Progress in Retinal and Eye Research 107 (2025) 101374

Contents lists available at ScienceDirect

Progress in Retinal and Eye Research

journal homepage: www.elsevier.com/locate/preteyeres

Artificial intelligence in ophthalmology: Progress, challenges, and ethical implications

Maria Cristina Savastano [a,b,1], Clara Rizzo [c,1,*], Claudia Fossataro [a], Daniela Bacherini [c], Fabrizio Giansanti [c], Alfonso Savastano [d,e], Giovanni Arcuri [f], Stanislao Rizzo [a,b,g,2], Francesco Faraldi [h,2]

[a] Ophthalmology Unit, "Fondazione Policlinico Universitario A. Gemelli IRCCS", Rome, Italy
[b] Catholic University "Sacro Cuore,", Rome, Italy
[c] Department of Neurosciences, Psychology, Drug Research, and Child Health, Eye Clinic, University of Florence, AOU Careggi, 50139, Florence, Italy
[d] Libera Università Mediterranea Degennaro (L.U.M.), Casamassima, BA, Italy
[e] Ospedale Regionale Generale "F. Miulli", Acquaviva delle Fonti, BA, Italy
[f] Department of Health Technology, IRCCS Fondazione Policlinico A. Gemelli, 00168, Rome, Italy
[g] Neuroscience Institute, Italian National Research Council, CNR, Pisa, Italy
[h] Torino, Eye Clinic, ASL Torino 5, 10024, Turin, Italy

ARTICLE INFO

Keywords:
Artificial intelligence in ophthalmology
AI ethical considerations
Black-box problem
Patient safety
Data privacy
Liability
Trustworthy AI

ABSTRACT

The adoption of artificial intelligence (AI) in ophthalmology holds great promise for improving diagnostic accuracy, optimizing workflows, and enhancing patient care. However, regulatory, ethical, and technical challenges must be addressed to ensure its safe and effective implementation. Bias in AI can lead to disparities in healthcare delivery, while the "black-box problem" raises concerns about transparency and trust. Ethical principles must guide AI integration, particularly regarding patient safety, accountability, and liability. Privacy risks related to data collection and security are especially critical in ophthalmology, where large imaging datasets are essential. Additionally, AI-generated inaccuracies, or "hallucinations," pose potential risks to clinical decision-making. Cybersecurity threats targeting AI-powered healthcare systems further emphasize the need for robust protections. Despite these challenges, AI has the potential to improve access to ophthalmic care, particularly in underserved regions, as seen in AI-assisted diabetic retinopathy screening. However, financial and infrastructural barriers remain significant obstacles to widespread adoption. Addressing these issues requires collaboration among stakeholders, including regulators, healthcare providers, AI developers, and policymakers, to establish clear guidelines and promote trustworthy AI systems. This review explores key regulatory and ethical concerns and highlights strategies to ensure the responsible integration of AI into ophthalmology.

1. Introduction

The concept of artificial intelligence (AI) was first introduced by McCarthy and colleagues (McCarthy et al., 2006) in 1956, who defined it as intelligent machines capable of replicating human-like problem-solving abilities. These abilities include language use, abstract thinking and reasoning, concept formation, problem-solving, meaning discovery, generalization, and learning from experience to accomplish goals without being explicitly programmed for specific tasks (Bali et al., 2019; McCarthy et al., 2006). However, there remains no consensus has yet been reached on the precise definition of what constitutes AI.

AI integration is rapidly increasing across various medical specialties, but ophthalmology stands out as particularly well-suited for its adoption. This is largely due to the field's heavy dependence on imaging and data for different and specific tasks such as screening, diagnosis, prognosis, and treatment of eye conditions (Labib et al., 2024a). (Fig. 1)

The wide range of imaging techniques used in ophthalmology can provide several opportunities for the development of AI programs.

* Corresponding author.
E-mail address: clararizzo2@gmail.com (C. Rizzo).
[1] Co-first author: These authors contributed equally to this work.
[2] Co-last author: These authors contributed equally to this work.

https://doi.org/10.1016/j.preteyeres.2025.101374
Received 28 February 2025; Received in revised form 28 May 2025; Accepted 29 May 2025
Available online 3 June 2025

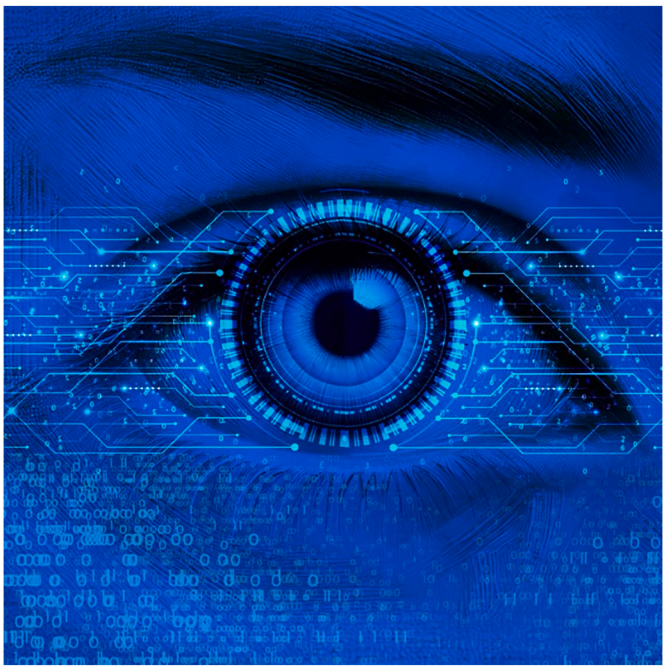

Fig. 1. ***AI-Powered Analysis in Ophthalmology***– An illustration of an eye as the central focus from which various AI-driven analytical processes emerge. (Graphic designer Donata Piccioli creation).

Furthermore, in this field, AI has already demonstrated excellent diagnostic performance in screening for common eye diseases such as diabetic retinopathy (DR) (Gulshan et al., 2016; Ting et al., 2017), glaucoma (Hemelings et al., 2021), and cataracts (Tham et al., 2022).

AI's role in ophthalmology could offer the potential to enhance diagnostic accuracy, enable early diagnosis, and improve the overall quality of care.

Much of the focus has been on how AI can address the growing demand for sustainable healthcare in the face of a rapidly expanding and aging global population (Wahl et al., 2018; Williams et al., 2021a; Xie et al., 2020).

This is especially crucial in low-income (LIC) and low-to middle-income countries (LMIC), where preventable causes of blindness, such as DR and age-related macular degeneration (AMD), could be mitigated through screening programs, home monitoring systems, or telemedicine. Furthermore, this is particularly important in remote areas where access to specialized ophthalmologists is limited (Cen et al., 2021) and AI-driven screening could standardize eye care and improve access. For instance, tools like IDx Technologies (Coralville, USA), which has received approval from the Food and Drug Administration (FDA), are already being utilized as screening solutions in clinical settings (Kras et al., 2020). The powerful capabilities of AI raise concerns that, without proper development and regulation, this technology could pose risks and potential harm to patients. These concerns are especially critical in healthcare, where data breaches or misuse of AI models could have severe consequences.

In this review we discuss the potential issues of AI implementation in ophthalmology practice such as ethical and accountability concerns, along with other concerns such as explainability, cybersecurity, and privacy that must be carefully addressed before AI systems can be widely adopted in clinical settings (Swaminathan and Daigavane, 2024a).

2. Terminology

The terms artificial intelligence, machine learning, and deep learning are often used interchangeably, but they represent distinct concepts. To clarify the key issues linked to AI adoption, it is helpful to define these entities more precisely (Fig. 2).

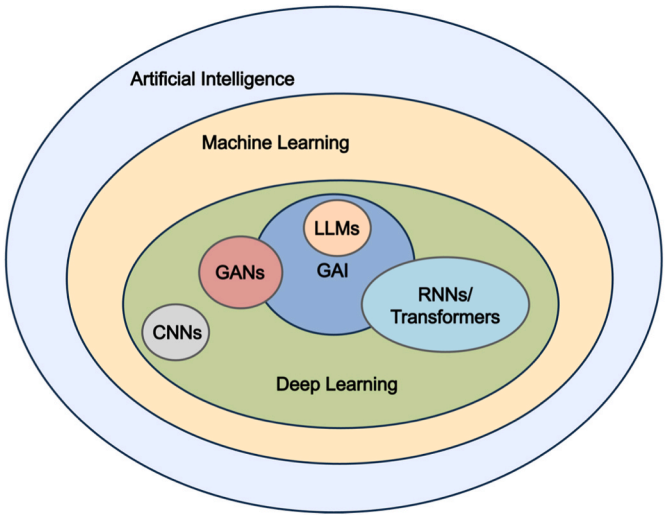

Fig. 2. ***Connections Between Key AI Entities*** – A conceptual diagram illustrating the relationships between various components of artificial intelligence, including Machine Learning (ML), Deep Learning (DL), Generative Adversarial Networks (GANs), Generative Artificial Intelligence (GAI), Recurrent Neural Networks (RNNs) and Trasformers, Convolutional Neural Networks (CNNs) and Large Language Models (LLMs). (MCS creation).

2.1. Machine learning (ML)

Most AI-based models are built upon machine learning (ML), a branch of AI in which algorithms learn patterns and structures from data without relying on predefined rules. ML includes three main approaches: unsupervised learning, where the model identifies hidden patterns in unlabeled data; supervised learning, where models are trained on labeled data to predict known outcomes; and reinforcement learning, where systems learn through feedback by optimizing actions based on rewards and penalties. The goal of these systems is to replicate the functioning of the human brain, allowing them to "learn" from exposure to data and thereby improve the accuracy of their outputs. The solutions it "learns" can subsequently be applied to new, unknown contexts for more efficient data analysis (Labib et al., 2024b; Schmidt-Erfurth et al., 2018).

ML has led to the development of artificial neural networks (ANNs), which are inspired by the brain's neural structure and designed to emulate certain human learning patterns. ANNs consist of interconnected artificial neurons organized in multiple layers. ANNs are particularly effective in classification and diagnostic tasks, such as screening for DR using biomarkers in color fundus photography. Over time, ANNs have evolved into deep neural networks (DNNs), a more advanced architecture that forms the foundation of deep learning (DL) (LeCun et al., 2015).

2.2. Deep learning (DL)

Deep learning (DL) is a subfield of ML based on deep neural networks (DNNs)—architectures with multiple hidden layers capable of automatically extracting and learning abstract data representations. DL models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have achieved remarkable success in image and sequential data analysis, respectively. CNNs use localized filters and shared weights to detect hierarchical features in images, making them particularly effective in medical imaging tasks. RNNs, by contrast, are designed to handle temporal data, leveraging memory of prior inputs for tasks like speech recognition and time-series forecasting. The performance of DL models scales with increased data availability and

computational power, enabling high precision across various domains. However, this complexity often comes at the cost of reduced interpretability and transparency compared to traditional ML approaches. In particular, DL-based systems typically require large annotated datasets for effective training, while ML models can operate with smaller datasets, offering better explainability but often at the expense of accuracy (Mai and Schmidt-Erfurth, 2024; Marey et al., 2024; O'Shea and Nash, 2015; Schmidt, 2019).

2.3. Generative adversarial networks (GANS)

In data-limited fields such as medical imaging, Generative Adversarial Networks (GANs) offer an effective solution for data augmentation. GANs consist of two neural networks—a generator that creates synthetic data samples and a discriminator that evaluates their realism—in a competitive framework. Through this adversarial process, the generator progressively learns to produce highly realistic data, which can be used to expand training sets and improve model generalization without predefined augmentation strategies. This approach enhances the robustness of CNN-based models and helps mitigate the limitations of small datasets. (Goodfellow et al., 2014; Tonti et al., 2024a).

2.4. Generative Artificial Intelligence (GAI)

Generative AI (GAI) refers to a class of algorithms capable of creating original content, such as text, images, videos, music, and even engaging in conversations. Generative AI has evolved from multiple architectures, especially Transformers for language and GANs for images. Transformers are a more recent and powerful alternative to RNNs, designed to overcome the limitations of RNNs by incorporating self-attention mechanisms (Vaswani et al., 2017). These mechanisms allow transformers to evaluate the significance of each word in a sequence relative to all others, bypassing the constraints of sequential processing. As a result, transformers can significantly improve efficiency and enhance the ability to capture long-term dependencies in data. Transformers have played a pivotal role in the rise of powerful Large Language Models (LLMs) such as ChatGPT (Ali et al., 2024; Marey et al., 2024).

Other critical components in the evolution of GAI include autoencoders, which compress input data into a latent space and reconstruct it, laying the groundwork for representation learning and denoising. Latent variable models (LVMs) introduce probabilistic reasoning into generative tasks, enabling more controlled and interpretable generation, as seen in Variational Autoencoders (VAEs)(Kong et al., 2022). Diffusion models represent a newer paradigm, generating data by iteratively denoising a random signal. These models have achieved state-of-the-art performance in high-quality image generation and are being explored for use in audio and video as well (Bengesi et al., 2024). Multi-modal encoders like CLIP (Contrastive Language–Image Pretraining) bridge vision and language by learning aligned representations of images and text. This capability supports generative systems that can interpret and produce multi-modal content, such as image generation from text prompts. CLIP has recently attracted growing attention in the field of medical imaging, being employed both as a pre-training strategy for aligning visual and textual data and as a core component in various clinical applications—such as automated report generation, diagnostic support, and visual question answering on medical images (Zhao et al., 2025).

2.5. Large language models (LLMs)

LLMs are a subset of GAI, primarily designed for language-related tasks. While GAI models have broad applications across various creative fields, LLMs are specifically built to handle tasks that involve text, such as understanding and generating coherent, contextually appropriate written content. Trained on massive datasets of text and code, LLMs can interpret text, commands, and queries, producing responses that approximate human conversation. Among them, ChatGPT has become the most prominent, gaining widespread use and recognition, even within scientific literature. GPT, which stands for "Generative Pretrained Transformer," is an AI model designed to understand and generate human-like text. The potential applications of ChatGPT span numerous fields, including journalism, marketing, education, and professional writing. However, its role in medicine, particularly in specialized areas like retinal care, remains uncertain (Bellanda et al., 2024a). In healthcare, potential applications include medical education, patient counseling, charting, clinical decision support, and academic writing, as the technology continues to advance (Dave et al., 2023).

3. Bias

Bias in AI refers to the systematic and unjust preferences or prejudices that AI systems may exhibit toward specific groups or individuals. Real-world populations are highly diverse, encompassing variations in age, race, comorbidities, socioeconomic status, geographic location, and other factors. This complexity poses significant challenges in developing accurate predictive models for healthcare. Traditional models often struggle to account for this variability, resulting in biases and inaccuracies (Thaler et al., 2024). Gonzalez et al. (González-Gonzalo et al., 2022) identified two primary biases in AI systems trained on inadequately designed or unrepresentative datasets. The first, known as source bias, occurs when models fail to generalize effectively to patients from different racial or ethnic backgrounds. For instance, retinal pigmentation varies in color distribution among ethnic groups, and such differences may result in misclassification. The second, context bias, arises when models rely on features that are merely correlated with the disease, rather than identifying the pathology itself. For example, if a training dataset includes primarily patients from a specific age group, the model might associate certain age-related characteristics with the disease, failing to capture its true defining features. When a model is trained on a dataset containing images that include both the disease of interest and unrelated features, the classifier may prioritize the correlated features over the relevant ones. As a result, the system may fail to extract the necessary disease-specific characteristics. These types of biases can have significant implications for patients in clinical settings where AI systems are implemented. Reduced model performance can lead to inaccurate automated diagnoses, ultimately resulting in inappropriate clinical management decisions and potentially compromising patient care. Furthermore, a critical limitation lies in the representativeness of the datasets, the quality and the data quantity used to train AI algorithms. Poor-quality or biased datasets can impair model performance and limit its ability to generalize across diverse patient populations leading to flawed predictions in real-world applications (Hashemian et al., 2024).

An important root cause of AI bias, particularly relevant in medical settings, is the domain shift between training data and real-world (test) data. ML models can generalize effectively only if training and test data originate from the same underlying distribution. However, deviations between these distributions frequently occur, representing a significant challenge in medical and biological applications (Stacke et al., 2021). Such domain shifts can result from temporal changes in patient populations, new disease variants, evolving diagnostic criteria, changes in clinical practices, or differences in data collection methods. For instance, during the COVID-19 pandemic, the distribution of test results, clinical parameters, and disease prevalence continuously changed due to factors such as new virus mutations and evolving testing strategies(Roland et al., 2022). Consequently, these shifts can degrade predictive performance over time, making previously reliable models ineffective or biased when applied in real clinical settings. Furthermore, standard ML approaches typically struggle to cope with these dynamic shifts, resulting in unreliable and over-optimistic performance estimates (Elsahar and Gallé, 2019). Thus, neglecting domain shifts can unintentionally reinforce biases or inaccuracies, thereby compromising the

clinical reliability of AI systems. In the following subsections, we will examine the main forms of bias that may affect the development, deployment, and performance of AI systems in Ophthalmology.

3.1. Bias related to inadequate representation of real-life population

To ensure that AI algorithms are effective and equitable, it is essential that the datasets used for their training reflect the diversity of real-world populations. The use of large, high-quality datasets is crucial to avoid the underrepresentation of specific patient characteristics such as ethnicity, age, and disease stages (Frank-Publig et al., 2024). However, many AI models, such as those for glaucoma, are often developed using datasets that inadequately represent different demographic groups. This lack of diversity can lead to models that perform well for some populations but poorly for others, creating disparities in healthcare. Since for example glaucoma manifests differently across populations, a model trained predominantly on one group may fail to accurately diagnose the disease in others (Tonti et al., 2024b). To address this issue, it is crucial to include diverse datasets in the training process to ensure that the models are applicable to a wide range of patients. In ophthalmology, ML models require well-labeled, representative, and high-quality datasets, but currently, these datasets are primarily available in a few countries, with limited global representation. While gathering global data is a distant goal, it is important to ensure equitable representation across continents, ethnicities, and countries in order to reduce biases and improve the generalizability of algorithms (Nakayama et al., 2022). Biases in training data not only limit the generalizability of models, but they can also perpetuate and amplify preexisting inequalities, raising significant ethical concerns. These biases particularly affect minority groups, immigrants, rural populations, and individuals with lower socioeconomic status, who are often underrepresented in databases and electronic health records (Boyd et al., 2023; Sonmez et al., 2024a).

3.2. Bias related to low data quality

Low data quality can lead to output errors and biases, particularly when data collection is not standardized, mining the efficient use of this data for AI model training and testing. This challenge has prompted numerous multicentric efforts to address the significant variability in examination processes, timing of laboratory tests, and image quality. To improve data usability, standardizing clinical data collection is essential to generate high-quality, consistent, and complete information that can support the development of reliable medical AI products. For example, health examination records should encompass all necessary items with complete results. Furthermore, low-quality image data often results in the loss of critical diagnostic information, which can negatively affect AI-driven image analyses. Therefore, implementing image quality assessments is crucial to filter out substandard images, thereby improving the performance of AI-based diagnostic models in real-world clinical settings (Feng et al., 2024; Li et al., 2023a).

3.3. Bias related to data scarcity

Real training datasets frequently suffer from limitations such as under-representation of certain patient demographics, and rare diseases (Norori et al., 2021; Sonmez et al., 2024b). Insufficient data sample sizes pose a significant challenge in AI research, especially for rare ocular diseases with low incidence rates. Researchers often struggle to gather enough data to develop reliable AI models for these conditions. One solution to this issue is the generation of synthetic images, which can address critical challenges in ophthalmology. By augmenting datasets with synthetic images, researchers can reduce the reliance on real ones, improving model performance across a broader range of scenarios. This approach also helps tackle generalizability issues, as models trained on limited datasets often struggle to perform well on underrepresented image types. **However, it is important to note that synthetic images—although useful—may not fully capture the complexity and heterogeneity of real-life clinical findings. As a result, exclusive or excessive reliance on artificial data could limit the model's ability to generalize effectively in real-world settings. This represents a minor but relevant concern in the development of AI systems.** Data augmentation must be performed carefully to ensure that the data is diverse, balanced, and representative while also respecting data privacy concerns. In particular, LLMs are able to address the challenges posed by limited data volumes. They enable data augmentation and make effective use of existing resources, driving advancements in medical image classification, rare disease diagnosis, and other healthcare applications. However, while LLMs can help mitigate the scarcity of medical data, they have notable limitations, including hallucinations and lack of standardization (Wong et al., 2024). Hallucinations occur when GAI models produce content that appears plausible but is actually incorrect. This issue can arise from various factors related to the quality, quantity, and inherent biases of the training data. Therefore, it is essential to address and identify hallucinations to ensure the accuracy of medical diagnoses, decision-making and the reliability of medical AI. On another note, to address the issue of data scarcity, large databases have already been created in some countries (Martins et al., 2022). Currently, there are extensive databases, such as electronic medical records and digital images, that enable the recognition of patterns in large volumes of data in a short time. One notable example is the *Intelligent Research in Sight* database, which was developed to store data from 17,363,018 patients across 7200 ophthalmologists in the United States. The primary goal of this initiative was to enhance individual patient care and inform public health policies (Chiang et al., 2018).

3.4. Identification of bias in the development of AI

Nakayama et al. (2024) identified seven critical steps in which bias may develop when applying AI in ophthalmology. These steps include data collection, defining the model task, data preprocessing and labeling, model development, model evaluation and validation, deployment, and post-deployment evaluation. Herein, we will discuss them separately.

3.4.1. Data collection

Bias can arise from unbalanced data sources, as representation plays a critical role in evaluating the data used to train AI algorithms. Strict inclusion criteria may inadvertently exclude essential subgroups, while overly broad criteria can lead to an increased number of false positives within the dataset. Additionally, underrepresented populations are often left out, and even datasets with broad representation can carry historical biases that misrepresent certain groups. Currently, as previously mentioned, ophthalmic datasets used in AI research are notably imbalanced, with retinal imaging datasets lacking adequate representation from middle- and low-income countries (Khan et al., 2021). Another form of bias that could affect data collection is spectrum bias, as mentioned by Gonzalez et al. (González-Gonzalo et al., 2022), which occurs when the dataset does not encompass the full spectrum of disease severity in the target population, resulting in suboptimal model performance. For example, ophthalmology datasets predominantly featuring mild cases of DR may lead to algorithms that are proficient at detecting mild cases but less effective in identifying more severe instances, or vice versa (Faes et al., 2020).

3.4.2. Defining the model task

Bias can occur during the phase of selecting the appropriate disease target, as the prevalence of ocular diseases varies significantly across geographic regions and is influenced by demographic and socioeconomic factors. Therefore, defining the disease target to a specific population and healthcare context is fundamental to ensure that the model works properly. Another factor to consider is data sparsity bias, which

arises when a dataset lacks enough cases representing specific combinations of exposures and outcomes (Greenland et al., 2016). Even large databases may not include an adequate sample of cases, risk factors, variables, or outcomes, leading to limitations in model training and performance.

3.4.3. Data preprocessing and labeling

In data preprocessing, it is crucial to be mindful of potential bias when handling missing data. This can result from various factors such as human and machine limitations, data collection errors, limited access to clinical screenings, or participants' refusal to engage in research. Additionally, when GAI techniques are used to augment datasets, they can perpetuate biases if the training data does not accurately represent diverse groups. In ophthalmology, datasets often lack essential information on comorbidities, demographics, or examination results. Labeling errors and inconsistencies also contribute to bias. Inaccurate or inconsistent labeling can diminish model performance and reliability. Labeling in ophthalmology is especially challenging due to different grading criteria and standards, as well as significant variability among expert evaluators, which can further introduce bias into the labels.

3.4.4. Model development

Bias can emerge during the model development process in several ways. One example is diagnostic suspicion bias, which occurs when prior knowledge of a patient's exposures or conditions influences the data. Another source of bias is data leakage, which happens when information from the training dataset inadvertently leaks into the validation cohort or when certain features reveal outcome information (Tampu et al., 2022). Lastly, model shortcuts can occur when models rely on non-clinical features for predictions, such as hospital location or the type of equipment used.

3.4.5. Model evaluation and validation

AI algorithms are typically tested on external datasets to assess their real-world performance. However, bias can arise due to several factors: one potential issue is the use of wrong evaluation metrics which may obscure the underperformance of models in certain patient subgroups, leading to widened outcome disparities. Another source of bias is the use of wrong evaluation methods. Evaluation bias occurs when the methods used to assess model performance are themselves biased. For example, using inadequate external datasets that do not accurately represent the target population can result in biased evaluations of the model's true performance.

3.4.6. Model deployment

Deploying AI systems in real-world settings requires understanding the data characteristics, social factors influencing data generation, and proper model task definitions. Bias can occur when defining the model threshold. Decision thresholds must align with the model's intended purpose and the healthcare context. For instance, in DR screening, a lower threshold increases sensitivity but may result in more false positives and unnecessary referrals. Bias can also arise from covariate and dataset shifts. Covariate shifts occur when the distribution of input variables differs between the training and deployment phases. In ophthalmology, factors such as changes in patient demographics, disease prevalence or equipment use can hinder the model's ability to generalize.

3.4.7. Post-deployment monitoring and recalibration

Continuous monitoring and evaluation of deployed AI systems are crucial to assess their clinical impact. Bias can occur during the updating of AI systems, as the procedures for updates may vary across different settings.

3.4.8. Authors' perspective

Drawing from our clinical and research experience in Ophthalmology, we have directly encountered many of the challenges and biases described above in real-world AI applications. The deployment of AI systems in settings such as diabetic retinopathy screening or AMD detection has revealed performance disparities, particularly when the models are applied to images from underrepresented populations or acquired with different imaging devices than those used during training. These limitations often stem from source bias, context bias, and domain shifts that are not accounted for during model development. A particularly recurrent issue in clinical practice is the limited generalizability of AI models. For example, models trained predominantly on datasets from high-income countries may misclassify or overlook disease manifestations in patients from other countries, due to differences in retinal pigmentation or disease presentation. We have also observed that low image quality, common in routine care, frequently results in unreliable predictions if not properly filtered during preprocessing. Furthermore, rare ophthalmic diseases continue to be a blind spot for most AI systems due to data scarcity—an issue that even synthetic data and large language models only partially address. To ensure clinically meaningful and equitable AI, we believe the focus must shift toward the creation and curation of diverse, high-quality, and globally representative datasets. Rigorous standardization in data collection and labeling is crucial, along with the incorporation of domain adaptation techniques and continuous post-deployment monitoring. Embedding explainability tools and defining context-appropriate decision thresholds are equally important, especially when dealing with sensitive tasks such as disease triaging and referral decisions. Most importantly, AI should not function in isolation but as part of a clinician-in-the-loop system, where human oversight and expertise remain central to safe and effective care delivery.

4. Ethical consideration for AI implementation

The foundational principles of medical ethics include beneficence, nonmaleficence, justice, and autonomy. Beneficence emphasizes the moral obligation of physicians to actively promote the well-being of their patients. This involves preventing harm, aiding individuals with disabilities, and taking action to help those in danger. Nonmaleficence, closely related, requires physicians to avoid causing harm and to carefully weigh the benefits and risks of medical interventions. Justice focuses on the fair distribution of healthcare resources, which can be achieved through various approaches, such as prioritizing those in greatest need, or merit-based distribution. Autonomy asserts the rights of individuals to make decisions about their own care. Within the framework of autonomy, aspects such as informed consent, truth-telling, and confidentiality become especially important in the context of AI, raising critical ethical questions about its use in healthcare (Kalaw and Baxter, 2024). A critical dimension closely tied to justice in AI-driven healthcare is fairness, which involves ensuring equitable access to high-quality care and equitable health outcomes across diverse patient populations (Marmot and Bell, 2012). Fairness requires addressing biases originating from the data itself, algorithmic design, and interactions between AI systems, healthcare providers, and patients. Data biases such as minority bias, where underrepresentation of certain groups in training datasets leads to inaccurate predictions, can exacerbate existing healthcare disparities. For example, cardiovascular risk models trained primarily on male patient data may inaccurately predict risk among female patients, leading to inadequate care (Appelman et al., 2015). Similarly, missing data bias arises when information from specific patient groups is systematically absent, hindering AI's ability to deliver reliable and equitable predictions. Algorithmic biases, including label bias and cohort bias, further perpetuate inequalities; Label bias occurs when AI models are trained using labels or classifications that reflect underlying healthcare disparities rather than objective truths (Ueda et al., 2024a). Cohort bias arises when AI systems are developed based on predefined or overly generalized patient groups, neglecting important variations or subgroups; for instance, overlooking the unique healthcare experiences of specific populations like LGBTQ +

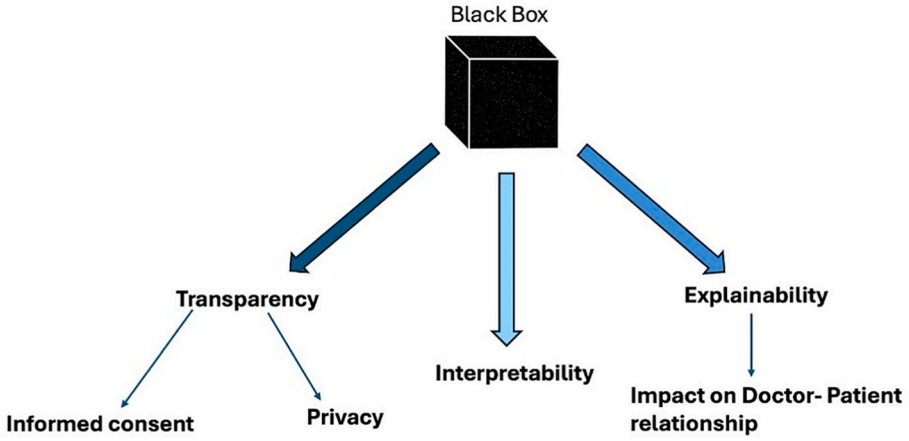

Fig. 3. ***The Black Box Phenomenon in AI*** – A conceptual illustration of the "black box" problem in artificial intelligence, highlighting challenges in transparency, interpretability, and explainability. The image depicts how AI processes input data and generates outputs without clear visibility into its decision-making process. Key concerns such as informed consent, data privacy, and the impact on the patient-doctor relationship are illustrated, emphasizing the ethical and practical implications of opaque AI systems in ophthalmology. (CR creation).

individuals, potentially resulting in inaccurate predictions and inadequate care (Meyer, 2003). Notably, algorithms employing biased proxies—such as healthcare costs to predict medical needs—have resulted in discriminatory outcomes against racial minorities (Obermeyer et al., 2019). Additionally, clinician interaction-related biases can also impact fairness. Automation bias, wherein healthcare professionals overly rely on AI recommendations even when incorrect, can disproportionately affect patient care quality for underrepresented groups ("How Should AI Be Developed, Validated, and Implemented in Patient Care?," 2019). Patient interaction-related biases, including privilege bias and informed mistrust, highlight disparities in access to AI technology and skepticism due to historical inequities and exploitation in healthcare (Rajkomar et al., 2018).

Recent guidance and legislation have emphasized key principles for the implementation of AI, including ethical considerations (Muralidharan et al., 2023). In response, health systems have established governance committees and processes aimed at ensuring the safe and equitable deployment of AI tools. Despite these efforts, there remains a critical gap as no standardized assessment tool currently exists to identify and address ethical challenges during the implementation of AI prediction models in healthcare practice (Ning et al., 2024). The ethical implications of AI in healthcare involve multiple dimensions, with significant concerns about bias and transparency. Inadequate or biased training datasets can unintentionally reinforce existing biases, resulting in discrimination and unfair treatment of patients. Another critical issue is the "black-box" nature of many AI models, which makes it difficult to understand or explain their decision-making processes. This lack of transparency is closely tied to concerns about patient privacy and informed consent, as individuals may not fully understand how their data is being used or how AI-driven decisions are being made.

4.1. Blackbox

AI systems that interpret data and make decisions have sparked significant debate, particularly around the "black-box problem." This issue highlights the opacity of these systems, making it difficult to understand, predict, or influence how they transform inputs into outputs (Veritti et al., 2024).

As a result, users often struggle to trust AI and are hesitant to rely on systems they cannot fully comprehend (Swaminathan and Daigavane, 2024b). Black-box models are especially common in fields like image analysis and processing, such as ophthalmology, where they have proven highly effective. Three concepts clarify the challenges of black-box models: transparency, interpretability, and explainability. Transparency refers to the degree to which the internal workings of an AI model can be observed or understood by humans, focusing on the structural and process-oriented aspects of the system. Interpretability, in contrast, relates to how well humans can explain and understand the relationship between inputs, internal processing, and outputs, with an emphasis on the model's behavior and outcomes (Schmidt-Erfurth et al., 2018). Meanwhile, explainability involves providing meaningful, post hoc explanations for predictions or decisions, enabling user understanding through external tools or generated outputs. Essentially, explainability is about creating actionable insights for systems that would otherwise remain opaque (Stead, 2018; Wong and Bressler, 2016).

The black-box phenomenon becomes especially critical when an AI model does not perform as expected or provides incorrect answers. DL models, in particular, rely heavily on large volumes of historical data, such as fundus photography in ophthalmology, to recognize features associated with specific conditions. However, when presented with a novel image, the AI may misdiagnose a patient if the training set is incomplete or inadequate and without transparency, it is impossible to understand why a failure occurred (Evans et al., 2022a). Additionally, the opaque decision-making process inherent in the "black-box" nature of AI systems raises serious concerns about trust and reliability among healthcare providers and patients. This lack of transparency can also harm the doctor-patient relationship, as clinicians may struggle to explain the AI's decisions to their patients, thereby undermining confidence in the system (Laï et al., 2020). As AI continues to advance, the need to explain the decision-making processes behind these systems has become increasingly important. Several methods have been made available to provide explanations for the black-box nature of DL algorithms. One method is the generation of saliency maps, or heat maps, using techniques like class activation mapping (CAM). The saliency map is typically represented as an image heatmap, showing the local morphology changes that would most significantly alter the network's predictions (Schmidt-Erfurth et al., 2018). As this method visually highlights areas in an image that are most important for the classification decision made by the model, this approach greatly aids clinical understanding. However, there is no consensus on the most suitable saliency map generation method for ophthalmic imaging data. Some evidence suggests that explainability methods that rely purely on visual assessments, such as the ones commonly used in ophthalmology, may be misleading or insufficient for truly explaining the relationship between model inputs and outputs (Liu and Bressler, 2020).

4.2. Black box and informed consent

In standard medical procedures, physicians are required to ensure that patients understand the nature of the procedure, its risks, and available alternatives. The same principle applies to AI-assisted diagnosis, such as DR detection. In the European Union, the General Data Protection Regulation (GDPR) asserts that patients own and control their data, and explicit consent is required for its use or sharing. As such, it is essential that patients are fully informed about how their data will be handled, with mandatory consent required when it is shared with AI developers. Patients must know who can access their data, how it will be stored, used, and how their privacy will be protected (Kelly et al., 2019). Transparency and informed consent are closely connected ethical principles in the use of patient data for AI applications in ophthalmology. Clinicians and researchers must prioritize educating patients, ensuring they fully understand and consent to the collection, storage, and use of their health data. This approach empowers patients to make informed decisions about their participation in AI-driven projects, fostering trust and supporting ethical practices (Swaminathan and Daigavane, 2024c). However, for informed consent to be comprehensive and understandable, the "black box" nature of ML systems poses a challenge to explainability. Patient privacy is another critical ethical consideration in AI-assisted healthcare. Upholding patient privacy rights and ensuring meaningful control over data usage is essential, requiring clear communication about how patient data is utilized in AI systems. Patients must fully understand what the system will do with their data, ensuring transparency and trust in how their personal health information is handled. This understanding is key to promoting ethical practices and protecting patient autonomy in the integration of AI technologies in healthcare (Karimian et al., 2022). (Fig. 3)

4.3. Ehical issues of LLMs

The integration of LLMs such as chatbots into healthcare, particularly in ophthalmology, goes beyond simple patient interactions, expanding into telemedicine platforms and offering services like medical video consultations, tele-reporting, administrative tele-consultancy, and tele-assistance for data transmission from ambulances to hospitals (Ittarat et al., 2023). This makes healthcare more accessible, especially in remote areas where physical consultations may not be feasible. Chatbots can also inform patients about ongoing research or clinical trials, assist in recruitment, and gather feedback after consultations or surgeries, providing ophthalmologists with valuable insights. With multilingual support, chatbots break down language barriers, ensuring that quality care is accessible to a broader range of patients. However, when incorporating LLMs into these systems, several ethical concerns arise, particularly in relation to the four main principles of medical ethics: beneficence, nonmaleficence, autonomy, and justice (Kalaw and Baxter, 2024). LLMs can compromise harm avoidance, as they may offer inaccurate or unsupported recommendations. For example, one study found that ChatGPT incorrectly recommended using silicone intubation and mitomycin-C in dacryocystorhinostomy without any supporting evidence (Ali, 2023). This undermines the principle of nonmaleficence by providing potentially harmful advice. Furthermore, LLMs raise concerns about patient autonomy and data privacy. Ethical and legal issues related to data collection, breaches, and the potential for misrepresentation are also significant. The lack of transparency in how data is processed can exacerbate these problems. Hallucination, a term used to describe AI-generated content that is nonsensical or inaccurate, further complicates the issue, as it can result in misinformation that might influence patient care. The issue of justice also arises in the use of LLMs in healthcare, as the absence of accountability mechanisms raises concerns about their ethical use, particularly when sensitive patient data is involved (Bressler, 2023). Furthermore, chatbots, which process personal and sensitive information, must adhere to strict data security protocols to protect patient data from unauthorized access or misuse. Informed consent is crucial for using ophthalmology chatbots. Patients should be informed about the chatbot's purpose, capabilities, limitations, and how their data will be used. Consent should be obtained before initiating any interaction, and patients should have the option to opt out and withdraw consent at any time. Confidentiality is also fundamental in maintaining patient trust. Ophthalmology chatbots should implement secure data transmission methods, such as encryption, to ensure that patient information is protected. They should also provide clear disclaimers that chatbots are not substitutes for in-person consultations and that patients should seek professional medical advice when necessary. This ensures that patients are aware of the limitations of chatbot interactions. The lack of transparency in these black-box models makes it difficult to understand the rationale behind their decisions, which is critical in clinical contexts. Healthcare professionals may be reluctant to trust AI-generated suggestions without clear explanations of how they were derived. Despite these concerns, LLMs in healthcare can be a valuable tool, but human involvement and supervision remain essential (Yang et al., 2024).

4.4. Authors' perpective

As researchers and clinicians engaged in the integration of AI within ophthalmology, we have come to recognize that ethical considerations are foundational to the safe and effective deployment of AI technologies. While the four core medical ethics principles—beneficence, nonmaleficence, justice, and autonomy—have long guided clinical decision-making, they take on new complexity in the context of AI systems, particularly as these tools become increasingly autonomous, opaque, and influential in care delivery. In our experience, one of the most urgent ethical challenges lies in addressing fairness and transparency. AI systems often perform well in controlled environments, yet in real-world clinical settings, we have observed that these same systems can perpetuate or even amplify existing healthcare disparities. Additionally, black-box algorithms pose significant difficulties for clinicians and patients alike—raising concerns not only about accountability but also about patient trust, informed consent, and the preservation of clinician–patient communication. The ethical limitations of LLMs further complicate the landscape. While we recognize their promise in extending access to care and supporting tele-ophthalmology, we have also witnessed firsthand their tendency to hallucinate, provide unsupported clinical recommendations, and handle sensitive data in ways that challenge both privacy and accountability standards. These limitations underscore the need for human oversight and transparent disclosure about AI's capabilities and constraints, especially when interacting directly with patients. We believe that future AI systems must be designed with ethical safeguards embedded at every stage: from data governance and bias mitigation, to model explainability and patient education. Only by aligning technological advancement with ethical responsibility can we ensure that AI contributes to a more equitable, trustworthy, and humane healthcare system.

5. Privacy

Privacy concerns in data collection, storage, and usage remain a significant issue for healthcare providers, especially as AI systems require large, high-quality datasets to achieve optimal performance. In fields like ophthalmology, the demand for extensive, well-curated imaging datasets presents unique challenges. As patient information accumulates over time and can potentially be linked with other datasets, complete anonymization becomes increasingly difficult, making it nearly impossible to fully protect privacy. Privacy concerns are particularly pressing in retinal imaging, where true anonymization is challenging due to the unique, identifying characteristics of the retinal vasculature, which can act as a "fingerprint" for patients (Akram et al., 2020). For example, fundus photographs can reveal details such as age, gender, and cardiovascular risks (Poplin et al., 2018). Furthermore, gaps

Fig. 4. ***AI and Privacy in Healthcare*** – An artistic representation of a robotic hand delicately holding a locket, symbolizing the critical balance between technological advancement and patient privacy. The locket signifies the protection of sensitive medical data, while the robotic hand represents AI and automation in healthcare. (Graphic designer Donata Piccioli creation).

in existing regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, make the situation more complex. For instance, data re-identification can occur through methods like data triangulation, which raises concerns and contributes to hesitations among healthcare providers about adopting AI technologies (Tseng et al., 2021a). To mitigate these risks, data protection experts and ethics committees call for robust "de-identification" or "de-personalization" of medical images. They stress that while full anonymity may be unachievable, appropriate safeguards should be put in place to reduce re-identification threats (Schmidt-Erfurth et al., 2018). Another key issue involves GDPR Article 17, the "right to be forgotten," which allows patients to request the removal of their data from AI models. However, even when data is deleted, trained models may still produce outputs that contain patient-specific information. Prolonged data retention and failures to erase data upon request heighten the risks of unauthorized access, privacy violations, and cyberattacks (Fig. 4).

Managing data sharing in an ethically responsible way is another critical issue. AI development often requires consolidating data from multiple institutions, especially when dealing with rare diseases. This means that sensible information must be shared outside the institution where it was originally collected, raising the risk of privacy breaches. Additionally, collaborations between healthcare organizations and corporations in the pharmaceutical and technology sectors increase concerns about the commercialization of clinical data (Tom et al., 2020). Exclusive contracts or licenses that restrict the sharing of data for broader patient benefit may undermine fairness and equity in healthcare. To address these privacy concerns, several innovative solutions are being explored (Liu and Bressler, 2020). One strategy to address privacy concerns is through the use of synthetic imaging data, which can be provided by GANs. In ophthalmology, GANs have been successfully employed to produce synthetic fundus photographs and autofluorescence images, allowing DL models to be trained without the need for real patient data.

In addition to these approaches, researchers are exploring advanced privacy-enhancing technologies (PETs) to complement and strengthen existing safeguards. These include differential privacy, which introduces calibrated statistical noise into model outputs to prevent the identification of individual data points, even if they have access to additional background information. This technique ensures that the presence or absence of a single patient does not significantly affect the outcome of a model, thereby enabling privacy-preserving analytics on sensitive datasets. It is particularly valuable when institutions wish to release aggregate statistics or train models on population data without exposing individuals. Beyond privacy protection, differential privacy has also been shown to improve generalization, reduce overfitting, and support fairness by smoothing out outlier effects (Zhu et al., 2022). Secure multi-party computation (SMPC), on the other hand, enables multiple institutions to collaboratively compute functions over their combined data without revealing their individual inputs to each other. This is especially relevant in clinical settings where direct data sharing is restricted due to legal, ethical, or regulatory barriers. SMPC protocols use cryptographic primitives such as homomorphic encryption and secret sharing to break sensitive data into encrypted fragments that are processed jointly. In healthcare AI, SMPC is increasingly used in distributed training of diagnostic models, risk prediction systems, and genomic analyses, allowing for large-scale collaboration without compromising patient confidentiality. While computationally intensive, recent optimizations have significantly improved the scalability of SMPC for practical medical use (Zhou et al., 2024). Cryptographic methods, including zero-knowledge proofs (ZKPs), provide mechanisms to verify the correctness of a computation or model inference without revealing the underlying data or model itself. In medical AI, these tools can be used to ensure data integrity and verifiable compliance with access rules, particularly when models are deployed in cloud-based or cross-border systems. ZKPs and related cryptographic protocols also support traceability and accountability in decentralized environments by enabling institutions to audit data use without violating privacy (Taherdoost et al., 2025). These techniques are especially valuable in trust-sensitive contexts such as clinical trials, insurance claims, and public health monitoring. Another promising tool is machine unlearning, which supports the selective removal of individual records or entire subsets from trained models. This is particularly relevant in response to legal requests for data deletion, such as those arising from the GDPR's "right to be forgotten." Unlike conventional retraining from scratch—which is costly and inefficient—machine unlearning enables localized forgetting through strategies like statistical pruning, gradient updates, or partitioned model rollback. This allows developers to correct for outdated, erroneous, or ethically problematic data after deployment. In addition to regulatory compliance, machine unlearning contributes to model transparency, user autonomy, and risk mitigation, and is increasingly recognized as a key enabler of responsible and trustworthy AI in medicine (Hine et al., 2024). Together, these technologies reinforce the protection of sensitive patient data in modern decentralized AI development environments. By enabling privacy-by-design across all stages of the AI lifecycle—from data acquisition to model deployment—they help ensure that ethical, legal, and social standards are met while maintaining the collaborative potential of AI across institutions, countries, and sectors.

Another promising approach is federated learning (FL), a ML framework that enables the training of AI models across multiple independent, decentralized systems, fostering collaboration while maintaining the privacy of sensitive data. More recently, swarm learning has emerged as a blockchain-based peer-to-peer data security solution. Unlike FL, which depends on a central analytics server, swarm learning fully decentralizes data (Warnat-Herresthal et al., 2021). These emerging AI technologies that emphasize decentralization offer enhanced privacy protection by keeping data within individual institutions while still delivering comparable results to traditional centralized methods (Lo et al., 2021).

5.1. Federated learning (FL)

In FL, a global algorithm is trained by aggregating model parameters

from multiple institutions without sharing raw data. During federated analysis, the algorithmic code is sent to each data site for local processing, and the results are then aggregated at a central location for further analysis. FL is also expected to be particularly useful in the future for diagnosing, predicting, and treating rare or geographically uncommon diseases, such as ocular tumors or inherited retinal diseases, where low incidence rates and small datasets present challenges. Connecting multiple institutions globally could improve clinical decision-making regardless of patients' locations and demographic backgrounds. While FL offers improved privacy preservation and collaborative training, there are still numerous limitations within the framework (Nguyen et al., 2022). First, the robustness of the FL model depends heavily on the quality of data provoided by each participating site. Since original data is kept private at local sites and not directly accessed by the entire federation, there may be the unintentional inclusion of poor-quality data that could compromise the model's performance. Second, explainability remains crucial for real-world acceptance of DL models. Although explainability analyses, such as heat map generation and saliency analysis, are possible, further studies are needed to validate these methods within the FL framework to increase clinician and patient acceptance. Third, while original data remains private, the transfer of model updates in FL is still vulnerable to privacy attacks. Attackers may use model parameters to reconstruct original datasets (reconstruction attack) or infer if specific data records were used in model training (inference attack). Another issue in FL arises from the heterogeneity of medical data across different institutions. Institutions contributing to FL often have varied amounts of data, different imaging protocols, and diverse patient populations. Additionally, the problem of bias could be exacerbated, as each participant brings their own bias into the global model and may even generate new biases. These biases could also be related to differences in devices. Another challenge with FL is communication. Federated networks, especially those involving many devices, may suffer from slower transmission speeds compared to a local one, particularly when local models are uploaded to a central server. Additionally, continuous communication between participants and the global server requires reliable network bandwidth to support the extensive download and upload processes.

5.2. Author' perspective

In our direct experience with AI development and clinical implementation in Ophthalmology, privacy remains one of the most pressing and complex challenges—both technically and ethically. The increasing reliance on large-scale imaging datasets, particularly high-resolution fundus and OCT images, creates inherent risks for patient re-identification, even when traditional anonymization techniques are applied. We have observed that the retinal vasculature, akin to a biometric identifier, renders complete anonymization nearly impossible. This tension between the need for numerous longitudinal data and the imperative to protect patient confidentiality represents a major barrier to scaling AI solutions in real-world practice. Moreover, existing regulatory frameworks, such as HIPAA and GDPR, while foundational, are often insufficient in guiding the nuanced requirements of AI-era privacy. For instance, the implementation of the "right to be forgotten" is especially challenging when it comes to AI models already trained on sensitive data. In our research settings, we have seen that even when data is deleted from storage, model outputs may still contain traces of that information—a critical issue for compliance and ethical integrity. Emerging privacy-preserving technologies such as federated learning, differential privacy, and secure multi-party computation offer promising pathways. We have begun experimenting with federated frameworks to enable inter-institutional collaboration without centralizing data. However, these techniques come with their own limitations: federated models remain susceptible to reconstruction and inference attacks, and heterogeneity in imaging protocols across centers can introduce new layers of bias. Furthermore, while synthetic imaging generated by GANs can supplement datasets in low-resource contexts, we caution that such data should be used judiciously, as it may not fully reflect the variability and complexity of real clinical scenarios. To move forward responsibly, we believe AI development in Ophthalmology must adopt privacy safeguards at every stage—from data acquisition and preprocessing to model training and deployment. Only through transparent, secure, and ethically guided data governance can we unlock the true potential of AI while safeguarding one of healthcare's most foundational principles: patient privacy.

6. Patient safety

Patient safety is a fundamental concern in the application of AI technology within ophthalmology. One of the primary risks lies in the potential of GAI to produce inaccurate or misleading results, often referred to as "hallucinations." These errors can jeopardize patient safety by spreading health-related misinformation and may lead to more severe consequences if these systems are integrated into clinical decision-making, reporting, or training processes for diagnostic algorithms and medical professionals (Jin et al., 2023). Unlike clinicians, who are adept at assessing risks, considering outcomes, and balancing the impact of false-positive or false-negative results on a patient's condition, AI systems can fail when faced with unfamiliar data. In such cases, even a reliable AI system may produce erroneous predictions that a busy clinician might accept without further scrutiny, potentially overlooking alternative diagnoses or treatments. Another limitation of current AI tools is their design, which often focuses on analyzing specific regions of interest or single disorders. This narrow scope renders them ineffective in addressing multiple conditions or complex multisystem complaints simultaneously. Real-world data, characterized by variability due to regional and population differences, equipment disparities, and inconsistent data curation, further complicates the situation. AI systems frequently struggle to interpret this variability or recognize critical contextual changes in the data that should influence predictions, leading to overlooked secondary pathologies or confidently incorrect conclusions that disregard individual patient contexts. Over time, as algorithms age, their inability to adapt to changes in diseases, medical practices, or medications risks perpetuating outdated or inappropriate approaches. Established AI systems are particularly rigid when it comes to incorporating medical advancements, as the absence of prior data for retraining limits their capacity to integrate new practices, ultimately hindering innovation in healthcare (Ahmed et al., 2023). Equally concerning is the potential harm associated with data breaches. While anonymized datasets have been instrumental in advancing technology, they also present significant risks. Breaches of patient privacy violate the ethical principle of beneficence, which requires healthcare professionals to "do no harm," and can lead to numerous consequences. These include threats to employment opportunities, insurance coverage complications, and identity theft, such as the misuse of Social Security numbers or financial information (Shi et al., 2020). The task of ensuring the complete removal of identifiable information from large datasets is extremely difficult, and any failure in this regard can result in a loss of anonymity or misuse of data, leading to adverse outcomes such as mental stress and compromised patient welfare (Abdullah et al., 2021; Tom et al., 2020).

6.1. Authors' perspective

From our clinical and research engagement with AI systems in ophthalmology, patient safety has emerged as a central dimension in evaluating the real-world viability of these technologies. While AI offers exciting diagnostic efficiencies, we have observed that its integration into clinical workflows must be approached with caution, particularly given the risk of "hallucinations" or confidently incorrect outputs generated by generative AI models. In real-world settings, clinicians often work under time constraints, and the risk of over-relying on AI-

generated suggestions—especially when they appear accurate—is nontrivial. We have encountered cases where AI tools misinterpret subtle retinal features or miss comorbid conditions simply because the model was trained to focus narrowly on a single pathology. These limitations, when unrecognized, may lead to clinical errors, inappropriate management, or delayed care—each of which has direct implications for patient safety. Furthermore, many current models lack adaptability. As medical knowledge, diagnostic tools, and treatment evolve, static algorithms risk becoming outdated, yet they may still be deployed without periodic updating. This raises concerns not only about clinical performance but also about the ethical obligation to provide up-to-date care. Another critical concern pertains to data privacy breaches and the potential harm they can cause. While de-identified data is a cornerstone of AI training, the residual risk of re-identification persists—especially with imaging data such as fundus photos, which contain unique biometric signatures. The consequences of a breach extend beyond clinical settings, potentially affecting a patient's psychological well-being, insurability, or financial security. To uphold patient safety, AI systems must be continuously monitored post-deployment and currently employed in workflows that emphasize human oversight. It is essential to create mechanisms that allow clinicians to verify, override, or question AI-driven conclusions without workflow disruption. Moreover, periodic reassessment of algorithmic performance, especially in the face of evolving clinical practices, should be institutionalized as part of AI governance frameworks.

7. Legal implications of application of AI in clinical practice

The integration of AI in healthcare raises significant concerns about both liability and accountability, particularly regarding AI-induced errors. Liability refers to the legal responsibility of a party when harm occurs due to an AI system's actions. In contrast, accountability encompasses the moral, ethical, or legal obligation to explain and justify the decisions or actions taken by AI systems. As such, accountability is about identifying who is responsible for ensuring that AI systems meet ethical standards and avoid causing harm. Unlike liability, which often results in financial consequences, accountability covers broader legal and ethical responsibilities. One key issue is the lack of transparency of many AI models. The black box phenomenon, as previously mentioned, makes it difficult to understand how decisions are made, complicating the identification of responsibility in cases of misdiagnosis or wrong treatment recommendations (Lang et al., 2022). From a legal perspective, the relationship between clinician trust and accountability is crucial. When a physician relies on an AI system for diagnosis or treatment suggestions without understanding how it arrived at those recommendations, it can create complications if an error occurs. In healthcare, particularly in fields like ophthalmology, where AI assists in diagnosing conditions such as DR or glaucoma, understanding the AI's decision-making process is essential for assigning accountability and liability. If an AI system fails to detect a condition or suggests an incorrect course of action, determining who is legally responsible becomes complicated. Currently, a common legal approach is to treat medical AI as a confirmatory tool, meaning physicians must verify every output to ensure it meets the standard of care (Li et al., 2023b). If a clinician follows an AI recommendation outside the standard of care without verifying it independently and an error results, the clinician may be held liable for malpractice (Smith and Fotheringham, 2020). The standard of care for a physician is the level of treatment and practice considered appropriate by the medical community, based on current knowledge, guidelines, and patient-specific factors, used to assess the reasonableness and competence of medical actions. However, this approach can limit the effectiveness of AI in healthcare, as it might discourage clinicians from relying on AI systems that could outperform human experts in certain situations. The "responsibility gap" arises when it is unclear who should be held accountable for errors — whether it's the AI developer, the healthcare provider using the system, or the healthcare institution that deployed it. Experts have suggested various approaches for assigning liability (Smith and Heath Jeffery, 2020). For instance, AI manufacturers might bear liability for autonomous systems when used as intended, while healthcare providers might be responsible if the system is used off-label or beyond its prescribed purpose. Alternatively, healthcare providers might be held accountable for negative outcomes, as they are responsible for making the final decision based on AI input. As AI systems become more autonomous, however, the question arises: should healthcare providers still bear full responsibility? Another approach is shared responsibility, where different stakeholders, such as AI developers and healthcare providers, share responsibility for the use and functioning of the AI system. Risk pooling has also been proposed, wherein all parties — AI developers, healthcare providers, and institutions — contribute to an insurance pool that compensates patients for harm caused by AI errors (Smith and Fotheringham, 2020). This would enable collective risk management and quicker compensation, avoiding lengthy legal battles. The legal landscape is still developing, and current regulations are not fully equipped to handle the complexities posed by AI in healthcare. As AI becomes increasingly integrated into clinical practice, clearer legal frameworks will be necessary to address liability issues effectively. Ethical frameworks will also play a crucial role, as they can help differentiate between the responsibility to ensure the AI system functions safely and the moral or legal liability when harm occurs. Accountability remains a central aspect of these discussions, with all parties — clinicians, healthcare organizations, and AI developers — responsible for ensuring that AI systems do not harm patients. Companies like IDx, which develop AI diagnostic tools, have taken steps to address the responsibility gap by accepting liability for errors in their AI platforms and purchasing liability insurance for their products (Abràmoff et al., 2020). However, when AI systems are used off-label or in unintended ways, the responsibility may shift to the healthcare provider (Evans et al., 2022b). As AI technology evolves, so must regulatory frameworks and legal definitions surrounding liability and responsibility. Clear guidelines are needed to ensure the safe and effective use of AI in healthcare while

Fig. 5. ***Legal Implications of AI in Clinical Practice*** – An illustration of a judge's gavel surrounded by lines of software code, symbolizing the intersection of AI and legal accountability in healthcare. This image highlights key issues such as liability in the use of AI-driven diagnostics and decision-making tools in clinical settings. (Graphic designer Donata Piccioli creation).

maintaining public trust. The allocation of responsibility for AI-related errors will remain a key legal issue, requiring ongoing updates to legislation and closer collaboration among healthcare providers, AI developers, and regulators (Fig. 5).

7.1. Authors' perspective

Through our experience working with AI technologies in ophthalmology, we have come to recognize that legal uncertainty remains one of the greatest barriers to responsible AI adoption in clinical practice. While the promise of AI is substantial, it brings with it a growing complexity in determining who is accountable when things go wrong—a dilemma that clinicians, developers, and institutions alike are now facing. We have encountered firsthand the tension between technological advancement and legal responsibility, particularly when using AI models that function as diagnostic support tools. In practice, clinicians are still expected to validate AI outputs and align them with the established standard of care. However, as models grow more autonomous and sophisticated, the action of constant oversight diminishes—raising questions about whether clinicians should continue to bear the full legal burden for AI-driven decisions they did not fully understand or control. The "responsibility gap" is especially pronounced in black-box models, where opacity obstructs transparency and traceability. In our view, this lack of explainability makes it nearly impossible to assign clear legal liability when AI systems misclassify diseases or fail to detect relevant findings. Additionally, reliance on AI recommendations in time-constrained settings may lead clinicians to accept flawed suggestions, further blurring the line between use and misuse, intent and negligence. We believe that a shared responsibility model, supported by well-defined legal and ethical frameworks, is urgently needed. Such a model would appropriately allocate liability among AI developers, healthcare institutions, and clinical users, depending on the context and nature of the system's use. As the legal landscape catches up with technological innovation, we emphasize the importance of transparency, documentation, and continuous post-deployment monitoring.

8. Ownership of AI-generated intellectual property

The issue of ownership surrounding new data generated by AI in healthcare has become a significant concern due to the lack of regulatory guidelines. As AI systems continue to evolve and generate new insights, there is ambiguity regarding who owns the newly created material. This problem extends also to the data sets used to train AI algorithms, which may remain unstandardized, raising ethical concerns regarding data ownership and deployment (Bai et al., 2024). As such, the question of AI-generated health data ownership in ophthalmology remains a critical concern. A notable example is Eye2Gene, an AI algorithm designed to diagnose inherited retinal diseases (IRDs). Eye2Gene uses patient retinal images to predict causative genes, and its development involves training on datasets from multiple hospitals, followed by validation on external datasets (Nguyen et al., 2023). However, there is still no clear indication of who holds ownership of these datasets. This lack of transparency highlights the need for clear policies and regulations to define data and algorithm ownership in the context of healthcare AI, ensuring ethical standards and the proper use of sensitive patient information (Veritti et al., 2024).

Beyond data ownership, emerging academic discussions have emphasized that the rise of AI-generated content (AIGC) introduces complex intellectual property challenges that extend well beyond traditional copyright law. AI systems are increasingly trained on massive repositories of digital content—often containing copyrighted material—without explicit licensing or compensation to the original creators. In this context, questions arise not only about the legality of such practices, but also about the redistribution of economic value between content originators and AI developers(Kou et al., 2024). In healthcare, these issues take on a unique urgency, as both training data (e.g., medical images, patient records) and generated outputs (e.g., diagnostic predictions, treatment recommendations) may be considered valuable intellectual assets with potential commercial applications. Furthermore, the ownership of AI-generated outputs remains legally ambiguous. While many jurisdictions do not recognize AI systems as legal authors or inventors, the lack of consensus on whether human stakeholders (e.g., developers, institutions, or data contributors) can claim exclusive rights to AIGC introduces uncertainty in the commercialization and reuse of such material. For example, in a clinical setting, it is unclear whether a hospital contributing patient data to train an AI model has any claim over the model's outputs, such as diagnostic predictions or novel biomarker discoveries derived from aggregate analysis. The use of protected or proprietary datasets without appropriate governance can undermine incentives for data sharing and research collaboration, particularly if contributors perceive that their assets are being exploited without recognition or benefit. This is especially problematic in healthcare, where patient data is not only sensitive but often subject to strict regulatory frameworks (e.g., GDPR, HIPAA). There is an increasing need for transparent governance models that address both data provenance and value distribution. Another emerging issue concerns the possible erosion of innovation incentives. If AI systems are able to freely generate derivative works based on proprietary clinical content, the original innovators—be they researchers, healthcare institutions, or clinicians—may be disincentivized to continue producing high-quality data or content. This could result in long-term negative impacts on medical research, innovation, and public health outcomes. To mitigate these risks, it is essential to establish frameworks that formally recognize contributions to training datasets, ensure equitable attribution, and, where appropriate, provide financial or reputational compensation. In conclusion, while the Eye2Gene case exemplifies immediate concerns around dataset transparency and consent, it also serves as a broader illustration of unresolved questions about intellectual property in AIGC. As AI systems become increasingly integrated into healthcare delivery and research, robust legal, ethical, and economic frameworks are urgently needed to define ownership, protect creators' rights, and ensure equitable distribution of benefits derived from AI-generated medical knowledge.

8.1. Authors' perspective

In our experience working with AI systems within ophthalmology, we are encountering growing uncertainty surrounding the question of who owns AI-generated insights, predictions, and outputs. While the value of AI-driven tools is evident—particularly in cases like Eye2Gene, which utilize patient images to predict genetic causes of inherited retinal diseases—the ownership of the data used, the algorithms developed, and the knowledge produced remains deeply ambiguous. One critical concern is that many AI tools are trained on multicenter datasets contributed by hospitals, research institutions, and clinicians, often without a clear agreement on how outputs—be they diagnostic models or novel biomarkers—should be credited, shared, or monetized. We have seen that this lack of transparency and formal governance can discourage data sharing, undermine trust among collaborators, and ultimately slow the pace of innovation in healthcare AI. The challenges extend further when AI systems generate content with potential commercial or research value. In our view, the absence of legal recognition for AI-generated outputs as intellectual property, paired with unclear attribution rights for human contributors risks creating a vacuum in which valuable outputs are produced without accountability or fair distribution of benefits. This is particularly concerning in ophthalmology, where the datasets required for high-performing models are not only costly to collect but also sensitive under GDPR and HIPAA frameworks. We also believe the current ambiguity risks disincentivizing innovation. If researchers, institutions, or clinicians perceive that their contributions are repurposed or commercialized without acknowledgment or compensation, it could hinder future data contributions and

weaken academic collaboration. Conversely, failing to protect AI-generated outputs might limit their integration into regulated clinical workflows, due to unresolved questions about authorship and legal liability. To address these issues, we advocate for the development of transparent, equitable governance frameworks that define ownership and usage rights over training datasets and AI-generated outputs, establish attribution and benefit-sharing mechanisms for contributors, and ensure compliance with data protection laws while supporting responsible innovation.

9. Cybersecurity

There has been a growing trend of cyberattacks targeting AI-powered systems, with the healthcare sector predicted to face more attacks than other industries. Gonzalez et al. (González-Gonzalo et al., 2022) identified two key factors that amplify this risk: financial interests and technical vulnerabilities. Financial motivations are closely linked to healthcare fraud, which can involve both large corporations and individuals. These fraudulent activities often manipulate patient data for insurance claims, clinical decisions, or drug and device approvals, resulting in significant global financial losses. At the same time, technical vulnerabilities within healthcare infrastructures create opportunities for external breaches, leading to risks such as blackmail, ransomware, and malicious data manipulation. The increasing use of malware to compromise personal information has eroded trust in AI algorithms, with patients becoming increasingly concerned about the safety of their data. This concern has grown as more personal information, such as facial recognition data, is collected in healthcare and ophthalmology. While legislation like the HIPAA in the United States offers important protections for patient data, the privatization of AI in healthcare has raised new concerns. In fact, reduced oversight in private AI systems can lower algorithm quality and jeopardize patient safety (Labib et al., 2024a). As AI becomes more prevalent in healthcare, safeguarding patient rights and strengthening privacy protections are essential. Measures such as encryption, access controls, and anonymization should be rigorously implemented to reduce the risks of unauthorized access, data breaches, and misuse. GAI, particularly LLMs, introduces additional cybersecurity challenges due to its reliance on vast clinical datasets for training. This data is vulnerable to attacks that can expose sensitive information through model outputs. For instance, a data breach involving OpenAI's ChatGPT in March affected 1.2 million users, exposing names, addresses, payment details, and chat histories due to a bug in the open-source code (Wang et al., 2024). Moreover, the financial burden of implementing robust security systems to protect data adds another layer of complexity. In addition to these risks, AI systems used in healthcare are susceptible to sophisticated attacks like adversarial examples, where imperceptible perturbations in input data can mislead models into making incorrect predictions (Sarker, 2023).

In simple terms, these attacks involve making tiny changes to medical images or other inputs—so small that the human eye can't notice them—that can still trick the AI into giving a wrong result. For example, a slightly modified retinal scan could be incorrectly classified by the AI as showing diabetic retinopathy, or could fail to detect it when it's actually there. This poses serious risks to clinical safety, potentially leading to incorrect diagnoses or treatments. Researchers are working on methods to make AI systems more resistant to such attacks, a concept known as 'adversarial robustness'. However, these solutions are still being developed and often require significant computing power, making them hard to apply in clinical practice today. Improving adversarial robustness is therefore critical to ensure safe and reliable AI use in ophthalmology and other medical fields. Another major threat is the risk of backdoor attacks, in which models are deliberately manipulated during training so that they behave normally under standard conditions but produce malicious outputs when triggered by specific inputs (Grosse et al., 2022). In this case, an attacker can "poison" the training data by inserting hidden patterns—such as a specific pixel arrangement in a fundus image—that later causes the AI to give a wrong result only when that pattern is present. In all other cases, the AI works normally, making it very hard to detect the issue. In healthcare, this could be used to fraudulently influence diagnostic results or medical decisions, for example by always flagging certain images as showing a disease, regardless of their true content. This could lead to mistreatment or insurance fraud, and is especially dangerous when AI models are trained on external or unverified datasets. Jailbreak attacks also undermine AI system integrity by exploiting LLMs to bypass safety filters and elicit harmful outputs (Tshimula et al., 2024). These attacks work by cleverly phrasing a prompt (input) so that the model ignores its safety rules. For instance, a user might pretend to be a doctor asking for emergency advice and trick the AI into giving restricted medical information. In a clinical context, such attacks could lead AI systems to provide inaccurate or unauthorized recommendations, which could be harmful if followed by healthcare professionals or patients. These kinds of vulnerabilities are particularly concerning when using LLMs in medical chatbots, virtual assistants, or clinical documentation tools. Finally, prompt injection attacks have emerged as a novel class of vulnerabilities specific to LLM-based systems. These attacks manipulate model behavior by inserting hidden prompts into user inputs or external content, effectively hijacking the model's intent and overriding its safeguards. For example, an attacker might hide special instructions in a medical note or website text that a clinical AI tool reads. The model might then follow these hidden instructions instead of behaving as expected—possibly leaking confidential patient data or giving inappropriate advice. In a healthcare setting, this could have serious consequences if LLMs are integrated into electronic health records or used in automated decision-making. To address these challenges, blockchain technology has emerged as a promising solution. Blockchain integrates decentralized and secure features to enhance transparency, trust, and accountability in AI systems. By recording data in a distributed ledger composed of linked and verified blocks, blockchain ensures the integrity of patient data while enabling efficient and transparent data sharing. In ophthalmology, this technology has revolutionized data management by providing a secure and transparent method for handling large volumes of sensitive information. Integrating blockchain with AI systems enhances privacy protections, maintains data integrity, and facilitates international collaboration across healthcare networks (Gurnani et al., 2023). Alongside blockchain, other important cybersecurity strategies have been developed to protect AI systems (Ramos-Cruz et al., 2024). One example is the cybersecurity mesh, a flexible security model that allows different parts of a healthcare system—such as clinics, devices, or servers—to be protected individually while still working together. This helps detect threats earlier and limits the spread of cyberattacks. Another useful tool is the intrusion detection and prevention system (IDS/IPS), which uses AI to monitor activity in real time and stop suspicious behavior before damage occurs. There are also lightweight, efficient tools like Bloom filters that help scan large amounts of data quickly to find irregularities or threats. Additionally, cryptographic techniques such as digital signatures and hash trees help verify that clinical data has not been tampered with, without needing to decrypt it. Finally, the "zero trust" approach means that every access to data—by a person, device, or software—is verified, regardless of where it comes from. This reduces the risk of unauthorized access or data leaks. Malicious cyberattacks on AI systems in healthcare threaten not only patient privacy but also the overall quality of care, resulting in financial losses and diminished trust in AI technologies. These risks can ultimately hinder the successful integration of AI into clinical practice. Employing blockchain or other advanced security measures is therefore critical to safeguarding sensitive patient information, maintaining transparency, and ensuring the secure and effective use of AI in healthcare.

9.1. *Authors' perspective*

Cybersecurity is main concenrn in the complete integration of AI

systems in the clinical practice of ophthalmology. As healthcare becomes more digitized and reliant on AI, we have observed an increased vulnerability to sophisticated cyber threats, from data breaches to adversarial and backdoor attacks, many of which could directly compromise patient safety and erode institutional trust. In particular, the growing integration of LLMs and generative AI in medical applications has introduced novel risks, such as prompt injection, jailbreak manipulation, and model inversion, which we have found to be poorly understood outside of specialized AI communities. These threats are especially concerning when clinical tools rely on external datasets or are deployed across interconnected systems, making them difficult to secure uniformly. In our view, many institutions remain underprepared to defend against these advanced forms of attack. To address this, we believe healthcare systems must adopt proactive strategies—such as the implementation of blockchain for immutable data tracking or cybersecurity mesh frameworks that decentralize and isolate system vulnerabilities. Importantly, cybersecurity efforts must go hand-in-hand with clinician education to ensure not only technical robustness but also operational awareness of the risks AI systems can pose. In our view, building AI systems that are not only intelligent but also resilient and secure must be treated as a core responsibility shared by developers, institutions, and regulators alike.

10. Job displacement

With the rapid development of AI, concerns about job displacement have emerged alongside its advantages and disadvantages. One notable advantage is AI's potential to make healthcare more accessible, efficient, and affordable, such as in AI-assisted DR screening, which addresses the shortage of ophthalmologists (Resnikoff et al., 2012) in low-income countries, where the prevalence of diabetes may be high (Kazi et al., 2020; Wahl et al., 2018; Williams et al., 2021b). AI's diagnostic capabilities also stand out, enhancing accuracy and efficiency in medical practice, as demonstrated in image-based diagnostics for conditions like DR. AI systems have shown performance levels that may surpass human ophthalmologists and detect subtle patterns in data that humans cannot, enabling earlier disease detection and novel insights into disease mechanisms (Korot et al., 2020a; Poplin et al., 2018). Despite these advances, AI models face limitations when dealing with image variability and patient-specific differences outside their training data, necessitating clinician oversight to interpret findings and provide personalized care.

AI's narrow capabilities mean it cannot independently diagnose or encompass all possible conditions, highlighting the indispensable need for physician involvement (Blasi et al., 2001; Crincoli et al., 2023).

Clinicians remain essential in integrating AI insights with patient preferences and comorbidities, ensuring AI supports rather than replaces decision-making (Tonti et al., 2024c).

Ethical considerations also arise, including accountability, transparency, and the impact of AI on the doctor-patient relationship, which itself has therapeutic value (Tseng et al., 2021b).

Patients often prefer human interaction over AI-based systems, particularly for personalized understanding and assurance. However, if integrated with human patient-doctor interaction, AI's ability to process large datasets might allow clinicians to focus more on direct patient interactions, leading to a more effective and compassionate care (Labib et al., 2024c). While AI can assist with tasks more efficiently than humans, its present role is to complement, not replace, clinical judgment, enhancing efficiency and patient care when integrated with human expertise.

10.1. AI in surgical fields: opportunities and challenges of replacing doctors

AI applications in surgery are still in their early stages, with research and development efforts primarily focusing on clinical decision support tools, assistive technologies, and systems for improved outcomes prediction and surgical case selection. A previous article reported that while 47 % of jobs are at risk of automation, the risk for physicians and surgeons is only 0.4 % (Korot et al., 2020b). In ophthalmology, most advancements have concentrated on static images, such as fundus photography and optical coherence tomography (OCT). However, areas like video analysis and action recognition within surgical workflows still represent significant unmet needs for AI applications in healthcare. The creation of robust image and video libraries is essential for developing AI-based software and tools for assistive or autonomous surgical applications. For these technologies to enhance surgery effectively, they must be able to track and monitor surgical instruments, devices, and critical anatomical landmarks in real-time (Leiderman et al., 2024a). In cataract surgery, some researchers have employed DL methods to automate trainee performance analysis using structured assessment tools or idealized performance benchmarks (Hira et al., 2022; Tabuchi et al., 2022; Wang et al., 2022). However, accurately assessing step-skill levels, overall expertise, or progress in surgical training remains a significant challenge for ML. Posterior segment ophthalmic surgeries present additional obstacles to AI development, including the need for a broader dynamic range of image acquisition due to vitreoretinal surgery's lighting conditions, partial visualization of the posterior segment during operations, and the complex optics of surgical viewing systems for vitreous surgery (Leiderman et al., 2024b). Progress in future AI-guided intraocular robotic surgical systems could augment surgeons' natural visual and physical abilities, aid in decision-making processes, and even enable fully autonomous procedures through real-time feedback. By integrating AI-generated information with robotic systems, surgeons' abilities to manipulate intraocular anatomy could be enhanced. Features such as tremor dampening, collision avoidance, and exclusion zones around sensitive anatomical structures could be implemented. Ultimately, although at present we have not achieved this yet, a fully AI-guided intraocular robotic system holds the theoretical potential to achieve fully autonomous surgical procedures.

10.2. Role of chatbots (LLMs) in job replacement: opportunities and challenges

The rise of LLMs is revolutionizing healthcare by allowing physicians and patients to interact with AI systems. AI language models, such as ChatGPT, have garnered significant attention for their ability to process large amounts of information and provide human-like responses. As ophthalmology integrates advanced technologies like image analysis and telemedicine, LLMs are likely to play a key role in future eye care. In addition to ChatGPT, other AI models like Google's Bard AI, which excels in accurate text summarization and can access real-time information, are contributing to the AI landscape in healthcare (Chotcomwongse et al., 2024). While ChatGPT's performance in patient management has yielded mixed results (Bellanda et al., 2024b), it has shown potential in providing diagnostic and management recommendations based on patient data. Studies have tested ChatGPT's ability to provide medical advice, with varying results depending on the complexity of the case and the benchmarks used for comparison (Anguita et al., 2024; Bernstein et al., 2023; Cappellani et al., 2024; Caranfa et al., 2023; Ferro Desideri et al., 2023; Nanji et al., 2024; Patil et al., 2024). For instance, ChatGPT's ability to triage and counsel patients in ophthalmology and retinal care has been explored, with patients interacting with the AI to describe symptoms and receive guidance on appropriate care (Lyons et al., 2024a; Waisberg et al., 2023; Zandi et al., 2024). However, issues like miscommunication and the AI's ability to accurately interpret symptoms remain challenges. LLMs have improved in terms of response capabilities and eloquent dialogue, offering enhanced human-AI interaction. By leveraging a wide range of internet sources, LLMs can respond to a variety of inquiries, including problem-solving, summarization, and information creation. In ophthalmology, LLM chatbots can assist in answering patient queries about eye diseases, treatments, and research,

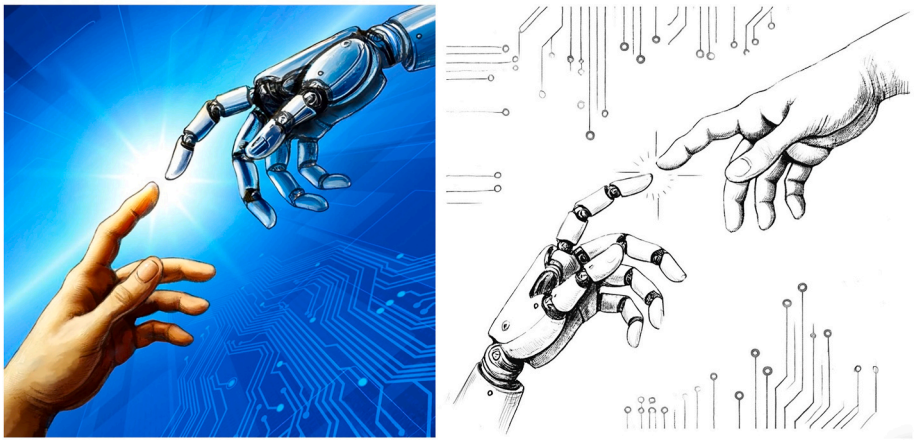

Fig. 6. ***Collaboration Between AI and Humans in Healthcare*** – An illustration of a human hand reaching out to touch a robot hand, symbolizing the essential collaboration between AI and human professionals in healthcare. It highlights the potential for AI to support, rather than replace, human healthcare providers. (Graphic designer Donata Piccioli creation).

enhancing patient engagement by providing reliable medical information. While many studies have demonstrated the value of LLMs in offering recommendations for ophthalmic diseases, such as DR, myopia, and surgical retinal conditions, some generated responses have been incomplete or inaccurate (Chotcomwongse et al., 2024). LLMs chatbots can also serve as diagnostic support tools, helping patients gather initial information about symptoms before seeing a physician. However, LLMs are less proficient in providing specialty-specific knowledge. GPT-4, for instance, has demonstrated faster diagnostic capabilities compared to specialists in some cases, but its performance in diagnosing rare diseases, like uveitis, has been less reliable (Rojas-Carabali et al., 2024). Futhermore while studies have shown that ChatGPT can achieve high triage accuracy in diagnosing ophthalmic symptoms (Knebel et al., 2024; Lyons et al., 2024b), concerns remain about the potential for harmful advice, particularly when the AI provides vague or incorrect treatment recommendations. While LLMs like ChatGPT have the potential to revolutionize patient care, their tendency to hallucinate or fabricate information presents a major concern (Cappellani et al., 2024; Chen et al., 2025). These fabricated responses could mislead patients and healthcare providers, raising ethical issues regarding the reliability and safety of AI in medical practice. Another significant concern with LLMs is the potential risk to patient privacy. As these models may require access to sensitive data such as medical history and ocular images to generate responses, it is essential to ensure that patient consent is obtained and that data are securely stored and encrypted to prevent breaches (Chotcomwongse et al., 2024). Finally, in rapidly evolving fields like ophthalmology, addressing the limitations of LLMs like ChatGPT, especially its inability to access real-time information (The training data of earlier versions of LLMS such as ChatGPT-3.5 has been reported to stop at September 2021), is crucial (Gurnani and Kaur, 2024). The continuous development of new therapies, surgical techniques, and clinical trials means that AI systems must incorporate mechanisms for real-time updates and validation to ensure accuracy and safety. As AI continues to evolve, it is essential for ophthalmologists to supervise and collaborate with LLMs, ensuring that their outputs are rigorously tested and validated to avoid patient harm (Fig. 6).

Fig. 7. ***Global Standards for AI Governance*** – An illustration representing the growing international focus on creating regulatory frameworks for AI, such as the EU's AI Act, which classifies AI systems based on risk levels and provides guidelines for their implementation. (Graphic designer Donata Piccioli creation).

10.3. Authors' perspective

As ophthalmologists and researchers working at the intersection of clinical care and AI innovation, we recognize that the issue of job displacement by AI is often discussed as a direct threat to the medical profession and a potential pathway to the complete substitution of physicians. While AI tools, including diagnostic algorithms for diabetic retinopathy or surgical performance assessment models, have shown impressive capabilities, they remain highly dependent on curated data, specific conditions, and rigorous human oversight. Rather than replacing clinicians, these tools have thus far functioned best when augmenting our capacity to analyze data, triage care, or monitor performance—particularly in under-resourced settings or high-volume environments. We have seen that in surgical fields, including cataract and vitreoretinal surgery, AI still faces significant technical and ethical barriers before it can meaningfully impact real-time decision-making or autonomy. The development of intraocular robotic systems powered by AI is promising, but remains largely theoretical at this stage, and cannot replicate the clinical judgment and dexterity required in complex ophthalmic procedures. Similarly, while LLMs like ChatGPT offer novel possibilities in patient interaction, triage support, and medical summarization, our experience and ongoing evidence reveal considerable

limitations in clinical accuracy, interpretability, and safety. The risk of hallucinations, vague recommendations, and outdated knowledge underscores the importance of strict human validation, especially in fast-evolving subspecialties like ophthalmology. Moreover, patients still express a clear preference for human interaction, especially when it comes to complex or emotionally sensitive care decisions. If used responsibly, AI may actually improve the quality of patient care by freeing clinicians from repetitive administrative tasks and allowing them to focus more on direct communication and personalized care. However, the ethical, legal, and emotional implications of shifting aspects of diagnosis, counseling, or surgery to machines must be continuously evaluated. In our view, the future of ophthalmology is not one in which physicians are replaced, but rather one in which they must evolve into collaborators and supervisors of intelligent systems.

11. Global standards for AI governance

There is currently no universally established global standard for AI, though each country is working on its own regulations. A global standard for AI governance would involve a universally accepted set of principles, guidelines, and frameworks for the development, deployment, and regulation of AI systems across various sectors. The rapid development of AI models has been met with a slower regulatory response, leading some to suggest halting further advancements until stricter regulations are established (Sonmez et al., 2024c). Organizations like the European Union are advancing with frameworks such as the "AI Act," which categorizes AI models based on their risk levels and provides tailored guidelines (Quaranta et al., 2023). (Fig. 7)

Healthcare digitalization frameworks will likely need to incorporate AI-specific provisions, particularly for GAI. International regulatory bodies, including the FDA and the European Commission, are progressing toward establishing standardized frameworks for AI-based medical devices, with an emphasis on patient safety, iterative software updates, and post-market monitoring (González-Gonzalo et al., 2022). Standardization in AI is the process of creating uniform guidelines and protocols that ensure AI systems are developed, deployed, and assessed consistently, reliably, and ethically across industries. This standardization includes establishing approval processes tailored to AI's unique aspects, such as data handling, adaptability, liability, defining specific AI applications, conducting benchmark tests, creating user guidelines, and validating models' effectiveness through clinical trials. These trials, with meaningful patient outcomes, should support initial approvals and be made publicly accessible. To accelerate these evaluations, regulatory bodies should promote collaboration with healthcare institutions, insurance providers, and patient advocacy groups. Engaging physicians in the creation of healthcare-specific AI regulations through dedicated task forces will also help ensure patient protection. As a matter of fact, ensuring that AI tools undergo thorough clinical evaluations will be critical to supporting regulatory decisions that guarantee their safety and efficacy for the target population. As AI technology evolves, continuous updates and standardized post-market surveillance are essential to maintaining system performance and ensuring patient safety. Deployed AI models require ongoing monitoring and rapid adaptation to new tasks and environments. Given AI's continuous learning capabilities, regulatory pathways must be flexible enough to accommodate updates without necessitating a full review for every change. This can be achieved through a combination of internal and external audits, ensuring AI systems remain up-to-date and perform as expected. Regular assessments of AI algorithms will help align them with clinical needs and identify any deviations from established standards. Future efforts should also prioritize the development of ethical guidelines and regulatory frameworks to address the challenges posed by unexplainable AI models in healthcare. Ensuring transparency and adherence to ethical principles is crucial for maintaining trust and safety in clinical practice. Legal frameworks must evolve to tackle AI's unique complexities, balancing innovation with trasparency. Transparency in regulatory and approval processes can for example be strengthened through public databases or registries of approved AI systems. These should include clear definitions, approval statements, clinical evidence, and deployment details to enhance oversight and public confidence and to ensure transparency and safety. While the FDA provides an open database for approved medical devices, the European Commission's EUDAMED database is not yet publicly available, though efforts are underway to improve its accessibility (González-Gonzalo et al., 2022). At the same time, privacy concerns must be carefully addressed. The security of patient data remains a significant issue, and AI systems must comply with data protection regulations like the GDPR to prevent breaches. In cases where privacy concerns limit access to clinical data, regulatory bodies could play a crucial role in facilitating information-sharing to support the approval process (Nair et al., 2024).

11.1. Authors' perspective

As clinicians and researchers actively engaged in AI development and implementation in ophthalmology, we believe that the absence of unified global standards for AI governance represents one of the most critical gaps in the safe and equitable integration of these technologies into clinical practice. In our experience, regulatory inconsistencies across countries create significant barriers to international collaboration and complicate the approval and deployment process for AI-based medical systems. While regional initiatives—such as the European Union's AI Act or evolving FDA guidelines—are commendable, they remain fragmented and insufficient to address the complex, cross-border nature of modern AI innovation. The lack of harmonization also poses challenges for clinical validation: many AI systems are trained in specific geographic or demographic contexts, and without shared regulatory benchmarks, their generalizability and safety across diverse populations cannot be ensured. Equally important is the creation of public registries that document approved AI tools, their intended use, clinical performance, and regulatory status—an initiative that could improve transparency, accountability, and public trust. Moreover, as AI systems increasingly incorporate sensitive patient data, we believe that global governance must include robust privacy protections aligned with evolving frameworks like the GDPR. In situations where data-sharing restrictions pose a barrier to AI validation, regulatory agencies should facilitate secure, ethical mechanisms for federated analysis and cross-border collaboration.

12. Financial implications of AI

The financial implications of adopting AI on a large scale must be carefully evaluated, as they can present significant barriers to the global adoption of this technology. Securing adequate funding is essential to support the adoption and maintenance of AI services. However, sustainable development of AI requires a careful balance between profit and expenses. For healthcare providers, the costs associated with adopting AI-enabled technologies must align with manufacturing prices and reimbursement models. Key costs include development, data collection, research, and validation, alongside expenses for software upgrades, IT infrastructure, long-term hardware maintenance, operator training, incorporating new patient data, and ensuring regulatory compliance (Chen et al., 2021; Ruamviboonsuk et al., 2021). In less developed countries, financial limitations further complicate AI implementation. Restricted resources make it challenging to purchase programs, educate patients, and acquire standardized, high-quality data, including images and interpretation protocols (Anton et al., 2022). Moreover, privacy regulations and state-specific rules, while designed to protect patient confidentiality, can create barriers to the adoption of AI. Data-sharing initiatives require anonymization and patient consent, adding to the complexity of implementation. Incorporating robust techniques to safeguard patient data involves significant investments in technical solutions for data storage, management, and analysis (Ting

Fig. 8. ***Environmental Impact of AI Implementation in Clinical Practice*** – An illustration of a robot hand holding a growing plant in its palm, symbolizing the environmental considerations tied to the deployment of AI technologies in healthcare. It highlights the importance of sustainable AI practices in clinical settings to minimize the ecological damage. (Graphic designer Donata Piccioli creation).

et al., 2019). These efforts demand substantial time, financial resources, and expertise. Since the benefits are not immediately apparent, investing in data-sharing initiatives becomes a difficult decision.

13. Environmental impact of AI implementation in clinical practice

The environmental impact of the application of AI in ophthalmology must also not be overlooked. The use of AI in healthcare can significantly contribute to carbon emissions throughout its lifecycle, from programming and development to deployment, due to the high energy demands of AI systems (Richie, 2022). Additionally, the extraction of minerals and metals for AI hardware has substantial environmental consequences, as the process often involves unsustainable practices, resulting in habitat destruction, biodiversity loss, and water pollution (Sovacool et al., 2020). AI systems, especially DL models, require immense computational power and energy consumption during both training and inference phases, leading to an increased demand for energy-intensive data centers. Moreover, the rapid technological advancements in AI have led to an increase in electronic waste (e-waste), as outdated devices and components are discarded. E-waste, containing harmful substances like lead, cadmium, and mercury, poses significant risks to both human health and the environment if not properly disposed of (Ueda et al., 2024b). These environmental challenges—carbon emissions, resource extraction, energy consumption, and e-waste—highlight the need for sustainable AI practices in healthcare to balance technological innovation with environmental responsibility (Fig. 8).

14. Conclusion

In conclusion, while AI holds significant promise for transforming ophthalmology, its integration into clinical practice is far from straightforward. Several challenges persist, including bias, data privacy concerns, lack of standardization in validation procedures, and the risk of over-reliance on automated systems. The increasing use of synthetic images—although useful to reduce dependence on real datasets—may limit the model's exposure to the complexity of real-life clinical findings, raising concerns about generalizability. These issues highlight the urgency of establishing strong regulatory and ethical frameworks. Responsible implementation will require close collaboration among developers, regulators, clinicians, and policy makers to ensure that AI tools are not only effective, but also safe, transparent, and equitable for all patients.

CRediT authorship contribution statement

Maria Cristina Savastano: Writing – review & editing, Writing – original draft, Conceptualization. **Clara Rizzo:** Writing – review & editing, Writing – original draft. **Claudia Fossataro:** Writing – review & editing. **Daniela Bacherini:** Writing – review & editing, Data curation. **Fabrizio Giansanti:** Writing – review & editing. **Alfonso Savastano:** Supervision. **Giovanni Arcuri:** Writing – review & editing, Data curation. **Stanislao Rizzo:** Writing – review & editing, Supervision. **Francesco Faraldi:** Writing – review & editing, Supervision.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of conflicting interests

None of the authors has any conflict of interest to declare.

Data availability

Data will be made available on request.

References

Abdullah, Y.I., Schuman, J.S., Shabsigh, R., Caplan, A., Al-Aswad, L.A., 2021. Ethics of artificial intelligence in medicine and ophthalmology. Asia-Pac. J. Ophthalmol. 10, 289–298. https://doi.org/10.1097/APO.0000000000000397.

Abràmoff, M.D., Tobey, D., Char, D.S., 2020. Lessons learned about autonomous AI: finding a safe, efficacious, and ethical path through the development process. Am. J. Ophthalmol. 214, 134–142. https://doi.org/10.1016/j.ajo.2020.02.022.

Ahmed, M.I., Spooner, B., Isherwood, J., Lane, M., Orrock, E., Dennison, A., 2023. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. Cureus. https://doi.org/10.7759/cureus.46454.

Akram, M.U., Abdul Salam, A., Khawaja, S.G., Naqvi, S.G.H., Khan, S.A., 2020. RIDB: a Dataset of fundus images for retina based person identification. Data Brief 33, 106433. https://doi.org/10.1016/j.dib.2020.106433.

Ali, M.J., 2023. ChatGPT and lacrimal drainage disorders: performance and scope of improvement. Ophthalmic Plast. Reconstr. Surg. 39, 221–225. https://doi.org/10.1097/IOP.0000000000002418.

Ali, S., Ravi, P., Williams, R., DiPaola, D., Breazeal, C., 2024. Constructing dreams using generative AI. Proc. AAAI Conf. Artif. Intell. 38, 23268–23275. https://doi.org/10.1609/aaai.v38i21.30374.

Anguita, R., Makuloluwa, A., Hind, J., Wickham, L., 2024. Large language models in vitreoretinal surgery. Eye (Lond. 1990) 38, 809–810. https://doi.org/10.1038/s41433-023-02751-1.

Anton, N., Doroftei, B., Curteanu, S., Catălin, L., Ilie, O.-D., Târcoveanu, F., Bogdănici, C. M., 2022. Comprehensive review on the use of artificial intelligence in ophthalmology and future research directions. Diagnostics 13, 100. https://doi.org/10.3390/diagnostics13010100.

Appelman, Y., Van Rijn, B.B., Ten Haaf, M.E., Boersma, E., Peters, S.A.E., 2015. Sex differences in cardiovascular risk factors and disease prevention. Atherosclerosis 241, 211–218. https://doi.org/10.1016/j.atherosclerosis.2015.01.027.

Bai, P., Beversluis, C., Song, A., Alicea, N., Eisenberg, Y., Layden, B., Scanzera, A., Leifer, A., Musick, H., Chan, R.V.P., 2024. Opportunities to apply human-centered design in health care with artificial intelligence–based screening for diabetic retinopathy. Int. Ophthalmol. Clin. 64, 5–8. https://doi.org/10.1097/IIO.0000000000000531.

Bali, J., Garg, R., Bali, R.T., 2019. Artificial intelligence (AI) in healthcare and biomedical research: why a strong computational/AI bioethics framework is required? Indian J. Ophthalmol. 67, 3–6. https://doi.org/10.4103/ijo.IJO_1292_18.

Bellanda, V.C.F., Santos, M.L.D., Ferraz, D.A., Jorge, R., Melo, G.B., 2024a. Applications of ChatGPT in the diagnosis, management, education, and research of retinal

diseases: a scoping review. Int. J. Retina Vitr. 10, 79. https://doi.org/10.1186/s40942-024-00595-9.
Bellanda, V.C.F., Santos, M.L.D., Ferraz, D.A., Jorge, R., Melo, G.B., 2024b. Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review. Int. J. Retina Vitr. 10, 79. https://doi.org/10.1186/s40942-024-00595-9.
Bengesi, S., El-Sayed, H., Sarker, M.K., Houkpati, Y., Irungu, J., Oladunni, T., 2024. Advancements in generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. IEEE Access 12, 69812–69837. https://doi.org/10.1109/ACCESS.2024.3397775.
Bernstein, I.A., Zhang, Y. (Victor), Govil, D., Majid, I., Chang, R.T., Sun, Y., Shue, A., Chou, J.C., Schehlein, E., Christopher, K.L., Groth, S.L., Ludwig, C., Wang, S.Y., 2023. Comparison of ophthalmologist and Large Language model chatbot responses to online patient eye care questions. JAMA Netw. Open 6, e2330320. https://doi.org/10.1001/jamanetworkopen.2023.30320.
Blasi, Z.D., Harkness, E., Ernst, E., Georgiou, A., Kleijnen, J., 2001. Influence of context effects on health outcomes: a systematic review. The Lancet 357, 757–762. https://doi.org/10.1016/S0140-6736(00)04169-6.
Boyd, A.D., Gonzalez-Guarda, R., Lawrence, K., Patil, C.L., Ezenwa, M.O., O'Brien, E.C., Paek, H., Braciszewski, J.M., Adeyemi, O., Cuthel, A.M., Darby, J.E., Zigler, C.K., Ho, P.M., Faurot, K.R., Staman, K.L., Leigh, J.W., Dailey, D.L., Cheville, A., Del Fiol, G., Knisely, M.R., Grudzen, C.R., Marsolo, K., Richesson, R.L., Schlaeger, J.M., 2023. Potential bias and lack of generalizability in electronic health record data: reflections on health equity from the National Institutes of Health Pragmatic Trials Collaboratory. J. Am. Med. Inf. Assoc. 30, 1561–1566. https://doi.org/10.1093/jamia/ocad115.
Bressler, N.M., 2023. What artificial intelligence chatbots mean for editors, authors, and readers of peer-reviewed ophthalmic literature. JAMA Ophthalmol. 141, 514. https://doi.org/10.1001/jamaophthalmol.2023.1370.
Cappellani, F., Card, K.R., Shields, C.L., Pulido, J.S., Haller, J.A., 2024. Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients. Eye (Lond. 1990) 38, 1368–1373. https://doi.org/10.1038/s41433-023-02906-0.
Caranfa, J.T., Bommakanti, N.K., Young, B.K., Zhao, P.Y., 2023. Accuracy of vitreoretinal disease information from an artificial intelligence chatbot. JAMA Ophthalmol. 141, 906. https://doi.org/10.1001/jamaophthalmol.2023.3314.
Cen, L.-P., Ji, J., Lin, J.-W., Ju, S.-T., Lin, H.-J., Li, T.-P., Wang, Yun, Yang, J.-F., Liu, Y.-F., Tan, S., Tan, L., Li, D., Wang, Yifan, Zheng, D., Xiong, Y., Wu, H., Jiang, J., Wu, Z., Huang, D., Shi, T., Chen, B., Yang, J., Zhang, X., Luo, L., Huang, C., Zhang, G., Huang, Y., Ng, T.K., Chen, H., Chen, W., Pang, C.P., Zhang, M., 2021. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nat. Commun. 12, 4828. https://doi.org/10.1038/s41467-021-25138-w.
Chen, E.M., Chen, D., Chilakamarri, P., Lopez, R., Parikh, R., 2021. Economic challenges of artificial intelligence adoption for diabetic retinopathy. Ophthalmology 128, 475–477. https://doi.org/10.1016/j.ophtha.2020.07.043.
Chen, J.S., Reddy, A.J., Al-Sharif, E., Shoji, M.K., Kalaw, F.G.P., Eslani, M., Lang, P.Z., Arya, M., Koretz, Z.A., Bolo, K.A., Arnett, J.J., Roginiel, A.C., Do, J.L., Robbins, S.L., Camp, A.S., Scott, N.L., Rudell, J.C., Weinreb, R.N., Baxter, S.L., Granet, D.B., 2025. Analysis of ChatGPT responses to ophthalmic cases: can ChatGPT think like an ophthalmologist? Ophthalmol. Sci. 5, 100600. https://doi.org/10.1016/j.xops.2024.100600.
Chiang, M.F., Sommer, A., Rich, W.L., Lum, F., Parke, D.W., 2018. The 2016 American academy of ophthalmology IRIS® registry (intelligent research in sight) database. Ophthalmology 125, 1143–1148. https://doi.org/10.1016/j.ophtha.2017.12.001.
Chotcomwongse, P., Ruamviboonsuk, P., Grzybowski, A., 2024. Utilizing Large Language Models in ophthalmology: the current landscape and challenges. Ophthalmol. Ther. 13, 2543–2558. https://doi.org/10.1007/s40123-024-01018-6.
Crincoli, E., Sacconi, R., Querques, G., 2023. Reshaping the use of artificial intelligence in Ophthalmology: sometimes you need to go backwards. Retina. https://doi.org/10.1097/IAE.0000000000003878.
Dave, T., Athaluri, S.A., Singh, S., 2023. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front. Artif. Intell. 6, 1169595. https://doi.org/10.3389/frai.2023.1169595.
Elsahar, H., Gallé, M., 2019. To annotate or not? Predicting performance drop under domain shift. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Presented at the Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 2163–2173. https://doi.org/10.18653/v1/D19-1222.
Evans, N.G., Wenner, D.M., Cohen, I.G., Purves, D., Chiang, M.F., Ting, D.S.W., Lee, A.Y., 2022a. Emerging ethical considerations for the use of artificial intelligence in ophthalmology. Ophthalmol. Sci. 2. https://doi.org/10.1016/j.xops.2022.100141.
Evans, N.G., Wenner, D.M., Cohen, I.G., Purves, D., Chiang, M.F., Ting, D.S.W., Lee, A.Y., 2022b. Emerging ethical considerations for the use of artificial intelligence in ophthalmology. Ophthalmol. Sci. 2, 100141. https://doi.org/10.1016/j.xops.2022.100141.
Faes, L., Liu, X., Wagner, S.K., Fu, D.J., Balaskas, K., Sim, D.A., Bachmann, L.M., Keane, P.A., Denniston, A.K., 2020. A clinician's guide to artificial intelligence: how to critically appraise machine learning studies. Transl. Vis. Sci. Technol. 9, 7. https://doi.org/10.1167/tvst.9.2.7.
Feng, X., Xu, K., Luo, M.-J., Chen, H., Yang, Y., He, Q., Song, C., Li, R., Wu, Y., Wang, H., Tham, Y.C., Ting, D.S.W., Lin, H., Wong, T.Y., Lam, D.S., 2024. Latest developments of generative artificial intelligence and applications in ophthalmology. Asia-Pac. J. Ophthalmol. 13, 100090. https://doi.org/10.1016/j.apjo.2024.100090.
Ferro Desideri, L., Roth, J., Zinkernagel, M., Anguita, R., 2023. Application and accuracy of artificial intelligence-derived large language models in patients with age related macular degeneration. Int. J. Retina Vitr 9, 71. https://doi.org/10.1186/s40942-023-00511-7.
Frank-Publig, S., Birner, K., Riedl, S., Reiter, G.S., Schmidt-Erfurth, U., 2024. Artificial intelligence in assessing progression of age-related macular degeneration. Eye (Lond. 1990). https://doi.org/10.1038/s41433-024-03460-z.
González-Gonzalo, C., Thee, E.F., Klaver, C.C.W., Lee, A.Y., Schlingemann, R.O., Tufail, A., Verbraak, F., Sánchez, C.I., 2022. Trustworthy AI: closing the gap between development and integration of AI systems in ophthalmic practice. Prog. Retin. Eye Res. 90, 101034. https://doi.org/10.1016/j.preteyeres.2021.101034.
Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. https://doi.org/10.48550/ARXIV.1406.2661.
Greenland, S., Mansournia, M.A., Altman, D.G., 2016. Sparse data bias: a problem hiding in plain sight. BMJ i1981. https://doi.org/10.1136/bmj.i1981.
Grosse, K., Lee, T., Biggio, B., Park, Y., Backes, M., Molloy, I., 2022. Backdoor smoothing: demystifying backdoor attacks on deep neural networks. Comput. Secur. 120, 102814. https://doi.org/10.1016/j.cose.2022.102814.
Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J.L., Webster, D.R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316, 2402. https://doi.org/10.1001/jama.2016.17216.
Gurnani, B., Kaur, K., 2024. Leveraging ChatGPT for ophthalmic education: a critical appraisal. Eur. J. Ophthalmol. 34, 323–327. https://doi.org/10.1177/11206721231215862.
Gurnani, B., Kaur, K., Morya, A.K., 2023. Adoption, implementation, definitions, and future of blockchain technology in ophthalmology. Indian J. Ophthalmol. 71, 1025–1026. https://doi.org/10.4103/ijo.IJO_1802_22.
Hashemian, H., Peto, T., Ambrósio Jr, R., Lengyel, I., Kafieh, R., Muhammed Noori, A., Khorrami-Nezhad, M., 2024. Application of artificial intelligence in ophthalmology: an updated comprehensive review. J. Ophthalmic Vis. Res. 19. https://doi.org/10.18502/jovr.v19i3.15893.
Hemelings, R., Elen, B., Barbosa-Breda, J., Blaschko, M.B., De Boever, P., Stalmans, I., 2021. Deep learning on fundus images detects glaucoma beyond the optic disc. Sci. Rep. 11, 20313. https://doi.org/10.1038/s41598-021-99605-1.
Hine, E., Novelli, C., Taddeo, M., Floridi, L., 2024. Supporting trustworthy AI through machine unlearning. Sci. Eng. Ethics 30, 43. https://doi.org/10.1007/s11948-024-00500-5.
Hira, S., Singh, D., Kim, T.S., Gupta, S., Hager, G., Sikder, S., Vedula, S.S., 2022. Video-based assessment of intraoperative surgical skill. Int. J. Comput. Assist. Radiol. Surg. 17, 1801–1811. https://doi.org/10.1007/s11548-022-02681-5.
How Should AI Be Developed, 2019. Validated, and implemented in patient care? AMA J. Ethics 21, E125–E130. https://doi.org/10.1001/amajethics.2019.125.
Ittarat, M., Cheungpasitporn, W., Chansangpetch, S., 2023. Personalized care in eye health: exploring opportunities, challenges, and the road ahead for chatbots. J. Personalized Med. 13, 1679. https://doi.org/10.3390/jpm13121679.
Jin, K., Yuan, L., Wu, H., Grzybowski, A., Ye, J., 2023. Exploring large language model for next generation of artificial intelligence in ophthalmology. Front. Med. 10, 1291404. https://doi.org/10.3389/fmed.2023.1291404.
Kalaw, F.G.P., Baxter, S.L., 2024. Ethical considerations for large language models in ophthalmology. Curr. Opin. Ophthalmol. 35, 438–446. https://doi.org/10.1097/ICU.0000000000001083.
Karimian, G., Petelos, E., Evers, S.M.A.A., 2022. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. AI Ethics 2, 539–551. https://doi.org/10.1007/s43681-021-00131-7.
Kazi, A.M., Qazi, S.A., Ahsan, N., Khawaja, S., Sameen, F., Saqib, M., Khan Mughal, M.A., Wajidali, Z., Ali, S., Ahmed, R.M., Kalimuddin, H., Rauf, Y., Mahmood, F., Zafar, S., Abbasi, T.A., Khoumbati, K.-U.-R., Abbasi, M.A., Stergioulas, L.K., 2020. Current challenges of digital health interventions in Pakistan: mixed methods analysis. J. Med. Internet Res. 22, e21691. https://doi.org/10.2196/21691.
Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D., 2019. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 17, 195. https://doi.org/10.1186/s12916-019-1426-2.
Khan, S.M., Liu, X., Nath, S., Korot, E., Faes, L., Wagner, S.K., Keane, P.A., Sebire, N.J., Burton, M.J., Denniston, A.K., 2021. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. Lancet Digit. Health 3, e51–e66. https://doi.org/10.1016/S2589-7500(20)30240-5.
Knebel, D., Priglinger, S., Scherer, N., Klaas, J., Siedlecki, J., Schworm, B., 2024. Assessment of ChatGPT in the prehospital management of ophthalmological emergencies – an analysis of 10 fictional case vignettes. Klin. Monatsblätter Augenheilkd. 241, 675–681. https://doi.org/10.1055/a-2149-0447.
Kong, X., Jiang, X., Zhang, B., Yuan, J., Ge, Z., 2022. Latent variable models in the era of industrial big data: extension and beyond. Annu. Rev. Control 54, 167–199. https://doi.org/10.1016/j.arcontrol.2022.09.005.
Korot, E., Wagner, S.K., Faes, L., Liu, X., Huemer, J., Ferraz, D., Keane, P.A., Balaskas, K., 2020a. Will AI replace ophthalmologists? Transl. Vis. Sci. Technol. 9, 2. https://doi.org/10.1167/tvst.9.2.2.
Korot, E., Wagner, S.K., Faes, L., Liu, X., Huemer, J., Ferraz, D., Keane, P.A., Balaskas, K., 2020b. Will AI replace ophthalmologists? Transl. Vis. Sci. Technol. 9, 2. https://doi.org/10.1167/tvst.9.2.2.
Kou, H., Rong, K., Xie, D., Yang, B., 2024. Artificial intelligence generated content, intellectual property rights, and digital tax. https://doi.org/10.2139/ssrn.4910480.

Kras, A., Celi, L.A., Miller, J.B., 2020. Accelerating ophthalmic artificial intelligence research: the role of an open access data repository. Curr. Opin. Ophthalmol. 31, 337–350. https://doi.org/10.1097/ICU.0000000000000678.
Labib, K.M., Ghumman, H., Jain, S., Jarstad, J.S., 2024a. A review of the utility and limitations of artificial intelligence in retinal disorders and pediatric ophthalmology. Cureus. https://doi.org/10.7759/cureus.71063.
Labib, K.M., Ghumman, H., Jain, S., Jarstad, J.S., 2024b. A review of the utility and limitations of artificial intelligence in retinal disorders and pediatric ophthalmology. Cureus. https://doi.org/10.7759/cureus.71063.
Labib, K.M., Ghumman, H., Jain, S., Jarstad, J.S., 2024c. Applications of artificial intelligence in ophthalmology: glaucoma, cornea, and oculoplastics. Cureus. https://doi.org/10.7759/cureus.73522.
Laï, M.-C., Brian, M., Mamzer, M.-F., 2020. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J. Transl. Med. 18, 14. https://doi.org/10.1186/s12967-019-02204-y.
Lang, M., Bernier, A., Knoppers, B.M., 2022. Artificial intelligence in cardiovascular imaging: "unexplainable" legal and ethical challenges? Can. J. Cardiol. 38, 225–233. https://doi.org/10.1016/j.cjca.2021.10.009.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Leiderman, Y.I., Gerber, M.J., Hubschman, J.-P., Yi, D., 2024a. Artificial intelligence applications in ophthalmic surgery. Curr. Opin. Ophthalmol. 35, 526–532. https://doi.org/10.1097/ICU.0000000000001033.
Leiderman, Y.I., Gerber, M.J., Hubschman, J.-P., Yi, D., 2024b. Artificial intelligence applications in ophthalmic surgery. Curr. Opin. Ophthalmol. 35, 526–532. https://doi.org/10.1097/ICU.0000000000001033.
Li, Z., Wang, L., Wu, X., Jiang, J., Qiang, W., Xie, H., Zhou, H., Wu, S., Shao, Y., Chen, W., 2023a. Artificial intelligence in ophthalmology: the path to the real-world clinic. Cell Rep. Med. 4, 101095. https://doi.org/10.1016/j.xcrm.2023.101095.
Li, Z., Wang, L., Wu, X., Jiang, J., Qiang, W., Xie, H., Zhou, H., Wu, S., Shao, Y., Chen, W., 2023b. Artificial intelligence in ophthalmology: the path to the real-world clinic. Cell Rep. Med. 4, 101095. https://doi.org/10.1016/j.xcrm.2023.101095.
Liu, T.Y.A., Bressler, N.M., 2020. Controversies in artificial intelligence. Curr. Opin. Ophthalmol. 31, 324–328. https://doi.org/10.1097/ICU.0000000000000694.
Lo, J., Yu, T.T., Ma, D., Zang, P., Owen, J.P., Zhang, Q., Wang, R.K., Beg, M.F., Lee, A.Y., Jia, Y., Sarunic, M.V., 2021. Federated learning for microvasculature segmentation and diabetic retinopathy classification of OCT data. Ophthalmol. Sci. 1, 100069. https://doi.org/10.1016/j.xops.2021.100069.
Lyons, R.J., Arepalli, S.R., Fromal, O., Choi, J.D., Jain, N., 2024a. Artificial intelligence chatbot performance in triage of ophthalmic conditions. Can. J. Ophthalmol. 59, e301–e308. https://doi.org/10.1016/j.jcjo.2023.07.016.
Lyons, R.J., Arepalli, S.R., Fromal, O., Choi, J.D., Jain, N., 2024b. Artificial intelligence chatbot performance in triage of ophthalmic conditions. Can. J. Ophthalmol. 59, e301–e308. https://doi.org/10.1016/j.jcjo.2023.07.016.
Mai, J., Schmidt-Erfurth, U., 2024. Role of artificial intelligence in retinal diseases. Klin. Monatsbl. Augenheilkd. 241, 1023–1031. https://doi.org/10.1055/a-2378-6138.
Marey, A., Arjmand, P., Alerab, A.D.S., Eslami, M.J., Saad, A.M., Sanchez, N., Umair, M., 2024. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt. J. Radiol. Nucl. Med. 55, 183. https://doi.org/10.1186/s43055-024-01356-2.
Marmot, M., Bell, R., 2012. Fair society, healthy lives. Public Health 126, S4–S10. https://doi.org/10.1016/j.puhe.2012.05.014.
Martins, T.G.D.S., Schor, P., Mendes, L.G.A., Fowler, S., Silva, R., 2022. Use of artificial intelligence in ophthalmology: a narrative review. Sao Paulo Med. J. 140, 837–845. https://doi.org/10.1590/1516-3180.2021.0713.r1.22022022.
McCarthy, J., Minsky, M., Rochester, N., Shannon, C., 2006. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI Mag. 27, 12–14.
Meyer, I.H., 2003. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol. Bull. 129, 674–697. https://doi.org/10.1037/0033-2909.129.5.674.
Muralidharan, V., Burgart, A., Daneshjou, R., Rose, S., 2023. Recommendations for the use of pediatric data in artificial intelligence and machine learning ACCEPT-AI. npj Digit. Med. 6, 166. https://doi.org/10.1038/s41746-023-00898-5.
Nair, P.P., Keskar, M., Borghare, P.T., Methwani, D.A., Nasre, Y., Chaudhary, M., 2024. Artificial intelligence in dry eye disease: a narrative review. Cureus. https://doi.org/10.7759/cureus.70056.
Nakayama, L.F., Kras, A., Ribeiro, L.Z., Malerbi, F.K., Mendonça, L.S., Celi, L.A., Regatieri, C.V.S., Waheed, N.K., 2022. Global disparity bias in ophthalmology artificial intelligence applications. BMJ Health Care Inform 29, e100470. https://doi.org/10.1136/bmjhci-2021-100470.
Nakayama, L.F., Matos, J., Quion, J., Novaes, F., Mitchell, W.G., Mwavu, R., Hung, C.J.-Y.J., Santiago, A.P.D., Phanphruk, W., Cardoso, J.S., Celi, L.A., 2024. Unmasking biases and navigating pitfalls in the ophthalmic artificial intelligence lifecycle: a narrative review. PLOS Digit. Health 3, e0000618. https://doi.org/10.1371/journal.pdig.0000618.
Nanji, K., Yu, C.W., Wong, T.Y., Sivaprasad, S., Steel, D.H., Wykoff, C.C., Chaudhary, V., 2024. Evaluation of postoperative ophthalmology patient instructions from ChatGPT and Google Search. Can. J. Ophthalmol. 59, e69–e71. https://doi.org/10.1016/j.jcjo.2023.10.001.
Nguyen, Q., Woof, W., Kabiri, N., Sen, S., Daich Varela, M., Cabral De Guimaraes, T.A., Shah, M., Sumodhee, D., Moghul, I., Al-Khuzaei, S., Liu, Y., Hollyhead, C., Tailor, B., Lobo, L., Veal, C., Archer, S., Furman, J., Arno, G., Gomes, M., Fujinami, K., Madhusudhan, S., Mahroo, O.A., Webster, A.R., Balaskas, K., Downes, S.M., Michaelides, M., Pontikos, N., 2023. Can artificial intelligence accelerate the diagnosis of inherited retinal diseases? Protocol for a data-only retrospective cohort study (Eye2Gene). BMJ Open 13, e071043. https://doi.org/10.1136/bmjopen-2022-071043.
Nguyen, T.X., Ran, A.R., Hu, X., Yang, D., Jiang, M., Dou, Q., Cheung, C.Y., 2022. Federated learning in ocular imaging: current progress and future direction. Diagnostics 12, 2835. https://doi.org/10.3390/diagnostics12112835.
Ning, Y., Liu, X., Collins, G.S., Moons, K.G.M., McCradden, M., Ting, D.S.W., Ong, J.C.L., Goldstein, B.A., Wagner, S.K., Keane, P.A., Topol, E.J., Liu, N., 2024. An ethics assessment tool for artificial intelligence implementation in healthcare: care-ai. Nat. Med. 30, 3038–3039. https://doi.org/10.1038/s41591-024-03310-1.
Norori, N., Hu, Q., Aellen, F.M., Faraci, F.D., Tzovara, A., 2021. Addressing bias in big data and AI for health care: a call for open science. Patterns 2, 100347. https://doi.org/10.1016/j.patter.2021.100347.
Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Sci. Technol. Humanit. 366, 447–453. https://doi.org/10.1126/science.aax2342.
O'Shea, K., Nash, R., 2015. An Introduction to Convolutional Neural Networks. ArXiv E-Prints.
Patil, N.S., Huang, R., Mihalache, A., Kisilevsky, E., Kwok, J., Popovic, M.M., Nassrallah, G., Chan, C., Mallipatna, A., Kertes, P.J., Muni, R.H., 2024. The ability of artificial intelligence chatbots ChatGPT and Google Bard to accurately convey preoperative information for patients undergoing ophthalmological surgeries. Retina. https://doi.org/10.1097/IAE.0000000000004044.
Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., Webster, D.R., 2018. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat. Biomed. Eng. 2, 158–164. https://doi.org/10.1038/s41551-018-0195-0.
Quaranta, M., Amantea, I., Grosso, M., 2023. Obligation for AI systems in healthcare: prepare for trouble and make it double? Rev. Socionetwork Strateg. 17. https://doi.org/10.1007/s12626-023-00145-z.
Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G., Chin, M.H., 2018. Ensuring fairness in machine learning to advance health equity. Ann. Intern. Med. 169, 866–872. https://doi.org/10.7326/M18-1990.
Ramos-Cruz, B., Andreu-Perez, J., Martínez, L., 2024. The cybersecurity mesh: a comprehensive survey of involved artificial intelligence methods, cryptographic protocols and challenges for future research. Neurocomputing 581, 127427. https://doi.org/10.1016/j.neucom.2024.127427.
Resnikoff, S., Felch, W., Gauthier, T.-M., Spivey, B., 2012. The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners. Br. J. Ophthalmol. 96, 783–787. https://doi.org/10.1136/bjophthalmol-2011-301378.
Richie, C., 2022. Environmentally sustainable development and use of artificial intelligence in health care. Bioethics 36, 547–555. https://doi.org/10.1111/bioe.13018.
Rojas-Carabali, W., Sen, A., Agarwal, A., Tan, G., Cheung, C.Y., Rousselot, A., Agrawal, Rajdeep, Liu, R., Cifuentes-González, C., Elze, T., Kempen, J.H., Sobrin, L., Nguyen, Q.D., de-la-Torre, A., Lee, B., Gupta, V., Agrawal, Rupesh, 2024. Chatbots vs. Human experts: evaluating diagnostic performance of chatbots in uveitis and the perspectives on AI adoption in ophthalmology. Ocul. Immunol. Inflamm. 32, 1591–1598. https://doi.org/10.1080/09273948.2023.2266730.
Roland, T., Böck, C., Tschoellitsch, T., Maletzky, A., Hochreiter, S., Meier, J., Klambauer, G., 2022. Domain shifts in machine learning based covid-19 diagnosis from blood tests. J. Med. Syst. 46, 23. https://doi.org/10.1007/s10916-022-01807-1.
Ruamviboonsuk, P., Chantra, S., Seresirikachorn, K., Ruamviboonsuk, V., Sangroongruangsri, S., 2021. Economic evaluations of artificial intelligence in ophthalmology. Asia-Pac. J. Ophthalmol. 10, 307–316. https://doi.org/10.1097/APO.0000000000000403.
Sarker, I.H., 2023. Multi-aspects AI -based modeling and adversarial learning for cybersecurity intelligence and robustness: a comprehensive overview. Secur. Priv. 6, e295. https://doi.org/10.1002/spy2.295.
Schmidt, R., 2019. Recurrent neural networks (RNNs): a gentle introduction and overview. https://doi.org/10.48550/arXiv.1912.05911.
Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B.S., Waldstein, S.M., Bogunović, H., 2018. Artificial intelligence in retina. Prog. Retin. Eye Res. 67, 1–29. https://doi.org/10.1016/j.preteyeres.2018.07.004.
Shi, M., Jiang, R., Hu, X., Shang, J., 2020. A privacy protection method for health care big data management based on risk access control. Health Care Manag. Sci. 23, 427–442. https://doi.org/10.1007/s10729-019-09490-4.
Smith, H., Fotheringham, K., 2020. Artificial intelligence in clinical decision-making: rethinking liability. Med. Law Int. 20, 131–154. https://doi.org/10.1177/0968533220945766.
Smith, M., Heath Jeffery, R.C., 2020. Addressing the challenges of artificial intelligence in medicine. Intern. Med. J. 50, 1278–1281. https://doi.org/10.1111/imj.15017.
Sonmez, S.C., Sevgi, M., Antaki, F., Huemer, J., Keane, P.A., 2024a. Generative artificial intelligence in ophthalmology: current innovations, future applications and challenges. Br. J. Ophthalmol. 108, 1335–1340. https://doi.org/10.1136/bjo-2024-325458.
Sonmez, S.C., Sevgi, M., Antaki, F., Huemer, J., Keane, P.A., 2024b. Generative artificial intelligence in ophthalmology: current innovations, future applications and challenges. Br. J. Ophthalmol. 108, 1335–1340. https://doi.org/10.1136/bjo-2024-325458.
Sonmez, S.C., Sevgi, M., Antaki, F., Huemer, J., Keane, P.A., 2024c. Generative artificial intelligence in ophthalmology: current innovations, future applications and challenges. Br. J. Ophthalmol. 108, 1335–1340. https://doi.org/10.1136/bjo-2024-325458.

Sovacool, B.K., Ali, S.H., Bazilian, M., Radley, B., Nemery, B., Okatz, J., Mulvaney, D., 2020. Sustainable minerals and metals for a low-carbon future. Sci. Technol. Humanit. 367, 30–33. https://doi.org/10.1126/science.aaz6003.

Stacke, K., Eilertsen, G., Unger, J., Lundstrom, C., 2021. Measuring domain shift for deep learning in histopathology. IEEE J. Biomed. Healthc. Inf. 25, 325–336. https://doi.org/10.1109/JBHI.2020.3032060.

Stead, W.W., 2018. Clinical implications and challenges of artificial intelligence and deep learning. JAMA 320, 1107. https://doi.org/10.1001/jama.2018.11029.

Swaminathan, U., Daigavane, S., 2024a. Unveiling the potential: a comprehensive review of artificial intelligence applications in ophthalmology and future prospects. Cureus. https://doi.org/10.7759/cureus.61826.

Swaminathan, U., Daigavane, S., 2024b. Unveiling the potential: a comprehensive review of artificial intelligence applications in ophthalmology and future prospects. Cureus. https://doi.org/10.7759/cureus.61826.

Swaminathan, U., Daigavane, S., 2024c. Unveiling the potential: a comprehensive review of artificial intelligence applications in ophthalmology and future prospects. Cureus. https://doi.org/10.7759/cureus.61826.

Tabuchi, H., Morita, S., Miki, M., Deguchi, H., Kamiura, N., 2022. Real-time artificial intelligence evaluation of cataract surgery: a preliminary study on demonstration experiment. Taiwan J. Ophthalmol. 12, 147–154. https://doi.org/10.4103/tjo.tjo_5_22.

Taherdoost, H., Le, T.-V., Slimani, K., 2025. Cryptographic techniques in artificial intelligence security: a bibliometric review. Cryptography 9, 17. https://doi.org/10.3390/cryptography9010017.

Tampu, I.E., Eklund, A., Haj-Hosseini, N., 2022. Inflation of test accuracy due to data leakage in deep learning-based classification of OCT images. Sci. Data 9, 580. https://doi.org/10.1038/s41597-022-01618-6.

Thaler, A., Ong, J., Al-Aswad, L.A., 2024. Upholding artificial intelligence transparency in ophthalmology: a call for collaboration between academia, industry, and government for patient care in the 21st century. Asia-Pac. J. Ophthalmol. 13, 100093. https://doi.org/10.1016/j.apjo.2024.100093.

Tham, Y.-C., Goh, J.H.L., Anees, A., Lei, X., Rim, T.H., Chee, M.-L., Wang, Y.X., Jonas, J. B., Thakur, S., Teo, Z.L., Cheung, N., Hamzah, H., Tan, G.S.W., Husain, R., Sabanayagam, C., Wang, J.J., Chen, Q., Lu, Z., Keenan, T.D., Chew, E.Y., Tan, A.G., Mitchell, P., Goh, R.S.M., Xu, X., Liu, Y., Wong, T.Y., Cheng, C.-Y., 2022. Detecting visually significant cataract using retinal photograph-based deep learning. Nat. Aging 2, 264–271. https://doi.org/10.1038/s43587-022-00171-6.

Ting, D.S.W., Cheung, C.Y.-L., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I.Y., Lee, S.Y., Wong, E.Y.M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N.C., Finkelstein, E.A., Lamoureux, E.L., Wong, I.Y., Bressler, N.M., Sivaprasad, S., Varma, R., Jonas, J.B., He, M.G., Cheng, C.-Y., Cheung, G.C.M., Aung, T., Hsu, W., Lee, M.L., Wong, T.Y., 2017. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 318, 2211. https://doi.org/10.1001/jama.2017.18152.

Ting, D.S.W., Peng, L., Varadarajan, A.V., Keane, P.A., Burlina, P.M., Chiang, M.F., Schmetterer, L., Pasquale, L.R., Bressler, N.M., Webster, D.R., Abramoff, M., Wong, T.Y., 2019. Deep learning in ophthalmology: the technical and clinical considerations. Prog. Retin. Eye Res. 72, 100759. https://doi.org/10.1016/j.preteyeres.2019.04.003.

Tom, E., Keane, P.A., Blazes, M., Pasquale, L.R., Chiang, M.F., Lee, A.Y., Lee, C.S., AAO Artificial Intelligence Task Force, 2020. Protecting data privacy in the age of AI-enabled ophthalmology. Transl. Vis. Sci. Technol. 9, 36. https://doi.org/10.1167/tvst.9.2.36.

Tonti, E., Tonti, S., Mancini, F., Bonini, C., Spadea, L., D'Esposito, F., Gagliano, C., Musa, M., Zeppieri, M., 2024a. Artificial intelligence and advanced technology in glaucoma: a review. J. Personalized Med. 14, 1062. https://doi.org/10.3390/jpm14101062.

Tonti, E., Tonti, S., Mancini, F., Bonini, C., Spadea, L., D'Esposito, F., Gagliano, C., Musa, M., Zeppieri, M., 2024b. Artificial intelligence and advanced technology in glaucoma: a review. J. Personalized Med. 14, 1062. https://doi.org/10.3390/jpm14101062.

Tonti, E., Tonti, S., Mancini, F., Bonini, C., Spadea, L., D'Esposito, F., Gagliano, C., Musa, M., Zeppieri, M., 2024c. Artificial intelligence and advanced technology in glaucoma: a review. J. Personalized Med. 14, 1062. https://doi.org/10.3390/jpm14101062.

Tseng, R.M.W.W., Gunasekeran, D.V., Tan, S.S.H., Rim, T.H., Lum, E., Tan, G.S.W., Wong, T.Y., Tham, Y.-C., 2021a. Considerations for artificial intelligence real-world implementation in ophthalmology: providers' and patients' perspectives. Asia-Pac. J. Ophthalmol. 10, 299–306. https://doi.org/10.1097/APO.0000000000000400.

Tseng, R.M.W.W., Gunasekeran, D.V., Tan, S.S.H., Rim, T.H., Lum, E., Tan, G.S.W., Wong, T.Y., Tham, Y.-C., 2021b. Considerations for artificial intelligence real-world implementation in ophthalmology: providers' and patients' perspectives. Asia-Pac. J. Ophthalmol. 10, 299–306. https://doi.org/10.1097/APO.0000000000000400.

Tshimula, J.M., Ndona, X., Nkashama, D.K., Tardif, P.-M., Kabanza, F., Frappier, M., Wang, S., 2024. Preventing jailbreak prompts as malicious tools for cybercriminals: a cyber defense perspective. https://doi.org/10.48550/ARXIV.2411.16642.

Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., Matsui, Y., Nozaki, T., Nakaura, T., Fujima, N., Tatsugami, F., Yanagawa, M., Hirata, K., Yamada, A., Tsuboyama, T., Kawamura, M., Fujioka, T., Naganawa, S., 2024a. Fairness of artificial intelligence in healthcare: review and recommendations. Jpn. J. Radiol. 42, 3–15. https://doi.org/10.1007/s11604-023-01474-3.

Ueda, D., Walston, S.L., Fujita, S., Fushimi, Y., Tsuboyama, T., Kamagata, K., Yamada, A., Yanagawa, M., Ito, R., Fujima, N., Kawamura, M., Nakaura, T., Matsui, Y., Tatsugami, F., Fujioka, T., Nozaki, T., Hirata, K., Naganawa, S., 2024b. Climate change and artificial intelligence in healthcare: review and recommendations towards a sustainable future. Diagn. Interv. Imaging 105, 453–459. https://doi.org/10.1016/j.diii.2024.06.002.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł.U., Polosukhin, I., 2017. Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Veritti, D., Rubinato, L., Sarao, V., De Nardin, A., Foresti, G.L., Lanzetta, P., 2024. Behind the mask: a critical perspective on the ethical, moral, and legal implications of AI in ophthalmology. Graefes Arch. Clin. Exp. Ophthalmol. 262, 975–982. https://doi.org/10.1007/s00417-023-06245-4.

Wahl, B., Cossy-Gantner, A., Germann, S., Schwalbe, N.R., 2018. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob. Health 3, e000798. https://doi.org/10.1136/bmjgh-2018-000798.

Waisberg, E., Ong, J., Zaman, N., Kamran, S.A., Sarker, P., Tavakkoli, A., Lee, A.G., 2023. GPT-4 for triaging ophthalmic symptoms. Eye (Lond. 1990) 37, 3874–3875. https://doi.org/10.1038/s41433-023-02595-9.

Wang, T., Xia, J., Li, R., Wang, R., Stanojcic, N., Li, J.-P.O., Long, E., Wang, J., Zhang, X., Li, J., Wu, X., Liu, Z., Chen, J., Chen, H., Nie, D., Ni, H., Chen, R., Chen, W., Yin, S., Lin, D., Yan, P., Xia, Z., Lin, S., Huang, K., Lin, H., 2022. Intelligent cataract surgery supervision and evaluation via deep learning. Int. J. Surg. 104, 106740. https://doi.org/10.1016/j.ijsu.2022.106740.

Wang, Y., Liu, C., Zhou, K., Zhu, T., Han, X., 2024. Towards regulatory generative AI in ophthalmology healthcare: a security and privacy perspective. Br. J. Ophthalmol. 108, 1349–1353. https://doi.org/10.1136/bjo-2024-325167.

Warnat-Herresthal, S., Schultze, H., Shastry, K.L., Manamohan, S., Mukherjee, Saikat, Garg, V., Sarveswara, R., Händler, K., Pickkers, P., Aziz, N.A., Ktena, S., Tran, F., Bitzer, M., Ossowski, S., Casadei, N., Herr, C., Petersheim, D., Behrends, U., Kern, F., Fehlmann, T., Schommers, P., Lehmann, C., Augustin, M., Rybniker, J., Altmüller, J., Mishra, N., Bernardes, J.P., Krämer, B., Bonaguro, L., Schulte-Schrepping, J., De Domenico, E., Siever, C., Kraut, M., Desai, M., Monnet, B., Saridaki, M., Siegel, C.M., Drews, A., Nuesch-Germano, M., Theis, H., Heyckendorf, J., Schreiber, S., Kim-Hellmuth, S., , COVID-19 Aachen Study (COVAS), Balfanz, P., Eggermann, T., Boor, P., Hausmann, R., Kuhn, H., Isfort, S., Stingl, J.C., Schmalzing, G., Kuhl, C.K., Röhrig, R., Marx, G., Uhlig, S., Dahl, E., Müller-Wieland, D., Dreher, M., Marx, N., Nattermann, J., Skowasch, D., Kurth, I., Keller, A., Bals, R., Nürnberg, P., Rieß, O., Rosenstiel, P., Netea, M.G., Theis, F., Mukherjee, Sach, Backes, M., Aschenbrenner, A.C., Ulas, T., , Deutsche COVID-19 Omics Initiative (DeCOI), Angelov, A., Bartholomäus, A., Becker, A., Bezdan, D., Blumert, C., Bonifacio, E., Bork, P., Boyke, B., Blum, H., Clavel, T., Colome-Tatche, M., Cornberg, M., De La Rosa Velázquez, I.A., Diefenbach, A., Dilthey, A., Fischer, N., Förstner, K., Franzenburg, S., Frick, J.-S., Gabernet, G., Gagneur, J., Ganzenmueller, T., Gauder, M., Geißert, J., Goesmann, A., Göpel, S., Grundhoff, A., Grundmann, H., Hain, T., Hanses, F., Hehr, U., Heimbach, A., Hoeper, M., Horn, F., Hübschmann, D., Hummel, M., Iftner, T., Iftner, A., Illig, T., Janssen, S., Kalinowski, J., Kallies, R., Kehr, B., Keppler, O.T., Klein, C., Knop, M., Kohlbacher, O., Köhrer, K., Korbel, J., Kremsner, P.G., Kühnert, D., Landthaler, M., Li, Y., Ludwig, K.U., Makarewicz, O., Marz, M., McHardy, A.C., Mertes, C., Münchhoff, M., Nahnsen, S., Nöthen, M., Ntoumi, F., Overmann, J., Peter, S., Pfeffer, K., Pink, I., Poetsch, A.R., Protzer, U., Pühler, A., Rajewsky, N., Ralser, M., Reiche, K., Ripke, S., Da Rocha, U.N., Saliba, A.-E., Sander, L.E., Sawitzki, B., Scheithauer, S., Schiffer, P., Schmid-Burgk, J., Schneider, W., Schulte, E.-C., Sczyrba, A., Sharaf, M.L., Singh, Y., Sonnabend, M., Stegle, O., Stoye, J., Vehreschild, J., Velavan, T.P., Vogel, J., Volland, S., Von Kleist, M., Walker, A., Walter, J., Wieczorek, D., Winkler, S., Ziebuhr, J., Breteler, M. M.B., Giamarellos-Bourboulis, E.J., Kox, M., Becker, M., Cheran, S., Woodacre, M.S., Goh, E.L., Schultze, J.L., 2021. Swarm Learning for decentralized and confidential clinical machine learning. Nature 594, 265–270. https://doi.org/10.1038/s41586-021-03583-3.

Williams, D., Hornung, H., Nadimpalli, A., Peery, A., 2021a. Deep learning and its application for healthcare delivery in low and middle income countries. Front. Artif. Intell. 4, 553987. https://doi.org/10.3389/frai.2021.553987.

Williams, D., Hornung, H., Nadimpalli, A., Peery, A., 2021b. Deep learning and its application for healthcare delivery in low and middle income countries. Front. Artif. Intell. 4, 553987. https://doi.org/10.3389/frai.2021.553987.

Wong, I.N., Monteiro, O., Baptista-Hon, D.T., Wang, K., Lu, W., Sun, Z., Nie, S., Yin, Y., 2024. Leveraging foundation and large language models in medical artificial intelligence. Chin. Med. J. (Engl.). https://doi.org/10.1097/CM9.0000000000003302.

Wong, T.Y., Bressler, N.M., 2016. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. JAMA 316, 2366. https://doi.org/10.1001/jama.2016.17563.

Xie, Y., Gunasekeran, D.V., Balaskas, K., Keane, P.A., Sim, D.A., Bachmann, L.M., Macrae, C., Ting, D.S.W., 2020. Health economic and safety considerations for artificial intelligence applications in diabetic retinopathy screening. Transl. Vis. Sci. Technol. 9, 22. https://doi.org/10.1167/tvst.9.2.22.

Yang, Z., Wang, D., Zhou, F., Song, D., Zhang, Y., Jiang, J., Kong, K., Liu, X., Qiao, Y., Chang, R.T., Han, Y., Li, F., Tham, C.C., Zhang, X., 2024. Understanding natural language: potential application of large language models to ophthalmology. Asia-Pac. J. Ophthalmol. 13, 100085. https://doi.org/10.1016/j.apjo.2024.100085.

Zandi, R., Fahey, J.D., Drakopoulos, M., Bryan, J.M., Dong, S., Bryar, P.J., Bidwell, A.E., Bowen, R.C., Lavine, J.A., Mirza, R.G., 2024. Exploring diagnostic precision and triage proficiency: a comparative study of GPT-4 and bard in addressing common ophthalmic complaints. Bioengineering 11, 120. https://doi.org/10.3390/bioengineering11020120.

Zhao, Z., Liu, Y., Wu, H., Wang, M., Li, Y., Wang, S., Teng, L., Liu, D., Cui, Z., Wang, Q., Shen, D., 2025. CLIP in medical imaging: a survey. Med. Image Anal. 102, 103551. https://doi.org/10.1016/j.media.2025.103551.

Zhou, I., Tofigh, F., Piccardi, M., Abolhasan, M., Franklin, D., Lipman, J., 2024. Secure multi-party computation for machine learning: a survey. IEEE Access 12, 53881–53899. https://doi.org/10.1109/ACCESS.2024.3388992.

Zhu, T., Ye, D., Wang, W., Zhou, W., Yu, P.S., 2022. More than privacy: applying differential privacy in key areas of artificial intelligence. IEEE Trans. Knowl. Data Eng. 34, 2824–2843. https://doi.org/10.1109/TKDE.2020.3014246.

Glossary

(AI): Artificial intelligence
(AIGC): AI-generated content
(AMD): Age-related macular degeneration
(ANNs): Artificial neural networks
(CAM): Class activation mapping
CLIP: (Contrastive Language–Image Pretraining)
(CNNs): Convolutional Neural Networks
(DL): Deep Learning
(DNNs): Deep neural networks
(DR): Diabetic retinopathy
(FDA): Food and Drug Administration -United States
(FL): Federated learning
(GAI): Generative Artificial Intelligence
(GANs): Generative Adversarial Networks
(GDPR): General Data Protection Regulation
(HIPAA): Health Insurance Portability and Accountability Act
(IRDs): Inherited retinal diseases
(IDS/IPS): Intrusion detection and prevention system
(LLMs): Large Language Models
(LVMs): Latent variable models
(LIC): Low-income
(ML): Machine Learning
(LMIC): Middle-income countries
(OCT): Optical coherence tomography
(RNN): Recurrent Neural Networks
(SMPC): Secure multi-party computation
(VAEs): Variational Autoencoders
(ZKPs): Zero-knowledge proofs