Large language models and their impact in ophthalmology

Bjorn Kaijun Betzler, Haichao Chen*, Ching-Yu Cheng, Cecilia S Lee, Guochen Ning, Su Jeong Song, Aaron Y Lee, Ryo Kawasaki, Peter van Wijngaarden, Andrzej Grzybowski, Mingguang He, Dawei Li, An Ran Ran, Daniel Shu Wei Ting, Kelvin Teo, Paisan Ruamviboonsuk, Sobha Sivaprasad, Varun Chaudhary, Ramin Tadayoni, Xiaofei Wang, Carol Y Cheung, Yingfeng Zheng, Ya Xing Wang†, Yih Chung Tham†, Tien Yin Wong†*

The advent of generative artificial intelligence and large language models has ushered in transformative applications within medicine. Specifically in ophthalmology, large language models offer unique opportunities to revolutionise digital eye care, address clinical workflow inefficiencies, and enhance patient experiences across diverse global eye care landscapes. Yet alongside these prospects lie tangible and ethical challenges, encompassing data privacy, security, and the intricacies of embedding large language models into clinical routines. This Viewpoint highlights the promising applications of large language models in ophthalmology, while weighing up the practical and ethical barriers towards their real-world implementation. This Viewpoint seeks to stimulate broader discourse on the potential of large language models in ophthalmology and to galvanise both clinicians and researchers into tackling the prevailing challenges and optimising the benefits of large language models while curtailing the associated risks.

Lancet Digit Health 2023; 5: e917–24

*Joint first authors

†Joint last authors

Centre for Innovation and Precision Eye Health (Prof C-Y Cheng PhD, Y C Tham PhD) **and Department of Ophthalmology** (Prof C-Y Cheng, Y C Tham), **Yong Loo Lin School of Medicine, National University of Singapore, Singapore** (B K Betzler MBBS); **Tsinghua Medicine, Tsinghua University, Beijing, China** (H Chen MBBS, Prof T Y Wong PhD); **Singapore Eye Research Institute, Singapore National Eye Centre, Singapore** (Prof C-Y Cheng, D S W Ting PhD, K Teo PhD, Y C Tham, Prof T Y Wong); **Ophthalmology and Visual Science Academic Clinical Program, Duke-NUS Medical School, Singapore** (Prof C-Y Cheng, D S W Ting, K Teo, Y C Tham); **University of Washington School of Medicine, Department of Ophthalmology, Seattle, WA, USA** (C S Lee MD, A Y Lee MD); **Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China** (G Ning PhD); **Department of Ophthalmology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, South Korea** (S J Song PhD); **Division of Public Health, Department of Social Medicine, Graduate School of Medicine, Osaka University, Osaka, Japan** (Prof R Kawasaki PhD); **Artificial Intelligence Center for Medical Research and Application, Osaka University Hospital, Osaka, Japan** (Prof R Kawasaki); **Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Melbourne, VA, Australia** (P van Wijngaarden PhD); **Ophthalmology, University of Melbourne Department of Surgery, East Melbourne, Melbourne, VA, Australia**

Introduction

Since its release in November, 2022, ChatGPT (Open AI, San Francisco, CA, USA) has impressed with its ability to carry out human-like conversations and provide nuanced answers to questions on a wide range of topics. ChatGPT is a chatbot, built on the GPT-3.5 (generative pretrained transformer-3.5) and GPT-4 families of large language models (LLMs), which are a subtype of deep learning systems (table 1). The advanced LLM chatbots in use today have performed well in various academic tasks (table 2 and appendix p 7)[13–17] and can be thought of as personal virtual assistants to anyone with an internet connection—able to receive free-text prompts, understand their semantics, respond accordingly, and track the context of an ongoing personalised conversation.

For patients and health-care providers, the potential applications of LLM technology in medicine are numerous.[18] Examples include facilitating virtual consultations or organising appointments, writing clinical memos or discharge summaries, suggesting treatment options, and helping patients to better self-organise and manage their health information, with a higher degree of personalisation and enhanced scalability and efficiency compared with other existing electronic health-care systems.

In ophthalmology, a continued heavy reliance on tertiary-level health services is increasingly unsustainable with growing and ageing populations. The diverse applications of artificial intelligence (AI), deep learning, and new digital models of care[19–22] to augment, and even disrupt, current systems of eye care are a topic of intense discussion. In this imaging-extensive medical specialty, deep learning algorithms have performed well in detecting diabetic retinopathy,[23,24] glaucoma,[25] age-related macular degeneration,[26] ocular surface diseases,[27,28] and other cases of visual impairment.[29] Although these existing algorithms provide value in terms of the diagnosis and stratification of eye diseases, LLM technology could have benefits that lean towards tackling deficiencies in clinical workflows or supporting a patient's journey through the tertiary eye care system.

The potential applications of LLM technology in ophthalmology signify a new domain in digital eye care, which we examine in this Viewpoint. We categorise these applications into two main areas: improving the patient experience in eye care and optimising the delivery of care by providers. Recognising that innovation often brings challenges, we conclude by highlighting the barriers and limitations associated with these applications of LLMs and sharing potential solutions.

Improving patients' experiences and streamlining their journeys

To understand how LLMs might improve a patient's experience at a tertiary eye centre, we start this Viewpoint by reviewing the current workflow in most ophthalmology facilities (figure 1). Patients usually first present to their community optometrist or primary-care physician, who then makes referrals to eye hospitals. Patients attend appointments for preliminary tests conducted by allied health staff, and nearly all patients receive a detailed examination and consultation by specialist ophthalmologists. This examination is followed by visits to the pharmacy or clinical procedure rooms as appropriate. Subsequently, patients are followed up at the tertiary centre for treatment, surgery, regular screening, or long-term interval observation. This system suffers from several problems,[30,31] including over-referral by primary-care services; long waiting lists for appointments; long waiting times during appointments; convoluted and fragmented pathways from the patient's perspective; and patients remaining indefinitely in the care of tertiary hospitals. This section discusses several applications through which LLMs could potentially overcome or mitigate these challenges.

Facilitating triage and appointment prioritisation

LLMs could be a useful tool for the remote triaging of patients by supplying information and responding to

	Deep learning	GPT	ChatGPT
Architecture and definition	A subset of machine learning; typically uses artificial neural networks to perform tasks; there are many architectures, which can be broadly split into semisupervised and unsupervised algorithms	A specific implementation of deep learning that uses LLMs that are based on transformer architecture for language generation tasks	A variant of GPT, with a chatbot interface to facilitate human interaction with the GPT LLM; this allows the model to generate text in a conversational manner
Input data	Varies, but models are able to decipher complex patterns from unstructured data; neural networks can use image inputs; tabular, text, or video data can also be used	Text data, in sequence-to-sequence tasks	Text input from a user or conversational data
Knowledge coverage	Limited to training data, but requires large amounts of training data	Broad coverage, pretrained on large amounts of text data from the internet, with GPT-4 having additional live access to public web information	Similar to GPT, with live access to public web information, and continuously improved with reinforcement learning from conversational data during user interaction
Efficiency and AI alignment*	One task (or multiple similar tasks) for one model with high alignment	Multiple tasks with low alignment	Multiple tasks with high alignment
Interactivity	None	Low	High (in a human-like manner)
General applications	Can be used for a wide range of tasks, such as image recognition, speech recognition, and language translation	Primarily used for text generation tasks, such as extracting information from texts, summarising and organising texts, translation, and answering questions	Similar functions to GPT, but specifically designed for generating text in a conversational manner; this is personalised with contextual information from previous responses

LLMs are large model sizes, such as GPT-3 and beyond, which are trained with a massive amount of linguistic data. AI=artificial intelligence. GPT=generative pretrained transformer. LLM=large language model. *A highly aligned AI system achieves the full range of intended objectives or desired behaviour (as defined by the programmer), while staying away from undesired behaviour.

Table 1: **Comparisons between deep learning models, GPT, and ChatGPT**

	GPT-4 (2023)[1]	LLaMA (2023)[2]	PaLM (2022)[3]	BLOOM (2022)[4]	Chinchilla (2022)[5]
Knowledge-based common-sense reasoning (based on the ARC challenge; %)[6]	96·3%	57·8%	65·9%	32·9%	NR
Context-based common-sense reasoning (based on WinoGrande; %)[7]	87·5%	77%	85·1%	NR	74·9%
Sentence completion (based on HellaSwag; %)[8]	95·3%	84·2%	83·8%	NR	80·8%
Multitask language understanding (based on MMLU; %)[9]	86·4%	63·4%	69·3%	NR	67·6%
Code generation (based on HumanEval; %)[10]	67·0%	79·3%	NR	55·5%	NR
Reading comprehension and arithmetic (based on DROP; %)[11]	80·9%	NR	70·8%	NR	NR
Grade-school mathematics (based on GSM8K; %)[12]	92%	69·7%	58%	NR	NR

ARC=AI2 reasoning challenge. DROP=discrete reasoning over paragraphs. GPT=generative pretrained transformer. GSM=grade school math. MMLU=massive multitask language understanding. NR=not reported.

Table 2: **Performance of common large language models in reasoning, coding, and mathematics**

(P van Wijngaarden); **Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland** (Prof A Grzybowski PhD); **Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China** (Prof M He PhD); **College of Future Technology, Peking University, Beijing, China** (D Li PhD); **Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China** (A Ran Ran PhD, C Y Cheung PhD); **Department**

questions on topics ranging from common ocular symptoms (eg, blurred vision, ocular pain, and redness) to specific eye diseases (eg, glaucoma, cataract, and diabetic retinopathy). In terms of their implementation, LLMs could be linked to mobile applications via an application programming interface (API) or be linked to eye hospital websites. Through such interfaces, LLMs could aid in remotely requesting specific information and chief complaints from patients, including their ocular symptoms, medical history, and other relevant details, and subsequently coordinate an appointment with the appropriate care provider (figure 2). This triaging is applicable in situations when patients cannot physically go to a hospital or clinic, or when quick advice is needed to establish the need for professional medical attention.

An LLM can adapt its recommendations on the basis of context and new information provided by individual patients. Depending on the severity of a case or the likelihood of sight-threatening complications, LLMs could prioritise patients for same-day, same-week, or later in-person appointments. In the event of urgent cases, immediate tele-consultations could be prompted and facilitated by LLMs. Given that non-emergent conditions account for almost half of all eye-related emergency department visits,[32] interventions to improve the triaging of these patients could allow emergency services to be more dedicated to truly emergent ophthalmic issues. Furthermore, a large proportion of new referral cases to tertiary eye centres are currently attributed to visually inconsequential cataracts, dry eyes, or correctable refractive error—which can generally be managed at primary-care centres without consultations in tertiary settings.[33]

The strength of LLMs resides in their ability to facilitate dynamic, two-way communication, enriched by personalised context, a grasp of common-sense knowledge, and human-like cognitive abilities, such as chain-of-thought reasoning. This adaptability enables LLMs to modify their recommendations on the basis of fresh information provided by patients. In addition, LLM technology can facilitate preliminary decision making for risk stratification. Typically, such stratification is the domain of ophthalmologists or specialised ophthalmic nurses, as opposed to primary-care physicians. Altogether, when deployed appropriately, LLMs could help to improve triaging.

Personalising patient visits

LLMs can also be harnessed to streamline administrative tasks tailored to a patient's initial visit or subsequent

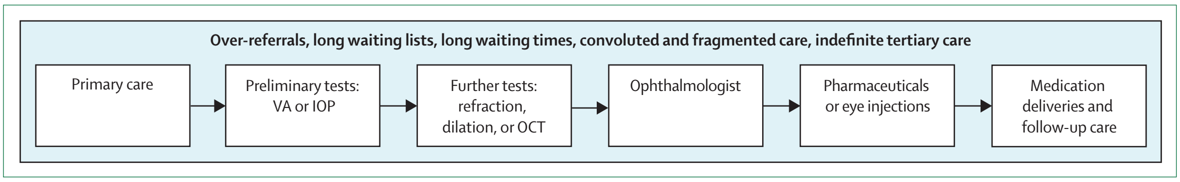

***Figure 1:* Outline of the current clinical workflow in tertiary eye centres**
IOP=intraocular pressure. OCT=optical coherence tomography. VA=visual acuity.

of Ophthalmology, Rajavithi Hospital, Bangkok, Thailand (P Ruamviboonsuk MD)**; NIHR Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital, London, UK** (S Sivaprasad MD)**; Department of Surgery, McMaster University, Hamilton, ON, Canada** (V Chaudhary MD)**; Université Paris Cité, AP-HP, Lariboisière, Saint Louis, and Rothschild Foundation Hospitals, Paris, France** (Prof R Tadayoni PhD)**; Key Laboratory for Biomechanics and Mechanobiology of Ministry of Education, Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing, China** (X Wang PhD)**; State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China** (Y Zheng PhD)**; Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou, China** (Y Zheng)**; Beijing Institute of Ophthalmology, Beijing Ophthalmology and Visual Science Key Lab, Beijing Tongren Hospital, Capital Medical University, Beijing, China** (Y X Wang PhD)

Correspondence to: Dr Yih Chung Tham, Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, 119228, Singapore
thamyc@nus.edu.sg

See **Online** for appendix

For more on **Bard** see https://bard.google.com/

For more on **Bing Chat** see https://www.microsoft.com/en-us/edge/features/bing-chat

hospital visits (figure 2). This streamlining encompasses delineating visit sequences, intelligently scheduling appointments, coordinating medication deliveries, and automating form completion. The transition between nurses, optometrists, ophthalmologists, and different clinic rooms during a visit is frequently confusing and time-consuming. Rather than relying on physical queue slips or multiple in-person interactions with health-care staff for administrative matters, an LLM system can provide patients with sequential, personalised guidance, offering clarifications when needed. When scheduling appointments, instead of calling multiple agents for appointments, patients might only need to interact with the chatbot interface and type in their requests. Essentially, LLM technology has the potential to serve as an efficient, cost-effective virtual assistant, guiding patients seamlessly through their eye care experience.

In the context of cataract surgery, which is the most common ocular surgery, LLMs could promote the transition towards virtual, risk-based preoperative medical evaluations.[34] A pilot study by the Kellogg Eye Center[35] has shown that the use of a preoperative risk assessment questionnaire, administered via a virtual consultation, is associated with safe and efficient outcomes, with fewer case delays, and no statistically significant differences in intraoperative complications or same-day cancellations. Building on this evidence, LLM technology could conceivably assist in administering preoperative questionnaires. This assistance might further streamline the preoperative process for low-risk cataract surgeries, while continuing to offer safe, high-value care.

Enhancing patient engagement in eye care with LLMs

LLM technology offers promising avenues to bolster health literacy and adherence in eye care (figure 2).

For example, chatbot interactions could be used to explain diagnoses or care plans at various difficulty levels, tailored to the individual. A substantial number of patients inadvertently or deliberately deviate from medical advice due to misunderstandings, forgetfulness, or neglect,[36] leading to adverse effects on visual outcomes. For instance, poorly controlled intraocular pressure in glaucoma or uncontrolled blood glucose in diabetic retinopathy can lead to asymptomatic progression of the diseases.[37,38] Poor contact lens hygiene and wearing habits can lead to complications ranging from lens discomfort to infective keratitis.[39] Non-adherence with post-surgical topical corticosteroids leads to excessive postoperative inflammation of the eye.[40] Similarly, patients with meibomian gland dysfunction often neglect regular and proper lid hygiene practice.[41] Although no single approach guarantees universal patient adherence, LLMs can potentially address multiple facets of this challenge. They can potentially automate reminders with enhanced frequency; evaluate and reinforce a patient's grasp of their treatment plan; reiterate the rationale behind prescribed medications or eyedrops; offer a platform for patients to voice concerns between in-person consultations; and provide translations in various languages, while ensuring context and nuance are retained.

Although LLMs cannot replace the expertise and guidance of a trained optometrist or ophthalmologist, their capacity for engaging in intuitive, personalised dialogues with patients is noteworthy (eg, ChatGPT, Bard, and Bing Chat). These capabilities can strengthen the patient–provider relationship between clinical visits, particularly in a time characterised by a surge in ophthalmology patients and health-care staffing constraints.

Optimising eye care delivery by providers

Although LLM technology can be a valuable tool to support patients with eye conditions, there are also conceivable ways it can support care providers—including ophthalmologists, optometrists, ophthalmic nurses, and allied health staff.

Streamlining medical record documentation

A promising avenue for the application of LLMs lies in assisting with the documentation of electronic medical records (EMRs), including crafting discharge summaries, consultation notes, operative dictations, and clinic letters (figure 3). Although the potential of using LLMs to enhance health-care practices has been met with both enthusiasm and caution,[42] the idea of LLM-assisted documentation is gaining traction. LLMs, such as ChatGPT, have been trained on vast human language datasets and acquired a nuanced grasp of medical terminology, including terms specific to ophthalmology. The benefits of this feature could be maximised by combining the speech-to-text capabilities of other AI software with the transcription capabilities of LLMs (eg, OpenAI's Whisper—a speech recognition system that can transcribe and translate audio files in many languages). Given that optometrists and ophthalmologists often allocate a substantial portion of their time to

For more on the **Whisper AI** see https://openai.com/research/whisper examinations and documentation,[43] an LLM-assisted verbal narration system could offer a streamlined approach, allowing for more efficient documentation during patient examinations.

Imagine a scenario in which LLM software functions similarly to modern smart speakers—actively recording patient symptoms, concerns, and specific details from ocular examinations. At the clinician's behest, the LLM could craft notes that align with the eye clinic's documentation protocols, incorporate pertinent billing codes, and even propose subsequent investigations or follow-up appointments. As an explorative exercise, we used ChatGPT-3.5 to emulate clinical communication scenarios characteristic of a tertiary ophthalmology service (appendix pp 2–5). The responses generated by ChatGPT were mostly articulate and encompassing, although based on its performance at the time, the responses would probably still require amendments by an ophthalmologist to further ensure accuracy and patient-specific relevance. It is also worth noting that ChatGPT's recommendations seemed to lean towards a more generalised approach, occasionally missing the subtleties of individualised treatment. Nonetheless, this method holds promise for reducing the time spent on manual documentation. As LLMs evolve and are trained on more comprehensive EMR datasets, we can anticipate even more refined and efficient LLM-assisted documentation processes.

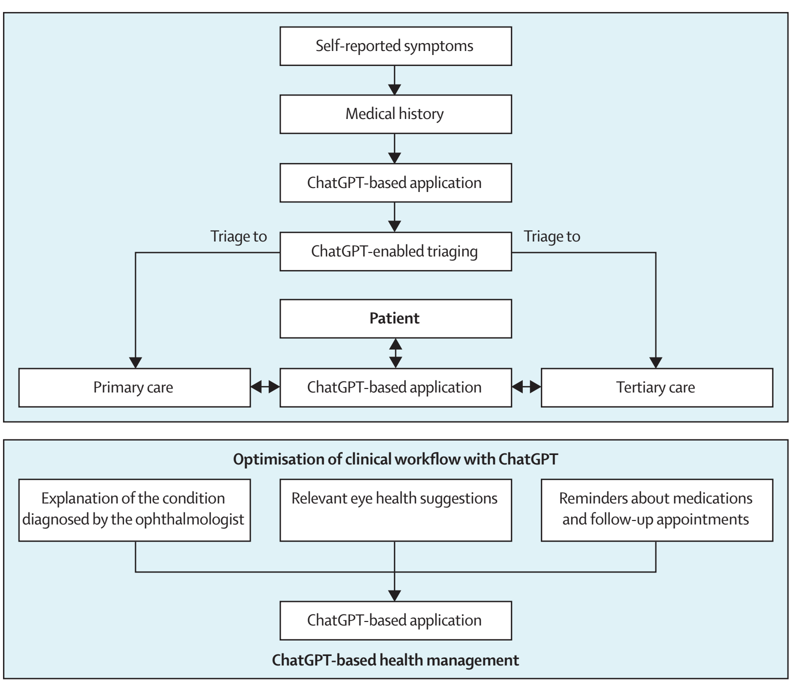

Figure 2: **Use of ChatGPT to enhance delivery of patient-centric health services**

Enhancing ophthalmic education with LLMs

LLMs present a transformative potential for enhancing education for eye care professionals (figure 4). One of their strengths lies in swiftly summarising extensive texts, be these academic papers or comprehensive ophthalmology literature, offering concise overviews. LLMs can distil key points from clinical guidelines or highlight relevant articles on specific subjects. Furthermore, LLMs can be used to simulate real-world interactions, excelling in role-playing and verbal simulations. On this note, they can also function as an immediate feedback tool to refine educators' communicative techniques.

Although LLMs cannot replace hands-on training for clinical procedures, they can serve as invaluable post-training refreshers, summarising crucial steps in patient counselling or procedures themselves. For instance, ophthalmic nurses administering nurse-led intravitreal injections[44,45] might review essential pointers before the procedure, whereas residents could revisit the crucial stages of surgeries. By generating case studies and simulating clinical scenarios, LLMs can help providers improve their decision-making skills and patient interaction acumen.

In clinical settings, LLMs could be integrated into EMR systems to provide a quick summary of and key clinical details for each patient before their consultation. The provider could then use this information to get an immediate overview of the patient's profile, and check for specifics as required. This strategy might be particularly useful in clinics with high patient volumes.

Two studies from 2023[46,47] suggest that LLMs exhibit proficiency in answering ophthalmology-related queries, even in the absence of specialised training in

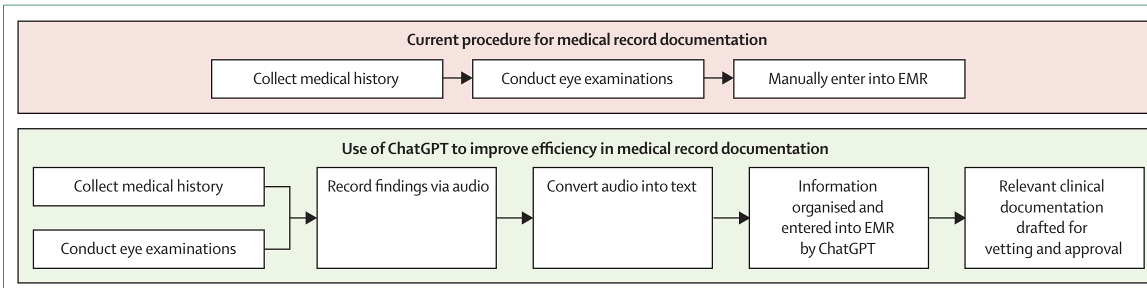

Figure 3: **Use of ChatGPT to enhance efficiency in medical record documentation**
EMR=electronic medical record.

ophthalmology. Likewise, when evaluating ophthalmology board-style questions, Cai and colleagues[47] showed that ChatGPT-4.0 and Bing Chat performed comparably to human respondents. They also observed areas in which LLMs particularly excelled, such as when answering questions about test types and single-step reasoning questions, and more challenging areas for LLMs, such as image interpretation[47]—a domain still in its infancy.

Beyond serving as topic refreshers, LLMs can craft multiple-choice questions for examination topics, assisting in examination preparation. An ophthalmology trainee, for instance, could feed ChatGPT an article from EyeWiki, an eye encyclopaedia website, prompting the LLM to generate test questions for them to review or use for mock self-assessments. In the appendix (p 6), we showcase examples of ChatGPT formulating pertinent questions and providing accurate answers with materials extracted from EyeWiki.

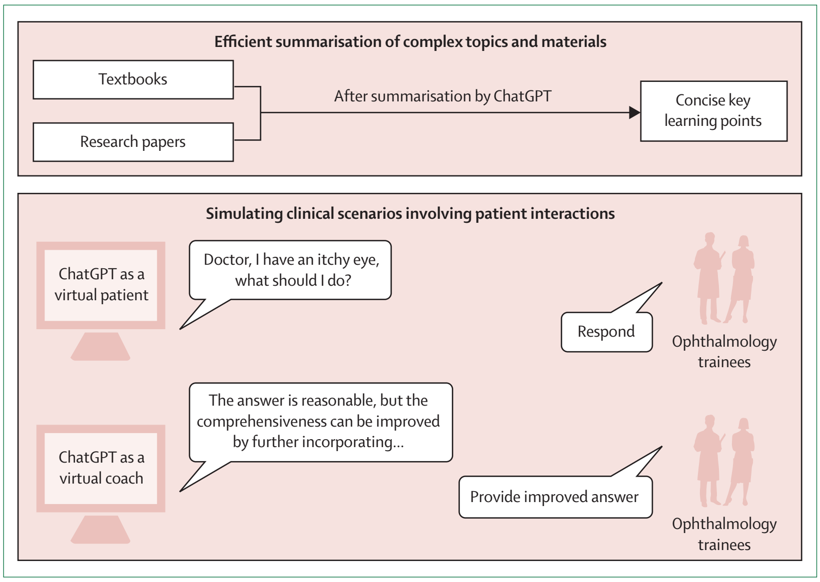

Figure 4: **Use of ChatGPT to facilitate medical education and learning**

Implementation challenges and possible solutions

Despite the wide-ranging benefits LLMs might bring to eye care, concerns have also been raised about their application.[48,49]

Constraints of LLMs in ophthalmological examinations and procedures

Much of ophthalmological practice hinges on meticulous physical examinations of the eyes, an aspect unattainable through mere text-based interactions with an LLM. Key procedures, such as assessing symptoms, gauging visual acuity, observing eye movements, performing tonometry, and conducting fundoscopy, necessitate direct observation and cannot be entirely replicated virtually.

Although LLMs are not intrinsically tailored for clinical procedures, their integration with API plugins and complementary software tools can substantially enhance their applicability. For instance, the seamless fusion of ChatGPT with the Argil plugin facilitates the creation of images derived from textual prompts.[50] By synergising LLMs with ocular photo-based deep learning algorithms, the potential for automated quantification and textual interpretation of ocular imaging data emerges. This collaborative methodology could facilitate the communication of AI-assisted potential diagnoses to ophthalmologists, expanding the scope and promise of LLMs in the realm of ophthalmology.

Privacy and security concerns

A pressing concern about the integration of LLMs into clinical settings revolves around cybersecurity and data privacy, especially when software needs to be trained on EMR data or is directly embedded into a live EMR system.[51] For an LLM to be clinically pertinent, it would inevitably require access to comprehensive patient medical histories, including previous eye conditions, ocular images, examination records, surgical histories, medications, and allergies. This access requirement leads to the pivotal issue of patient consent and acceptance.

Medical data protection in ophthalmology must adhere to various legal and regulatory requirements, such as the Health Insurance Portability and Accountability Act in the USA or the General Data Protection Regulation in the EU. LLMs in ophthalmology clinics would need access to eye data management systems such as the Zeiss Forum or the Heidelberg Eye Explorer HEYEX 2. Ocular images could potentially be considered as patient identifiers, and the security of ophthalmic imaging with a new LLM tool would need to be reviewed before implementation.

Ensuring adherence to regulations, along with the implementation of robust access controls, encryption measures, data backups, regular audits, and timely notifications of data breaches, could prove intricate and demand substantial resources. A collaborative approach, involving health-care providers, technology vendors, regulatory agencies, and researchers is imperative. Such collaboration aims to gain a clear understanding of how LLM algorithms will interact with ocular EMR data and the feasibility of instituting safeguards to uphold individual data privacy, and to ensure alignment with global data protection mandates.

To address the safety apprehensions associated with the integration of LLM systems into EMRs, blockchain technology could be a promising solution. Distinct from traditional centralised databases, blockchain technology operates on a decentralised and distributed ledger framework.[52,53] Every interaction between an EMR and an LLM can be securely chronicled as distinct blocks within this chain. By harnessing cryptographic algorithms and dispersing data across a multitude of nodes, blockchain technology holds the potential to substantially mitigate the risk of data breaches during interactions between EMRs and LLMs. Furthermore, the intrinsic transparency of blockchain technology facilitates the immediate identification and tracing of discrepancies in EMR data, ensuring timely identification and rectification of errors.

False responses by LLMs

Another notable concern about LLMs is the potential for false or misleading responses, colloquially termed hallucinations.[18] Although these models have shown impressive alignment with US Medical Licensing Exam questions,[54] they can produce factual or contextual medical errors, which can be deceptively convincing to patients. For instance, ChatGPT, in its current form, derives its knowledge primarily from the vast expanse of public text on the internet. This means it does not have a comprehensive understanding and can only mimic general cognitive knowledge on the basis of prevalent linguistic patterns. ChatGPT is not specifically tailored for intricate technical or medical tasks.[13] A profound grasp of eye anatomy, physiology, and diseases might elude the current capabilities of LLMs.

To truly excel as an auxiliary tool in eye care, LLMs would need to be trained on specialised ophthalmological research literature and clinical guidelines. However, a substantial portion of this material is not publicly accessible, placing it outside the purview of current LLM training iterations. Hence, the onus would be on clinicians to proofread the outputs, and ensure they are grounded in appropriate facts.

In addition, previous deep learning applications in ophthalmology, such as those used for the interpretation of fundus images, were built on narrowly defined models with clear outcome measures, such as diabetic retinopathy and cataract detection. In contrast, with the rapid evolution of LLMs, assessing their expansive intelligence and setting definitive clinical performance standards presents a more intricate challenge. Thorough and robust evaluation is essential to ascertain the reliability and safety of these tools in clinical settings.

Other capability limitations of LLMs

Another notable limitation of LLMs is their potential inability to keep pace with the latest advancements in the diagnosis and treatment of eye diseases. Given the dynamic nature of medical knowledge, LLMs might lag if they do not incorporate the latest research findings or clinical guidelines. Furthermore, when offering clinical recommendations outside of a clinical setting (where there is no access to patients' EMRs), LLMs might not be able to provide tailored advice, compromising the precision of their suggestions. Lastly, language support remains a concern. LLMs might not cater to all languages, especially in terms of specialised medical vocabulary. This limitation poses potential risks of misinterpretation and raises broader issues about accessibility and health equity.

Ethical considerations

In the past year, several major disciplines and professions—not limited to ophthalmology—have begun to consider what LLMs mean for them. Given how fast LLM technology is moving, there are ethical and legal concerns of liability should medical errors arise from any of the practical drawbacks described above.[55] Until accuracy and safety standards deemed acceptable by the medical community are put in place, any prompts (ie, text entered into ChatGPT) related to the medical use of LLMs should be restricted and ideally should contain explicit warnings. Arguably, many of the bioethics concerns associated with LLMs mirror those prevalent in existing AI applications in medicine—encompassing data ownership, consent, bias, and privacy. However, there are other ethical issues that have been brought to the forefront by LLMs[56]—including misinformation, medical deepfakes, the imperative of informing patients when AI analyses their medical data, and the potential inequities stemming from overly rapid technological advancements.[56,57] Conversely, however, LLMs also present promising avenues for democratising knowledge and empowering patients.

Conclusions

LLMs hold great promise in the field of ophthalmology, offering transformative avenues to enhance clinical workflows and care paradigms. Yet before we can integrate these models into existing health systems, it is imperative that we address pressing concerns about their robustness and reliability. Although we remain optimistic about LLMs, drawing parallels with other digital ophthalmology tools such as telemedicine and ocular photo-based deep learning, we emphasise the crucial need for accuracy assessment, governance, and the establishment of protective measures when integrating them into clinical and care settings.

Contributors

TYW, YXW, and YCT conceived and planned this Viewpoint. BKB, HC, and YCT drafted the manuscript. All authors contributed to the intellectual development of this Viewpoint and provided critical feedback on the manuscript. The final version of this Viewpoint has been seen and approved by all authors.

Declaration of interests

PvW declares financial interests as a cofounder of Enlighten Imaging, an early-stage medical technology start-up company devoted to hyperspectral retinal imaging and image analysis, including the development of artificial intelligence systems; research grant support from Roche and Bayer; and honoraria from Roche, Bayer, Novartis, and Mylan. RT declares consulting fees from Novartis, AbbVie, Allergan, Bayer, Alcon, Roche Genentech, Thea, Apellis, Iveric Bio, and Oculis; honoraria from Optic 2000; support for attending meetings from Novartis, AbbVie, Allergan, and Bayer; participation on advisory boards for Novartis, AbbVie, Allergan, Bayer, Apellis, Iveric Bio, Oculis, and Roche Genentech; a leadership role as the President Elect of Euretina; stock in Oculis; and equipment and materials from Zeiss. SS declares grants from Bayer and Boehringer Ingleheim; consulting fees from Roche, AbbVie, Apellis, Bayer, Biogen, Boehringer Ingelheim, Novartis, Janssen Pharmaceuticals, Optos, Ocular Therapeutix, and OcuTerra; support for attending meetings and travel from Bayer and Roche; participation on data safety monitoring boards or advisory boards for

Novo Nordisk and Bayer; leadership in RCOphth, Macular Society, and Editor of EYE; stock in Eyebiotech; and equipment and materials from Boehringer Ingleheim. AYL reports grants to their affiliated institutions from Santen, Novartis, Carl Zeiss Meditec, Microsoft, NVIDIA, the National Institutes of Health (NIH; grant numbers NIH/NEI K23EY029246 and NIH OT2OD032644), Research to Prevent Blindness, and the Latham Vision Innovation Award. AYL reports consulting fees from USDA, Genentech, Verana Health, Johnson & Johnson, Gyroscope, Janssen Research and Development, and Regeneron; payment for lectures from Topcon; and payment from Alcon Research for various educational events and activities. ARR declares a grant from the Health and Medical Research Fund (grant number 19201991). CSL reports grants to their institution from the NIH (grant numbers NIH/NIA R01AG060942, NIH/NIA U19AG066567, and NIH OT2OD032644), Research to Prevent Blindness, and the Latham Vision Innovation Award; and consulting fees from Boehringer Ingelheim. TYW declares consulting fees from Aldropika Therapeutics, Bayer, Boehringer Ingelheim, Genentech, Iveric Bio, Novartis, Plano, Oxurion, Roche, Sanofi, and Shanghai Henlius; funding from the National Key R&D Program, China (grant number 2022YFC2502800); and being an inventor, patent holder, and cofounder of the start-up companies EyRiS and Visre. All other authors declare no competing interests.

Acknowledgments

YCT is supported by a grant from the National Medical Research Council of Singapore (grant number NMRC/MOH/HCSAINV21nov-0001). TYW receives funding from the National Key R&D Program, China (grant number 2022YFC2502800). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of this Viewpoint.

References

1 Open AI. GPT-4 technical report. March, 2023. https://ui.adsabs.harvard.edu/abs/2023arXiv230308774O (accessed May 2, 2023).
2 Touvron H, Lavril T, Izacard G, et al. Llama: open and efficient foundation language models. *arXiv* 2023; published online Feb 27. https://doi.org/10.48550/arXiv.2302.13971 (preprint).
3 Chowdhery A, Narang S, Devlin J, et al. Palm: scaling language modeling with pathways. *arXiv* 2022; published online April 5. https://doi.org/10.48550/arXiv.2204.02311 (preprint).
4 Scao TL, Fan A, Akiki C, et al. Bloom: a 176b-parameter open-access multilingual language model. *arXiv* 2022; published online Nov 9. https://doi.org/10.48550/arXiv.2211.05100 (preprint).
5 Hoffmann J, Borgeaud S, Mensch A, et al. Training compute-optimal large language models. *arXiv* 2022; published online March 29. https://doi.org/10.48550/arXiv.2203.15556 (preprint).
6 Clark P, Cowhey I, Etzioni O, et al. Think you have solved question answering? Try arc, the ai2 reasoning challenge. *arXiv* 2018; published online March 14. https://doi.org/10.48550/arXiv.1803.05457 (preprint).
7 Sakaguchi K, Bras RL, Bhagavatula C, Choi Y. Winogrande: an adversarial winograd schema challenge at scale. *Commun ACM* 2021; **64:** 99–106.
8 Zellers R, Holtzman A, Bisk Y, Farhadi A, Choi Y. Hellaswag: Can a machine really finish your sentence? *arXiv* 2019; published online May 19. https://doi.org/10.48550/arXiv.1905.07830 (preprint).
9 Hendrycks D, Burns C, Basart S, et al. Measuring massive multitask language understanding. *arXiv* 2020; published online Sept 7. https://doi.org/10.48550/arXiv.2009.03300 (preprint).
10 Chen M, Tworek J, Jun H, et al. Evaluating large language models trained on code. *arXiv* 2021; published online July 7. https://doi.org/10.48550/arXiv.2107.03374 (preprint).
11 Dua D, Wang Y, Dasigi P, Stanovsky G, Singh S, Gardner M. DROP: a reading comprehension benchmark requiring discrete reasoning over paragraphs. *arXiv* 2019; published online March 1. https://doi.org/10.48550/arXiv.1903.00161 (preprint).
12 Cobbe K, Kosaraju V, Bavarian M, et al. Training verifiers to solve math word problems. *arXiv* 2021; published online Ot 27. https://doi.org/10.48550/arXiv.2110.14168 (preprint).
13 Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. *arXiv* 2023; published online April 12. https://doi.org/10.48550/arXiv.2303.13375 (preprint).
14 Yasunaga M, Leskovec J, Liang P. Linkbert: pretraining language models with document links. *arXiv* 2022; published online March 29. https://doi.org/10.18653/v1/2022.acl-long.551 (preprint).
15 Yasunaga M, Bosselut A, Ren H, et al. Deep bidirectional language-knowledge graph pretraining. *Adv Neural Inf Process Syst* 2022; **35:** 37309–23.
16 Taylor R, Kardas M, Cucurull G, et al. Galactica: a large language model for science. *arXiv* 2022; published online Nov 16. https://doi.org/10.48550/arXiv.2211.09085 (preprint).
17 Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *arXiv* 2022; published online Dec 26. https://doi.org/10.48550/arXiv.2212.13138 (preprint).
18 Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med* 2023; **388:** 1233–39.
19 Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res* 2019; **72:** 100759.
20 Gunasekeran DV, Tham YC, Ting DSW, Tan GSW, Wong TY. Digital health during COVID-19: lessons from operationalising new models of care in ophthalmology. *Lancet Digit Health* 2021; **3:** e124–34.
21 Tham YC, Husain R, Teo KYC, et al. New digital models of care in ophthalmology, during and beyond the COVID-19 pandemic. *Br J Ophthalmol* 2022; **106:** 452–57.
22 Li JO, Liu H, Ting DSJ, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. *Prog Retin Eye Res* 2021; **82:** 100900.
23 Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; **316:** 2402–10.
24 Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017; **318:** 2211–23.
25 Hood DC, De Moraes CG. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 2018; **125:** 1207–08.
26 Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol* 2017; **135:** 1170–76.
27 Zhang Z, Wang Y, Zhang H, et al. Artificial intelligence-assisted diagnosis of ocular surface diseases. *Front Cell Dev Biol* 2023; **11:** 1133680.
28 Fang X, Deshmukh M, Chee ML, et al. Deep learning algorithms for automatic detection of pterygium using anterior segment photographs from slit-lamp and hand-held cameras. *Br J Ophthalmol* 2022; **106:** 1642–47.
29 Tham YC, Anees A, Zhang L, et al. Referral for disease-related visual impairment using retinal photograph-based deep learning: a proof-of-concept, model development study. *Lancet Digit Health* 2021; **3:** e29–40.
30 Khou V, Ly A, Moore L, et al. Review of referrals reveal the impact of referral content on the triage and management of ophthalmology wait lists. *BMJ Open* 2021; **11:** e047246.
31 Hribar MR, Huang AE, Goldstein IH, et al. Data-driven scheduling for improving patient efficiency in ophthalmology clinics. *Ophthalmology* 2019; **126:** 347–54.
32 Channa R, Zafar SN, Canner JK, Haring RS, Schneider EB, Friedman DS. Epidemiology of eye-related emergency department visits. *JAMA Ophthalmol* 2016; **134:** 312–19.
33 Shah R, Edgar DF, Khatoon A, et al. Referrals from community optometrists to the hospital eye service in Scotland and England. *Eye (Lond)* 2022; **36:** 1754–60.
34 Chen CL, McLeod SD, Lietman TM, et al. Preoperative medical testing and falls in medicare beneficiaries awaiting cataract surgery. *Ophthalmology* 2021; **128:** 208–15.
35 Cuttitta A, Joseph SS, Henderson J, et al. Feasibility of a risk-based approach to cataract surgery preoperative medical evaluation. *JAMA Ophthalmol* 2021; **139:** 1309–12.
36 Martin LR, Williams SL, Haskard KB, Dimatteo MR. The challenge of patient adherence. *Ther Clin Risk Manag* 2005; **1:** 189–99.

37 Keenum Z, McGwin G Jr, Witherspoon CD, Haller JA, Clark ME, Owsley C. Patients' adherence to recommended follow-up eye care after diabetic retinopathy screening in a publicly funded county clinic and factors associated with follow-up eye care use. *JAMA Ophthalmol* 2016; **134:** 1221–28.

38 Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA* 2014; **311:** 1901–11.

39 Yee A, Walsh K, Schulze M, Jones L. The impact of patient behaviour and care system compliance on reusable soft contact lens complications. *Cont Lens Anterior Eye* 2021; **44:** 101432.

40 Lu AQ, Rizk M, O'Rourke T, et al. Safety and efficacy of topical vs intracanalicular corticosteroids for the prevention of postoperative inflammation after cataract surgery. *J Cataract Refract Surg* 2022; **48:** 1242–47.

41 Chuckpaiwong V, Nonpassopon M, Lekhanont K, Udomwong W, Phimpho P, Cheewaruangroj N. Compliance with lid hygiene in patients with meibomian gland dysfunction. *Clin Ophthalmol* 2022; **16:** 1173–82.

42 Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare* 2023; **11:** 887.

43 Baxter SL, Gali HE, Huang AE, et al. Time requirements of paper-based clinical workflows and after-hours documentation in a multispecialty academic ophthalmology practice. *Am J Ophthalmol* 2019; **206:** 161–67.

44 Raman V, Triggol A, Cudrnak T, Konstantinos P. Safety of nurse-led intravitreal injection of dexamethasone (Ozurdex) implant service. Audit of first 1000 cases. *Eye* 2021; **35:** 388–92.

45 Teo AWJ, Rim TH, Wong CW, et al. Design, implementation, and evaluation of a nurse-led intravitreal injection programme for retinal diseases in Singapore. *Eye* 2020; **34:** 2123–30.

46 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.

47 Cai LZ, Shaheen A, Jin A, et al. Performance of generative large language models on ophthalmology board-style questions. *Am J Ophthalmol* 2023; **254:** 141–49.

48 Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine* 2023; **90:** 90.

49 Nature Medicine. Will ChatGPT transform healthcare? *Nat Med* 2023; **29:** 505–06.

50 Argil. How to generate ChatGPT images using Argil plugin. 2023. https://www.argil.ai/blog/how-to-generate-chatgpt-images-using-argil-plugin (accessed May 2, 2023).

51 Hasal M, Nowaková J, Ahmed Saghair K, Abdulla H, Snášel V, Ogiela L. Chatbots: security, privacy, data protection, and social aspects. *Concurr Comput* 2021; **33:** e6426.

52 Witte JH. The Blockchain: a gentle four page introduction. *arXiv* 2016; published online Dec 6. https://doi.org/10.48550/arXiv.1612.06244 (preprint).

53 Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Blockchain vehicles for efficient medical record management. *NPJ Digit Med* 2020; **3:** 1.

54 Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digit Health* 2023; **2:** e0000198.

55 Parviainen J, Rantala J. Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. *Med Health Care Philos* 2022; **25:** 61–71.

56 Cohen IG. What should ChatGPT mean for bioethics? *Am J Bioeth* 2023; **13:** 1–9.

57 Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. *Lancet Digit Health* 2023; **5:** e333–35.