# Foundation models in ophthalmology: opportunities and challenges

*Mertcan Sevgi*[a,b,e], *Eden Ruffell*[a,c,d,e], *Fares Antaki*[a,b,f], *Mark A. Chia*[a,b,e] *and Pearse A. Keane*[a,b,e]

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024

**Purpose of review**
Last year marked the development of the first foundation model in ophthalmology, RETFound, setting the stage for generalizable medical artificial intelligence (GMAI) that can adapt to novel tasks. Additionally, rapid advancements in large language model (LLM) technology, including models such as GPT-4 and Gemini, have been tailored for medical specialization and evaluated on clinical scenarios with promising results. This review explores the opportunities and challenges for further advancements in these technologies.

**Recent findings**
RETFound outperforms traditional deep learning models in specific tasks, even when only fine-tuned on small datasets. Additionally, LMMs like Med-Gemini and Medprompt GPT-4 perform better than out-of-the-box models for ophthalmology tasks. However, there is still a significant deficiency in ophthalmology-specific multimodal models. This gap is primarily due to the substantial computational resources required to train these models and the limitations of high-quality ophthalmology datasets.

**Summary**
Overall, foundation models in ophthalmology present promising opportunities but face challenges, particularly the need for high-quality, standardized datasets for training and specialization. Although development has primarily focused on large language and vision models, the greatest opportunities lie in advancing large multimodal models, which can more closely mimic the capabilities of clinicians.

**Keywords**
foundation models, large language models, large vision models, medical artificial intelligence, multimodal models

## INTRODUCTION

The advancement of artificial intelligence in ophthalmology has been significantly propelled by deep learning techniques, which analyse eye images to detect patterns and anomalies, thereby significantly enhancing diagnostic and prognostic capabilities [1]. Despite their effectiveness, a main constraint for such artificial intelligence models is that they typically require vast amounts of labelled data for model training, often not obtainable.

More recently, the widespread use of foundation models in healthcare has marked a significant step towards general medical artificial intelligence (GMAI). Coined by researchers at Stanford, 'foundation model' refers to deep learning models that provide a versatile base of knowledge adaptable for various specific purposes, unlike traditional artificial intelligence systems designed for specific tasks (Fig. 1) [2,3]. Enabled by innovative approaches such as self-supervised learning (SSL) and vision transformers (ViTs), foundation models are trained on large, diverse, unlabelled datasets and fine-tuned with smaller labelled datasets for specific medical use cases. This approach allows foundation models to achieve the same accuracy as traditional models but with fewer labelled examples [4].

In this review, we examine the current status and opportunities of foundation models in

[a]Institute of Ophthalmology, University College London, [b]Moorfields Eye Hospital NHS Foundation Trust, [c]Institute of Health Informatics, [d]Centre for Medical Image Computing, University College London, [e]NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust, London, UK and [f]The CHUM School of Artificial Intelligence in Healthcare, Montreal, Quebec, Canada

Correspondence to Pearse A. Keane, MD, Institute of Ophthalmology, University College London, London, UK. E-mail: p.keane@ucl.ac.uk

**Curr Opin Ophthalmol** 2024, 35:000–000
DOI:10.1097/ICU.0000000000001091

1040-8738 Copyright © 2024 Wolters Kluwer Health, Inc. All rights reserved.



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

## KEY POINTS

- Large language models are useful in education and research, but for clinical use, they need better domain-specific knowledge, fewer confabulations, and rigorous evaluation.
- Large vision models can revolutionize global ophthalmology with their advanced diagnostic and prognostic capabilities, addressing health data inequality without the need for retraining.
- Open-source foundation models like RETFound enable researchers to utilize small labelled datasets to discover novel applications, driving innovation in ophthalmology.
- Federated learning has the potential to support collaborative training of foundation models without exchanging sensitive data, enhancing model robustness and generalizability whilst sharing computational resources.
- Large multimodal models are akin to clinicians in how they integrate multimodal data to achieve the most accurate clinical understanding. Although they hold great potential for utility, they currently lack the necessary datasets for full implementation.

ophthalmology, including large language models (LLMs), large vision models (LVMs) and large multimodal models (LMMs). Notably, we discuss RETFound, the first LVM for ophthalmology and its current use cases and future opportunities. We also explore the key next challenges to advance foundation models in ophthalmology.

## LARGE LANGUAGE MODELS

Among the most well known types of foundation models are LLMs, known for their capacity to understand and generate human-like text [5,6]. Training these LLMs involves processing extensive text data from diverse sources, often encompassing billions of words, which enables them to grasp nuanced and complex language patterns [5].

To enhance these models, newer iterations require larger and more up-to-date text corpora for training, in addition to vast amounts of computational power [7]. Within ophthalmology, we often lack the resources and data to train a domain-specific LLM from scratch. Instead, it is more beneficial to specialize existing LLMs using methods such as fine-tuning, prompt engineering, and retrieval-augmented generation (RAG) as seen in Fig. 2 [8–11].

Fine-tuning involves updating the internal parameters of the LLM using small labelled datasets, to equip the model with specific domain knowledge. Within the context of ophthalmology, an LLM can be fine-tuned using validated question and answer (QnA) pairs. This process can involve providing accurate responses to clinical management questions, such as the treatment protocols, for example, acute angle-closure glaucoma [12]. When updating

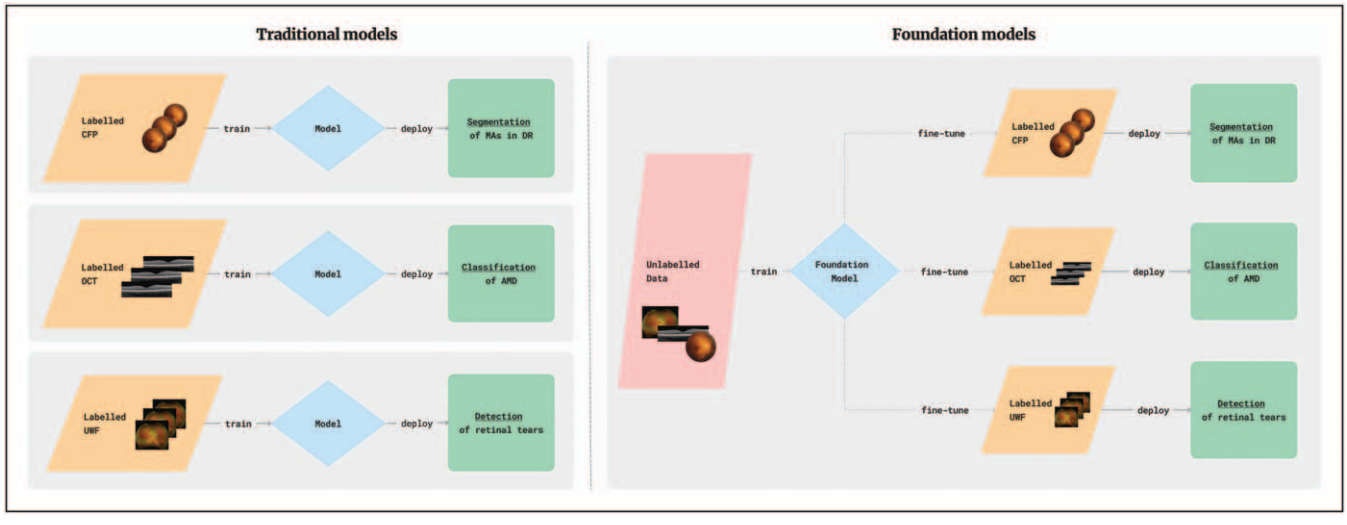



**FIGURE 1.** Schematic representation of training traditional deep learning models and foundation models. The differences between training traditional deep learning (DL) models and foundation models (FM) are highlighted. Traditional DL models typically require labelled datasets and are trained for specific tasks. In contrast, foundation models are usually trained once on unlabelled data and subsequently fine-tuned for a variety of tasks and modalities, such as segmentation, classification, and object detection. CFP, colour fundus photo; DN, diabetic retinopathy; MA, microaneurysm; OCT, optical coherence tomography; UWF, ultra-wide field. Adapted from [4], licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) Licence.



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024

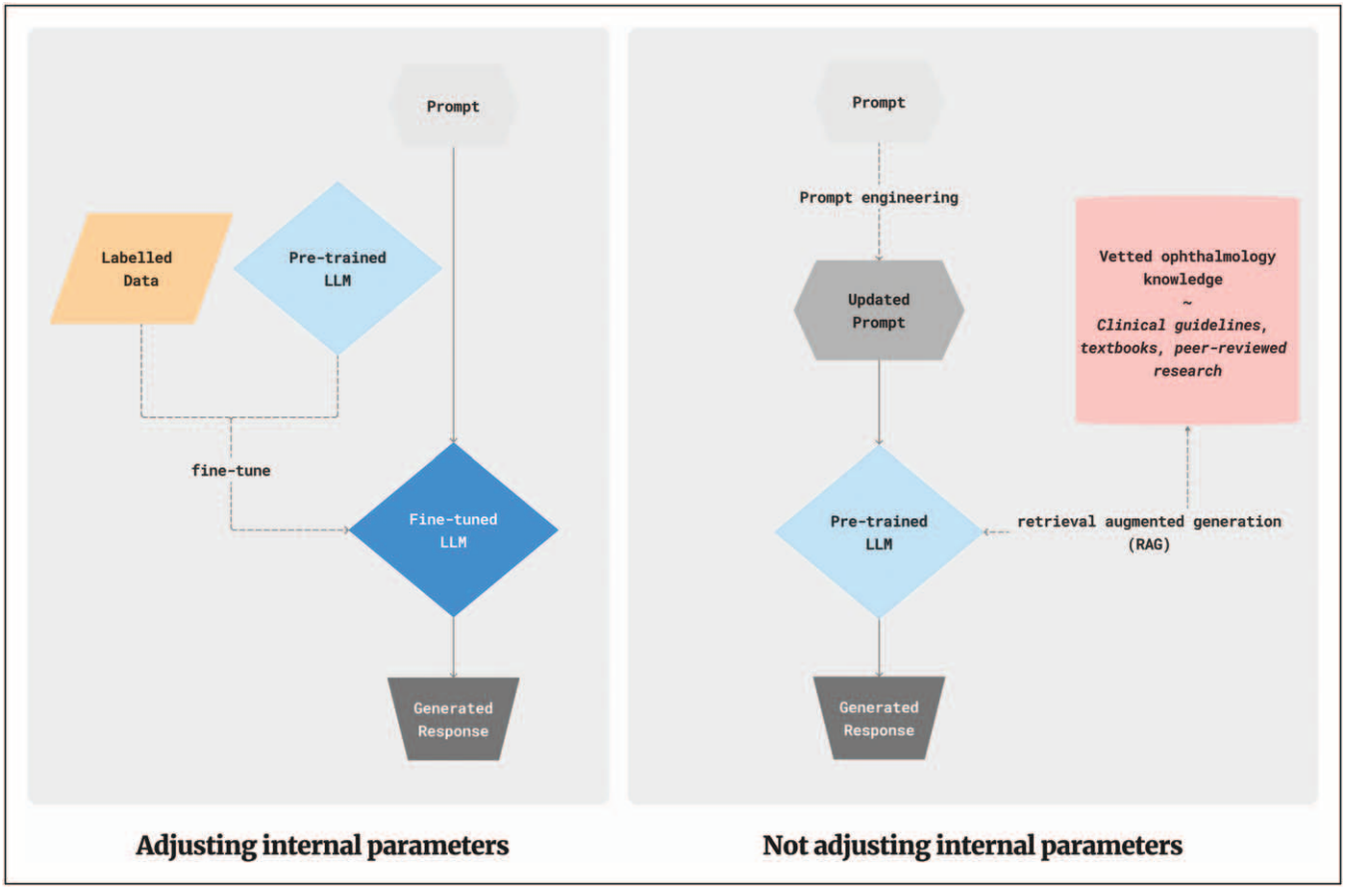



**FIGURE 2.** Methods of specializing large language models. The various techniques used to tailor LLMs for specific applications, including fine-tuning, prompt engineering, and retrieval-augmented generation (RAG) are illustrated. Fine-tuning involves adjusting the internal model parameters to improve performance on a specific task, while prompt engineering and RAG do not alter the model parameters but instead enhance the model's output through different approaches.

the internal parameters, fine-tuning can be approached in two main ways, with a frozen or unfrozen backbone. Fine-tuning with a frozen backbone means that the core parameters of the model remain unchanged or 'frozen', and only the top parameters, known as the head, are updated [13]. This approach is computationally efficient and reduces the risk of overfitting, which is when an artificial intelligence model replicates the training data without learning the underlying patterns. However, with a frozen backbone, the ability to adapt to highly specialized tasks is limited because the core parameters cannot be modified. Conversely, fine-tuning with an unfrozen backbone involves updating the entire model, including the core parameters. This allows for greater adaptability and often results in higher accuracy on domain-specific tasks, as the model can fine-tune all its parameters end-to-end. The trade-off is that this approach requires more computational resources, takes longer to train, and has a higher risk of overfitting because of the extensive parameter updates [13]. Both of these approaches typically require technical coding skills, however, recent advancements on platforms such as Google's Vertex AI allow for fine-tuning LLMs without the need for coding knowledge [10].

Other methods of specialization focus on optimizing the outputs of the LLM rather than altering its parameters. These approaches enable the model to act as an agent, capable of retrieving domain-specific information [14]. Given the propensity for LLMs to confabulate information or provide outdated knowledge with opaque reasoning processes, RAG has emerged as a promising solution that allows LLMs to pull from external databases when generating a response [11]. Such external databases could include peer-reviewed articles, clinical trials, guidelines, and electronic health records (EHR). This approach enhances the accuracy and credibility of LLM outputs, especially for knowledge-intensive tasks seen in healthcare, by allowing continuous

1040-8738 Copyright © 2024 Wolters Kluwer Health, Inc. All rights reserved.


Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024

updates and integration of domain-specific information. For example, practical use cases of LLMs with RAG include creating virtual teachers that generate Q&A from clinical guidelines, and summarize research articles [15▪]. Platforms like OpenAI allow users to create their own custom GPT models. These models can use RAG to incorporate additional documents and perform web searches for more accurate and comprehensive responses [16].

Similar to RAG, another approach to LLM specialization is prompt engineering. This technique steers the model's responses by designing and structuring input prompts to guide the model towards producing more accurate and relevant outputs [17]. By providing specific contexts or framing questions in particular ways, prompt engineering helps the model leverage its existing knowledge more effectively, ensuring that generated responses align closely with the desired domain-specific information [14,18]. For prompt engineering, there are several different prompting strategies that can be applied, depending on what is required [17,19]. For example, the simplest approach is zero-shot or direct prompting is when the prompt used to interact with the LLM does not contain examples or demonstrations, for example, 'Give me a list of causes of red eye', Another type of prompt engineering is role prompting, which involves assigning the LLM a specific role, for instance 'You are a clinical triage assistant, please explain to the patient the causes of red eye'. In this way, the prompt provides the LLM with situational context and a better understanding of what style of language it should use for its audience [17]. Through methods like one-shot or few-shot prompting, an LLM is provided with the ideal output response structure, to help guide its responses if they are required to be in a specific format such as a clinical letter or report. Finally, chain-of-thought (CoT) prompting encourages the LLM to explain its reasoning step-by-step, which is particularly important in healthcare to understand the reasoning behind the model's clinical responses [17]. For instance, 'Explain your thought process for diagnosing the cause of red eye in a patient presenting with these symptoms'.

Before specializing a foundation model, such as an LLM, it is common to first use a linear probe to evaluate the quality of the patterns and features learned by the foundation model. A linear probe involves training a simple classifier, such as logistic regression, on top of the pretrained foundation model to perform a specific task. This method helps determine how well the pretrained features can support the task with minimal additional learning. If a linear probe performs well, it suggests that the model's pretrained knowledge is robust and effective for the given task.

These specialized LLMs often outperform traditional machine learning models, even when using very small labelled datasets, making for a more effective use of the limited data available in hospitals [4]. Many popular LLMs such as Gemini and GPT-4, have undergone general medical specialization using these techniques, exemplified by GPT-4 Medprompt and Med-Gemini. These achieve medical specialization through a combination of self-training via web search, fine-tuning with customised encoders and prompt engineering using CoT prompting [18,20▪▪].

Initially, most evaluations of LLMs in ophthalmology have focused on assessing the performance of out-of-the-box models through a variety of specific tasks, without medical specialization [21]. These tasks include answering patient queries, drafting operative notes, triaging clinical cases, analysing imaging reports, and responding to clinical examination questions [22–27,28▪]. More recent research has begun to outline and explore the benefits of specializing LLMs specifically to ophthalmology encompassing use-cases in medical education, workflow improvement and clinical assistance [12,29,30]. LLMs for medical education could summarize clinical guidelines and academic articles, as well as acting as a virtual patient to which questions can be asked to aid learning [15▪,16]. Meanwhile, report generation and streamlining EHR documentation are ways in which LLMs could improve workflow management [29,31].

For clinicians, LLMs could also aid in summarizing clinical content, assist with making referrals and provide suggestions based on the medical history. As reliance on tertiary-level eye care services increases, these types of LLMs are likely to be increasingly useful [32]. For example, they may be integrated into face-to-face clinical encounters or into virtual clinic services, complementing tele-ophthalmology efforts with the potential to transform patient care in underserved areas [33]. In low-income and middle-income countries (LMICs), there is a significant shortage of clinicians, which exacerbates the challenges of providing adequate eye care [34]. Therefore, implementing LLMs in these regions could yield substantial benefits by augmenting the limited human resources, improving access to eye care, and enhancing the overall quality of patient care [29,35]. Depending on their use cases, such as a patient-facing application for diagnosing diseases, LLMs could be considered as medical devices. In these cases, the deployment of such LLMs would require regulatory approval, necessitating rigorous



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

testing and validation to ensure their safety and efficacy in clinical settings [36].

## LARGE VISION MODELS

In addition to language processing, foundation models can be trained to have vision capabilities, which can be useful for analysing and interpreting medical images. These models, known as LVMs, are especially valuable in ophthalmology, a field heavily reliant on imaging. However, while imaging data is routinely collected in ophthalmology, it is often unlabelled as it is not typically accompanied by formal reports, as seen in radiology. By employing techniques such as SSL, we can harness this extensive amount of unlabelled image data to train LVMs effectively [37▪▪].

Developed by researchers at Moorfields Eye Hospital and UCL, RETFound serves as the first LVM for ophthalmology. It was trained separately on colour fundus photography (CFP) and optical coherence tomography (OCT), resulting in two versions of the model. When fine-tuned for specific downstream tasks, using small labelled datasets, RETFound was employed for diagnosis and prognosis of ocular diseases and the prediction of systemic diseases, where it outperformed other supervised and self-supervised models. As an open-source model, RETFound is accessible for other groups to explore and fine-tune for new tasks, such as image segmentation [38].

One potential research avenue in ophthalmology is the development of LVMs trained on 3D images, a technique being explored in other medical specialties [39▪▪]. At present, LVMs like RETFound tend to only utilize a select number of 2D B-scans from an OCT volume during training, such as the central B-scan [40▪▪]. Therefore, creating a model that utilizes the entire OCT volume has the potential to provide greater spatial information and thus more accurate diagnosis. However, within ophthalmology, there is a notable lack of pretrained 3D models to test this hypothesis [41].

Similarly, research is being conducted to better understand the benefits of training LVMs on multiple imaging modalities, with VisionFM and EyeFound as being notable examples [30,42]. Both models were trained on large datasets, including modalities such as OCT, CFP, fundus fluorescein angiography, in addition to B-ultrasound, external eye photos and slit lamp photos. Incorporating these modalities is valuable as it covers a broad spectrum of ophthalmic conditions across multiple sub-specialties. However, whilst these models report high accuracy in diagnosing ocular diseases showing promising results, it is important to note that these models are currently detailed in preprint, and further development and rigorous validation are necessary to fully confirm their capabilities.

## LARGE MULTIMODAL MODELS

Despite the promise of LLMs and LVMs, to truly assist clinicians, it is essential to develop models that can integrate various types of data, such as text and images. Developing such LMMs marks a step towards GMAI, mirroring how clinicians combine multiple data modalities for a comprehensive understanding. Consequently, more recent foundation model research has emphasized developing LMMs with multimodal capabilities.

One such model is Google's Med-Gemini, an LMM adapted from Gemini on diverse medical text and imaging data from various medical specialties, including CFP images from a EyePACS diabetic retinopathy dataset. Using these capabilities, Med-Gemini was developed to support a range of different clinical tasks including disease classification, image report generation, and polygenic risk prediction [20▪▪,39▪▪]. Med-Gemini also performs visual question answering (VQA), where the model responds to visual prompts such as fundus photos and engages in dialogue with the user regarding the management of conditions. For instance, in ophthalmology, the LMM can diagnose a condition based on the medical history provided through text and clinical images as demonstrated in Fig. 3 [43].

During evaluation on ophthalmology tasks, Med-Gemini was shown to outperform the out-of-the-box Gemini Ultra in several experiments, but for the task of diabetic retinopathy lesion detection, it did not surpass the accuracy of the supervised learning model. Similar projects have looked to evaluate out-of-the-box LMMs like Google's Gemini Pro, on tasks such as on interpreting OCT B-scans, however, these models have similarly failed to reach high accuracy compared with supervised models [44].

Another LMM with ophthalmology capabilities is OphGLM, fine-tuned from ChatGLM using fundus image–text pairs generated from medical–patient conversations [45]. This model was utilized for a variety of downstream tasks, including generating image descriptions, explaining causes and symptoms, and guiding diagnosis and examination procedures for multiple ocular diseases. Despite showing promise, no formal metrics for the model's full pipeline performance were provided; only examples of output dialogues were shared. Therefore, further clinical evaluation is required to determine its potential value within ophthalmology.

Moreover, it is notable that neither Med-Gemini nor OphGLM were initially trained on ophthalmology

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

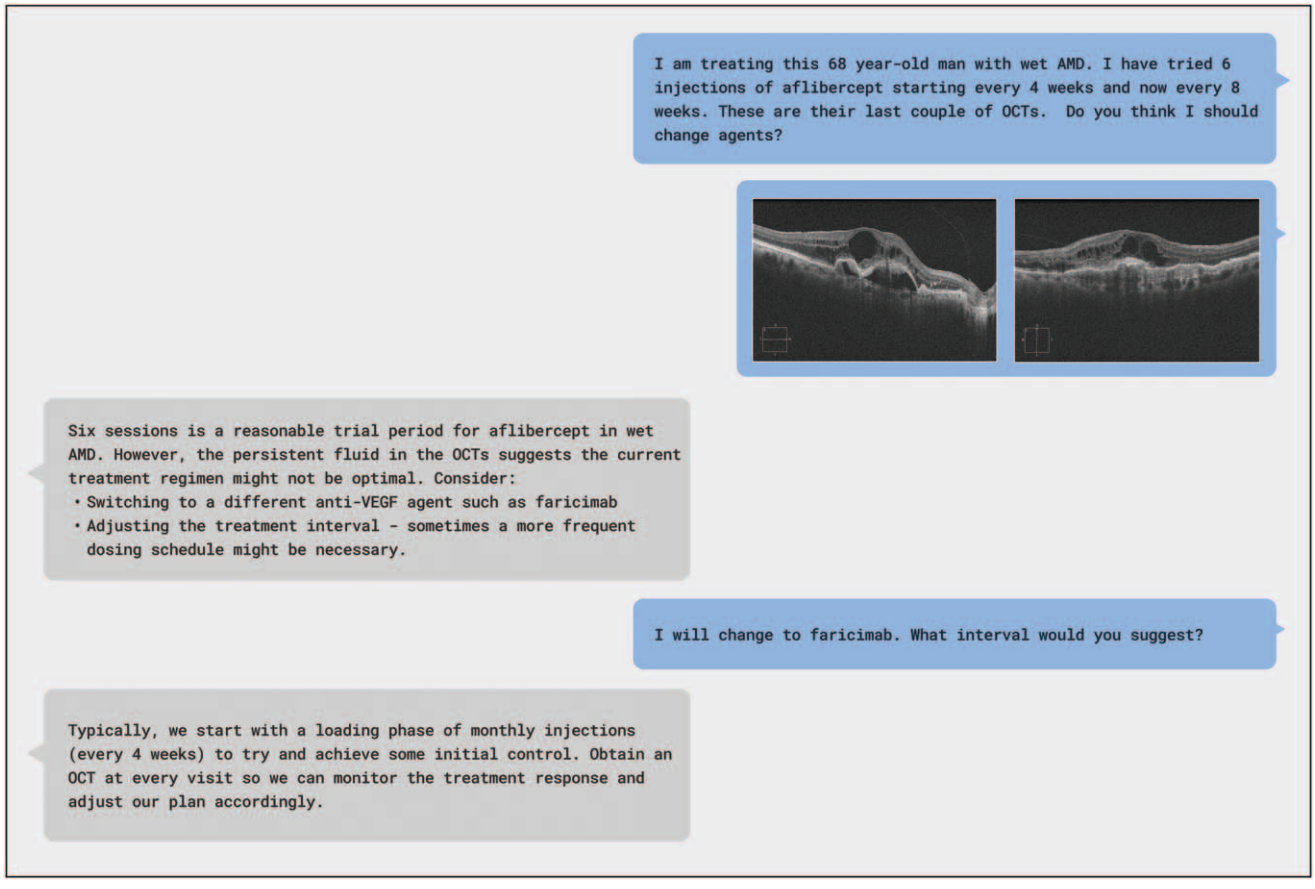



**FIGURE 3.** Visual question and answer example scenario involving an ophthalmologist using a large multimodal language model for treating wet age-related macular degeneration. The LMM interprets OCT (optical coherence tomography) images of a patient with wet age-related macular degeneration, offering guidance on treatment adjustment. The model also responds to follow-up questions. Images are from [43] licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) Licence. LMM, large multimodal models.

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024

data, and only later fine-tuned. Consequently, at their core, these foundation models are not specifically designed as ophthalmology-focused LMMs. We, therefore, speculate that better results could be achieved if LMMs were trained on, at least partially, ophthalmology-specific image–text pairs, during both initial training and fine-tuning. This approach, however, would require large, high-quality, ophthalmology-specific datasets, which are currently lacking in the field, as noted by the authors from the OphGLM paper [45].

## CHALLENGES

Despite these promising steps, the development of foundation models in ophthalmology are confronted with significant constraints. One of the main challenges is the availability of vast ophthalmology-specific datasets needed for foundation model training, as collecting and maintaining high-quality data is costly and time-consuming [21]. This often leads researchers to rely on existing foundation models, such as Gemini, which are trained on general information but can be fine-tuned with ophthalmology-specific data [39▪▪]. Models like RETFound, however, incorporate ophthalmic data during both the training and fine-tuning phases with promising results, which demonstrates the substantial advantages of having ophthalmology-specific data integrated throughout the training process [40▪▪].

Another significant challenge in developing foundation models for ophthalmology is the immense computational resources required. Building foundation models from scratch requires expensive GPUs, and with each new version, the costs rise because of the increased computational demands [46]. This often makes foundation model training feasible only for large institutions and companies that can afford the necessary infrastructure and



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024

resources, again leading researchers to rely on existing pretrained general purpose foundation models, like Gemini and GPT-4.

Given these challenges, it is important to promote the development and use of open-source ophthalmology foundation models, such as RETFound, which can be fine-tuned by other researchers and hospitals using their own datasets. This approach not only democratizes access to advanced artificial intelligence tools but also allows researchers to leverage their limited amounts of labelled data to tailor models to local populations, enhancing their effectiveness and relevance in diverse clinical settings [47].

Another promising solution to overcome these challenges is federated learning, which enables collaborators to share data and resources, thereby mitigating high costs and computational demands [48]. By distributing computational tasks among research collaborators, federated learning helps to alleviate major computational constraints. Additionally, federated learning supports the use of datasets from hospitals to train models without directly sharing data, thereby expanding the pool of available training data without encountering information governance issues. Although this technique has been successfully used to train deep learning models in ophthalmology, using data from multiple centres, there is currently a lack of research around the plausibility of training foundation models using federated learning. Therefore, the next challenge could be to leverage federated learning to build ophthalmology foundation models [49▪]. We can also consider using platforms such as FedEYE and Bitfount to provide user-friendly federated learning environments, allowing centres without technical expertise to contribute their data for model training [50,51].

A final hurdle to developing foundation models for ophthalmology is that of regulation. However, recommendations for artificial intelligence as a medical device (AIaMD) are less clear compared with medical devices and drugs [52]. AIaMD is already used in ophthalmology, with research underway to explore its applications in ophthalmic imaging [52]. As we move forward, we need to consider different regulatory requirements for foundation models, especially for integrating language models and vision models [53,54]. Copyright infringement is currently a heated topic in the media [55]. Ensuring the provenance of the data used to train clinical LLMs and establishing clear regulations will be crucial [56].

## CONCLUSION

In summary, foundation models in ophthalmology present an array of promising opportunities and notable challenges. The advancements in LLMs are particularly exciting because of their ability to mimic human-like conversation and provide valuable insights despite current constraints in specific domains such as ophthalmology. The ongoing efforts to improve these models by training on broader medical data and employing fine-tuning methods are encouraging, as they aim to reduce confabulations and enhance model reliability.

Similarly, the potential of vision models in ophthalmology is immense. These models offer advanced diagnostic and prognostic capabilities, and their adaptability to various tasks without the need for retraining is poised to revolutionize global ophthalmology. This is especially crucial in addressing health data inequality. The advent of open-source models like RETFound allows researchers worldwide to experiment and discover novel applications, further driving innovation in the field. Moreover, federated learning offers a promising solution for researchers to train foundation model models collaboratively without exchanging sensitive data, helping to build larger, more diverse foundation models. This collaborative approach not only expands the pool of training data but also enhances the robustness and generalizability of the models, while reducing the computational burden on a single institution.

Lastly, LMMs hold significant promise by integrating various modalities in a manner akin to a clinician's approach, moving closer to achieving GMAI. This capability could fundamentally transform healthcare delivery and the interaction between patients, healthcare professionals, and artificial intelligence systems. However, to make this progress, there is a pressing need for high-quality, standardized datasets specifically for model training and medical specialization. As these technologies continue to evolve, they are likely to play a pivotal role in shaping the future of ophthalmology and healthcare at large, offering more precise, accessible, and efficient care for patients globally.

### Acknowledgements

*None.*

### Financial support and sponsorship

*None.*

*Funding: E.R. is supported by the UCL UKRI Centre for Doctoral Training in AI-enabled healthcare systems Studentship (EP/S021612/1) F.A. is supported by the Fonds de recherche du Québec – Santé (FRQS). M.C. is supported by a General Sir John Monash Scholarship. P. K. is supported by a UK Research & Innovation Future Leaders Fellowship (MR/T019050/1).*



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

## Conflicts of interest

*There are no conflicts of interest.*

## REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest


1. De Fauw J, Ledsam JR, Romera-Paredes B, *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018; 24:1342–1350.
2. Chia MA, Antaki F, Zhou Y, *et al.* Foundation models in ophthalmology. Br J Ophthalmol 2024; 108:1341–1348; doi: 10.1136/bjo-2024-325459. [Epub ahead of print]
3. Bommasani R, Hudson DA, Adeli E, *et al.* On the opportunities and risks of foundation models. *arXiv [cs.LG]*. 2021. Available at: http://arxiv.org/abs/2108.07258. [Accessed 3 June 2024]
4. Ross A, McGrow K, Zhi D, *et al.* Foundation models, generative AI, and large language models: essentials for nursing. Comput Inform Nurs 2024; 42:377–387.
5. Brown TB, Mann B, Ryder N, *et al.* Language models are few-shot learners. *arXiv [cs.CL]*. 2020; 14165. Available at: http://arxiv.org/abs/2005. [Accessed 3 June 2024]
6. OpenAI, Achiam J, Adler S, *et al.* GPT-4 Technical Report. *arXiv [cs.CL]*, 2023. Available at: http://arxiv.org/abs/2303.08774. [Accessed 3 June 2024]
7. OpenAI. Models: gpt 4 turbo and gpt 4. 2024. Available at: https://platform. penai.com/docs/models/gpt-4-turbo-and-gpt-4. [Accessed 17 May 2024].
8. Martineau K. What is retrieval-augmented generation? *IBM Research Blog*. 2023. Available at: https://research.ibm.com/blog/retrieval-augmented-generation-RAG. [Accessed 28 January 2024]
9. OpenAI. Prompt Engineering. 2024. Available at: https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results. [Accessed 17 May 2024]
10. VertexAI. Tune models overview. *Google Cloud*. 2024. Available at: https://cloud.google.com/vertex-ai/generative-ai/docs/models/tune-models. [Accessed 16 May 2024]
11. Gao Y, Xiong Y, Gao X, *et al.* retrieval-augmented generation for large language models: a survey. *arXiv [cs.CL]*. 2023; 10997. Available at: http://arxiv.org/abs/2312. [Accessed 3 June 2024]
12. Tan TF, Elangovan K, Jin L, *et al.* Fine-tuning Large Language Model (LLM) Artificial Intelligence Chatbots in Ophthalmology and LLM-based evaluation using GPT-4. *arXiv [cs.AI]*. 2024; 10083. Available at: http://arxiv.org/abs/2402. [Accessed 3 June 2024]
13. TensorFlow. Transfer learning & fine-tuning. *TensorFlow*. Available at: https://www.tensorflow.org/guide/keras/transfer_learning. [Accessed 4 June 2024]
14. Mehandru N, Miao BY, Almaraz ER, *et al.* Evaluating large language models as agents in the clinic. NPJ Digit Med 2024; 7:84.
15. ■ Sevgi M, Antaki F, Keane PA. Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities. Br J Ophthalmol 2024; 108:1354–1361.

This article is a valuable resource, showcasing how clinicians without coding expertise can leverage custom GPT features to engage in prompt engineering and use RAG for medical education in ophthalmology.

16. OpenAI. Introducing GPTs. 2024. Available at: https://openai.com/index/introducing-gpts/. [Accessed 3 June 2024]
17. Google. Prompt engineering for generative AI. *Google for Developers*. 2024 https://developers.google.com/machine-learning/resources/prompt-eng. [Accessed 3 June 2024]
18. Nori H, Lee YT, Zhang S, *et al.* Can generalist foundation models outcompete special-purpose tuning? Case study in medicine. *arXiv [cs.CL]*. 2023. Available at: http://arxiv.org/abs/2311. 16452. [Accessed 3 June 2024]
19. promptingguideai. Prompting techniques. 2024. Available at: https://www.promptingguide.ai/techniques. [Accessed 4 June 2024]
20. ■■ Saab K, Tu T, Weng W-H, *et al.* Capabilities of Gemini Models in medicine. *arXiv [cs.AI]*. 2024; 18416. Available at: http://arxiv.org/abs/2404.

This Med-Gemini paper outlines the functionality of the model as a clinical conversational tool, evaluated through text-based clinical benchmarks. It details various medical specializations, including web search retrieval and prompt engineering strategies.

21. Wong M, Lim ZW, Pushpanathan K, *et al.* Review of emerging trends and projection of future developments in large language models research in ophthalmology. Br J Ophthalmol 2023; 108:1362–1370.
22. Tsui JC, Wong MB, Kim BJ, *et al.* Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot. Eye (Lond) 2023; 37:3692–3693.
23. Knebel D, Priglinger S, Scherer N, *et al.* Assessment of ChatGPT in the prehospital management of ophthalmological emergencies - an analysis of 10 fictional case vignettes. Klin Monbl Augenheilkd 2023; 241:675–681.
24. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. Semin Ophthalmol 2023; 38:503–507.
25. Liu X, Wu J, Shao A, *et al.* Uncovering language disparity of ChatGPT on retinal vascular disease classification: cross-sectional study. J Med Internet Res 2024; 26:e51926.
26. Antaki F, Touma S, Milad D, *et al.* Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci 2023; 3:100324.
27. Milad D, Antaki F, Milad J, *et al.* Assessing the medical reasoning skills of GPT-4 in complex ophthalmology cases. Br J Ophthalmol 2024; 108:1398–1405.
28. ■ Antaki F, Milad D, Chia MA, *et al.* Capabilities of GPT-4 in ophthalmology: an analysis of model entropy and progress towards human-level medical question answering. Br J Ophthalmol 2023; 108:1371–1378.

This article evaluates the performance of GPT-4 in the ophthalmology question-answering domain, demonstrating that GPT-4 achieves higher accuracy than GPT-3.5 and performs comparably to historical human performance on board-style exams.

29. Betzler BK, Chen H, Cheng C-Y, *et al.* Large language models and their impact in ophthalmology. Lancet Digit Health 2023; 5:e917–e924.
30. Yaghy A, Yaghy M, Shields JA, Shields CL. Large language models in ophthalmology: potential and pitfalls. Semin Ophthalmol 2024; 39:289–293.
31. Jin K, Yuan L, Wu H, *et al.* Exploring large language model for next generation of artificial intelligence in ophthalmology. Front Med 2023; 10:1291404.
32. Royal College of Ophthalmologists. Facing workforce shortages and backlogs in the aftermath of COVID-19: the 2022 census of the ophthalmology consultant, trainee and SAS workforce. March 2023. Available at: https://www.rcophth.ac.uk/wp-content/uploads/2023/03/2022-Ophthalmology-census-Facing-workforce-shortages-and-backlogs-in-the-aftermath-of-COVID-19.pdf. [Accessed 3 June 2024]
33. Biswas S, Davies LN, Sheppard AL, *et al.* Utility of artificial intelligence-based large language models in ophthalmic care. Ophthalmic Physiol Opt 2024; 44:641–671.
34. Resnikoff S, Lansingh VC, Washburn L, *et al.* Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs? Br J Ophthalmol 2020; 104:588–592.
35. Wang X, Sanders HM, Liu Y, *et al.* ChatGPT: promise and challenges for deployment in low- and middle-income countries. Lancet Reg Health West Pac 2023; 41:100905.
36. Medicines and Healthcare products Regulatory Agency (MHRA). Software and artificial intelligence (AI) as a medical device. *GOV.UK*. 2024. Available at: https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device. [Accessed 20 May 2024]
37. ■■ Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. Nat Biomed Eng 2022; 6:1346–1352.

This review highlights self-supervised learning methods for medical applications, covering tasks like electronic health records and medical images, and suggests these methods can accelerate medical artificial intelligence development.

38. Holland R, Leingang O, Bogunović H, *et al.* Metadata-enhanced contrastive learning from retinal optical coherence tomography images. *arXiv [cs.CV]*. 2022. Available at: http://arxiv.org/abs/2208.02529. [Accessed 3 June 2024]
39. ■■ Yang L, Xu S, Sellergren A, *et al.* Advancing multimodal medical capabilities of Gemini. *arXiv [cs.CV]*. 2024. Available at: http://arxiv.org/abs/2405.03162. [Accessed 3 June 2024]

This project by Google outlines Med-Gemini's functionalities as a multimodal model. It has been evaluated on a comprehensive set of clinically relevant benchmarks. Specifically in ophthalmology, it has been tested on diabetic retinopathy classification tasks. Med-Gemini outperforms the out-of-the-box Gemini in these evaluations.

40. ■■ Zhou Y, Chia MA, Wagner SK, *et al.* A foundation model for generalizable disease detection from retinal images. Nature 2023; 622:156–163.

This article introduces RETFound, the first foundation model in ophthalmology utilizing self-supervised learning SSL techniques and ViT architecture, trained on 1.6 million unlabelled retinal images to enhance model performance and generalizability across various clinical applications.

41. Rasel RK, Wu F, Chiariglione M, *et al.* Assessing the efficacy of 2D and 3D CNN algorithms in OCT-based glaucoma detection. Sci Rep 2024; 14:11758.
42. Qiu J, Wu J, Wei H, *et al.* VisionFM: a multi-modal multi-task vision foundation model for generalist ophthalmic artificial intelligence. *arXiv [eess.IV]*. 2023. Available at: http://arxiv.org/abs/2310.04992. [Accessed 3 June 2024]
43. Gholami P, Lakshminarayanan V. Optical coherence tomography image retinal database. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2019-02-20. Available at: https://doi.org/10.3886/E108503V1-45001. [Accessed 3 June 2024]
44. Antaki F, Chopra R, Keane PA. Vision-language models for feature detection of macular diseases on optical coherence tomography. JAMA Ophthalmol 2024; 142:573–576.
45. Gao W, Deng Z, Niu Z, *et al.* OphGLM: training an ophthalmology large language-and-vision assistant based on instructions and dialogue. *arXiv [cs.CV]*. 2023. 12174. Available at: http://arxiv.org/abs/2306. [Accessed 3 June 2024]


Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

**46.** Maslej N, Fattorini L, Perrault R, *et al. The AI Index 2024 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI (April 2024). Available at: https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf. [Accessed 3 June 2024]
**47.** The STANDING Together collaboration. Recommendations for Diversity, Inclusivity, and Generalisability in Artificial Intelligence Health technologies and Health Datasets (1.0). Zenodo. 2023 https://doi.org/10.5281/zenodo.10048356
**48.** Rieke N, Hancox J, Li W, *et al.* The future of digital health with federated learning. NPJ Digit Med 2020; 3:119.
**49.** Zhuang W, Chen C, Lyu L. When foundation model meets federated learning:
■ motivations, challenges, and future directions. *arXiv [cs.LG]*. 2023; 15546. Available at: http://arxiv.org/abs/2306. [Accessed 3 June 2024]
This paper explores the synergy between foundation model and FL, highlighting mutual benefits. FL enhances foundation model by providing more data and distributing the training load, while foundation model's pretrained capabilities improve FL's performance and efficiency. The interplay facilitates collaborative, inclusive AI development, and addresses challenges like data diversity, overfitting, and privacy. The paper aims to inspire future research to advance scalable, privacy-preserving AI systems.
**50.** Yan B, Cao D, Jiang X, *et al.* FedEYE: A scalable and flexible end-to-end federated learning platform for ophthalmology. Patterns (N Y) 2024; 5:100928.
**51.** BitFount. RETFound. 2024. Available at: https://www.bitfount.com/post/retfound-eye-oct-ai. [Accessed 17 May 2024]
**52.** Ong AY, Hogg HDJ, Kale AU, *et al.* AI as a medical device for ophthalmic imaging in Europe, Australia, and the United States: protocol for a systematic scoping review of regulated devices. JMIR Res Protoc 2024; 13:e52602.
**53.** Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023; 6:120.
**54.** Stade EC, Stirman SW, Ungar LH, *et al.* Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. Npj Ment Health Res 2024; 3:12.
**55.** Metz C. OpenAI says New York Times Lawsuit against it is 'Without Merit'. *The New York Times*, 9 January 2024. Available at: https://www.nytimes.com/2024/01/08/technology/openai-new-york-times-lawsuit.html. [Accessed 22 May 2024]
**56.** Ong JCL, Chang SY-H, William W, *et al.* Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health 2024; 6: e428–e432.

Downloaded from http://journals.lww.com/co-ophthalmology by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 09/29/2024



Copyright © 2024 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.