AMERICAN ACADEMY OF OPHTHALMOLOGY®

Check for updates

Editorial

Emerging Ethical Considerations for the Use of Artificial Intelligence in Ophthalmology

Nicholas G. Evans, PhD - *Lowell, Massachusetts*
Danielle M. Wenner, PhD - *Pittsburgh, Pennsylvania*
I. Glenn Cohen, JD - *Cambridge, Massachusetts*
Duncan Purves, PhD - *Gainesville, Florida*
Michael F. Chiang, MD - *Bethesda, Maryland*
Daniel S.W. Ting, MD, PhD - *Singapore, Republic of Singapore*
Aaron Y. Lee, MD, MSCI - *Seattle, Washington*

Rapid developments in artificial intelligence (AI) promise improved diagnosis and care for patients, but raise ethical issues.[1–5] Over 6 months, in consultation with the American Academy of Ophthalmology Committee on Artificial Intelligence, we analyzed potential ethical concerns, with a focus on applications of AI in ophthalmology that are deployed or will be deployed in the near future.[6] We identified 3 pressing issues: (1) transparency, paradigmatically through the explanation or interpretation of AI models; (2) attribution of responsibility issues for particular harms arising from the use or misuse of AI; and (3) scalability of use cases and screening infrastructure.

Transparency

The ability to understand why a machine learning model has produced a particular result is an oft-cited ethical principle for AI.[4,5,7–10] We distinguish between AI models that are interpretable, or governed by models that are directly understandable by humans, and AI models that are too complex for any human to comprehend (sometimes called "black box" models), requiring post hoc explainability for how results are produced.[4] Recent work has shown that lack of transparency is associated with decreased accuracy of AI algorithms.[11,12] Issues of transparency may arise, for example, in diagnosing diabetic retinopathy, glaucoma, age-related macular degeneration, and retinopathy of prematurity (ROP).[1]

Transparency also may be important when an AI model does not perform as expected or gives a false answer. Given a novel image to analyze, for example, AI may misdiagnose a patient based on an incomplete or inadequate training set. Machine learning and especially deep learning platforms need to be trained on large amounts of historical data (e.g., fundus photography) to learn which features of an image are associated with a particular condition. When a novel image is presented that is atypical, such as if a diabetic retinopathy AI model is given an image harboring central retinal vein occlusion, the AI model may provide false or even nonsense answers. Without transparency, it may be impossible to explain why a particular failure occurred. Even if the general explanation is that the training set is insufficiently broad, what data are missing or needed may be opaque.

Transparency is arguably secondary to the capacity for AI to improve patient outcomes and public health. Machine learning systems in ophthalmology have been tested, but to date only 1 trial has demonstrated improved patient outcomes.[13] Experiences in other specialties, such as a 2017 trial of using automated interpretation of cardiotocographs in labor, have found no improvement in clinical outcomes as a result of AI.[14] Thus, transparency may be insufficient to justify the use of AI if it fails to improve patient outcomes.

The setting in which medical AI is implemented is a major determinant of the risks and benefits.

The degree to which transparency is obligatory may also depend on the medical specialty. In some cases, accurate, empirically verified results may be sufficient. In infectious disease, for example, broad-spectrum antibiotics may be tried in the absence of detailed information of a pathogen.[8] However, ophthalmology is highly explainable in diagnostic terms, with strict definitions for most diseases. Deferring to AI may present a significant decrease in confidence in the diagnostic process, especially when only modest increases in verifiability are achieved. The degree to which this arises, and how this trade-off between transparency and confidence varies by specialty, needs further investigation.

Lack of sufficient transparency may exacerbate other issues in the use of medical AI. Although human physicians can reflect on and justify their actions to colleagues, an AI model's mistakes are predetermined through training. Errors may propagate from a single point of failure if they become the diagnostic standard across, say, an entire hospital network. Patients may seek a second opinion, but if an algorithm is widely distributed, the same system may be performing the diagnosis at a separate clinic. Future AI models may be able to revise their predictions in response to new data, but this presents its own challenges, especially if these revisions lack transparency. Excessive trust in AI may

https://doi.org/10.1016/j.xops.2022.100141
ISSN 2666-9145/22

be worse for patient outcomes than if AI were approached more skeptically.[15]

Sometimes, the benefits of AI may outweigh transparency concerns. Consider ROP, a leading cause of childhood blindness worldwide. The clinical benefit of screening is well established, but is hampered globally by cost and human labor requirements.[1] Artificial intelligence may provide a low-cost screening option in resource-scarce settings, where even modest improvements in testing and treatment could have a significant impact, given the steep long-term costs of ROP.[16] Although challenges translating diagnosis to treatment in low-income settings remain,[17] the large potential benefits and low cost justify the use of AI.

Explainable AI may obviate some of these transparency concerns. However, Rudin[8] has noted that explainability may be a misnomer. Instead, the focus should be on creating models that are inherently interpretable, rather than attempting to generate solutions for unexplainable AI. For the foreseeable future, then, a tension exists between deploying black-boxed AI immediately or waiting for explainable AI, where delays may come at the cost of improvements to patient outcomes.

Responsibility

Ethical frameworks may distinguish between the responsibility for ensuring AI performs in a certain way and the moral or legal liability when harms occur. Herein, we deal with only ethical responsibilities and not, for example, legal liability, although these are related issues. In health care, a responsibility gap arises when responsibility cannot be easily attributed to 1 or more actors, including hospitals, health and malpractice insurers, individual physicians and nurses, and so on. In ophthalmology, 1 private company, IDx, has accepted responsibility for errors in their AI platform, effectively attempting to close the responsibility gap through claiming responsibility for AI outcomes, enshrining this in legal terms by purchasing liability and malpractice insurance on behalf of the platform.[5]

Companies are responsible for ensuring that AI algorithms function appropriately and safely when used as indicated, but may not be for off-label uses. In their consideration of the legal aspects of AI, for example, IDx claims their principles require that creators "assume liability for harm caused by the diagnostic output of the device when used properly and on-label."[5] Responsibility for ensuring appropriate off-label use thus may seem to fall to the provider, but the fragile nature of these models means even strong associations between patient outcomes and off-label postmarket AI use may be undermined if subtle changes in patient characteristics cause the algorithm to produce flawed results.[13,18] Whether providers can responsibly determine appropriate use based on these unknown variations is unclear.

Responsibility issues may become more acute in future adaptive AI that update their weightings of factors associated with a diagnosis in response to new data. Here, responsibility for appropriate use might include managing which data are retained by the system. For these adaptive regimes, evaluating performance for on-label and off-label conditions will require continuous postmarket monitoring, rather than the current premarket approval approach for pharmaceuticals or other devices.

Allocating responsibility at the level of governance and regulation is an additional challenge. Others have argued that regulation of AI should focus on continuous monitoring[19] with a system view that sees new AI as part of a larger network of actors and institutions and evaluates its performance in the context of that network.[15] The obligation to promote benefits and reduce harms is held jointly by, and distributed between, the creators and users of an AI platform. However, implementing this in practice would require overhauling the institutions that govern medical innovation and practice.

One preliminary approach would require large, adaptive clinical trials of human adjudication versus AI diagnosis. This approach could validate AI performance in a variety of contexts to improve outcomes, adapt to other potential uses, and develop trust in the system. In 2018, engineers at Google demonstrated that image adjudication by retinal specialists improved algorithmic outcomes for the diagnosis of diabetic retinopathy.[20] In the same year, IDx reported that their autonomous AI-based diagnostic platform exceeds human reference standards.[13] Last year, 2 AI-assisted ROP diagnosis packages were approved for use[21] as part of China's developing medical AI landscape.[18] When specialist opinion can be linked to correct surrogate outcomes or risk of poor outcomes, these trials become an intermediate step toward demonstrating the efficacy of AI, improving patient outcomes, enhancing trust, and providing a broader context for AI use.

Scalability and Implementation

One promise of AI is to automate high-volume screening. Consider a near-future hypothetical. In the United Kingdom, the English National Health Service Diabetic Eye Screening Program screened more than 2 million patients from 2015 through 2016 for diabetic retinopathy.[22] We could imagine a case in which this service incorporates AI diagnosis, an implementation that could place most cases of diabetic retinopathy in the country under a single algorithm.

Two failure modes exist for mass AI-driven diagnostics. First, standard errors in diagnostics matter at scale: a sensitivity of 99.9% for a test that applies to a condition affecting hundreds of millions of patients still entails hundreds of thousands of false-negative results.[2] Importantly, transitioning to AI could redistribute false-positive or false-negative results in a population. This raises concerns of justice if, for example, AI misdiagnoses disproportionately impact disadvantaged groups, as has occurred with pulse oximeters[23] and radiograph datasets,[24] resulting in a form of health poverty in which individuals, groups, or populations are unable to benefit from AI because of a scarcity of representative data, and may even be harmed by it at the population level.[25] The degree to which this may occur with ophthalmologic AI applications is an empirical question. However, we do know that racial bias in ophthalmologic clinical trials is an ongoing concern,[26] and this trend could continue into AI development if it remains unchecked.

However, the distribution of harms using AI might be traded against the distribution of services through the deployment of AI, such that:

1. Some patients have worse outcomes than others because of the distribution of risk by AI; yet
2. Those patients have better outcomes than they would otherwise have had because
   a. The AI model is ultimately less biased than physician treatment alone or
   b. The benefits of access to services outweigh the potential harms of bias or
   c. Both.

Consider the proliferation of telemedicine during the COVID-19 pandemic, particularly for individuals who otherwise may experience delayed diagnosis or treatment.[27,28] Artificial intelligence-assisted diagnostics could make it easier to diagnose patients remotely and at local points of care using, for example, new innovations such as slit-lamp biomicroscopes used with smartphones[29] and AI-based interpretation of results. A potential trade-off arises between errors caused by AI when a physician cannot directly access the patient and the benefits of receiving early diagnosis. In rare or emergent cases (such as a pandemic) where the risk of travel to a medical facility presents additional risk, AI may provide preliminary guidance on whether to seek care inside a clinical setting.[30] Moreover, even if AI does produce worse outcomes than a physician diagnosis, AI may be justified to the extent that delayed or missed diagnosis is worse.

However, the social benefit of AI to telemedicine relies in part on the extent to which inequalities of access to information technology can be remedied. Telemedicine is unevenly adopted by providers, may not be supported by insurers, and depends on reliable internet access. However, smartphone penetration may be higher than access to specialist medical care in some if not many areas, and thus favorable tradeoffs may exist through local AI-driven solutions. Like other emerging technologies.

A second failure mode is a systemic failure that affects all or most users simultaneously. These very low-probability, very high-consequence events could arise, for example, in the case of a continual learning AI system intended to improve with additional data,[31] but that, through sustained machine error, ultimately diverges radically from its original parameters and begins assigning false results. Depending on how submissions to the AI platform are structured, adversarial uses could arise in which intentionally doctored images are submitted to achieve the same effect.

Protection from systemic failures is unlikely to be achieved through self-governance and will require regulatory action to guard against. Adding ongoing cyber security and fault tree testing to the approval requirements is 1 solution, but 2 challenges arise. First, premarket regulation typically does entail continuous monitoring of the system; study of results by human analysts and quality control tests against the algorithm to prevent system failures may become dysfunctional on a large scale. Second, the Food and Drug Administration regulates only medical devices, of which IDx is one, but some AI models (such as the Apple Watch pulse oximeter [Apple, Inc]) may constitute a general wellness product designed to be sold directly to consumers.[32] Addressing both challenges may reduce the possibility of low-probability and high-consequence events, but represent trade-offs in system efficiency and resource use around AI in medicine.

In response, the Food and Drug Administration and similar agencies in other countries may require reform to accommodate the challenges presented by AI. Alternatively, the mismatch between the current regulatory structure and the potential impacts of AI in medicine may mean that the Food and Drug Administration is ultimately not well-suited for regulating AI. In the latter case, a new agency may be required, or governance could occur through a different mechanism entirely, for example, through government payment choices in national health insurance schemes.

In conclusion, artificial intelligence presents a range of novel opportunities to improve medical care and to make health care more widely accessible to patients. However, the use of AI raises many ethical concerns, even in cases where it augments the capabilities of human physicians and technicians. These issues are in part endogenous to AI, and are in part a function of the regulatory, social, and political circumstances in which it is developed and implemented. Realizing the full benefits of AI will require reaching a consensus on which trade-offs are acceptable as this technology is implemented at scale.

Acknowledgments

The authors thank the American Academy of Ophthalmology Task Force on Artificial Intelligence, Dr Ron Pelton, and the American Academy of Ophthalmology Ethics Committee for their review and comments on this article.

Footnotes and Disclosures

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): I.G.C.: Consultant – Otsuka, Dawnlight; Bioethics council – Bayer

D.S.W.T.: Patent – Deep learning system for retinal diseases; Equity owner – EyRIS Ptd Ltd.

M.F.C.: Financial support – Novartis, Genentech; Equity owner – InTeleretina, LLC

A.Y.L.: Financial support – Santen, Regeneron, Carl Zeiss Meditec, Novartis, Genentech, Topcon, Verana Health

Supported by the National Science Foundation (grant nos.: 1734521 [N.G.E.], 1917712 [D.P.], and 1622679 [M.F.C.]); the Greenwall Foundation Faculty Scholars Program (N.G.E.); the Davis Educational

Foundation (N.G.E.), the United States Air Force Office of Scientific Research (grant no. FA9550-21-1-0142); the National Eye Institute, National Institutes of Health, Bethesda, Maryland (grant no.: K23EY029246 [A.Y.L.]); National Institutes of Health, Bethesda, Maryland; Research to Prevent Blindness, Inc., New York, New York (A.Y.L.); the United States Food and Drug Administration (A.Y.L.); the Greenwall Foundation Faculty Scholars Program (D.M.W.); the Collaborative Research Program for Biomedical Innovation Law, a scientifically independent collaborative research program supported by the Novo Nordisk Foundation (grant no.: NNF17SA0027784 [I.G.C.]); Diagnosing in the Home: The Ethical, Legal, and Regulatory Challenges and Opportunities of Digital Home Health (I.G.C.); and the Gordon and Betty Moore Foundation (grant no.: 9974 [I.G.C.]). M.F.C. is now supported by the National Eye Institute. The sponsors and funding organizations had no role in the design or conduct of this research.

Daniel S.W. Ting and Aaron Y. Lee are editors of this journal, and were recused from the peer-review process of this article and had no access to information regarding its peer-review.

Correspondence:
Nicholas G. Evans, PhD, 883 Broadway Street, Dugan 200F, Lowell, MA 01853. E-mail: Nicholas_evans@uml.ed.

References

1. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167–175.
2. Lee A. Machine diagnosis. *Nature*. 2019 Apr 10. https://doi.org/10.1038/d41586-019-01112-x. Online ahead of print.
3. Lin D, Lin H. Translating artificial intelligence into clinical practice. *Ann Transl Med*. 2020;8(11):715.
4. Emanuel EJ, Wachter RM. Artificial intelligence in health care: will the value match the hype? *JAMA*. 2019;321(23): 2281–2282.
5. Abràmoff MD, Tobey D, Char DS. Lessons learned about autonomous AI: finding a safe, efficacious, and ethical path through the development process. *Am J Ophthalmol*. 2020;214:134–142.
6. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. 2019;1(9):389–399.
7. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep*. 2019;49(1):15–21.
8. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206–215.
9. Gunning D. Explainable artificial intelligence (XAI). Paper presented at: DARPA/I20 program update; November 2017; Washington DC.
10. Hoffman RR, Klein G, Mueller ST. Explaining explanation for "explainable Ai.". *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2018;62(1):197–201.
11. Floridi L, Cowls J, Beltrametti M, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*. 2018;28(4): 689–707.
12. Shah MP, Merchant SN, Awate SP. Abnormality detection using deep neural networks with robust quasi-norm autoencoding and semi-supervised learning. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Washington, DC.: IEEE; 2018:568–572.
13. Abràmoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1(1):1–8.
14. Brocklehurst P, Field D, Greene K, et al. Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial. *Lancet*. 2017;389(10080):1719–1729.
15. Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ Digit Med*. 2020;3(1):1–4.
16. Frick KD, Mashfeghi DM. A case for universal eye screening. *Retina Today*; 2015. Available at: http://retinatoday.com/2015/06/a-case-for-universal-eye-screening/. Accessed 02.06.20.
17. Brown R, Evans NG. The social value of candidate HIV cures: actualism versus possibilism. *J Med Ethics*. 2017;43(2): 118–123.
18. Li R, Yang Y, Wu S, et al. Using artificial intelligence to improve medical services in China. *Ann Transl Med*. 2020;8(11):711.
19. Babic B, Gerke S, Evgeniou T, Cohen IG. Algorithms on regulatory lockdown in medicine. *Science*. 2019;366(6470): 1202–1204.
20. Beede E, Baylor E, Hersch F, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI: Association for Computing Machinery; 2020: 1–12.
21. CN Pharm Co., Ltd.. New progress in AI research and development. Two Tangwang fundus image aided diagnosis software approved for market [in Korean]. Available at: http://www.cnpharm.com/c/2020-08-11/748522.shtml. Accessed 23.08.20.
22. Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003–2016. *Acta Diabetol*. 2017;54(6): 515–525.
23. Sjoding MW, Dickson RP, Iwashyna TJ, et al. Racial bias in pulse oximetry measurement. *N Engl J Med*. 2020;383(25): 2477–2478.
24. Larrazabal AJ, Nieto N, Peterson V, et al. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci U S A*. 2020;117(23):12592–12594.
25. Ibrahim H, Liu X, Zariffa N, et al. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health*. 2021;3(4):e260–e265.
26. Coney JM. Racial bias in clinical trials: what you need to know. Retina Today. Available at: https://retinatoday.com/articles/2021-mar/racial-bias-in-clinical-trials-what-you-need-to-know. March 2021 Accessed 05.07.21.
27. Evans NG. The ethics of social distancing. *The Philosopher's Magazine*. 2020;89(2):96–103.
28. Saleem SM, Pasquale LR, Sidoti PA, Tsai JC. Virtual ophthalmology: telemedicine in a COVID-19 era. *Am J Ophthalmol*. 2020;216:237–242.
29. Spaide T, Wu Y, Yanagihara RT, et al. Using deep learning to automate Goldmann applanation tonometry readings. *Ophthalmology*. 2020;127(11):1498–1506.
30. Evans NG, Sekkarie MA. Allocating scarce medical resources during armed conflict: ethical issues. *Disaster Mil Med*. 2017;3(1):5. https://doi.org/10.1186/s40696-017-0033-z. eCollection 2017.

31. Lee CS, Lee AY. Clinical applications of continual learning machine learning. *Lancet Digit Health*. 2020;2(6):e279−e281.

32. Minssen T, Gerke S, Aboy M, et al. Regulatory responses to medical machine learning. *J Law Biosci*. 2020;7(1):lsaa002. https://doi.org/10.1093/jlb/lsaa002. eCollection Jan-Dec 2020.