Stanford | Health Policy
Freeman Spogli Institute and Stanford School of Medicine

HAI
Stanford University
Human-Centered
Artificial Intelligence

Stanford Law School

Facilitating Responsible Governance of Healthcare AI Tools

Michelle M. Mello, JD, PhD[1]
Stanford University

Testimony presented to the U.S. Senate Committee on Finance, February 8, 2024.

Chairman Wyden, Ranking Member Crapo, and Members of the Committee, thank you for the opportunity to speak with you today.

I have the extraordinary privilege of being part of a group of ethicists, data scientists, and physicians at Stanford University—long a leading hub of AI innovation—that is directly involved in governing how healthcare AI tools are used in patient care. I have studied patient safety, healthcare quality regulation, and data ethics for more than two decades. I apply that expertise in our team's evaluations of all AI tools proposed for use in Stanford Health Care facilities, which care for over 1 million patients per year, and our recommendations about whether and how they can be used safely and effectively.

I would like to share the three most important things we've learned so far.

First, while hospitals are starting to recognize the need to vet AI tools before use, **most healthcare organizations don't have robust review processes yet**. Some, like Stanford, have plentiful resources to drawn on; others don't. All need help. Although as a lawyer I know that more law isn't always the answer, in this case there is much that Congress could do to help.

Second, to be effective, **governance can't focus only on the algorithm**. **It must also encompass how the algorithm is integrated into clinical workflow**. By "workflow," I mean how physicians, nurses, and other staff interact with each other, the AI tool, the patient, and other systems. Currently, conversations about regulating healthcare AI mostly focus on the AI tool

---

[1] Professor of Health Policy, Department of Health Policy, Stanford University School of Medicine; Professor of Law, Stanford Law School; Affiliate Faculty, Stanford Institute for Human-Centered Artificial Intelligence; Institute Faculty, Freeman Spogli Institute for International Studies, Stanford University.

itself—for example, is its output biased? How often does it make wrong predictions or misclassify things?

These things matter. But it is equally important to consider how medical professionals will interact with the tool. A key area of inquiry is the expectations placed on physicians and nurses to evaluate whether AI output is accurate for a given patient, given the information readily at hand and the time they will realistically have. For example, large-language models like ChatGPT are employed to compose summaries of clinic visits and doctors' and nurses' notes, and to draft replies to patients' emails. Developers trust that doctors and nurses will carefully edit those drafts before they're submitted—but will they? Research on human-computer interactions shows that humans are prone to automation bias: we tend to overrely on on computerized decision support tools and fail to catch errors and intervene where we should.[2]

Therefore, regulation and governance should address not only the algorithm, but also how the adopting organization will use and monitor it. To take a simple analogy, if we want to avoid motor vehicle accidents, we can't just set design standards for cars. Road safety features, driver's licensing requirements, and rules of the road all play important roles in keeping people safe.

Third, because the success of AI tools depends on the adopting organization's ability to support them through vetting and monitoring, **the federal government should establish standards for organizational readiness and responsibility to use healthcare AI tools**, **as well as for the tools themselves**. As countless historical examples of medical innovations have shown, having good intentions isn't enough to protect against harm. The community needs some guardrails and guidance.

I believe there is a right and a wrong way to do this. It would be a mistake to enshrine in legislation detailed standards for healthcare AI tools and how they can be used. In light of how quickly things are moving in the field, we have to have the humility to acknowledge that we don't know what the best standards will be two years from now. Regulation needs to be adaptable or else it will risk irrelevance—or worse, chilling innovation without producing any countervailing benefits. **The wisest course now is for the federal government to foster a consensus-building process that brings experts together to create national consensus standards and processes for evaluating proposed uses of AI tools.**

[2] Mello MM, Guha N. Understanding liability risk from using healthcare artificial intelligence tools. *N Engl J Med*. 2024;390(3):271-278. Full text available from the author (mmello@law.stanford.edu).

It can also begin requiring that entities regulated by federal agencies adhere to those standards and processes. Through its operation of and certification processes for Medicare, Medicaid, the Veterans Affairs Health System, and other health programs, Congress and federal agencies can require that participating hospitals and clinics have a process for vetting any AI tool that affects patient care before deployment and a plan for monitoring it afterwards. As an analogue, the Centers for Medicare and Medicaid Services (CMS) uses The Joint Commission, an independent, not-for-profit organization, to inspect healthcare facilities for purposes of certifying their compliance with the Medicare Conditions of Participation. The Joint Commission recently developed a voluntary certification standard for the Responsible Use of Health Data which focuses on how patient data will be used to develop algorithms and pursue other projects. A similar certification could be developed for facilities' use of AI tools.

The initiative currently underway to create a network of "AI assurance labs,"[3] and consensus-building collaboratives like the 1,400-member Coalition for Health AI, can be pivotal supports for these facilities. Such initiatives can develop consensus standards, provide technical resources, and perform certain evaluations of AI models, like bias assessments, for organizations that don't have the resources to do it themselves. Adequate funding will be crucial to their success.

It's important to recognize that some aspects of AI review will need to be done locally, as individual healthcare organizations are best positioned to identify and address problems that could result from how they embed AI tools within clinical workflow.[4] Here, too, regulatory requirements can ensure that organizations invest in making it happen—just as the federal regulations known as "the Common Rule" did for ethical review of human subjects research.

We have developed such a review process at Stanford. For each AI tool proposed for deployment in Stanford hospitals, data scientists evaluate the model for bias and clinical utility. Ethicists interview patients, clinical care providers, and AI tool developers to learn what matters to them and what they're worried about.[4] We find that with just a small investment of effort, we can spot potential risks, mismatched expectations, and questionable assumptions that we and the AI designers hadn't thought about. In some cases, our recommendations may halt deployment;

---

[3] Shah NH, Halamka JD, Saria S, et al. A nationwide network of health AI assurance laboratories. *JAMA*. 2024;331(3):245–249.
[4] Mello MM, Shah NH, Char DS. President Biden's executive order on artificial intelligence—implications for healthcare organizations. *JAMA*. 2024;331(7):17-18.

in others, they strengthen planning for deployment. We designed this process to be scalable and exportable to other organizations.

I will close with one final point from other research I have conducted on AI: **don't forget health insurers**. Just as with healthcare organizations, real patient harm can result when insurers use algorithms to make coverage decisions. For instance, members of Congress have expressed concern about Medicare Advantage plans' use of an algorithm marketed by NaviHealth in prior-authorization decisions for post-hospital care for older adults. In theory, human reviewers were making the final calls while merely factoring in the algorithm output; in reality, they had little discretion to overrule the algorithm. This is another illustration of why humans' responses to model output—their incentives and constraints—merit oversight.

CMS recently took the important step of addressing these practices through a Final Rule requiring Medicare Advantage plans to make medical necessity determinations "based on the circumstances of the specific individual…as opposed to using an algorithm or software that doesn't account for an individual's circumstances." The Final Rule further specifies that determinations must be reviewed by a medical professional. But ambiguity remains around what it means to merely "use" algorithms, as opposed to allowing them to drive decisions; and what it means to "account for" individual circumstances or to have algorithm results "reviewed by" a human.[5] How much freedom must human reviewers have to overrule algorithm recommendations? Must algorithms include information on social determinants of health or patients' social supports to "account for" individual circumstances? Must insurers disclose the prediction algorithm? Additional clarity about regulators' expectations would be very helpful.

In summary, Congress can support healthcare organizations and health insurers navigating the uncharted territory of AI tools by imposing some guardrails while allowing the rules to evolve with the technology. Specifically, Congress should:

1. Require that healthcare organizations have robust processes for determining whether planned uses of AI tools meet certain standards, including undergoing ethical review.

2. Fund a network of AI assurance labs to develop consensus-based standards and ensure that lower-resourced healthcare organizations have access to necessary expertise and infrastructure to evaluate AI tools.

---

[5] Mello MM, Rose S. Denial—artificial intelligence tools and health insurance coverage decisions. *JAMA Health Forum* (forthcoming; available from the author (mmello@law.stanford.edu)).

3. Require developers of AI tools to disclose information that a consensus-based organization such as an assurance lab determines is essential to evaluating the safety and ethics of AI tools.

4. Work with CMS to provide further guidance to Medicare Advantage plans about permissible and impermissible uses of algorithms in coverage decisions.

5. Ensure that relevant federal agencies, including but not limited to CMS, the Department of Veterans Affairs, and the Food and Drug Administration, have clear grants of authority to adopt standards for all types of healthcare AI and require entities within their purview to adhere to them. Clarity and specificity are essential here, because courts are increasingly holding that on matters of vast social and economic significance like AI, Congress must speak clearly when it intends to give authority to agencies.

Thank you, and I welcome your questions.