

U.S. Food and Drug Administration
Center for Devices and Radiological Health
Document No. 2026-N-7874

Considerations for the Regulation of Generative AI-Enabled Medical Devices: Discussion Paper and Request for Feedback

October 19, 2026

Introduction

We, a group of scholars affiliated with the Stanford Institute for Human-Centered Artificial Intelligence and Stanford University School of Medicine, submit the following comments in response to your [request for feedback](#).¹

As practicing physicians, legal scholars, technologists, ethicists, and policy professionals working to advance the governance of healthcare AI,² we are heartened by the discussion paper, and support the efforts presented to consider how best to regulate genAI-enabled medical devices. We offer our reflections on the risk assessment, evaluation, and ongoing monitoring of genAI devices based on our interdisciplinary academic research and our experiences operating a responsible AI lifecycle governance system at Stanford Health Care.

We share considerations for operationalizing the two-axis risk assessment framework and our thoughts on how to ensure that the proposed competency-based approach to evaluation and monitoring can better align with current developer practices and the science of AI benchmarking, in an effort to both maximize patient safety and support innovation. While we do not draw a firm line between general-purpose and purpose-built tools in our submission, the distinction merits consideration.

We make seven targeted recommendations:

On the Two-Axis Risk Framework

1. Require evidence that human oversight is an effective risk control for the task(s) the device performs.
2. Specify how the consequences axis is determined and consider measures of typical use.

¹ Authors write in their individual capacity as experts in the field. Views expressed do not represent official positions of Stanford University.

² Stanford HAI's Healthcare AI Policy Steering Committee conducts interdisciplinary, evidence-based research and convenes multistakeholder discussions to develop tangible recommendations for policymakers that help ensure healthcare AI can benefit patients, doctors, and developers alike. See: Stanford HAI, "Healthcare AI Policy," accessed October 14, 2026, <https://hai.stanford.edu/policy/policy-efforts/healthcare-ai-policy>.

On Premarket Evaluation

3. Strengthen benchmarking to address construct validity and prevent metric optimization at the expense of utility.
4. Establish real-world prospective testing as a supplement or alternative to benchmarking.

On Postmarket Evaluation

5. Place greater accountability for postmarket monitoring on developers and equip healthcare institutions with the ability to support continued monitoring.
6. Use hybrid monitoring, with machine-based monitoring tools to flag issues and humans as the final arbiters.
7. Ensure that postmarket monitoring occurs by requiring a prespecified monitoring plan.

1. Feedback on the two-axis risk framework for genAI-enabled devices

The proposed two-axis framework for assessing the risks posed by genAI-enabled devices has strong intuitive appeal. We particularly support the intent to assess a device’s directive functions (i.e., its categorization as non-directive versus action-directing) based on how the tool is likely to be used. We agree that developers may otherwise use strategic wording to avoid being placed in a more stringent risk category.

The discussion paper also rightly recognizes that multi-turn conversations between humans and genAI devices entail special risks. In addition to becoming more directive over the course of a multi-turn conversation, some chatbots, evidence suggests, engage in psychological manipulation³ and create significant dependencies⁴ as they become more intimately familiar with their user — even if they don’t become more “action-directing” in terms of their functions. This can undermine the user’s likelihood of seeking to validate its advice through external means. Manipulative chatbot behavior must therefore be a consideration in the risk framework.

Another strength of the framework is the recognition that overescalation to care is a harm worthy of consideration. At Stanford Health Care, clinicians and patients who participate in AI assessments often note patient anxiety due to overescalation as a substantial risk for some types of AI tools. Our governance process for using AI safely, responsibly, and equitably views this as a legitimate risk that needs to be weighed against the benefits of the tool.⁵

³ Nizan Geslevich Packin and Karni Chagal-Feferkorn, “This Is Not a Game: The Addictive Allure of Digital Companions,” *Seattle University Law Review* 48, no. 3 (2025): 693, <https://digitalcommons.law.seattleu.edu/sulr/vol48/iss3/6/>.

⁴ Myra Cheng et al., “Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence,” *Science* 391, no. 6792 (2026): eaec8352, <https://www.science.org/doi/10.1126/science.aec8352>.

⁵ Stanford Health Care, “Our Commitment to Using AI Safely, Responsibly, and Equitably,” accessed October 14, 2026, <https://stanfordhealthcare.org/campaigns/ai-education/responsible-use.html>.

We identify several key challenges to operationalizing the two-axis framework:

The risk distinction between advisory and action-taking devices will only hold if human reviewers are consistently able to meaningfully supervise AI outputs. The activity axis assumes they can, but in practice, this is not always true. The axis spans a broad range of heterogeneous tasks, and how well a human can check an output depends on the task and the reviewer. For example, a radiologist can readily verify an AI-generated radiology report if they have access to the original images. Yet a patient will struggle to verify an AI-generated summary of their complex medical record. Individuals also differ in their skepticism: Some will triangulate an AI output with other sources, while others will accept it uncritically.

Oversight also varies with setting and erodes over time. Physicians in one clinical setting may have the time, resources, expertise, and incentives to scrutinize AI output carefully, while physicians in another may not. In an emergency care setting, for example, AI output related to a specialist domain may only receive a short glance from an emergency medicine physician. Active supervision is also likely to fade over time as users become more familiar and possibly develop automation complacency. Further, the potential for automation bias is high in many areas of healthcare, where clinicians work under time pressure and manage workloads that create chronic exhaustion and burnout.

The framework also doesn't specify how placement on the consequence axis will be determined for multipurpose AI tools. GenAI devices — especially patient-facing tools with general purpose capabilities — can field thousands of different questions or prompts and thus have many different potential uses that go beyond the primary intended use. Consequently, the resulting outcomes of one AI tool may range from trivial to highly consequential. Will the determination of a tool's consequence (from “limited” to “severe”) be based on the most severe possible outcome or some measure of the typical outcome? How will the framework take into account the reality that the consequences of an AI tool may vary significantly across patient populations, including clinical subgroups and groups underrepresented in the data used to build the genAI device? Lastly, how will the framework take into account the fact that the risks associated with general purpose genAI devices can change over time as use changes?

Recommendation #1. Require evidence that human oversight is an effective risk control for the task(s) the device performs.

Human oversight should be treated as a risk control whose effectiveness must be demonstrated for the task(s) the device performs, rather than always assumed to be helpful. Where a developer relies on human oversight to reduce the risk of its genAI device, the FDA should require the developer to provide evidence that oversight will consistently and meaningfully occur in the intended settings and that reviewers can detect the errors the device is likely to make.

Such evidence could include real-world testing data on how often, how long, and in what ways AI tool users engage in activities that could be considered supervision or validation of AI output — and how this varies across different users in different clinical settings. For example, how long do healthcare professionals (HCPs) spend examining an output before acting on it? How often do they follow or reject the tool’s recommendation? How often do they click through to source material, where provided? Do they have access to the necessary information to detect errors? How do the answers to these questions differ between HCP-facing and patient-facing AI tools?

Recommendation #2. Specify how the consequences axis is determined and consider measures of typical use.

The FDA should clarify how genAI devices with wide-ranging functions are placed on the consequences axis, including whether placement reflects the worst-case use, the typical use, or a distribution across types of users and settings. We expand on this recommendation in our feedback on how the FDA should specifically approach premarket and postmarket monitoring for genAI devices in the following pages.

The framework must also take into account and weigh appropriately harms that are harder to see or measure when determining the severity of a tool’s consequences. Psychological harms, for example, have traditionally often been viewed as lower-risk than other, physical harms,⁶ so a genAI device that engages with users in distress could be determined to have “limited” consequences even though the potential for harm is significant. Frameworks such as the FDA’s fair-balance requirements for prescription drug promotion, which state that important risk information should not be minimized in comparison to efficacy claims, could be applied here.⁷

2. Feedback on the competency-based approach for premarket evaluation

Given that AI is not indication-based, and given the vast range of potential functions that many genAI devices may perform, the competency-based approach presents an attractive alternative to a more burdensome function-by-function review. However, we highlight several concerns about the effectiveness of a competency-based approach that leans heavily on benchmarks.

⁶ Julian Herpertz et al., “FDA-Authorized Software as a Medical Device in Mental Health: A Perspective on Evidence, Device Lineage, and Regulatory Challenges,” *NPJ Mental Health Research* 4, no. 62 (2025): 1–9, <https://doi.org/10.1038/s44184-025-00174-2>.

⁷ Hyosun Kim, “Trouble Spots in Online Direct-to-Consumer Prescription Drug Promotion: A Content Analysis of FDA Warning Letters,” *International Journal of Health Policy and Management* 4, no. 12 (2015): 813-821, <https://doi.org/10.15171/ijhpm.2015.157>.

First, intended use alone may be an incomplete basis for evaluating multipurpose genAI devices. The same device can perform functions related to hypertension, diabetes, and heart failure — and it is challenging to decide what its core competencies should be. When evaluation is anchored to the product’s stated intended use, a tool may be authorized for a narrow purpose while being deployed by its users in many other ways, and the evidence generated for the stated use may say little about the risks of those other uses. This also complicates two parts of the framework: what developers must evaluate before and after deployment, and where the FDA places a device on the consequence axis.

One way to address this gap without function-by-function review is through the risk management process developers already follow. ISO 14971 instructs developers to include reasonably foreseeable misuse in the hazard analysis and validation phases. The FDA does not require developers to follow this standard, but it references it as the gold standard and requires developers to document a risk management process under the Quality Management System Regulation (QMSR). The FDA could therefore provide recommendations in guidance documents on intended use and reasonably foreseeable misuse beyond the stated intended use, and keep that scoping current as real-world use cases emerge.

Benchmark validity is also a significant source of uncertainty in the proposed competency-based approach. The discussion paper acknowledges that public benchmarks may suffer from data contamination, saturation, and limited representativeness of real-world conditions. A more basic problem is construct validity: Benchmarks often do not measure what they claim to measure, and different benchmarks purporting to measure the same property frequently disagree.⁸ Consider a benchmark question created to measure bias that deliberately gives a model incomplete information, where the correct answer is that there is not enough information to answer. A model can score well simply by recognizing the trick, regardless of its underlying biases.

The problem is especially acute in medicine, where many of the skills we want to measure are inherently hard to measure. A competency-based approach risks resembling medical examinations with deterministic answers, which may not reflect real medical tasks.⁹ Performance may also not carry over when the format changes: Models that perform well on multiple-choice vignettes can fail when the same case is presented as a multi-turn conversation.¹⁰

⁸ Max Zhang et al., “Why Do Safety Guardrails Degrade Across Languages?,” arXiv preprint, arXiv:2605.17173 (2026), <https://arxiv.org/abs/2605.17173>; Olawale Salaudeen, “Validating Claims About AI: A Policymaker’s Guide,” *Stanford Institute for Human-Centered Artificial Intelligence*, September 24, 2025, <https://hai.stanford.edu/policy/validating-claims-about-ai-a-policymakers-guide>.

⁹ Inioluwa Deborah Raji et al., “It’s Time to Bench the Medical Exam Benchmark,” *NEJM AI* 2, no. 2 (2025): AIe2401235, <https://doi.org/10.1056/AIe2401235>.

¹⁰ Shrutika Singh et al., “The Pitfalls of Multiple-Choice Questions in Generative AI and Medical Education,” *Scientific Reports* 15 (2025): 42096, <https://doi.org/10.1038/s41598-025-26036-7>.

Weak construct validity may result in measuring convenient outcomes, rather than the most appropriate ones. As the bias benchmark example above shows, an AI model can pass a test by recognizing the structure of the test, without delivering what is needed for good outcomes. The results can still be used by a developer to support the claim that the model works. There are indications that this is already happening in the mental health context. For example, the competencies chosen for evaluation within the field (e.g., warmth of response) tend to be the easiest to measure, but these often miss the mark of what is needed in the real world in order to produce a good outcome (e.g., appropriate response to suicidality).¹¹

Raising the bar for premarket evidence, however, carries its own tension: If validation requirements are too demanding, clinical genAI tools could take years to reach authorization, by which point the underlying technology may be obsolete, and the process could become so burdensome that developers choose not to build these tools for clinical use at all.

Finally, the clinician credentialing model underlying the competency-based approach has limits of its own that are likely to be magnified when applied to AI.¹² In medicine, competency assessments are informative because test performance tracks other traits, such as intelligence, diligence, and professionalism. But the assessment does not capture the years of training and experience that support a clinician's practical judgment (phronesis) real-world care requires. An AI system lacks that experience and the ability to move beyond the domains in which it was tested, so the test is effectively all there is. Competency assessment may therefore offer weaker assurance for AI than it does for physicians.

Recommendation #3: Strengthen benchmarking to address construct validity and prevent metric optimization at the expense of utility.

Benchmarks are useful evidence only if they measure what matters. Developers should be required to provide evidence of construct validity, showing that the competency being tested corresponds to the skill needed for the real task and for producing good outcomes. The discussion paper already notes that adjudicators should be structurally independent from the sponsor, and we encourage the FDA to apply the same independence expectation to the process of selecting benchmark criteria, especially when sponsors propose their own. As mentioned above, developing benchmarks with high construct validity is expensive. To address this, the FDA could utilize existing processes of guidance documents and comment periods or explore further mechanisms to support the development of more robust vendor-neutral benchmarks.

¹¹ Matthew Flathers et al., "Benchmarking Language Models for Clinical Safety: A Primer for Mental Health Professionals," *Psychotherapy and Psychosomatics* (2026), <https://doi.org/10.1159/000553576>; John Torous and Andrea Cipriani, "A Paradigm Shift in Progress: Generative AI's Evolving Role in Mental Health Care," *JMIR Mental Health* 12 (2025): e82369, <https://doi.org/10.2196/82369>.

¹² MJMH (Kiki) Lombarts, "Competence-Based Education Misses the Essence of the Medical Profession," *Perspectives on Medical Education* 4, no. 6 (2015): 326-328, <https://pmc.ncbi.nlm.nih.gov/articles/PMC4673067/>.

To prevent these issues, benchmarks should be preregistered before testing. This complements the prespecification the discussion paper already expects by preventing developers from choosing favorable benchmarks or adjusting criteria after seeing results. However, the tools used to run these evaluations need the same scrutiny. LLM judges should be validated against a human expert standard, since research has found that they tend to favor their own outputs, and those of similar models, over others of comparable quality.¹³ Patient simulators should count as evidence only if sponsors can show that simulated users behave like real ones, for example in how often they accept, question, or push back on outputs, by comparing against a left-out sample of real interactions. They must also be grounded in real-world data and varied on the traits that determine whether a user can catch an error, such as health literacy, baseline trust or skepticism toward AI, and persistence when a device declines to answer. Simulator results lacking this validation should be labeled exploratory and not pooled with clinical confirmation evidence.

Recommendation #4: Establish real-world prospective testing as a supplement or alternative to benchmarking.

In contrast to competency exams, real-world usage would give data on performance in the intended use scenario. Sandboxes, such as Utah’s AI Sandbox,¹⁴ may provide the opportunity to collect real-world data in a monitored setting. This data could be used for submitting premarket performance to the FDA. It may be beneficial for the FDA to indicate what scope of premarket data would be acceptable from such a sandbox. Other methods of generating real-world premarket data in the intended use scenario should also be considered.

For patient-facing tools, premarket evaluation should also test how patients interpret and are affected by genAI device outputs, as the FDA has done for direct-to-consumer genetic tests.¹⁵

3. Feedback on postmarket monitoring approaches

Postmarket monitoring is one of the most important and challenging aspects of regulating genAI devices used in healthcare contexts. As already mentioned in Section 1 of our submission, the range of possible prompts a genAI tool may receive is effectively unbounded, performance can drift over time, and human oversight is fallible. While performance on predefined tasks can be assessed during premarket evaluation, real-world safety signals and harms are more difficult to

¹³ Roxana Daneshjou et al., “The Judge is Not Impartial: Self-Preference in Medical LLM Evaluation,” preprint, Research Square, September 17, 2026, <https://doi.org/10.21203/rs.3.rs-11022502/v1>.

¹⁴ Utah Office of Artificial Intelligence Policy, “Understanding Utah’s AI Sandbox,” Utah Department of Commerce, accessed October 19, 2026, <https://commerce.utah.gov/ai/regulatory-relief/>.

¹⁵ Scott D. Crawford et al., “Decoding FDA DTC Policy: The Complete Series,” SoundRocket (blog), November 2018, https://www.genomes2people.org/wp-content/uploads/2018/11/20181112_sound-rocket.pdf.

surface during limited pilot runs, especially when there are harms that may only appear after the patient leaves their care setting.

However, the responsibility of postmarket monitoring cannot rest on the healthcare institutions adopting the tools alone. Technical capacity and staffing across U.S. healthcare organizations varies widely. Some institutions, including our own, can perform sophisticated in-house monitoring of how genAI tools are used and what problems HCPs are reporting. Yet even those institutions are finding that the volume of deployed genAI tools is increasingly outstripping their ability to monitor them and that assessing the impacts of AI on patient outcomes is quite challenging.¹⁶

Institutions that don't have the resources or technical capacity to perform their own monitoring must rely wholly on genAI tool developers to provide them with monitoring information. However, developers often do not provide sufficient monitoring support to healthcare institutions. For example, they frequently omit information about the most important failure modes of their products and suggest only limited monitoring metrics that do not capture the clinical utility or risks of the tool. Further, shared information may be in the form of dashboards that do not enable healthcare institutions to conduct their own data analysis.

Recommendation #5. Place greater accountability for postmarket monitoring on developers and equip healthcare institutions with the ability to support continued monitoring.

Developers should assume a greater role in ensuring the safety and effectiveness of genAI devices across the total product lifecycle. Existing device requirements (including QMSR, Medical Device Reporting, and special controls) offer a starting point for conceptualizing obligations for developers.

Healthcare organizations play a critical role in local evaluation, particularly in surfacing problems that arise from how AI tools are placed into specific workflows. But because their monitoring capacity varies widely, there currently exists an uneven self-regulatory landscape and, consequentially, disparate patient protection. AI developers should be required to give deployers usable data access, documentation of known failure modes (with updates as knowledge evolves), and suggested monitoring metrics that help assess the product's clinical utility. They should also provide technological mechanisms for frontline users to submit feedback on the AI output and for deploying organizations to receive reports of such feedback both within their organization and across all deployers. This information would help healthcare organizations conduct monitoring, identify areas where local workflows could be changed to help catch and respond to AI errors before they become consequential, signal when local users

¹⁶ Zhongnan Fang et al., "Operationalizing Real-Time Monitoring of Clinical AI," *Stanford Institute for Human-Centered Artificial Intelligence*, May 14, 2026, <https://hai.stanford.edu/policy/operationalizing-real-time-monitoring-of-clinical-ai>.

find fault with the AI device so often that discontinuation of the tool should be considered, and identify specific problems that deployers train users to look out for when using the tool.

Additionally, monitoring should ideally aim to go beyond model outputs to include patient outcomes and changes in use over time. When the intended use is to deliver on a specified outcome, postmarket monitoring should not only focus on model output quality but also on downstream patient outcomes that health systems already capture, such as unplanned return visits, escalations in level of care, and readmissions after discharge. Because harms from erroneous output often appear after discharge, single-site monitoring will often undercount them: Roughly a third of return visits within three days of an emergency department discharge occur at a different hospital.¹⁷ Thus, these outcomes should be aggregated across sites through a common data model. The willingness of health systems that deploy the devices and electronic health record (EHR) vendors to provide developers with access to this real-world deployment and monitoring data will, in turn, greatly impact developers' ability to conduct safety monitoring. Requiring this data to be made available to developers may go beyond the FDA's authority and will likely require action from other regulatory bodies, or programs linking patient outcomes with payment incentives.¹⁸

Recommendation #6. Use hybrid monitoring, with machine-based monitoring tools to flag issues and humans as the final arbiters.

Machine-based screening tools may be needed for comprehensive postmarket monitoring, and, when appropriate, they should flag issues for human review rather than act on their own. Because monitoring of genAI is too onerous for many institutions to perform locally, the FDA should support lightweight tools, qualified through its Medical Device Development Tool (MDDT) program and hosted by a neutral party, that can run across health systems to detect outliers. When clusters of mistakes are substantiated on review, the finding should be reported to the FDA and the developer. Any automated monitoring should itself be validated against human performance in a way that external auditors can reproduce.

As mentioned in Section 1 of this submission, human review is not a safeguard by default. Until fully automated drift detection matures, reviewers' agree/disagree adjudications can serve as an interim proxy for drift, with periodic manual review for high-stakes uses.

Recommendation #7. Ensure that postmarket monitoring occurs by requiring a prespecified monitoring plan.

¹⁷ Rishi Duseja et al., "Revisit Rates and Associated Costs After an Emergency Department Encounter: A Multistate Analysis," *Annals of Internal Medicine* 162 (2015): 750-756. <https://doi.org/10.7326/m14-1616>.

¹⁸ Centers for Medicare & Medicaid Services, "ACCESS (Advancing Chronic Care with Effective, Scalable Solutions) Model," CMS Innovation Center, accessed October 19, 2026, <https://www.cms.gov/priorities/innovation/innovation-models/access>.

We see no alternative to placing heavy emphasis on postmarket monitoring to assure the safety of genAI tools because of their dynamic, evolving, and multifunctional nature. In doing so, the FDA should require a postmarket monitoring plan, with defined endpoints and thresholds, to be in place before authorization. Moving parts of the evidence burden into the postmarket period is consistent with the FDA’s “least burdensome” approach, but it is only defensible if postmarket commitments are specified up front rather than worked out after the device is in use. Many health systems already work this way: At SHC, each adopted AI tool has a monitoring plan, designed jointly by the proposing team and the health system’s Data Science Team, that complements a robust assessment conducted prior to adoption. The FDA can require developers to do the same.

The FDA must also be able to ensure that such a plan is carried out. In the drug space, the FDA has struggled to ensure that postmarket surveillance plans are actually followed by manufacturers, and its legal authorities and levers for enforcing them are limited.¹⁹ For genAI devices, one way to strengthen verification is to require that safety assessment and quality reporting be designed into the product itself, so that data collection, performance tracking, and feedback channels are built-in features rather than separate commitments a developer must remember to honor. This would give the FDA something concrete to verify. We therefore urge the FDA to consider what mechanisms are available to confirm that postmarket monitoring plans are actually implemented, and to take action if they either are not fully implemented or reveal problems that substantially alter the tool’s risk/benefit balance.

Overall, the regulatory approach outlined in the discussion paper is on the right track. The two-axis risk assessment framework and competency-based approach to evaluation are reasonable starting points for a technology that does not fit conventional device review. Their value will depend on how they are operationalized, especially with regard to determining the effectiveness of human oversight and benchmarking as part of monitoring strategies for genAI devices that have a myriad of potential uses and outcomes.

Postmarket monitoring is where this framework will succeed or fail. Because genAI tools are dynamic and multifunctional, the framework will lean heavily on what happens after authorization. Monitoring plans with defined endpoints and thresholds must be in place before authorization, and the FDA should consider how to verify implementation and act on these findings. Success also depends on shared responsibility across developers, deployers, and the FDA, and their willingness to share usable data, documentation of known failure modes, and

¹⁹ Daniel G. Dauner et al., “FDA’s Unimproved Enforcement of Postmarketing Requirements and Commitments: Implications for Providers and Patients,” *Research in Social and Administrative Pharmacy* 16, no. 6 (2020): 844-847, <https://doi.org/10.1016/j.sapharm.2019.11.004>.

feedback mechanisms. Finally, humans, not automated tools, should remain the final arbiters of what monitoring flags.

We appreciate the opportunity to work collaboratively with the FDA to arrive at frameworks that support innovation and patient access to useful tools, while ensuring that developers, clinicians, and healthcare systems face reasonable expectations when it comes to ensuring new genAI devices are safe and effective.

Sincerely,

Roxana Daneshjou, MD, PhD, FAAD

Assistant Professor of Biomedical Data Science and Dermatology and, by courtesy, Radiology, Stanford University School of Medicine

Dev Dash, MD, MPH

Clinical Assistant Professor of Emergency Medicine, Stanford University School of Medicine

Diliana Funtarova

Master's Candidate, Clinical Informatics Management, Stanford University

Curtis Langlotz, MD, PhD

Professor of Radiology, Medicine, and Biomedical Data Science, Stanford University School of Medicine; Senior Fellow, Institute for Human-Centered Artificial Intelligence (HAI); Senior Associate Vice Provost for Research; Stanford University

Kara Liu

PhD Candidate, Computer Science, Stanford University School of Engineering

David Magnus, PhD

Thomas A. Raffin Professor of Medicine and Biomedical Ethics, Pediatrics, and Medicine, and, by courtesy, of Bioengineering, Stanford University

Josh Makower, MD

Co-Founder and Director, Stanford Mussallem Center for Biodesign; Professor of Medicine and Bioengineering, Stanford University Schools of Medicine and Engineering

Nicole Martinez-Martin, JD, PhD

Assistant Professor of Pediatrics and, by courtesy, of Psychiatry and Behavioral Sciences, Stanford University

Caroline Meinhardt

Senior Policy Research Manager, Stanford HAI

Michelle Mello, JD, PhD

Professor of Health Policy, Stanford University School of Medicine; Professor of Law, Stanford Law School

Jared Moore

PhD Candidate, Computer Science, Stanford
University School of Engineering

Sandra Waugh Ruggles, PhD

Director of Policy Research, Stanford
Mussallem Center for Biodesign

Kevin Schulman, MD, PhD

Professor of Medicine, Stanford University
School of Medicine; Professor, by courtesy,
of Operations, Information and Technology,
Stanford Graduate School of Business

Nigam Shah, MBBS, PhD

Professor of Medicine and Biomedical Data
Science, Stanford University School of
Medicine

Artem Trotsyuk, PhD

Postdoctoral Research Fellow, Department
of Genetics, Stanford University School of
Medicine

Caroline Yee

Research Scholar, Stanford HAI