Int J Ophthalmol, Vol. 18, No. 7, Jul. 18, 2025 www.ijo.cn
Tel: 8629-82245172 8629-82210956 Email: ijopress@163.com

• Intelligent Ophthalmology •

Large language models in neuro-ophthalmology diseases: ChatGPT *vs* Bard *vs* Bing

Dong Hee Ha[1], *Ungsoo Samuel Kim*[2,3]

[1]Department of Ophthalmology, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul 06973, Republic of Korea
[2]Department of Ophthalmology, Chung-Ang University College of Medicine, Seoul 06974, Republic of Korea
[3]Department of Ophthalmology, Chung-Ang University Gwangmyeong Hospital, Gyeonggi-do 14353, Republic of Korea
Correspondence to: Ungsoo Samuel Kim. Gwangmyeong Hospital, Deokan-ro 110, Gwangmyeong-si, Gyeonggi-do 14353, Republic of Korea. ungsookim@cau.ac.kr
Received: 2024-03-26 Accepted: 2025-04-09

Abstract
● **AIM:** To investigate the capabilities of large language models (LLM) for providing information and diagnoses in the field of neuro-ophthalmology by comparing the performances of ChatGPT-3.5 and -4.0, Bard, and Bing.
● **METHODS:** Each chatbot was evaluated for four criteria, namely diagnostic success rate for the described case, answer quality, response speed, and critical keywords for diagnosis. The selected topics included optic neuritis, non-arteritic anterior ischemic optic neuropathy, and Leber hereditary optic neuropathy.
● **RESULTS:** In terms of diagnostic success rate for the described cases, Bard was unable to provide a diagnosis. The success rates for the described cases increased in the order of Bing, ChatGPT-3.5, and ChatGPT-4.0. Further, ChatGPT-4.0 and -3.5 provided the most satisfactory answer quality for judgment by neuro-ophthalmologists, with their sets of answers resembling the sample set most. Bard was only able to provide ten differential diagnoses in three trials. Bing scored the lowest for the satisfactory standard. A Mann-Whitney test indicated that Bard was significantly faster than ChatGPT-4.0 (Z=-3.576, P=0.000), ChatGPT-3.5 (Z=-3.576, P=0.000) and Bing (Z=-2.517, P=0.011). ChatGPT-3.5 and -4.0 far exceeded the other two interfaces at providing diagnoses and were thus used to find the critical keywords for diagnosis.
● **CONCLUSION:** ChatGPT-3.5 and -4.0 are better than Bard and Bing in terms of answer success rate, answer quality, and critical keywords for diagnosis in ophthalmology. This study has broad implications for the field of ophthalmology, providing further evidence that artificial intelligence LLM can aid clinical decision-making through free-text explanations.
● **KEYWORDS:** large language model; chatbot; ChatGPT; Bard; Bing; neuro-ophthalmology
DOI:10.18240/ijo.2025.07.05

Citation: Ha DH, Kim US. Large language models in neuro-ophthalmology diseases: ChatGPT *vs* Bard *vs* Bing. *Int J Ophthalmol* 2025;18(7):1231-1236

INTRODUCTION

Artificial intelligence (AI) and deep-learning methods have emerged as topics of interest in various fields of study. AI and natural language processing have steadily overcome the daunting challenges associated with the intricate and elusive nature of human languages, and such rapid progress has enabled the application of AI in diverse domains[1]. However, the field of medicine still remains as a challenge. When making medical decisions, clinicians often rely on uncertain, incomplete, heterogeneous, inaccurate, and missing datasets in arbitrarily high-dimensional spaces[2-3]. These characteristics pose great challenges to the application of AI in medicine.

Previous research suggested that AI holds potential for disease diagnosis, the development of personalized treatment plans, and clinical decision support[4-5]. These applications emphasize AI's growing role in enhancing efficiency and precision of healthcare. Specifically, in ophthalmology, deep learning-based AI has been applied to the interpretation of various diagnostic modalities such as fundus photographs, optical coherence tomography, and visual fields, for the detection of conditions including diabetic retinopathy, glaucoma-like disc, and age-related macular degeneration[6-9]. More recently, large language models (LLMs) have gained increasing attention in the medical field. LLMs integrated with chatbots such as ChatGPT-3.5 and -4.0 (OpenAI, CA, USA), Google Bard (Alphabet Inc., CA, USA), and Bing Chat (Microsoft Corporation, WA, USA) have been trained on enormous amounts of data to embody

Table 1 Differential diagnoses for Leber hereditary optic neuropathy and non-arteritic anterior ischemic optic neuropathy

Leber hereditary optic neuropathy	Non-arteritic anterior ischemic optic neuropathy
1) Inherited optic atrophy	1) Optic neuritis
2) Compressive optic atrophy	2) Compressive optic neuropathy
3) Vascular optic atrophy	3) Sphenoid sinus related disease
4) Inflammatory optic atrophy	4) Sarcoidosis
5) Infection optic atrophy	5) Vasculitis
6) Nutritional optic atrophy	6) Infectious optic neuropathy syphilis…
7) Toxic optic atrophy	7) Arteritic anterior ischemic optic neuropathy
8) Traumatic optic neuropathy	8) Toxic or nutritional optic neuropathy
9) Optic neuritis	9) Leber hereditary optic neuropathy
10) Mitochondria disorders other than LHON	10) Postviral demyelination

the generation of sophisticated text and answer complicated questions in a human-like manner. Despite this rising interest, LLMs in AI research in neuro-ophthalmology remains limited. Thus, the main aim of this study was to investigate the abilities of LLMs to provide information and diagnoses in the field of neuro-ophthalmology by comparing the performances of ChatGPT, Bard, and Bing.

MATERIALS AND METHODS

This study evaluates the capacities of three LLMs with chat interfaces, namely ChatGPT-3.5 and -4.0, Google Bard, and Bing Chat. Further, three topics of interest were selected for this study: optic neuritis (ON), non-arteritic anterior ischemic optic neuropathy (NAAION), and Leber hereditary optic neuropathy (LHON). An experienced neuro-ophthalmologist selected three representative neuro-ophthalmologic diseases and evaluated the responses produced by each chatbot based on scoring criteria created according to established textbooks guidelines. The scoring criteria for each question are detailed in the appendix section, presented in the form of tables.

We tested each LLM chatbot five times on sample questions covering general information and guidelines for treatment as well as evaluated the capacity of each chatbot to diagnose neuro-ophthalmological diseases. The basic settings used were the default conditions for ChatGPT and Bard and balanced results for Bing AI. After each answer to a query, the memory was cleared for all chatbots to prevent the previous queries from influencing the subsequent outputs. No additional training data or instructions were provided.

Each chatbot was evaluated on the basis of four criteria, namely diagnostic success rate for the described case, answer quality, response speed, and critical keywords for diagnosis. The critical keywords are terms that are essential for making a diagnosis with the LLM. When an LLM chatbot could not answer a question, it was evaluated as incorrect. If the diagnosis in question was included among the differential diagnoses and mentioned in the answer, the response was considered correct.

The diagnostic success rates for the described cases were measured by evaluating the rates of typical case scenarios presented and answered with the expected correct diagnoses. The answer quality was evaluated by checking whether each chatbot provided satisfactory standards for judgment by neuro-ophthalmologists. To evaluate the answer quality, we compared the results based on differential diagnoses from neuro-ophthalmology textbooks. Four sample questions were used to evaluate the answer quality for ten differential diagnoses for LHON, ten differential diagnoses for NAAION, five differences between ON and neuromyelitis optica, and five important messages for LHON patients during genetic counseling. The ten differential diagnoses each for LHON were selected from the *Will's Eye Handbook of Ocular Genetics*[10] and *Oxford Handbook of Ophthalmology*[11] (Table 1). The ten differential diagnoses each for NAAION were chosen from the *Clinical Neuro-Ophthalmology*: *A Practical Guide*[12] and *Oxford Handbook of Ophthalmology*[11] (Table 1). Based on a study by Srikajon *et al*[13] and the *Will's Eye Handbook of Ocular Genetics*[10], the five key differences between ON and neuromyelitis optica were selected as follows: 1) ophthalmologic exam findings and radiographic images; 2) scope of involvement; 3) associated autoantibodies; 4) response to treatment; 5) relapsing patterns.

The key differences among the aforementioned categories are described in detail as follows. The visual acuity was worse for acute ON attacks in the neuromyelitis optica spectrum disorder (NMOSD) than those found in multiple sclerosis (MS)[13]. Moreover, the NMOSD patients tended to have altitudinal visual field (VF) defects while MS patients tended to show arcuate-type VF defects. The NMOSD group had a trend toward lower retinal nerve fiber layer thickness compared to the MS group in the acute phase. Spinal magnetic resonance imaging (MRI) of longitudinal extensive transverse myelitis suggests NMOSD, whereas the presence of cerebrospinal fluid-oligoclonal bands and brain MRI fulfilling Swantom criteria are more indicative of MS[13]. Many NMOSD patients are anti-AQP4-Ab seropositive[13]. Early differential diagnosis between ON-NMOSD and ON-MS is essential in appropriately managing the patients, and misdiagnosis of ON-NMOSD as ON-MS may lead to inappropriate administration

Table 2 Answers by ChatGPT-4.0, -3.5, Bard, and Bing for four sample questions

Sample questions	ChatGPT 4.0	ChatGPT 3.5	Bard	Bing
What are 10 differential diagnoses for LHON	Optic neuritis Ischemic optic neuropathy Toxic optic neuropathy Nutritional optic neuropathy Compressive optic neuropathy Glaucoma Traumatic optic neuropathy Demyelinating diseases Infectious or inflammatory causes Hereditary or Genetic disorders	Optic neuritis Dominant optic atrophy Nutritional optic neuropathy Toxic optic neuropathy Ischemic optic neuropathy Compressive optic neuropathy LHON Primary open-angle glaucoma Retinitis pigmentosa Retinal detachment	Optic neuritis Autosomal dominant optic atrophy Wolfram syndrome Metabolic optic neuropathies (toxic, nutritional, and combinations) Chiasmal tumors Anterior ischemic optic neuropathy Multiple sclerosis NMOSD Thyroid eye disease Sarcoidosis	Retinal dystrophies or other retinal lesions Compressive lesions Dominant optic atrophy Toxic/nutritional neuropathies Optic neuritis Glaucoma Ischemic optic neuropathy Optic nerve tumors Neuromyelitis optica Multiple sclerosis
What are 10 differential diagnoses for NAAION	AAION Optic neuritis LHON Compressive optic neuropathy Toxic optic neuropathy Nutritional optic neuropathy Traumatic optic neuropathy Optic nerve glioma Optic nerve drusen CRAO	AAION Optic neuritis Compressive optic neuropathy Ischemic retinopathy Retinal artery or vein occlusion CRAO Giant cell arteritis Retinal detachment Ischemic optic neuropathy Papilledema	Optic neuritis Autosomal dominant optic atrophy Wolfram syndrome Metabolic optic neuropathies (toxic, nutritional, and combinations) Chiasmal tumors CRAO CRVO AMD Retinitis pigmentosa Thyroid eye disease	AAION Giant cell arteritis Optic neuritis Glaucoma Compressive optic neuropathy Inflammatory optic neuropathy Ischemic optic neuropathy Optic nerve drusen Papilledema Optic nerve head coloboma
What are 5 differences between optic neuritis and neuromyelitis optica	Clinical presentation Pathogenesis Imaging Serological testing Treatment and prognosis	Scope of involvement Associated autoantibody Risk of recurrence Spinal cord involvement Response to treatment	Age of onset Symptoms Cause Treatment Prognosis	Unilaterality Associated diseases Onset Symptoms Spinal cord involvement
What are 5 important messages for LHON patients when doing genetic counseling	Inheritance pattern Penetrance and gender bias Variability in expression Potential therapies and interventions Family planning	Inheritance pattern and risks Genetic testing and results Family planning and reproductive options Potential for vision loss Research and support services	Genetic disease Risk of development Treatment and prognosis Genetic testing and results Support groups	Inheritance Genetic testing Penetrance Treatment Support groups

LHON: Leber hereditary optic neuropathy; NMOSD: Neuromyelitis optica spectrum disorder; AAION: Arteritic anterior ischemic optic neuropathy; CRAO: Central retinal artery occlusion; CRVO: Central retinal vein occlusion; AMD: Age-related macular degeneration; NAAION: Non-arteritic anterior ischemic optic neuropathy.

of interferon-beta, natalizumab, and fingolimod, which may worsen NMOSD relapses[13].

The five important representative messages to deliver during genetic counseling for patients with LHON were chosen from *Will's Eye Handbook of Ocular Genetics*[10] as follows: 1) causes and inheritance patterns; 2) variable penetrance and expressions; 3) risk of vision loss and symptoms; 4) treatment; 5) family planning and genetic testing.

Response speed was defined as the time required to answer each sample question. Statistical analyses were conducted using SPSS (SPSS Inc., Chicago, IL, USA). A Mann-Whitney test was conducted to compare response speeds of chatbots. *P*-values less than 0.05 were considered statistically significant. Finally, the critical keywords needed to accurately infer a diagnosis were identified.

RESULTS

In terms of diagnostic success rates for the described cases, Bard was unable to arrive at a diagnosis. When Bing was asked to find a diagnosis for the NAAION text, NAAION came out as a possible answer within three trials. ChatGPT-3.5 was correct once at making a diagnosis for the ON text, three times for the NAAION text, and five times for the LHON text. ChatGPT-4.0 was correct in all attempts at providing diagnoses for ON, NAAION, and LHON texts.

All models provided differential diagnoses for the proposed diseases. ChatGPT-4.0 and -3.5 provided the most satisfactory answer quality for judgment by neuro-ophthalmologists, with the sets of answers that most resembled the sample set (Table 2). Bard was only able to provide ten differential diagnoses among three trials (Table 3). Clinically, Bard suggested more specific differential diagnoses, such as sarcoidosis and Wolfram syndrome for NAAION (Table 2). Bing scored the lowest for the satisfactory standard (Table 3).

A Mann-Whitney test was performed to evaluate whether the response speeds differed based on the chatbots used (Table 4, Figure 1). The results indicated that Bard was significantly faster than ChatGPT-4.0 (Z=-3.576, P=0.000), ChatGPT-3.5 (Z=-3.576, P=0.000), and Bing (Z=-2.517, P=0.011; Figure 1).

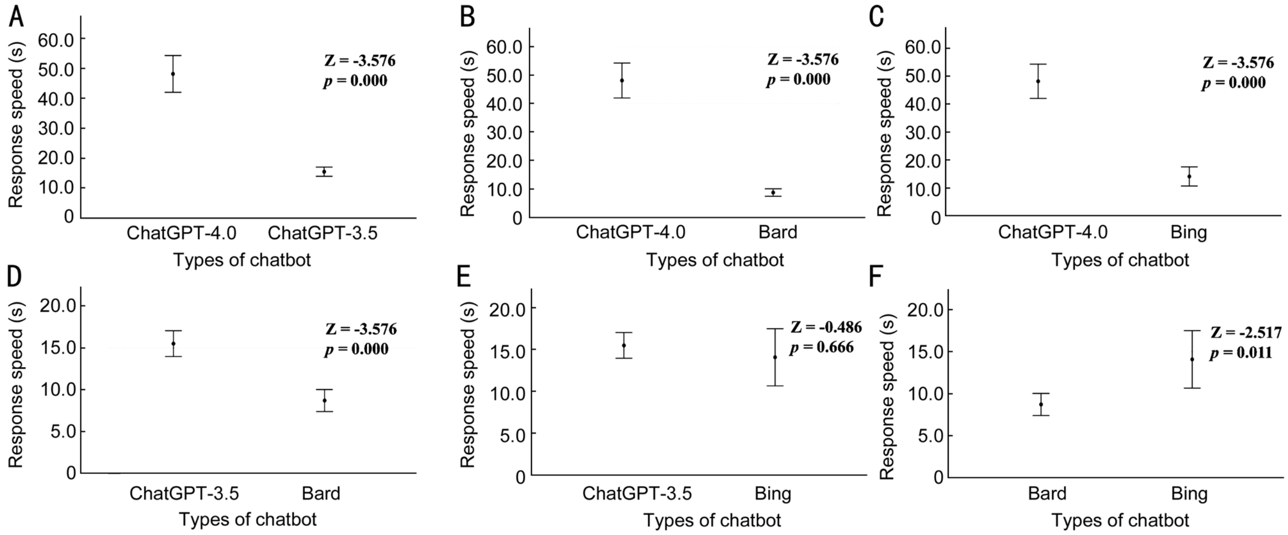

Figure 1 Comparison of ChatGPT-3.5 and -4, Bard, and Bing in response speed by boxplots.

Table 3 Number of correct answers by each chatbot for four sample questions across five independent trials

Parameters	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
ChatGPT-4.0	9/10	8/10	9/10	9/10	9/10
	6/10	3/10	3/10	3/10	3/10
	4/5	4/5	4/5	5/5	3/5
	5/5	5/5	5/5	5/5	5/5
ChatGPT-3.5	7/10	7/10	7/10	8/10	8/10
	7/10	6/10	3/10	4/10	3/10
	4/5	4/5	4/5	4/5	4/5
	5/5	5/5	5/5	5/5	5/5
Bard	4/10	2/10	2/10	2/10	1/10
	3/10	2/10	2/9	2/10	1/9
	3/5	3/5	3/5	2/5	3/5
	4/5	4/5	4/5	2/5	2/5
Bing	5/10	4/10	5/10	5/10	5/10
	5/10	5/10	5/10	4/10	2/10
	0/5	2/5	2/5	2/5	2/5
	4/5	3/5	3/5	2/5	2/5

Each row represents one trial, showing chatbot performance on the four sample questions.

ChatGPT-3.5 and Bing showed no significant difference in their response speeds.

ChatGPT-3.5 and -4.0 far exceeded the other two interfaces at making diagnoses and were thus used to find the critical keywords for diagnosis. ChatGPT-3.5 and -4.0 were the only chatbots that could provide critical terms for diagnoses of neuro-ophthalmological diseases. The critical terms for diagnosing ON, NAAION, and LHON were respectively "ocular pain when moving her eyes", "a 67-year-old man" in conjunction with "optic nerve head edema observed on fundus examination" and "an 18-year-old man", and "sudden visual loss" combined with "a maternal relative exhibiting similar symptom".

DISCUSSION

This study is a pilot attempt to compare LLM-driven chatbots and analyze the capability of each interface at providing accurate answers to sample questions covering general information and guidelines for the treatment of neuro-ophthalmological diseases. We also evaluated the capacity of each chatbot to diagnose representative cases of neuro-ophthalmological diseases, such as ON, NAAION, and LHON. ChatGPT-3.5 and -4.0 notably outperformed the other two interfaces in terms of success rates for the described cases at making diagnoses. Particularly, ChatGPT-4.0 showed outstanding performance by making correct diagnoses for all trials. In comparison, Bing was not suitably equipped to function as a diagnostic instrument but there was still a significant potential for refinement. ChatGPT-4.0 and -3.5 provided the most satisfactory answer quality for judgment by neuro-ophthalmologists, providing sets of answers that were very similar to the sample set. Notably, there were no cases in which the errors made by Bard or Bing were entirely unrelated to the topic. However, their responses were generally less detailed than those generated by ChatGPT, as a smaller proportion closely matched the reference answer set, as shown in Table 3. In terms of response speed, Bard was significantly faster than the other chatbots. ChatGPT was the only chatbot that could provide the critical keywords for diagnosis.

ChatGPT's exceptional performance highlighted it as the most satisfactory model out of the three chatbots. ChatGPT-4.0 provided the most in-depth and accurate information. The differences in the capabilities of ChatGPT and Bard are derived from the differences in their architectures and training methodologies[1]. Both ChatGPT and Bard utilize the Transformer neural network architecture. The attention, which is the core methodology of the Transformer neural network architecture developed by Google, allows the model

Table 4 Comparison of ChatGPT-3.5 and -4.0, Bard, and Bing in response speed

Parameters	General information			Guidelines for treatments			Diagnostic tool		
	Optic neuritis	NAAION	LHON	Optic neuritis	NAAION	LHON	Optic neuritis	NAAION	LHON
ChatGP-3.5	00:14.7	00:13.3	00:12.8	00:18.1	00:16.0	00:18.0	00:17.5	00:14.5	00:14.5
ChatGPT-4.0	00:39.5	00:39.2	00:37.7	00:50.1	00:57.5	00:55.4	00:48.5	00:58.3	00:47.4
Bard	00:06.3	00:08.3	00:09.8	00:07.5	00:08.1	00:06.8	00:10.2	00:10.9	00:10.7
Bing	00:17.6	00:14.3	00:15.2	00:14.5	00:08.7	00:08.5	00:09.1	00:18.1	00:20.7

NAAION: Non-arteritic anterior ischemic optic neuropathy; LHON: Leber hereditary optic neuropathy.

to concentrate on certain input sequence segments during prediction[14]. In comparison with traditional models, the Transformer architecture far exceeds other models in scenarios where the input sequence is complex and lengthy[1]. This mechanism enables ChatGPT and Bard to understand complex language patterns more effectively and generate more coherent and contextual texts[1]. In particular, reinforcement learning with human feedback incorporated into ChatGPT helps it in the process of learning how to follow instructions and in providing user friendly answers that fit human preferences[1].

Another important factor that determines the capability of a chatbot is its tokenization, which is the process of dividing the text into certain units called tokens. Every chatbot's language model has a particular maximum number of tokens[15]. The number of tokens utilized in a dataset plays an essential role in generating responses. OpenAI does not disclose the information in the public domain but GPT-4.0 contains 20T estimated tokens, which is seven times that of the 2.81T tokens in the Google Infinite Dataset used by Bard[16]. In terms of response accuracy, researchers have found ChatGPT is more accurate and more satisfactory in diverse fields because of Bard's draft system that provides users with multiple response options and allows them to select the most resonant answer[1].

There is no profound answer to why Microsoft Bing responded the fastest, but this may be due to the fact that the length of text provided by Bing is the shortest. Bing concentrates on content creation rather than generation of extensive essays[17]. Short answers and the tendency to focus on content creation could have resulted in the quickest response times noted.

With contextually proper responses to user prompts, LLMs can produce natural language text and converse in human-like language, allowing AI systems to generate human-like content[18]. Among the many AI-based chatbots available today, GPT-4.0 is one of the most developed multimodal models, with outstanding performance in generating human-like text based on user prompts. Compared with its predecessors such as GPT-3.5 and its variants, GPT-4.0 has a larger architectural model size[19]. Consequently, the increased model size enhances its natural language processing capabilities, enabling more valid and relevant responses and better reasoning[19].

The distinct nature of ophthalmology as a niche specialty makes it difficult to learn, thus posing many obstacles. From an educator's perspective, ophthalmology presents complexities when integrated with other broader medical themes[7] and can appear technically challenging to teach. From the learner's perspective, ophthalmology as a specialized field may not garner enough attention and be overshadowed by other disciplines. Neuro-ophthalmology specifically requires years of extensive training, given its integration of disciplines like neurology, neuroradiology, and neurosurgery. Yet, training programs in the USA are highly variable in quality and have not been subject to standardization or regulation[20]. LLMs can thus offer significant opportunities to advance ophthalmic education and primary care by improving accessibility to information. For example, medical students and ophthalmic residents can interact with LLMs to ask questions, clarify doubts, and gain deeper understanding of complex topics while cost-effectively supplementing traditional training methods. To illustrate, LLMs can produce clinical case studies, act as virtual test subjects, accelerate research outputs, and provide personalized feedback and assistance[18].

This study has broad implications for the field of ophthalmology, where large-scale medical AI models are being developed to aid clinical decision-making through free-text explanations, spoken recommendations, or image annotations[6,21]. Future research may concentrate on comparing these models with human experts or refining them to address complex and rare cases in neuro-ophthalmology. With ongoing advancements, AI has the potential to serve as a valuable complementary tool in clinical practice rather than a replacement for human judgment. Recognizing AI as a complementary tool is important due to ethical and practical issues associated with AI use in medicine, including data privacy, clinical accountability, and the necessity of medical oversight in automated decision-making.

Medical AI technologies have great potential but also pose limitations and challenges. There are possibilities that results may have been affected by the quality of input data, how the questions were formulated, or the nature of the clinical case simulation. The specialized medical information used to teach LLMs is also limited; thus, further research into LLMs can help medical providers in various clinical settings. It is essential to understand here that while LLMs offer numerous new opportunities, they should be used wisely and in conjunction with traditional educational and clinical methods. Misdiagnoses or over-reliance on such models without proper clinical context can be risky; while language models can assist with providing general information, relying on them for diagnoses is not advisable.

In conclusion, ChatGPT-3.5 and -4.0 may outperform Bard and Bing in terms of their answer success rates, answer quality, and critical keywords for diagnosis in ophthalmology. It is important to keep in mind that the specialized medical information used to teach LLMs is limited and thus the results may not be fully generalizable across all areas of medicine. Given the evolving nature of LLMs, ongoing updates and rigorous research are essential to ensure their reliable and safe integration into clinical practice. Further investigation is critical to advancing the utility of LLMs in healthcare; however, such efforts must be undertaken with caution and a clear understanding of the current limitations.

ACKNOWLEDGEMENTS

Authors' Contributions: Kim US conceptualized, investigated and curated data, and reviewed and edited the manuscript. Ha DH investigated and curated data, and wrote the original draft of the manuscript. All authors read and approved the final manuscript

Availability of Data and Materials: The datasets generated and/or analyzed during the current study are available in the ChatGPT-3.5 and -4.0, Bard, and Bing repository, https://chat.openai.com/, https://www.bing.com/ /ai, https://bard.google.com/chat.

Conflicts of Interest: Ha DH, None; **Kim US,** None.

REFERENCES

1 Ahmed I, Kajol M, Hasan U, *et al.* ChatGPT vs. Bard: a comparative study. *Engineering Reports* 2024;6(11):e12890.

2 Holzinger A, Dehmer M, Jurisica I. Knowledge discovery and interactive data mining in bioinformatics—state-of-the-art, future challenges and research directions. *BMC Bioinformatics* 2014;15(Suppl 6):I1.

3 Lee S, Holzinger A. *Knowledge discovery from complex high dimensional data*. Solving Large Scale Learning Tasks. Challenges and Algorithms: Essays Dedicated to Katharina Morik on the Occasion of Her 60th Birthday, 2016:148-167.

4 Alowais SA, Alghamdi SS, Alsuhebany N, *et al*. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ* 2023;23(1):689.

5 Bozkurt S, Fereydooni S, Kar I, *et al*. Investigating data diversity and model robustness of AI applications in palliative care and hospice: protocol for scoping review. *JMIR Res Protoc* 2024;13:e56353.

6 Ting DSW, Pasquale LR, Peng L, *et al*. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103(2):167-175.

7 Grzybowski A, Jin K, Zhou JX, *et al*. Retina fundus photograph-based artificial intelligence algorithms in medicine: a systematic review. *Ophthalmol Ther* 2024;13(8):2125-2149.

8 Restrepo D, Quion JM, Do Carmo Novaes F, *et al*. Ophthalmology optical coherence tomography databases for artificial intelligence algorithm: a review. *Semin Ophthalmol* 2024;39(3):193-200.

9 Lee GA, Kong GYX, Liu CH. Visual fields in glaucoma: where are we now? *Clin Exp Ophthalmol* 2023;51(2):162-169.

10 Levin AV, Zanolli M, Capasso JE. *Wills eye handbook of ocular genetics*. Thieme. 2016.

11 Denniston AKO, Murray PI. *Oxford Handbook of Ophthalmology*. Oxford University Press; 2018.

12 Schiefer U, Hart W, Hoyt WF, *et al*. *Clinical neuro-ophthalmology: a practical guide*. Springer Berlin Heidelberg; 2007.

13 Srikajon J, Siritho S, Ngamsombat C, *et al*. Differences in clinical features between optic neuritis in neuromyelitis optica spectrum disorders and in multiple sclerosis. *Mult Scler J Exp Transl Clin* 2018;4(3):2055217318791196.

14 Waswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. arXiv:1706.03762

15 Toraman C, Yilmaz EH, Şahi̇nuç F, *et al*. Impact of tokenization on language models: an analysis for Turkish. *ACM Trans Asian Low-Resour Lang Inf Process* 2023;22(4):1-21.

16 Thoppilan R, De Freitas D, Hall J, *et al*. LaMDA: language models for dialog applications. 2022:2201.08239. https://arxiv.org/abs/2201.08239v3.

17 Ventayen RJM. OpenAI ChatGPT, Google Bard, and Microsoft Bing: similarity index and analysis of artificial intelligence-based contents. 2023.

18 Abd-Alrazaq A, AlSaad R, Alhuwail D, *et al*. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ* 2023;9:e48291.

19 Koubaa A. GPT-4 vs. GPT-3.5: a concise showdown. Preprints 2023:2023030422.

20 Lee AG, Keltner JL. Fellowship training in neuro-ophthalmology in the United States. *J Neuroophthalmol* 2015;35(1):1-2.

21 Moor M, Banerjee O, Abad ZSH, *et al*. Foundation models for generalist medical artificial intelligence. *Nature* 2023;616(7956):259-265.