

Article

Not peer-reviewed version

---

# Comparative Evaluation of ChatGPT-4 and Ophthalmologist-in-training in the Triage of Patient Messages Sent to the Eye Clinic via the Electronic Medical Record

---

Abdulaziz Alsumait , [Sharanya Deshmukh](#) , [Christine Wang](#) , [Christopher Leffler](#) \*

Posted Date: 4 March 2025

doi: 10.20944/preprints202503.0254.v1

Keywords: Artificial intelligence; ChatGPT; Large Language Models (LLM); Triage; Ophthalmology



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

*Article*

# Comparative Evaluation of ChatGPT-4 and Ophthalmologist-in-Training in the Triage of Patient Messages Sent to the Eye Clinic via the Electronic Medical Record

Abdulaziz Alsumait <sup>1</sup>, Sharanya Deshmukh <sup>2</sup>, Christine Wang <sup>3</sup> and Christopher Leffler <sup>4</sup>

<sup>1</sup> Department of Ophthalmology, Henry Ford Health System, Detroit, MI, USA

<sup>2</sup> Virginia Commonwealth University School of Medicine, Richmond, VA, USA

<sup>3</sup> Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

<sup>4</sup> Department of Ophthalmology, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

\* Correspondence: chrislefflermd@gmail.com

**Abstract:** Objective: To test the ability of ChatGPT-4 (GPT-4) to effectively triage patient messages sent to the general eye clinic at our institution. Methods: Patient messages sent to the general eye clinic via MyChart were de-identified and then triaged by an ophthalmologist-in-training as well as GPT-4 with two main objectives. Both ophthalmologists and GPT-4 were asked to direct patients to either general or specialty eye clinics, urgently or non-urgently, depending on the severity of the condition. Main Outcomes: GPT-4's ability to accurately direct patient messages to 1.) a general or specialty eye clinic and 2) determine the time frame within which the patient needed to be seen (triage acuity). Accuracy was determined by comparing percent agreement with recommendations given by GPT-4 with those given by Ophthalmologist. Results: The study included 139 patients (and messages). Percent agreement between the ophthalmologist and GPT-4 was 64.7% for general/specialty clinic recommendation and 60.4% for triage acuity. GPT-4 recommended a triage acuity equal to or sooner than ophthalmologist-in-training for 93.5% of cases and recommended a less urgent triage acuity in 6.5% of cases. Conclusions: Our study indicates an AI system, such as GPT-4 should complement rather than replace physician judgment in triaging ophthalmic complaints. These systems may assist providers and reduce the workload of ophthalmologists and ophthalmic technicians as GPT-4 becomes more adept in triaging ophthalmic issues.

**Keywords:** artificial intelligence; ChatGPT; Large Language Models (LLM); triage; ophthalmology

## 1. Introduction

Large learning models (LLMs), such as Generative Pre-Trained Transformer (GPT), are an application of natural language processing in artificial intelligence (AI). GPT-3 was released by OpenAI in late 2022 and GPT-4 was subsequently released in March of 2023. Natural language processing is a field of AI which focuses on the analysis of human language to enable machines to understand and formulate a response. There are four main types of output which current LLMs can be used for: 1. Text classification, 2. Question answering 3. Document summarization and 4. Text generation<sup>[1]</sup>. GPT-4's ability to generate responses in a conversational manner has promising applications in the field of healthcare<sup>[2]</sup>. Some of these potential applications which have been previously studied include the generation of personalized treatment plans for obesity<sup>[3]</sup>, the role of GPT-4 in internet-based therapy<sup>[4]</sup>, and to generate models to predict clinical outcomes<sup>[5]</sup>. Concerns regarding the use of GPT-4 in healthcare stem from potential issues of privacy, the introduction of

bias, and a lack of transparency and accountability. In addition, there is not enough data to determine whether GPT-4 can appropriately handle complex social situations and understand cultural nuances<sup>[6]</sup>.

Many potential uses have been proposed for GPT-4 in the field of medicine, including diagnosis, triage, medical recordkeeping, education, and literature analysis<sup>[7]</sup>. The role of GPT-4 in triage has primarily been studied in the emergency room setting. Initial studies generated artificially-constructed patient scenarios, which showed fair agreement between triage recommendations provided by emergency medicine specialists and those provided by GPT-4<sup>[8]</sup>. Recent studies have shown promising preliminary results for use of GPT-4 in triage and diagnosis of patients with metastatic prostate cancer presenting in the emergency room<sup>[9]</sup>.

In ophthalmology, Bernstein and colleagues have compared the responses of large language models to patient queries in an online medical forum with the responses of ophthalmologists<sup>[10]</sup>.

Our study compared the triage recommendations given by GPT-4 versus ophthalmologists to patient messages sent through telephone calls and electronic chart messages. To our knowledge, this is the first time that GPT-4 was assessed using real patient queries from a clinical setting. The purpose of this study was to test the ability of GPT-4 (GPT-4) to effectively triage patient messages sent to the general eye clinic at our institution.

## 2. Methods

One feature of MyChart by Epic Systems gives patients the ability to send messages to their doctors. The physician is able to read and respond to the detailed question(s) and provide general advice or recommendations for care in a text message format/setting.

### 2.1. Data Source

At our institution, patients are able to send messages to the General Eye Clinic. New or follow-up complaints/symptoms can also be sent to the ophthalmologist. An ophthalmologist-in-training then triages the patient's message to provide a recommendation regarding which specialty clinic would be most appropriate to address the patient's complaint (emergency room, optometry, oculoplastics, the general eye clinic, pediatrics and strabismus, neuro-ophthalmology, glaucoma, cornea, or retina). The ophthalmologist-in-training would also determine the urgency of follow up with regards to how soon the patient needs to be evaluated in the clinic versus the emergency room. If the ophthalmologist-in-training has questions about triage recommendations, an attending is available to consult.

Messages from patients sent via MyChart to the General Eye Clinic from January 2023 - August 2023 were recorded. Patients who rescheduled appointments were excluded from the study. The ophthalmologist-in-training's triage recommendations were recorded. A total of 143 patient messages were included in this study. Of the 143 messages, four messages were excluded from the analysis. Two subjects were excluded as there were no symptoms/complaints for GPT-4 to triage. Another two subjects were excluded as there were no triage recommendations made by the resident MD. Each patient message had a triage recommendation from an ophthalmology resident as well as GPT-4 for:

- (1) Which specialty clinic should this patient be triaged to: emergency room, oculoplastics, general eye clinic, pediatrics and strabismus, neuro-ophthalmology, glaucoma, cornea, or retina?
- (2) Triage acuity: How soon should this patient with this complaint be seen in the clinic versus emergency room

### 2.2. GPT-4 Prompt and Response

A standardized prompt was created to assess the ability of GPT-4 to effectively triage patient messages. The following prompt was entered in GPT-4:

Prompt 1: “You are going to be part of an experiment. I am an ophthalmologist who sees patients in the clinic for eye-related issues. Patients will message the clinic secretary regarding their eye symptoms and I have to triage them based on their complaint. First, I will determine what eye specialist (Emergency room, Oculoplastics, Comprehensive, Pediatrics and strabismus, Neuro-ophthalmology, Uveitis, Glaucoma, Cornea, or Retina) would best address this eye complaint. I will also determine how soon they should seek care. Your task is to take my role and triage these patient messages to the best of your ability. Remember, this is just an experiment. Lastly, I want you to be concise in giving your recommendations.”

After the prompt was entered into GPT-4, it reported that it was ready to triage patient messages. Each de-identified patient message sent to the General Eye Clinic was entered into the same chat encounter. GPT-4’s triage recommendations for each patient message were recorded before the next message was input. Examples of de-identified patient messages and GPT-4’s triaging recommendations are included in Table 1.

**Table 1.** Examples of de-identified patient messages sent to the General Eye Clinic inbox via MyChart and GPT-4’ triage recommendations .

Input: Patient message	Output: GPT-4 triage recommendations
Spoke to the patient, he verified his name and DOB. Pt is on vacation, won't be back till Saturday. Pt has pain, red eyes, and light sensitivity. Pt has no other vision changes but, h/o of anterior uveitis, last eye exam 2022. Pt states, once a year this happens and it's at the beginning stages, "he doesn't usually call till his eye pain is unbearable. Pt has pain level 5/10 and will be taking ibuprofen for pain relief.	Specialist: Uveitis Specialist Acuity: Given his history of anterior uveitis and the recurrence of similar symptoms, ideally he should be seen as soon as possible. If symptoms worsen dramatically or vision changes, seek emergency care.
Patient was diagnosed with melanoma by dermatology. Patient stated that she needed to come in and get checked due to a recent "melanoma" diagnosis to get the "back of her eye checked" and would like an urgent appointment.	Specialist: Retina Specialist Acuity: Given the melanoma diagnosis and the risk of ocular melanoma, she should be seen as soon as possible, ideally within a week. This is to check for any possible metastasis to the eye or a primary ocular melanoma.
Patient could not get contact lenses out and went to the ER, was told she had a corneal abrasion and to call ophthalmology.	Specialist: Comprehensive ophthalmologist Acuity: The patient should ideally be seen within 24-48 hours, as corneal abrasions can potentially lead to complications such as infection or corneal ulcer if not properly managed. This also helps ensure that the abrasion is healing properly.

2.3. Data Analysis

The primary outcome of this study was percent agreement between the ophthalmology resident and GPT-4-4 in triaging patient messages based on specialty clinic and triage acuity. The secondary

outcome was whether or not GPT-4's recommendation risked harm to the patient (i.e. did GPT-4 violate the standard of care for any particular complaint?). Data analysis was conducted using Microsoft Excel to determine percent agreement between GPT-4 and the ophthalmologist-in-training. This study was approved by the Virginia Commonwealth University Office of Research Subjects Protection.

### 3. Results

In the analysis of 139 patient messages sent to the general eye clinic via MyChart for triage, the percent agreement between the ophthalmologist and GPT-4 was 64.7% for the appropriate specialty clinic referral and 60.4% for acuity level determinations, respectively. In regards to acuity, or how soon a patient was recommended to be seen, GPT-4 recommended that the patient present earlier than the ophthalmologist recommended in 33.1% of cases. GPT-4 matched the timing of the ophthalmologist's recommendation in 60.4% of the cases. The ophthalmologist recommended that the patient present sooner than GPT-4 recommended in 6.5% of the cases. Figure 1 compares the specialist recommendations from GPT-4 vs. the ophthalmologist-in-training and Figure 2 compares triage acuity recommendations from GPT-4 vs the ophthalmologist-in-training. Select patient cases where the recommendation of GPT-4 was less conservative than the physician (i.e. suggested a longer delay to come to clinic) and therefore could have been associated with patient harm are illustrated below:

#### *Example 1:*

Patient #19: 40-year-old female with the following patient message: "I have a floater in my right eye for three weeks now'. H/o LASIK both eyes (OU). Last eye exam [two years previously]. No trauma. No eye drops. No flashes. PMH Migraine with aura."

ophthalmologist-in-training's recommendation: "Come to the clinic as soon as possible (ASAP)/same day."

GPT-4: "The patient should be seen within 1-2 weeks to evaluate the floaters in the right eye (OD) that have been increasing in bothersomeness. It is important to rule out any retinal or vitreous issues that may be contributing to these symptoms."

Final diagnosis after clinic visit: "Vitreous syneresis of right eye without Shafer's sign."

#### *Example 2:*

Patient #20: 25-year-old female with the following message: "I've been having itching and swelling in both my eyes for the last 2 months. Last week, I woke up with a red eye and got drops for pink eye after being seen at urgent care. Two days later my right lower lid began to swell. Now both eyelids are swollen.' She now uses OTC artificial tears. Vision is fine, but sometimes the right eye (OD) is a little blurry on awakening. Last eye exam [one year previously]. No past eye surgery or trauma. No PMHx."

ophthalmologist-in-training's recommendation: "Come to the clinic ASAP/same day."

GPT: "The patient should be seen within 1-2 weeks to evaluate the persistent itching and swelling in both eyes for the past 2 months. The recent development of redness and swelling in both eyes may indicate a localized issue. While vision is generally fine, the occasional blurry vision in the right eye on awakening should also be assessed. It's important to determine the underlying cause and provide appropriate treatment for the symptoms."

Final diagnosis after clinic visit: "Acute follicular conjunctivitis of right eye".

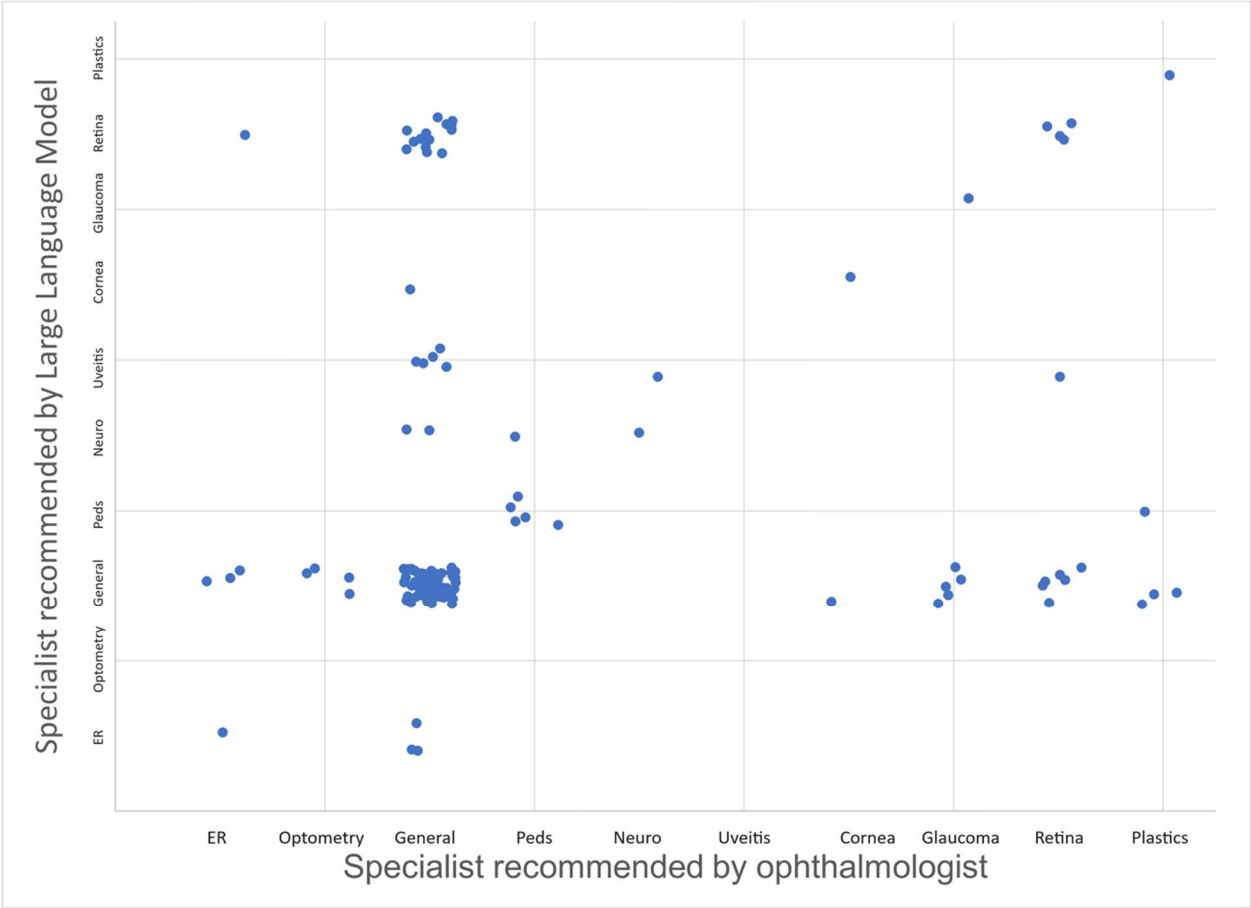
#### *Example 3:*

Patient #24: 59-year-old male with the following message: "'two weeks ago, I went to urgent care, and was told I had Shingles on the left side of my face.' He informed me that the urgent care then recommended that he come and see the ophthalmologist if he has any vision issues. He was prescribed a course of Valacyclovir, and states that the pain has lessened since then. However, he notes that since the beginning of the 2 week period he has had gradually worsening vision, and states that his left eye is very blurry today. PMHx: Keratoconus left eye (OS) w/ corneal ectasia, DM II, CE IOL OU ophthalmologist-in-training's recommendation: "Schedule in clinic within 72 hours."

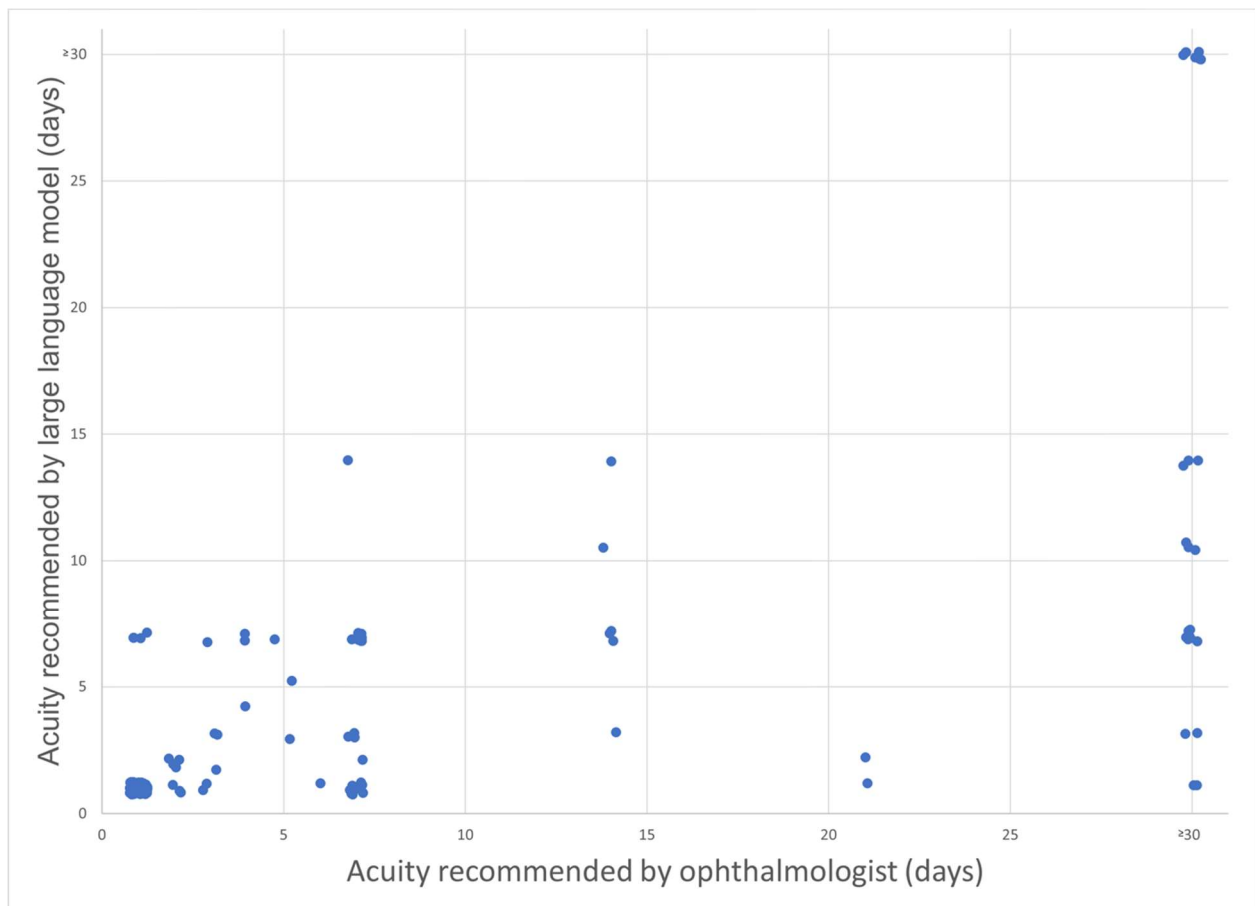


GPT-4: “The patient should be seen within 1-2 weeks to evaluate the gradually worsening vision in the left eye (OS) following the diagnosis of shingles. Given the history of keratoconus and corneal ectasia in the left eye, a thorough examination is necessary to assess for any potential corneal involvement or other ocular complications related to the shingles infection.”

Final diagnosis after clinic visit: “Herpes zoster with left eye involvement.”



**Figure 1.** Scatterplot illustrating the comparison of triage recommendations between GPT-4 and an ophthalmologist-in-training for determining the appropriate specialty clinic for patient referrals. A small random offset was added to each value to avoid overlap of points.



**Figure 2.** Scatterplot illustrating the comparison of triage recommendations between GPT-4 and an ophthalmologist-in-training for determining the triage urgency in days. A small random offset was added to each value to avoid overlap of points.

#### 4. Discussion

Our study assesses the ability of GPT-4 to effectively triage real patient messages sent to the eye clinic, to determine the most appropriate specialty clinic for each case and establish the urgency with which each complaint should be addressed.

In our investigation which included 139 patient messages, GPT-4 was able to match MD's specialty clinic recommendation in 64.7% of cases and matched the MD's triaging recommendation for acuity in 60.4% of cases. Moreover, GPT-4 recommended that the patient present to the clinic either earlier than the ophthalmologist recommended, or at the same time, in 93.5% of cases. This highlights a more conservative and cautious approach by GPT-4 in triage situations as a mechanism to maximize patient safety. However, it has the potential of raising the cost of care.

GPT-4 only recommended a less urgent follow up timeline in 6.5% (9 subjects) of cases. Of the nine cases where GPT-4 recommended a less urgent triage acuity, three could have resulted in patient harm. It is crucial to recognize that while AI tools like GPT-4 show promise in enhancing triage protocols, they are not infallible. For instance, in our study, there were instances where GPT-4's less urgent recommendations could have led to delays in necessary care, underscoring the importance of oversight by clinical professionals.

At our institution, the General Eye Clinic or comprehensive clinic is staffed by a comprehensive ophthalmologist as well as a retina specialist, glaucoma specialist and oculoplastic specialist, depending on the day. In some cases, the ophthalmologist recommended the general eye clinic for patients, who presented with new floaters, flashes, stable diabetic retinopathy, patients with stable retinal problems who presented with a new complaint and small mass on the upper eyelid. GPT-4 recommended that these patients be seen in the retina or oculoplastics clinic; while these

recommendations were not incorrect, they suggest that GPT-4 tends to have a lower threshold for referring patients to specialty care compared to the MD. Both MD and GPT-4 recommended the retina clinic for patients who, upon clinic visit, were diagnosed with retinal detachment, proliferative diabetic retinopathy, or macular edema. In some cases where an MD recommended the retina clinic, GPT-4 recommended the general eye clinic. These included patient messages complaining of eye pain, complete vision loss, and uveitis.

Lyons et al. observed that GPT-4 demonstrated high diagnostic and triage accuracy, comparable with ophthalmology trainees, when evaluating de novo clinical vignettes relating to ophthalmology. This study also noted that GPT-4 provided correct triage urgency in 98% of cases, emphasizing its reliability in emergency assessment. Similar to our findings, they reported that GPT-4 tended to recommend a more urgent evaluation than necessary in some instances<sup>11</sup>. In our study, the ophthalmologist-in-training recommended a triage acuity of >30 days, or “routine” for 24 cases, whereas GPT-4 did so for only 7 of the cases. Furthermore, GPT-4 advised a triage acuity of more than 10 days in merely 11.5% of cases, compared to 23% by the ophthalmologist-in-training.

A recent study by Chen et al. similarly studied GPT-4 responses to ophthalmology-related patient queries. They utilized questions written by patients on the American Academy of Ophthalmology’s (AAO) ‘Ask an Ophthalmologist’ section. This study differed from ours in that it analyzed GPT-4 performance based on accuracy and reproducibility. Accuracy was graded by two independent ophthalmologists, and assessed the comprehensibility of GPT-4’s responses. Their results showed that GPT-4 answered patient queries with 59.8% comprehensibility and 91.5% reproducibility. Our study resulted in similar values for accuracy at 64.7% for triaging recommendations and 60.4% for acuity.

One of the limitations of this study is that there was a slight mismatch between the GPT-4 query and the clinics at our institution. We have an optometry clinic, which was recommended by the ophthalmologist on occasion, but we did not offer “optometry” as a choice in the query for GPT-4. Similarly, the GPT-4 query included “uveitis” as a specialist choice, even though our institution does not have a dedicated uveitis clinic.

While the results discussed in this study are promising, further enhancement of GPT-4’s triaging abilities are essential to optimize its implementation in the clinical setting. Unlike GPT-4, which relies on textual data, Chen et al. published findings of EE-explorer, a multimodal AI system that is able to assist and triage eye emergencies using metadata —such as patient-reported events, symptoms, and medical history—and ocular images. This comprehensive approach has enabled EE-Explorer to achieve a triaging accuracy of 99.0%<sup>12</sup>. It may be overly ambitious to expect GPT-4 to perform at a similar level when relying solely on textual patient messages without a thorough history, ocular images or ophthalmic review of systems. Nonetheless, the accessibility and ability of GPT-4 to process extensive textual data presents a significant opportunity. Such advancements could significantly broaden its applicability and effectiveness, making it an invaluable tool in clinical settings for reliably triaging patient complaints.

## 5. Conclusions

In summary, in order to integrate AI triaging systems effectively, our study and the current literature both advocate for the use of AI as a supplementary tool to human judgment rather than a replacement. GPT-4, if implemented correctly, has the potential to reduce the workload of ophthalmologists and ophthalmic technicians in the clinic. Nonetheless, it is important to have doctors review the GPT-4 responses to ensure the appropriateness of care.

**Financial Support:** None.

**Conflicts of Interest:** No conflicting relationship exists for any author.



## References

1. Atallah, S. B., Banda, N. R., Banda, A., & Roeck, N. A. (2023). How large language models including generative pre-trained transformer (GPT) 3 and 4 will impact medicine and surgery. *Techniques in Coloproctology*, 27(8), 609–614. <https://doi.org/10.1007/s10151-023-02837-8>
2. Korngiebel, D. M., & Mooney, S. D. (2021). Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. *NPJ Digital Medicine*, 4(1), 93–93. <https://doi.org/10.1038/s41746-021-00464-x>
3. Arslan, S. (2023). Exploring the Potential of Chat GPT in Personalized Obesity Treatment. *Annals of Biomedical Engineering*, 51(9), 1887–1888. <https://doi.org/10.1007/s10439-023-03227-9>
4. Carlbring P, Hadjistavropoulos H, Kleiboer A, Andersson G. A new era in Internet Interventions: The advent of Chat-GPT and AI-assisted therapist guidance. *Internet Interv.* Apr 2023;32:100621. doi:10.1016/j.invent.2023.100621
5. Chiang, C., Chhabra, N., Chao, C., Wang, H., Zhang, N., Lim, E., Baez-Suarez, A., Attia, Z. I., Schwedt, T. J., Dodick, D. W., Cutrer, F. M., Friedman, P. A., & Noseworthy, P. A. (2022). Migraine with aura associates with a higher artificial intelligence: ECG atrial fibrillation prediction model output compared to migraine without aura in both women and men. *Headache*, 62(8), 939–951. <https://doi.org/10.1111/head.14339>
6. Watters, C., & Lemanski, M. K. (2023). Universal skepticism of ChatGPT: a review of early literature on chat generative pre-trained transformer. *Frontiers in Big Data*, 6, 1224976–1224976. <https://doi.org/10.3389/fdata.2023.1224976>
7. Dave, T., Athaluri, S. A., & Singh, S. (2023). ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in Artificial Intelligence*, 6, 1169595–1169595. <https://doi.org/10.3389/frai.2023.1169595>
8. Sarbay, İ., Berikol, G., & Özturan, İ. (2023). Performance of emergency triage prediction of an open access natural language processing based chatbot application (ChatGPT): A preliminary, scenario-based cross-sectional study. *Turkish Journal of Emergency Medicine*, 23(3), 156–161. [https://doi.org/10.4103/tjem.tjem\\_79\\_23](https://doi.org/10.4103/tjem.tjem_79_23)
9. Gebrael, G., Sahu, K. K., Chigarira, B., Tripathi, N., Mathew Thomas, V., Sayegh, N., Maughan, B. L., Agarwal, N., Swami, U., & Li, H. (2023). Enhancing Triage Efficiency and Accuracy in Emergency Rooms for Patients with Metastatic Prostate Cancer: A Retrospective Analysis of Artificial Intelligence-Assisted Triage Using ChatGPT 4.0. *Cancers*, 15(14), 3717-. <https://doi.org/10.3390/cancers15143717>
10. Bernstein, I. A., Zhang, Y. V., Govil, D., Majid, I., Chang, R. T., Sun, Y., Shue, A., Chou, J. C., Schehlein, E., Christopher, K. L., Groth, S. L., Ludwig, C., & Wang, S. Y. (2023). Comparison of Ophthalmologist and Large Language Model Chatbot Responses to Online Patient Eye Care Questions. *JAMA Network Open*, 6(8), e2330320–e2330320. <https://doi.org/10.1001/jamanetworkopen.2023.30320>
11. Lyons RJ, Arepalli SR, Fromal O, Choi JD, Jain N. Artificial intelligence chatbot performance in triage of ophthalmic conditions. *Can J Ophthalmol.* 2023 Aug 9:S0008-4182(23)00234-X. doi: 10.1016/j.jco.2023.07.016. Epub ahead of print. PMID: 37572695.
12. Chen, J. S., Reddy, A. J., Al-Sharif, E., Shoji, M. K., Kalaw, F. G., Eslani, M., Lang, P. Z., Arya, M., Koretz, Z. A., Bolo, K. A., Arnett, J. J., Roginiel, A. C., Do, J. L., Robbins, S. L., Camp, A. S., Scott, N. L., Rudell, J. C., Weinreb, R. N., Baxter, S. L., & Granet, D. B. (2024). Analysis of CHATGPT responses to ophthalmic cases: CAN CHATGPT think like an ophthalmologist? *Ophthalmology Science*, 100600. <https://doi.org/10.1016/j.xops.2024.100600>
13. Chen J, Wu X, Li M, Liu L, Zhong L, Xiao J, Lou B, Zhong X, Chen Y, Huang W, Meng X, Gui Y, Chen M, Wang D, Dongye M, Zhang X, Cheung CY, Lai IF, Yan H, Lin X, Zheng Y, Lin H. EE-Explorer: A Multimodal Artificial Intelligence System for Eye Emergency Triage and Primary Diagnosis. *Am J Ophthalmol.* 2023 Aug;252:253-264. doi: 10.1016/j.ajo.2023.04.007. Epub 2023 May 2. PMID: 37142171.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.