
- Fijačko N, Gosak L, Štiglic G, Picard CT, Douma MJ. Can ChatGPT pass the life support exams without entering the American Heart Association Course? Resuscitation [Internet]. 2023;185. https://www.resuscitationjournal.com/article/ S0300-9572(23)00045-X/fulltext.
- Zhu D, Chen J, Haydarov K, Shen X, Zhang W, Elhoseiny M. ChatGPT asks, BLIP-2 answers: automatic questioning towards enriched visual descriptions [Internet].

arXiv [Preprint] 2023 [cited 2023 Sep 1]. Available from: http://arxiv.org/abs/2303.06594.

- 13. Kitamura FC. ChatGPT is shaping the future of medical writing but still requires human judgment. Radiology. 2023;307:e230171.
- Babl FE, Babl MP. Generative artificial intelligence: can ChatGPT write a quality abstract? Emerg Med Australas. [Internet]. https://onlinelibrary.wiley.com/doi/ abs/10.1111/1742-6723.14233.
- 15. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health. 2023;5:e179–81.
- Littlefield B. PromptGenerator [Internet]. ChatGPT users Skool. 2023. https:// www.skool.com/chatgpt/promptgenerator?p=1e5ede93.
- Google. Bard Help. Bard Privacy Help Hub. https://support.google.com/bard/ answer/13594961?hl=en#human_review.
- OpenAI. OpenAI. Privacy policy. Available from: https://openai.com/policies/ privacy-policy.
- Attowooll J. NewsGP. 'Extremely unwise': warning over use of ChatGPT for medical notes. 2023. https://www1.racgp.org.au/newsgp/clinical/extremelyunwise-warning-over-use-of-chatgpt-for-m.
- Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Health. 2023;11:887.
- Chen L, Zaharia M, Zou J. How is ChatGPT's behavior changing over time? [Internet]. arXiv [Preprint] 2023 [cited 2023 Jul 23]. Available from: http:// arxiv.org/abs/2307.09009.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception of the work and drafted and approved the manuscript. Author OK also performed the major literature review and completed the first manuscript draft. Authors SB and WOC were the principal supervisors of this piece and provided major editorial guidance.

FUNDING

Open Access funding enabled and organized by CAUL and its Member Institutions.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41433-023-02772-w.

Correspondence and requests for materials should be addressed to Oliver Kleinig.

Reprints and permission information is available at http://www.nature.com/ reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023