FTC Hearing #7, on Algorithms, Artificial Intelligence, and Predictive Analytics

Panel: Understanding Algorithms, Artificial Intelligence, and Predictive Analytics Through Real World Applications

Comments delivered by Dr. Michael D. Abramoff, MD, PhD

The Robert C. Watzke Professor of Ophthalmology and Visual Sciences, Electrical and Computer Engineering & Biomedical Engineering, University of Iowa

Founder and CEO, IDx.

November 13, 2018

Priority Questions

1. Under what circumstances is it better to use artificial intelligence technologies rather than traditional algorithms, and vice versa? Is the selection of technology generally based on technical considerations or the purpose of the analysis, or are there other practical, policy, or ethical issues that add to the decision? *(Question to panel via email)*

   The use of AI is highly dependent on the intended use and deployment conditions. For specific types of medical applications, AI technology achieve higher performance on cognitively complex tasks than most humans, and there are many algorithmic approaches, including deep learning, that AI can be built with. AI technologies can improve accuracy and performance over traditional hard coded algorithms and perform better on highly complex optimization problems. When selecting AI technology, one must concurrently assess practical, policy, and ethical issues in addition to technical and use case considerations. For example, an AI that has superior accuracy may not address or mitigate unintended or unjustified bias, which presents ethical issues.(1)

2. How accurate are algorithms or artificial intelligence tools used for different purposes? If there is a wide range of accuracy, why is that so? Is accuracy related to the nature of the tool, the type of question being asked, or the data being used? *(Question to panel via email)*

   AI accuracy can be higher than individual experts in comparison to a true reference standard. The range of AI system accuracies is typically the result of the context in which they are measured, though the underlying true accuracy compared to the best possible reference standard is, of course, fixed. For example, measured accuracies differ depending on whether input (such as imaging) conditions are idealized or based on real-world inputs. When AI systems are tested on laboratory data, they may be indistinguishable from a perfect system. But in real-world settings, where there is the potential for human operator error, lower quality inputs, and infinite variance in input data, accuracy (specificity and sensitivity) is typically be lower, especially when compared to the best possible reference standard.

   AI accuracy should be determined based on the highest quality achievable reference standard for the field. In medicine, the reference standard should exhibit the least amount of 'diagnostic drift,' i.e. slowly worsening deviation of the reference standard from the best possible reference standard because it is passed along from clinician to clinician through a noisy process. The reference standard should include at least the same number of modalities than are used as input to the AI and should be aligned with clinical evidence and the preferred practice guidelines, clinical standards, etc. For example, in the case of IDx-DR the reference standard incorporated 3D and OCT based imaging, which is considered best practice, in addition to the fundus-based imaging even though fundus-based imaging is all that was available to the AI

system.

AI accuracy measures must also reflect the population for which its use is indicated. If training data or validation is only specific to a subset of the population, the AI will be brittle (i.e. has limited external validity) and risk the loss of accuracy when scaling to a broader, usually more diverse population for which the system is intended for use. Clinical validation of medical AI systems must show that the system can scale without losing sensitivity and specificity.

3. Your presentations suggest that from technical perspective, it is sometimes impossible to ascertain the basis for a result produced by artificial intelligence, but in other cases, the result is "explainable." What are the advantages and disadvantages of developing technologies for which the basis for the results can or cannot be determined? What criteria should determine when a "black box" system is acceptable, or when a result should be explainable? *(Question to panel via email)*

- We at IDx use the term "explainability" to distinguish medical AI systems designed to align with human clinical decision-making. Autonomous AI in medicine, where the result is not reviewed by a human expert, needs to be safe and trusted by patients and clinicians. In practice, this means autonomous AI systems in medicine must have an "explainable" architecture that is designed to align with decision-making that is familiar to physicians and that is directly tied to clinical evidence.
- The opposite of "explainable" AI is often referred to as "black box". These systems are either not aligned with known clinical decision-making criteria, or are imbued with limited clinical understanding post hoc.
- Under IDx's definition, "explainable" autonomous AI systems in medicine have the advantage of allowing validation of intermediate steps that are made or aggregated on the path to a clinical decision. For example, in looking at retinal images for diabetic retinopathy, an explainable design for autonomous AI in medicine allows developers and clinicians to examine how good the system is at detecting specific biomarkers – such as microaneurysms and exudates. These are the exact same, describable disease features that retinal specialists use for evidence-based clinical decision making.
- In addition, because the AI system is designed to use specific intermediate results, it will be robust to small perturbations in the input vector, that do not affect these intermediate results, thus avoiding the catastrophic failure that plagues non-explainable or post hoc explainable AI.(2-4)
- "Explainability" is part of a broader principle of transparency that is critical to the responsible implementation of AI in healthcare. Transparency isn't just a demand that should be placed upon an autonomous AI system's architecture. It should also, in principle, extend to the data that is used to develop these systems. A medical-grade AI system should be transparent with respect to the training process testing and validation methods. And, of course, transparency is critical for system-level testing, where AI systems are tested in pre-registered clinical trials – on populations and in care settings that aligned with their intended use.

4. What factors have facilitated the development or advancement of these technologies? Have certain resources or policies facilitated their development? *(Question from FTC website)*

- In medical AI, the largest challenge continues to be the scarcity of good quality data, given the ethics, risks and resources involved in obtaining patient data.

- Thus, for a long time, accuracy could not reach an acceptable level for patient safety. Not because we did not have the algorithms, but primarily because the input data was so noisy as it was primarily based on manual entry of doctors' individual impressions.

- With the development of low-cost sensors, such as CMOS, the quality of the input vectors itself increased dramatically, allowing for accuracy as we see it today – and then we realized that the reference standard, usually created by clinicians, was a problem.

- Regarding the resources that have facilitated development of AI, my personal research was supported for many years by NIH and worked closely with the FDA to advance the safety and validation of autonomous AI in medicine.

5. Are there policies or other factors that have impeded the development of these technologies, or might do so in the future? *(Question to panel via email)*

- While FDA authorization of IDx-DR, as the first autonomous AI diagnostic, took many years of discussion, the template is now available for any company, including us, to obtain authorization for autonomous AI.
- The lack of reimbursement models can be a challenge for any new medical technology. However, IDx-DR does have a pathway to payment, and we are currently seeking further coverage clarification from CMS as well as other commercial insurers.
- Interoperability of autonomous AI with electronic health records is crucial for increasing clinic efficiency, and the work by the National Coordinator for Health Information Technology at HHS to improve interoperability and coordinate should be accelerated as much as possible.
- While autonomous AI can iterate faster and easier than traditional medical devices, initial product review should still require a preregistered clinical trial with a "locked" down, deterministic AI that accounts for human factors and system-level validation, such as user training, product labeling, system workflow.

6. How are these technologies affecting competition, innovation, and consumer choices in the industries and business sectors in which they are used today? How might they do so in the future? *(Question from FTC website)*

Thanks to the robust FDA process, the trust in our product is high, which helps us and the competition that follows. We believe transparent and trustworthy AI systems will benefit consumer choices and competition.

In medicine, there are questions about who "owns" the medical data and what are appropriate uses of that data. Training data stewardship requires companies to maintain traceable authorizations and accountability for data privacy and protection. Resolving this will impact both companies and consumers.

Other Questions

1. What features distinguish products or services that use algorithms, artificial intelligence, or predictive analytics? In which industries or business sectors are they most prevalent? *(Question from FTC website)*

   In medicine, autonomous AI has enabled care to be moved from specialty clinics to primary care, where there is increased access for patients. Autonomous AI in healthcare has huge potential to improve outcomes and lower costs.

   While there are many characteristics that distinguish autonomous AI in medicine, we consider the more important question to be "what features of AI <u>align</u> with medical decision-making."

2. What are the advantages and disadvantages for consumers and for businesses of utilizing products or services facilitated by algorithms, artificial intelligence, or predictive analytics? *(Question from FTC website)*

   Carried out in a responsible manner, autonomous AI has the promise of lowering cost and increasing accessibility and quality in healthcare. However, consumers, patients, and physicians cannot make a determination of AI quality without being given visibility to the underlying architecture and rigorous testing.

3. How quickly are these technologies advancing? What are the implications of that pace of technological development from a policy perspective? *(Question from FTC website)*

   Autonomous AI in healthcare is now real, and there are host of ethical, legislative, and regulatory decisions that will need to be made as this technology gains usage, such as bias and data ownership. Furthermore, implementation of autonomous AI requires establishing a framework in itself including a quality management system, cybersecurity, data privacy controls, and post market monitoring.

4. How can regulators meet legitimate regulatory goals that may be raised in connection with these technologies without unduly hindering competition or innovation? *(Question from FTC website)*

   At IDx, we have gone through a robust regulatory process, creating a framework that can be followed by others. While the iterative nature of AI allows for a more streamlined framework, it is important to not bypass the standards and safety assurances the regulations were intended to implement – especially in medicine, where standards and practice patterns are based on clinical evidence. Patient safety for autonomous AI in medicine should be verified with preregistered clinical trials with the AI "locked down" and held by an algorithm integrity provider. Trial endpoints should prospectively define sensitivity, specificity, imageability, repeatability, reproducibility, human factors validation, and system level validation.

5. Are there tensions between consumer protection and competition policy with respect to these technologies? If so, what are they, and how should they be addressed? *(Question from FTC website)*

   The competitive landscape needs to be equal for all, not only the biggest players. The pathway we forged with FDA provides any company, of any size, with a clear, transparent, and reasonably

fast process moving forward. A fair competitive landscape for medical AI allows patients, physicians, data stewards, and AI companies to all simultaneously benefit from lower cost, higher quality and increased access.

6. What responsibility does a company utilizing these technologies bear for consumer injury arising from its use of these technologies? Can current laws and regulations address such injuries? Why or why not? *(Question from FTC website)*

In medicine, it depends on the claim the company is making – "assistive" or "autonomous" – and for the application/intended use including specified conditions of deployment. There needs to be enforcement of these claims. Autonomous AI, and thereby the company behind it, must assume responsibility for the output because it is, by definition, intended to be relied upon. For example, like a physician grading exams, IDx maintains medical malpractice insurance in case issues of liability arise.

7. For algorithmic or artificial intelligence analyses that do not involve classifying people (e.g., analysis of medical images), how is an accurate or transparent result defined? Is transparency important in that context? *(Question to panel via email)*

Transparency is critical. Medical AI developers need to explain how their AI conforms to the medical standards of care, including individual validation on the features important to the physician in making clinical decisions, which is based on clinical evidence. Simply correlating AI output to current standard of care output does not take into account the underlying reasoning and therefore risks straying from the clinical reasoning.

8. Should there be different ways of evaluating algorithms and artificial intelligence for different purposes, e.g., classification of people; detection of images that are readily identified, e.g., street signs; detection and diagnosis of conditions on the basis of data less readily categorized, e.g., medical images and other data? *(Question to panel via email)*

There should be categories and sub-categories within, which depend on the specific conditions of deployment. For example, in medicine, there should be categorizations across both risk level and autonomy level, depending on the intended use of the system, and different validation and regulatory requirements for each.

9. What is your view on the use by companies and others of proprietary algorithms or artificial intelligence, for which the bases for outcomes could be determined, but for the fact that the tools are proprietary? *(Question to panel via email)*

There needs to be oversight of AI that could harm the consumer. In medicine, while reviewing agencies need to have full access to proprietary algorithms and system, this does not mean their proprietary nature needs to be disclosed outside the reviewing agency. For example, this is standard for the FDA.

10. Are there circumstances in which non-proprietary, open decision-making tools should be required, and if so, what are they? *(Question to panel via email)*

We have considered this early on, and support open source projects, but the requirements for preregistered clinical trials and maintaining appropriate quality and security for the AI are such that open source models are difficult to fit into the required controls. In addition, presently this space is too new to place such requirements. If there is regulatory oversight for medical or high-risk AI, the regulator agencies or specialty groups can determine necessary requirements under the specific circumstances.

11. Is guidance available for algorithms or artificial intelligence tools that do not raise issues of legality with respect to their outcomes, but might raise issues of social good or best

practices? Is there a need for such guidance beyond that which might be available for any type of decision making? *(Question to panel via email)*

We have been closely following the FDA Pre-Certification Pilot Program. The intentions (lessening the regulatory burden and accelerating market access) are appropriate for assistive AI, but we caution its applicability to autonomous AI, where there is no physician or expert review of the results, and where patient safety is entirely dependent on the autonomous AI. This requires more rigorous, preregistered clinical testing and validation in the context in which it will be used. In my view, the path we forged with FDA is a transparent, fast, and risk appropriate path to market for autonomous AI. A new framework should not come at the cost of bypassing evidence-based medical or regulatory standards and should take into account the special controls and real-world implementation of autonomous AI.

12. Is it possible for best practices to address all types of artificial intelligence/machine learning, given that technology may have different levels of explainability and transparency built in? *(Question to panel via email)*

There may be universal best practices, such as the need for transparency and managing bias. However, best practices are almost entirely context driven. Within each field, there should be categorizations across both risk level and autonomy level, with different validation and best practice requirements for each. What is appropriate for autonomous medicine or autonomous vehicles, may not be appropriate for another setting, such as the legal system.

13. Would it be better to have a range of best practices to address the different types of artificial intelligence, or should there be just one global set of best practices? *(Question to panel via email)*

AI is a technology, not a tool, and the tools that can be built with it differ vastly with respect to autonomy level, acceptability of harm, and evidence-based best practices.

14. Some commentators are concerned about "human control" of artificial intelligence. In your view, will artificial intelligence remain in human control? What does it mean, practically speaking, for artificial intelligence to be in human control? *(Question to panel via email)*

In medicine, AI is part of a larger healthcare "system" rather than a product, so the clinical validation should ensure that the level of "human control" either maintains or improves safety and efficacy after introducing the AI system.

15. What technical, procedural, or policy mechanisms should be put in place to facilitate human control? What checks and balances could be put in place to allow for human interventions in AI when there are potential problems? *(Question to panel via email)*

- The use cases in medicine are such that somewhere in the diagnostic or management loop there is a physician or other provider.(1)
- A safe regulatory framework validates the interaction of the AI system and the human intervention in the setting where the AI is intended to be used.
- Statistically, it is impossible to guarantee the safety of AI in medicine, but this is true also human clinical decisions.
- The reference standard used in testing and validating AI rests on the best possible reference standard, and not on decisions by individual clinicians, where significant intra- and inter-observer variability and "diagnostic drift" cannot be accounted for.

- Autonomous AI systems are held to a higher standard than individual human experts.

  In other words, public trust relies on autonomous AI being held to a standard that humans may not be able to achieve.
- For medical AI, a full and compliant Quality Management System, including clinical validation and continuous efficacy monitoring, should be implanted to maintain safety and human control through the ongoing stages of AI implementation.

References

1.   Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. N Engl J Med. 2018;378(11):981-3.
2.   Shah A, Lynch S, Niemeijer M, Amelon R, Clarida W, Folk J, et al., editors. Susceptibility to misdiagnosis of adversarial images by deep learning based retinal image analysis algorithms. Proceedings - International Symposium on Biomedical Imaging; 2018.
3.   Lynch S, Abramoff MD. Catastrophic Failure in Image-Based Convolutional Neural Network Algorithms for Detecting Diabetic Retinopathy. IOVS. 2017;58(ARVO E-abstract 3776).
4.   Finlayson SG, Chung HW, Kohane IS, Beam AL. Adversarial Attacks Against Medical Deep Learning Systems. arXiv:170809843. 2018.