Acta Medica Alanya

RESEARCH ARTICLE

ARAŞTIRMA MAKALESİ

Acta Medica Alanya

2024;8(3): 221-227

DOI: 10.30565/medalanya.1531790

Is ChatGPT a Useful Tool for Ophthalmology Practice?

ChatGPT Oftalmoloji Pratiğinde Faydalı Bir Araç Mıdır?

Fuat Yavrum[1]*, Dilara Ozkoyuncu Kocabas[2]

1.Department of Ophthalmology, Alanya Alaaddin Keykubat University Faculty of Medicine, Antalya, Türkiye
2.Department of Ophthalmology, TOBB University of Economics & Technology, Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Aim: This study aimed to assess ChatGPT-3.5's performance in ophthalmology, comparing its responses to clinical case-based and multiple-choice (MCQ) questions.
Methods: ChatGPT-3.5, an AI model developed by OpenAI, was employed. It responded to 98 case-based questions from "Ophthalmology Review: A Case-Study Approach" and 643 MCQs from "Review Questions in Ophthalmology" book. ChatGPT's answers were compared to the books, and statistical analysis was conducted.
Results: ChatGPT achieved an overall accuracy of 56.1% in case-based questions. Accuracy varied across categories, with the highest in the retina section (69.5%) and the lowest in the trauma section (38.2%). In MCQ, ChatGPT's accuracy was 53.5%, with the weakest in the optics section (32.6%) and the highest in pathology and uveitis (66.7% and 63.0%, respectively). ChatGPT performed better in case-based questions in the retina and pediatric ophthalmology sections than MCQ.
Conclusion: ChatGPT-3.5 exhibits potential as a tool in ophthalmology, particularly in retina and pediatric ophthalmology. Further research is needed to evaluate ChatGPT's clarity and acceptability for open-ended questions.

Key Words: Artificial intelligence, ChatGPT, Large language model, Ophthalmology

ÖZ

Amaç: ChatGPT-3.5'in performansını göz hastalıkları alanında değerlendirmek, klinik vaka bazlı sorular ve çoktan seçmeli sorulara (ÇSS) verdiği yanıtların doğruluk oranını karşılaştırmaktır.
Yöntem: Çalışmada OpenAI tarafından geliştirilen bir yapay zeka modeli olan ChatGPT-3.5 kullanıldı. Modelden, "Ophthalmology Review: A Case-Study Approach" kitabından 98 vaka bazlı soruya ve "Review Questions in Ophthalmology" kitabından 643 ÇSS'ye yanıt vermesi istendi. ChatGPT'nin cevapları kitaplarla karşılaştırıldı ve istatistiksel analizi yapıldı.
Bulgular: ChatGPT, vaka bazlı sorularda genel olarak %56,1 doğruluk oranı gösterdi.. Doğruluk oranı kategoriler arasında en yüksek retina bölümünde (%69,5) ve en düşük travma bölümünde (%38,2) idi. ÇSS'de ChatGPT'nin genel doğruluk oranı %53,5 olarak gözlendi, bunların en düşüğü optik bölümünde (%32,6) ve en yükseği patoloji ve üveit bölümlerinde (%66,7 ve %63) idi. ChatGPT özellikle retina ve pediatrik oftalmoloji bölümlerindeki vaka bazlı sorularda ÇSS'ye kıyasla daha iyi performans gösterdi.
Sonuç: ChatGPT-3.5, özellikle retina ve pediatrik oftalmoloji alanlarında göz hastalıkları için potansiyel bir yardımcı araç olarak görülmektedir. ChatGPT'nin açık uçlu sorular için netlik ve kabul edilebilirliğini değerlendirmek için daha fazla araştırma yapılması gerekmektedir.

Anahtar Sözcükler: Yapay zeka, ChatGPT, Büyük dil modeli, Oftalmoloji

RECEIVED: 12/08/2024 ACCEPTED: 23/11/2024 PUBLISHED (ONLINE): 30/12/2024

*Corresponding Author: Fuat Yavrum. Department of Ophthalmology, Alanya Alaaddin Keykubat University
Faculty of Medicine, Antalya, Türkiye. Phone: + 90 505 583 9717 / E-mail: fuatyavrum@gmail.com

ORCID: 0000-0002-0708-5508

To cited: Yavrum F, Kocabas Ozkoyuncu D. Is ChatGPT a Useful Tool for Ophthalmology Practice? Acta Med. Alanya 2024;8(3) : 221-227 doi: 10.30565/medalanya.1531790

Introduction

Artificial intelligence (AI) based tools have recently been gaining popularity in medicine, including medical education, public health, and disease treatment and management. Deep learning algorithms (DLAs), a branch of AI, have been widely integrated into clinical practice. These algorithms obtain results via neural networks in a somewhat similar manner to the human brain. [1]

Chat Generative Pre-trained Transformer (ChatGPT), an AI-based chatbot developed by OpenAI (San Francisco, CA, USA), combines DLA and neural networks. This large language model (LLM) enables users to obtain text responses based on extensive textual datasets in various languages as human-like conversations.[2]

ChatGPT-3 has garnered significant attention worldwide since its release in November 2022. Like previous versions, it has also found its place in medicine.[3] While it was initially preferred for scientific writing, such as article abstracts or book chapters. [4] It has since found diverse uses in analyzing data generated from medical exams.

ChatGPT noted a significant improvement in answering the medical questions in the United States Medical Licensing Exam (USMLE).[5,6] Additionally, Cai et al.[7] demonstrated that the latest version of ChatGPT had a similar ability to human respondents in finding solutions in the Basic Science and Clinical Science Self-Assessment Program. These studies focused on multiple-choice questions (MCQs). However, clinical case-based learning, another popular educational method, is essential to achieve sufficient competency for clinical practice.

In ophthalmology, DLAs have proven a promising tool for diagnosing and screening common retinal diseases such as diabetic retinopathy and age-related macular degeneration.[8,9]

However, since chatbots have not yet reached impressive accuracy, their performance with ophthalmological questions requires improvement. Therefore, this study compared the performance of ChatGPT-3,5 with MCQ and clinical case-based questions in ophthalmology.

Material and Methods

AI

This study used the ChatGPT GPT-3.5 models. ChatGPT uses multiple mechanisms, including self-attention, training data, and fine-tuning, to produce natural language responses to text input on a user interface accessed at https://chat.openai.com/. This iterative deployment could not browse other databases or Internet searches at the time of this study. All responses are generated in situ, based on the abstract relationship between user-inputted words in the neural network. The ChatGPT version used in this study contained only information indexed from its last update until January 1, 2022.

Obtaining Data

The Ophthalmology Review: A Case-Study Approach book was chosen for the case-based questions.[10] It comprises 98 case-based questions divided into 11 categories: cornea and external disease, lens, glaucoma, retina, uveitis, tumors, posterior segment complications, trauma, neuro-ophthalmology, pediatric ophthalmology, and orbit/oculoplastics. Each question has seven sections:

- Section 1: The history and findings on examination of a typical patient;
- Section 2: Relevant diagnostic testing and interpretations;
- Section 3: The diagnosis and differential diagnosis;
- Section 4: Medical management;
- Section 5: Surgical management;
- Section 6: Recommendations for rehabilitation and follow-up care;
- Section 7: Suggested Reading.

The Review Questions in Ophthalmology book was chosen for the MCQs, which contains 1062 MCQs spanning 12 chapters: fundamentals, embryology and anatomy, optics, neuro-ophthalmology, pediatric ophthalmology and strabismus, plastics, pathology, uveitis, glaucoma, cornea, lens/cataract, and retina and vitreous.[11]

Model testing was conducted for two months, from February to April 2023, using version 3.5 of ChatGPT. Both books provide questions with their corresponding answers and explanations. Since the questions are not publicly accessible, they have not been indexed in any search engine or included in the ChatGPT database.

Each question was entered into the ChatGPT interface. A new ChatGPT session was launched for every question to prevent crossover learning and memory retention. The recorded responses were compared to the answers in the book. Due to inadequate interpretation by ChatGPT, 419 MCQs comprising clinical, radiological, or graphical images were excluded from the analysis. All 98 case-based questions were processed from the Ophthalmology Review: A Case-Study Approach book. These questions comprise five parts:

1. What is the most likely diagnosis of this case?

2. What is the most likely differential diagnosis of this case?

3. How should this case be medically managed?

4. How should this case be surgically managed?

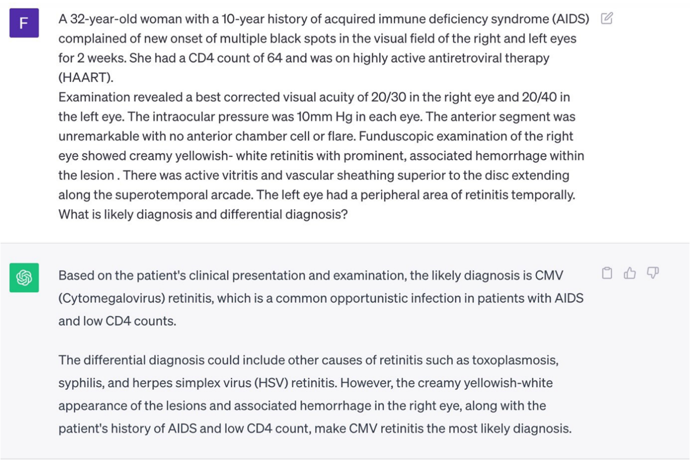
5. How should this case be rehabilitated and followed up?

Parts 1 (diagnosis) and 5 (rehabilitation and follow-up) were evaluated as "Correct" (1) or "Incorrect" (0). All other parts were converted to percentiles (i.e., the ability of ChatGPT to generate two correct answers in five differential diagnoses corresponded to 40%; Figure 1).

The 643 remaining MCQs that contained only the question stem without answer choices were processed from the Review Questions in Ophthalmology book. The responses were encoded as "Correct" and "Incorrect" (Figure 2).

Statistical Analysis

The data were statistically analyzed using the Statistical Package for the Social Sciences (SPSS) for Windows software (version 26.0). Categorical variables are presented as numbers (n) and percentages (%) that indicate the accuracy of ChatGPT responses. Categorical variables were compared using the Chi-Square test. The data were examined at a 95% confidence level. A P-value of <0.05 was considered statistically significant.

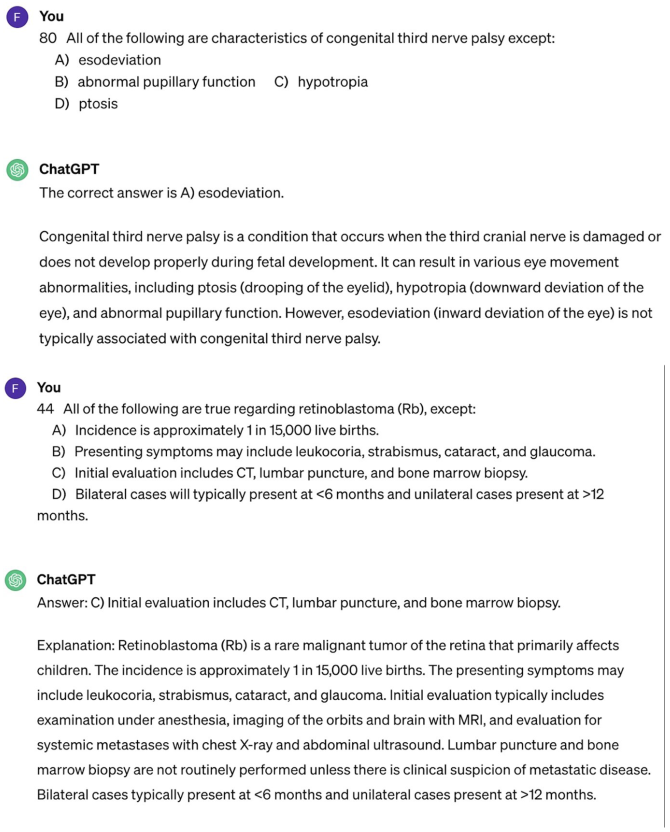


A 32-year-old woman with a 10-year history of acquired immune deficiency syndrome (AIDS) complained of new onset of multiple black spots in the visual field of the right and left eyes for 2 weeks. She had a CD4 count of 64 and was on highly active antiretroviral therapy (HAART).
Examination revealed a best corrected visual acuity of 20/30 in the right eye and 20/40 in the left eye. The intraocular pressure was 10mm Hg in each eye. The anterior segment was unremarkable with no anterior chamber cell or flare. Funduscopic examination of the right eye showed creamy yellowish- white retinitis with prominent, associated hemorrhage within the lesion . There was active vitritis and vascular sheathing superior to the disc extending along the superotemporal arcade. The left eye had a peripheral area of retinitis temporally.
What is likely diagnosis and differential diagnosis?

Based on the patient's clinical presentation and examination, the likely diagnosis is CMV (Cytomegalovirus) retinitis, which is a common opportunistic infection in patients with AIDS and low CD4 counts.

The differential diagnosis could include other causes of retinitis such as toxoplasmosis, syphilis, and herpes simplex virus (HSV) retinitis. However, the creamy yellowish-white appearance of the lesions and associated hemorrhage in the right eye, along with the patient's history of AIDS and low CD4 count, make CMV retinitis the most likely diagnosis.

Figure 1. A question inspired from the book "Ophthalmology Review: A Case-Study Approach" and ChatGPT's response.



You
80 All of the following are characteristics of congenital third nerve palsy except:
A) esodeviation
B) abnormal pupillary function C) hypotropia
D) ptosis

ChatGPT
The correct answer is A) esodeviation.

Congenital third nerve palsy is a condition that occurs when the third cranial nerve is damaged or does not develop properly during fetal development. It can result in various eye movement abnormalities, including ptosis (drooping of the eyelid), hypotropia (downward deviation of the eye), and abnormal pupillary function. However, esodeviation (inward deviation of the eye) is not typically associated with congenital third nerve palsy.

You
44 All of the following are true regarding retinoblastoma (Rb), except:
A) Incidence is approximately 1 in 15,000 live births.
B) Presenting symptoms may include leukocoria, strabismus, cataract, and glaucoma.
C) Initial evaluation includes CT, lumbar puncture, and bone marrow biopsy.
D) Bilateral cases will typically present at <6 months and unilateral cases present at >12 months.

ChatGPT
Answer: C) Initial evaluation includes CT, lumbar puncture, and bone marrow biopsy.

Explanation: Retinoblastoma (Rb) is a rare malignant tumor of the retina that primarily affects children. The incidence is approximately 1 in 15,000 live births. The presenting symptoms may include leukocoria, strabismus, cataract, and glaucoma. Initial evaluation typically includes examination under anesthesia, imaging of the orbits and brain with MRI, and evaluation for systemic metastases with chest X-ray and abdominal ultrasound. Lumbar puncture and bone marrow biopsy are not routinely performed unless there is clinical suspicion of metastatic disease. Bilateral cases typically present at <6 months and unilateral cases present at >12 months.

Figure 2. Examples of MCQ's inspired from the book “Review Questions in Ophthalmology” and ChatGPT's response. MCQ: Multiple choice question

Results

ChatGPT generated 56.1% correct responses to each part of the 98 case-based questions. The distribution of correct answer rates in the 11 categories is presented in Table 1. The correct answer rate was highest in the retina section (14 questions; 69.5%) and lowest in the trauma section (two questions; 38.2%). The correct answer rate was <50% for the tumors, posterior segment complication, trauma, neuro-ophthalmology, and orbit/oculoplastics sections. ChatGPT correct answer rates were highest in diagnosis and lowest in differential diagnosis.

ChatGPT correctly answered 344/643 MCQs (53.5%; Table 2). The correct answer rate was lowest in the optics section (32.6%) and highest in the pathology (66.7%) and the uveitis (63.0%) sections. ChatGPT achieved more than 50% correct answer rate in all sections, except the embryology and anatomy (40.9%) and the optics (32.6%).

Figure 3 represented as the comparison of each book according to sectors. The correct answer rates for case-book questions and MCQs did not differ significantly in each section except the retina and pediatric ophthalmology (p = 0.020 and p = 0.025, respectively).

Discussion

Since the release of ChatGPT 3.5, the conversational abilities of LLMs have come to the forefront, particularly in more human-like conversational features. This feature makes ChatGPT a prominent tool for assisting medical applications. However, clinical approaches require many abilities, such as situational assessment, interpretation, and theoretical knowledge. Additionally, clinical reasoning requires years of training and experience, making this complex cognitive process difficult for LLMs to master.

ChatGPT showed significant improvement in answering medical questions. Our study is the first to compare the quality of answers to ophthalmology open-ended questions and MCQs using ChatGPT. Case book questions are noteworthy in ophthalmology as a guide for clinical approach. While the correct answer rate was similar for each book, it was <60% in our study.

ChatGPT 3 achieved a passing score (>60%) for a third-year medical student on the National Board of Medical Examiners.[5] Additionally, ChatGPT 3 performed at or near the passing threshold for all three exams (Steps 1, 2, and 3) in the USMLE.[6] The performance of ChatGPT 3.5 on questions from each book seemed similar to its theoretical performance in different studies that compared its performance against human respondents.[7,12,13]

Antaki et al.[12] reported that ophthalmology residents who graduated in 2022 had an average score of 74% on the Basic and Clinical Science Course question bank and 63% on the OphthoQuestions. Additionally, ChatGPT Legacy and ChatGPT Plus performed worse than the human scores. Another MCQ study showed that ChatGPT could not correctly answer a sufficient number of questions from the OphthoQuestions practice question bank.[13] ChatGPT 3.5 achieved similar performance on the MCQs from the Review Questions in Ophthalmology book in our study to previous MCQ studies in ophthalmology or general medicine.[12,14]

Furthermore, ChatGPT has shown remarkable improvement in performance as its version was updated from ChatGPT Legacy to ChatGPT Plus.[12] The factors predicting answer accuracy were question difficulty, cognitive level, and examination section. The updated ChatGPT 3 Plus version was less affected by the examination section, but it performed poorly in neuro-ophthalmology, oculoplastics, and clinical optics.[12]

In our study, ChatGPT 3.5 performed worst in the MCQs on optics (32.6%) and embryology and anatomy (40.9%), consistent with ophthalmologists being more familiar than ChatGPT with these topics since they constitute the fundamentals of clinical practice. However, ChatGPT 3.5 performed worst in the case book questions on trauma (38.2%), followed by tumors (40.8%) and posterior segment complications (46.1%). However, it should be considered that these sections contained fewer than five questions. In addition, ChatGPT 3.5 performed poorly on the case book questions on neuro-ophthalmology (46.5%) and orbit/oculoplastics (46.7%), consistent with previous

Table 1. The distribution of answers among subfields for case-based questions book

	n		Diagnosis (%)	Differential diagnosis (%)	Medical management (%)	Surgical management (%)	Rehabilitation and follow-up (%)	Total (%)
Cornea and external disease	13	True	69.2	56	67	53	61.5	60.2
		False	30.8	44	33	47	38.5	
Lens	5	True	80.0	67	35	90	100	65.7
		False	20.0	33	65	10	0	
Glaucoma	12	True	58.3	24	61	50	58.3	54.0
		False	41.7	76	39	50	41.7	
Retina	14	True	78.6	53	68	87	64.3	69.5
		False	21.4	47	32	13	35.7	
Uveitis	3	True	66.7	49	58	50	66.7	52.3
		False	33.3	51	42	50	33.3	
Tumors	2	True	100.0	32	NA	50	0	40.8
		False	0	68	NA	50	100	
Posterior segment complications	4	True	100.0	13	62	68	75	46.1
		False	0	87	38	32	25	
Trauma	2	True	50.0	50	50	20	0	38.2
		False	50.0	50	50	80	100	
Neuro-Ophthalmology	20	True	50.0	37	60	64	55	46.5
		False	50.0	63	40	36	45	
Pediatric Ophthalmology	10	True	80.0	26	69	83	40	65.6
		False	20.0	74	31	17	60	
Orbit/Oculoplastics	13	True	61.5	37	42	53	76.9	46.7
		False	38.5	63	58	47	23.1	
Total	98	True	67.3	40	59	64	60.2	56.1
		False	32.7	60	41	36	39.8	

Table 2. The distribution of answers among subfields for multiple choice book

	Total n	True n (%)	False n (%)
Fundamentals	50	28 (56.0)	22 (44.0)
Embryology and Anatomy	44	18 (40.9)	26 (59.1)
Optics	46	15 (32.6)	31 (67.4)
Neuroophthalmology	63	34 (53.9)	29 (46.0)
Pediatric ophthalmolgy and Strabismus	67	34 (50.7)	33 (49.3)
Plastics	63	35 (55.6)	28 (44.4)
Pathology	6	4 (66.7)	2 (33.3)
Uveitis	54	34 (62.9)	20 (37.0)
Glaucoma	85	46 (54.1)	39 (45.9)
Cornea	63	38 (60.3)	25 (39.7)
Lens and Cataract	28	17 (60.7)	11 (39.3)
Retina and Vitreous	74	41 (55.4)	33 (44.6)
Total	643	344 (53.5)	299 (46.5)

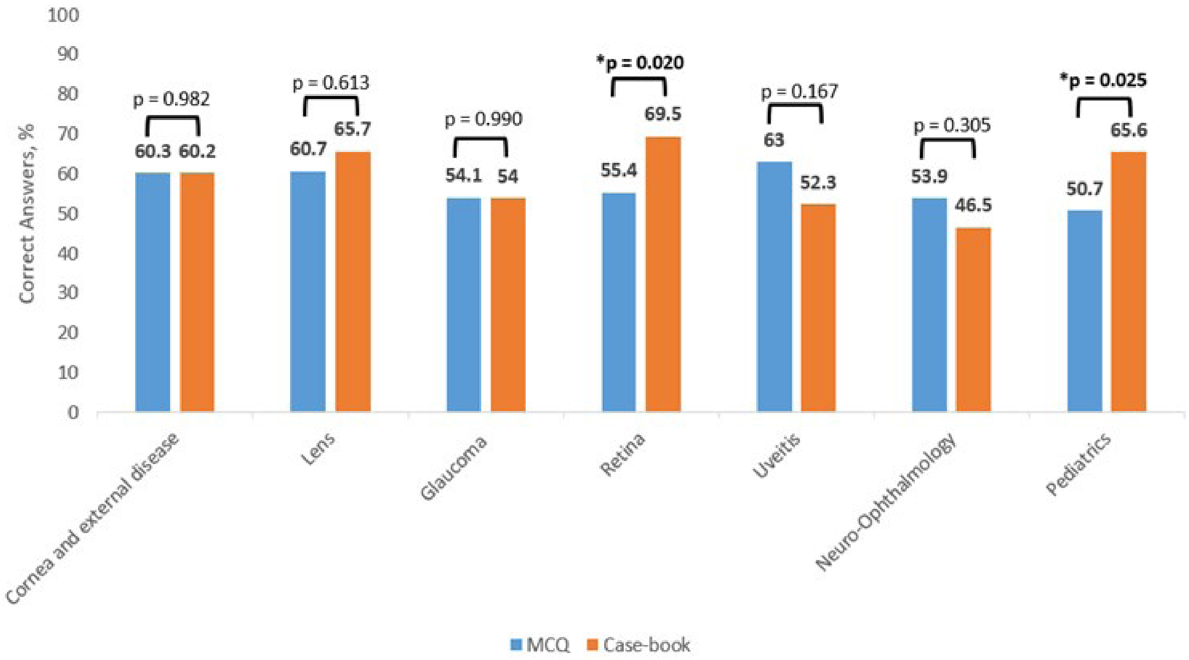


Figure 3. Pairwise comparison of correct answer rates for common subfields between two books.
* p values were obtained by Chi-Square test

MCQ studies.[12]

ChatGPT uses a vast amount of available data and resources on the internet. Like human respondents, ChatGPT is known to perform better on easier questions. Moshirfar et al.[15] demonstrated that ChatGPT 3.5 and ophthalmology professionals performed similarly on 467 ophthalmology StatPearls questions, although the performance gap increased with question difficulty. However, specialized domains, including optics, neuro-ophthalmology, and oculoplastics, are highly challenging and less familiar topics even within the ophthalmology community.[16] In addition, there is a knowledge cutoff of September 2023 for all versions of ChatGPT.

Notably, our study examined many more MCQs than previous studies. Our study found that ChatGPT 3.5 excelled in MCQs on pathology (66.7%), uveitis (63.0%), lens and cataracts (60.7%), and cornea (60.3%). Each section contained more than 25 questions except for pathology, which contained six. The correct answer rates for MCQs differ between our study and previous MCQ studies, possibly reflecting differences in the distribution and difficulty of questions in each section.[12,15] In addition, the latest version of ChatGPT outperforms previous versions. Moreover, Moshirfar et al. found that ChatGPT 4.0 was superior to humans on ophthalmology StatPearls questions.[15]

Furthermore, ChatGPT achieved the highest performance (69.5%) on the case book questions in the retina section. Indeed, its performance was significantly higher than our theoretical knowledge benchmark, the MCQs. Given its significantly higher performance on the case book questions on pediatric ophthalmology than the MCQs, this result indicates that ChatGPT 3.5 could be a tool for retinal diseases, as previously demonstrated with earlier versions, and for pediatric ophthalmology. It also achieved high performance on questions about lens (65.7%), pediatric ophthalmology (65.6%), and cornea and external disease (60.2%). In addition, it performed best in diagnosis and worst in differential diagnosis. Therefore, ChatGPT receives appropriate updates that are considered to be the results of user inputs. This outcome might be encouraging and notable in ophthalmology despite being an example of a specialized case book examination. However, ChatGPT provided longer and more vague explanations for open-ended questions than the case book answers.

Limitations

Our study was limited by its relatively small sample size of 98 case-book questions, even when considering their divisions. While our study included 643 MCQs, many more than previous studies, these comparisons may not represent the entire knowledge of ophthalmology, especially for clinical practice. Furthermore, our study only used ChatGPT 3.5, which might not reflect all available models at the time of publication. The last version (ChatGPT 4.0) was released in March 2023. However, it is less accessible and had a limited capacity of 25 messages during our study period. Our study did not assess the difficulty levels of the included questions. Additionally, unlike previous studies, our study did not compare answers between ChatGPT and human respondents.

In conclusion, our study suggests that ChatGPT may be suitable for future integration into clinical decision-making in ophthalmology, especially in the retina and pediatric ophthalmology sub-fields. Despite its promising performance, the existing limitations of ChatGPT may be overcome by enabling it to process images and incorporating other transformer models or domain-specific sources. Further studies are needed to assess the clarity and acceptability of LLM answers to open-ended questions.

Conflict of Interest: The authors declare no conflict of interest related to this article.

Funding sources: The authors declare that this study has received no financial support.

Ethics Committee Approval: No ethical board registration needed for this study.

ORCID and Author contribution:

Peer-review: Externally peer reviewed.

REFERENCES

1. Li JO, Liu H, Ting DSJ, Jeon S, Chan RVP, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. Prog Retin Eye Res. 2021;82:100900. doi: 10.1016/j.preteyeres.2020.100900.
2. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel). 2023;11(6):887. doi: 10.3390/healthcare11060887.
3. Introducing ChatGPT. https://openai.com/blog/chatgpt Accessed May 17, 2023.
4. Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus. 2023;15(2):e35179. doi: 10.7759/cureus.35179.
5. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med Educ. 2023;9:e45312. doi: 10.2196/57594.
6. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198. doi: 10.1371/journal.pdig.0000198.
7. Cai LZ, Shaheen A, Jin A, Fukui R, Yi JS, et al. Performance of Generative Large Language Models on Ophthalmology Board-Style Questions. Am J Ophthalmol. 2023;254:141-9. doi: 10.1016/j.ajo.2023.05.024.
8. Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. NPJ Digit Med. 2019;2:25. doi: 10.1038/s41746-019-0099-8.
9. Bogunovic H, Montuoro A, Baratsits M, Karantonis MG, Waldstein SM, et al. Machine Learning of the Progression of Intermediate Age-Related Macular Degeneration Based on OCT Imaging. Invest Ophthalmol Vis Sci. 2017;58(6):BIO141-50. doi: 10.1167/iovs.17-21789.
10. Singh K, Smiddy WE, Lee AG. Ophthalmology review : a case-study approach. Second edition. Thieme; 2018
11. Kenneth C. Chern, Michael A. Saidel. Review Questions in Ophthalmology. Third edition. Wolters Kluwer; 2014
12. 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(4):100324. doi: 10.1016/j.xops.2023.100324.
13. Mihalache A, Popovic MM, Muni RH. Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment. JAMA Ophthalmol. 2023;141(6):589-97. doi: 10.1001/jamaophthalmol.2023.1144.
14. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172-180. doi: 10.1038/s41586-023-06291-2.
15. Moshirfar M, Altaf AW, Stoakes IM, Tuttle JJ, Hoopes PC. Artificial Intelligence in Ophthalmology: A Comparative Analysis of GPT-3.5, GPT-4, and Human Expertise in Answering StatPearls Questions. Cureus. 2023;15(6):e40822. doi: 10.7759/cureus.40822.
16. Stunkel L, Mackay DD, Bruce BB, Newman NJ, Biousse V. Referral Patterns in Neuro-Ophthalmology. J Neuroophthalmol. 2020;40(4):485-93. doi: 10.1097/WNO.0000000000000846.