AUDIO: Recording in progress.

MILDRED CHO: Hi. I'm Mildred Cho, and I'm delighted to welcome you to our May ELSI Friday Forum, entitled Addressing Algorithmic Harms: Practices and Provocations for Health AI. For those of you familiar, the Forum is hosted by the Center for ELSI Resources and Analysis and held on the second Friday of every month for one hour starting at 12:00 noon Eastern Time. However, in June, as I'll tell you about in more detail at the end of today's Forum, we will be holding it on the third Friday of the month. More about that later. We also have a Zoom room reserved for more informal discussion immediately after the panel for 30 minutes. The Center for ELSI Resources and Analysis, or CERA, provides resources to support research on the ethical, legal, and social implications of genetics and genomics, and serves to connect scholars, scientists, policy‑makers, journalists, members of the public, and others. It is funded by the National Human Genome Research Institute at NIH and is managed by teams at Stanford and Columbia Universities in partnership with the Hastings Center and Harvard University. I encourage you to visit the online platform ELSIhub.org for the recording and the transcript of this forum and related references. Please use the link in the chat to access our latest ELSIhub Collection, curated by Sara Gerke. This reading list explores shortcomings in the current regulatory framework for AI‑based medical devices and key legal and ethical questions about bias, privacy, and liability. Please also go to the website to join the ELSI Scholar Directory, sign up for newsletters and other events like this one, get daily updates and news on Twitter.

So, some quick housekeeping. If you wish to use closed captioning, turn it on at the bottom of your screen. There's also the Q&A button at the bottom of your screen. You can register your enthusiasm for a question and elevate it up the list using the upvote button in the Q&A box. The chat box is available for further engagement. We'll post links to the references in today's discussion there as well. And as always, if you have questions, please e‑mail info@ELSIhub.org at any time.

So it's now my pleasure to introduce Kayte Spector‑Bagdady, an associate director at the Center for Bioethics and Social Sciences in Medicine, and an assistant professor of Obstetrics and Gynecology at the University of Michigan Medical School. She is a former practicing FDA attorney and associate director for President Obama's bio‑ethics commission. The overarching goal of Professor Spector's work is improving the governance of secondary research with health data and specimens. It's so great to have you here.

KAYTE SPECTOR-BAGDADY: Thank you very much. It's so great to be here. And it is now MY honor to introduce today's panelists. Professor Glenn Cohen is the James A. Attwood and Leslie Williams Professor of Law at Harvard Law School and current faculty director of the Petrie‑Flom Center. He is one of the world's leading experts on the intersection of bioethics and the law, as well as health law. A small sampling of the topics addressed by Glenn's current projects include big data, health information technologies, mobile health, reproduction/reproductive technology, and research ethics. He coleads the Regulatory Foundations, Ethics, and Law Program of Harvard Catalyst, the Harvard Clinical and Translational Science Center program. He also leads the Project on Precision Medicine, Artificial Intelligence, and the Law.

Professor Melissa McCradden is a bioethicist and assistant professor with the Department of Bioethics at The Hospital for Sick Children, SickKids. Melissa holds a Ph.D. in Neuroscience ‑‑ McMaster University ‑‑ and a M.HSc. in Bio‑ethics ‑‑ University of Toronto. Melissa's research focuses on novel technologies, including artificial intelligence, machine learning, precision health, and neurotechnologies. She is particularly interested in how to responsibly evaluate these new technologies from an evidence perspective, integrating research ethics and evidence‑based medicine.

It will certainly not be surprising to our ELSI Forum audience that technologies with great promise are inevitably paired with great challenge. These challenges include how to respect the people who engage in the research needed to validate advances, as well as use them in the future, limiting their harms and increasing their benefits, but also continuing to lean into our field's original conception of justice and demand that it does more than just protect from exploitation and not prevent access. Fortunately, in bioethics we do not purport to be tied to originalist conceptions of that which was respectful, helpful, and just in 1976, or reconceptualizing a regulatory framework still largely derived therefrom. Law and ethics are cyclical in this space and, where the law fails us now, ethics can be called upon to mitigate and improve across the spectrum of medical, scientific, and social progress. That is why this is such a critical conversation and Professors Cohen and McCradden are so well‑suited to lead us.

And with that, I will turn the floor to my colleague, Professor Cohen!

GLENN COHEN: Let me know, if you can't hear me, and ‑‑

KAYTE SPECTOR-BAGDADY: We can't here you. There you are. You have to restart.

GLENN COHEN: Thank you! We were told to give your best eight minutes of things we think you should think about. We divided and conquered. So here are my eight minutes. They are informed consent, liability, and privacy.

Informed consent. These, I think ‑‑ many people might have heard of PARO, the therapeutic seal robot. He's cute, he's cuddly, he's a little bit of a spy? He's a form of social robotics, which are experimentally being given to elderly patients in homes, including those with dementias. They both assist patients, but also give feedback and report back in some ways to physicians. That's one example of a more general problem, which is this idea that we're entering a future where artificial intelligence is all around us, but we are not aware always that we are dealing with an artificial intelligence. So imagine you're a patient who's been diagnosed with prostate cancer. Your physician is trying to decide between watch for waiting and a surgery. They recommend the surgical option. They spend time telling you about the risks and benefits ‑‑ eliminating the tumor, urinary incontinence. They don't mention, though, that part of the reason why they selected the therapy was based on an artificial intelligence analysis of your records, the records of millions of other patients, you know, tumor analysis, and all of this. Have you been harmed? Have you been wronged if someone hasn't told you? What if your physician doesn't tell you that they actually chose not to overrule an AI recommendation? Is THAT a problem?

So, come to think of it, how many of us know whether currently, the last time we had a medical encounter, what forms of artificial intelligence were involved in the decision‑making? This is I think a persistent question for bioethics and for the law: What right do we have to information about when AI is involved in our care? On the one hand, you could view this like a substitute surgeon scrubbing into a surgery, where we have legal cases telling us that if you fail to disclose that to a patient, you've violated their right to informed consent. On the other hand, we can think about all the other inputs in the original black box of the surgeon ‑‑ that is, things like medical school lectures, old issues of JAMA, the last 25 patients they saw, conversations with colleagues ‑‑ and we typically don't think that physicians have an obligation to disclose all of those inputs to reasoning to a patient. Is this more like one, more like the other? Does it matter whether the artificial intelligence agent we're talking about, ah, is interpretable or explainable as a form of AI? Does it matter whether it's been reviewed by FDA or others? So we can talk more about that. That's one bucket.

Second bucket! Liability and responsibility. There will inevitably be adverse events involving artificial intelligence ‑‑ even when artificial intelligence does what it does best, which is at a population level improve the quality of care or improve the access of care, in individual contexts there will be times when individuals are harmed... in that in a counterfactual world without the artificial intelligence recommendation, they would have done better. To what extent do we have a responsibility, and how do we parse that out, in particular with legal liability in this web of individuals involved? Developers. Hospital systems. So ideas about negative credentialing, for example, or an agency. Forms of tort liability. Physicians. And insurers. Is medical malpractice or tort law the optimal liability regime, as opposed to socializing the liability the way we've done with vaccine compensation funds, or, in some states, newborn injury funds? That is the idea that AI developers or adopters would pay in, and there would be a fund that was not a fault‑based ‑‑ basis for allocating resources to those who have been harmed. So decoupling compensation on one hand. Should a physician's liability depend on whether they can understand the AI? So is it fair to ask physicians to use non‑interpretable or non‑explainable AI, and nonetheless be responsible for the decisions it makes? How much of a right do physicians have to deviate from an AI, and how do insurers play into that?

Now, it's worth emphasizing the current setup of American tort law may actually be deterring the adoption of artificial intelligence. Under current law, a physician faces liability only when she or he (sic) does NOT follow the standard of care and an injury results. AI is particularly valuable when it gives us a personalized reason in a particular patient's case to deviate from that standard of care. But the liability regime is telling you you deviate at your own risk. And if you follow the standard of care, even if the deviation ‑‑ for example, a higher dosage ‑‑ would result in better outcomes for the patient, in the case where it does NOT, you are responsible. Whereas if you just stick to the standard of care, for the most part you will be, ah, free from liability ‑‑ or at least, a jury may not find you liable. So, on the one hand this is good news for physicians who don't want to adopt AI or want to be conservative. But on the bad news is, in a world where we think most of the value of artificial intelligence is that it's going to improve care, by causing physicians to deviate from the standards of care in appropriate cases, this is a huge disincentive for doing that. So we might view liability as at cross‑purposes with what the AI is doing.

Finally, I want to speak briefly about the relation of artificial intelligence to privacy. So AI is extremely data‑hungry. The larger the dataset, up to a certain point, the better the results you're going to get. It's quite difficult to reconcile the potential of big data with the need to protect individual privacy. We could engage in a form of data minimization ‑‑ so, limiting the upstream collection of personal health information, or imposing time limits on data retention. But we have to be kind of forthright about the fact this is going to be somewhat of an inhibition on the development of artificial intelligence, and in particular very good artificial intelligence. And here I think it's also worth kind of putting on the table the ways in which, you know, the more control we give to patients regarding when to opt out and when to remove their data DOES intersect a little bit with the bias of the dataset, because we have information showing us that actually the people who are least comfortable with sharing their data may also be the ones that are minoritized populations, who as a result we might have bigger gaps in the dataset. So this is just to say that this is not an easy problem to solve. Many people say, HIPAA! HIPAA is the answer to everything. But in fact! The American health privacy law, the practical health privacy law, HIPAA, is relatively kind of, I think, unhelpful in this area in the following respects: It, um, does a good job covering patient records and patient interactions with physicians, but it doesn't touch a huge amount of information that will be used in the future to predict health, and to predict health very well. Things like Fitbits and apps. Things like social media. Things like information from life insurers. None of these are touched by HIPAA. Indeed, we may enter a future where the best predictors and inferences of someone's health are based on non‑health care sources. Like Instagram filters and posting habits. Artificial intelligence analysis of that to predict depression, for example. So we really are in a future where if you say "I really care about privacy," one: that may cause some problems in terms of building these systems. And two: if that's what you care about, the existing federal protection in most of these areas is insufficient.

Lots more to say, but I'm going to stop there, and looking forward to your questions.

MELISSA McCRADDEN: Okay, I think that's my cue to give you my kinda eight‑minute spiel here, and hopefully I can do a good job of covering what is a really complex issue. So just for my disclosures, these are the sources of my grant funding. Now I'm gonna assume that everyone here is probably aware of this paper by Ziad Obermeyer and colleagues, which highlights the complications of using algorithms with respect to marginalized groups like Black and Indigenous individuals. If you haven't read this commentary, I think what she highlights so beautifully in her speech is that these are not just technical problems; these implicate the lives of real individuals, and this is an important issue that we really need to be looking beyond just the algorithms themselves.

And so she's called on us to beware the veneer of technical neutrality. And I think that's what I want to focus on a little bit more today.

So this has really come into question... historically, just based on looking at discrepancies in prediction patterns. However, THIS work has suggested we may be seriously underestimating the extent to which patterns based on, um, race and gender and differences in treatment may be implicated with seemingly benign data sources. So this paper found that using common imaging sources, AI could actually detect a patient's race. Which notably is something that humans can't. And so the comment on this was that it raises the stakes for what we might consider to be potential harms from algorithms.

So, in response ‑‑ or, not so much in response to this paper, but in response to, um, these issues in general ‑‑ my colleagues and I started exploring what are some principles that we can look to to help guide us in terms of what do we really need to know about algorithmic performance in order to look to issues of bias, look to issues of preventing harm. And so drawing from principles like justice, specifically racial justice, gender justice; looking to non‑maleficence, relevance, and institutional accountability ‑‑ we came up with the following recommendations. So one of which is disaggregation of model performance. And I hope that doesn't come as a surprise to many of you here. We recommend that there's actually two ways to go about looking at model performance. So one might be through the a priori selection of groups. So for instance, you might want to look at particular patient ethnic groups or particular diseases. And the other is a post‑methodology. This is really emphasized by my colleague's work on stratification, where we can explore how data is performing to identify clinical subgroups of patients on whom the data may perform less well than on the average. The other thing that's really important is granular data collection. And, um, you all have a bit of an advantage over us here in Canada, where you've started collecting, you know, race and ethnicity data. However, I also note that sometimes it's, you know, four racial categories, and I think we know that there's more than those. And another ‑‑ a really great paper that's come out recently by Sara Richardson on sex contextualism has also pointed out that to really move towards these approaches, we need to avoid binary... categorizing patients into binaries like male and female! Because we know that if we're actually interested in are things like hormones, different expressions of sexual characteristics, and different experiences that might influence how an individual's physiology responds or how they might psychologically respond, we need to look a little bit deeper than just these sort of surface‑level categorizations.

And, beyond actually just measuring how the model performs in order to move it into a translation space, we need to be looking at this on an ongoing basis. This has been termed algorithmovigilance, which might include routine auditing of algorithms that have reached the clinical space. Now, in addition to capturing these metrics on the performance itself, we also need to look to user guidance and training. So really enabling clinicians to be empowered to make decisions that center the best interests of patients, rather than, um... setting clinicians up for a strictly "follow or not‑follow" paradigm regarding artificial intelligence.

Now, in response to some of these issues of bias, the machine learning community has come up with a number of different methodologies to try and address bias. And so it's important to recognize that fairness, in this sense, is defined typically as the loss of accuracy, or perhaps different error rates between groups. And there's a number of different computational solutions. We might look to database methods ‑‑ so, increasing representation. We might look to model‑based methods ‑‑ so, changing how the algorithm functions as a whole. Or post hoc methods ‑‑ so different threshold selection. And what's really important to address about these is that these aren't just technical choices. They result, as Dr. Ruha Benjamin has stated, in real‑life consequences for individuals who may already be experiencing marginalization by the health care system.

So what we choose is really constrained by two things. One is on an epistemic level, where we need to know what we know and don't know about the causal nature of these unfairness patterns. Do these relate to differences in access? Does it relate to a lack of knowledge about true outcomes? Does it relate to different patterns that are associated with experiences of adverse childhood events, for example? And then second, we need to think about how we're gonna evaluate the model's real‑world performance! So it could be the case that if we say adjust for fairness at an algorithmic level, the model's predictions may or may not line up with what really happens to that patient, since that patient continues to live in a world that may be characterized BY barriers to access, or by their life influences over certain health outcomes.

So we need to make these decisions guided primarily by the desired intent of the model's use at the point of care.

So one of the examples that I wanna kind of take us through quickly is that of trisomy 21, or Down syndrome. So in this case, if we're trying to predict outcomes, say in the NICU or for intervention on cardiac procedures, we need to think about historical bias for these individuals. So, what data counts? We know that individuals with co‑morbidities are sometimes left out of the clinical research space, and so we might have less developed understanding of the context for those patients. We may also be affected by representation bias. So there's fewer numbers in general, which can result in a class imbalance within the dataset. There's also measurement bias, and this is something that is really beautifully highlighted in Chris Kaposy's book here, Choosing Down Syndrome, where he notes that the way we've chosen to measure persons with disabilities has been heavily driven by the medical system, where predictions of quality of life have been really influenced by a negative attitude toward people with disabilities. And when we change how those things are measured, such as by asking the individuals themselves or their families, we actually see a very different picture, generally one that is much more optimistic. And so the measurements that we're taking are affected by those biases. Additionally, when we're predicting outcomes we can run into things like the self‑fulfilling prophecy. So for example, if we routinely believe that people with trisomy 21 are likely to have less good outcomes, that may influence decisions regarding limitations of care or non‑escalation... that then result in a poor outcome for those patients.

We also have to think about which benchmark is the benchmark. There are regional practice differences in terms of whether patients with trisomy 21 or other genetic disorders are even offered interventions! And so, if we want to try and increase representation by pooling all of the data ON people with, um, with trisomy 21 together, we might be dealing with one hospital which treats trisomy 21 patients like all other patients and routinely offers all interventions that they would to ALL patients, and we might be dealing with other systems where they might not offer those similar interventions.

And so then how do we measure the correctness of the model output? If we are looking at a model that's trying to predict outcomes and there's a poor outcome in that patient, do we consider the model to have been wrong? Or might we think about the notion of a quality of opportunity, in that those patients were offered the opportunity to benefit from something like a surgical intervention? And it's important, 'cause there's similar considerations for trisomy 21 and others that have undergone significant sociohistorical change. So trisomy 21 was something that was NOT, um... for, those patients were NOT routinely offered interventions, you know, 20 years ago or so. We now know it's very different. And there's a similar change that's happening with trisomy 21 as well. (sic)

So one of the things that, um, that I think about sometimes is whether the use of machine learning can be reifying some of our present axioms about beliefs about things like outcomes, and that's a source of bias that I don't often see discussed, in the literature? So in thinking about how can we actually improve care might change the conversation.

So distributive justice, I think, is a useful principle to help us contextualize some of this. Distributive justice first points us toward the empirical characterization of the distribution of benefits and burdens. So from a machine learning perspective, we can think about things like accuracy or true positive, true negative rates, the error rates between groups. And that allows us to have an empirical foundation. Then we start talking about the actions that we can take. So do we need better representation? Is there an issue with the problem or the label selection here? What actions can be taken on those outputs? And are there ways to redress residual discrepancies?

So, as much as we talk about prioritizing accuracy... what if we thought about prioritizing justice, and not JUST accuracy?

Now, one of the important things here is that, ah, the role of clinical research has historically failed to help us do this sort of work. So, in the recent journal, American Journal of Bioethics, we had released our three‑stage pipeline that we use here at SickKids for the translational of machine learning products. One of the things pointed out by a number of the peer commentators was that historically, Institutional Review Boards have not done a great job of incorporating community or social harms into the review process. Canada has some guidelines in this respect regarding Indigenous patients, but overall it really hasn't been sufficient in the way that we would hope, and many people have noted this throughout history. So adapting local ethics review can enable some of these practices, like the ones we recommended regarding nuanced data collection and things ‑‑ can enable us to start measuring the distributive justice properties of algorithms as a local level sensitive to our own populations. And consistent with other equity‑based commitments within different organizations, we can then connect these with things like patient safety measures, to really ensure that where our models are working well for all patients, we're recording this, or where there are gaps that we can go ahead and address those.

So to try and move the clinical research space a little further down the path in this regard, my colleagues and I recently published this paper regarding what are called ethics methods. That's something I hope we can chat a little bit about during the Q&A, but really the idea would be to document and justify choices made with respect to algorithm design, development, or testing that implicate certain ethical commitments.

Really great example of how this is done is in this paper here, which we can certainly pop in the chat, where they compare different methodologies of addressing fairness. One of the important things to note is that they saw that there was a cost to accuracy at the expense of trying to improve equity. And I actually think these are really exciting conversations to have. What kinds of things can we actually do using machine learning to nudge us more toward equity‑driven care?

And, um, as we know, algorithmic bias is there because it's present in the current health care system. Colleagues have highlighted this in their paper, and a recent call from the American academy of pediatrics has noted we need to eliminate race‑based medicine. There's particular implications for machine learning here.

So just to note I'm not the first to point these out. Abeba Birhane ‑‑ this is the downside of preparing your slides in advance ‑‑ just the other day released a fabulous paper I hope is in the chat regarding the forgotten margins of AI ethics, where she and her colleagues note that though we talk about fairness, there's not a lot of talking about the actual harms to individuals and moving towards trying to address them. I think that's the next stage of the conversation I'm hoping we can have.

So just to say thank you so much. Really look forward to the discussion. And thank you to my collaborators who have informed this work.

KAYTE SPECTOR-BAGDADY: Great. Thank you so much, Melissa. We'll give Glenn a moment to come back. Those were great talks. We learned a lot. But let's learn more. Glenn, let's start with you. So you talked to us a little bit about liability. In your prepared... talk. But I was wondering if you could talk to us more about sort of, um, the law and regulatory policy of... those buckets as well. 'Cause I don't feel like we got as much on that.

GLENN COHEN: Sure, absolutely. So, what I like to say is that sometimes ‑‑ I mean, I'm not a system bioethicist; I'm a health lawyer. I talk to a lot of people who do AI in other areas, and they're often like, what do YOU have to add? It's just the same as driverless cars, right, or other areas. This is my best two minutes on why it's NOT. And here I would say there's a few special features about health care that are interesting and make it in some ways more interesting, more difficult. The first is that there's just a huge number of players between the person who develops the algorithm and the patient, in a way that's not with driverless cars. There's the developer. There's the hospital system that decides to purchase it or codevelop it. There's a physician on whom it's imposed or decides to use it or ignore its advice. There's a nurse who might be part of a union, and the union might have rules about deviation. There's a malpractice insurer, and there's a health reimbursement insurer, who will decide whether to pay for just what the algorithm recommends, or to pay for what the physician recommends even if it's at diversions of the algorithm. Each of those people have ethical and legal questions they face. And our policy solutions on one level? If we touch one level, we may actually exacerbate problems at the other. So you need to have holistic things. That's first cut.

Second cut is that the main regulator in the space people look to, FDA? Is not really that interested or experienced regulating software. And I don't blame them in that; in particular, artificial intelligence evolves ‑‑ we don't lock it, if it is an adaptive algorithm ‑‑ in a way that's quite different from small molecules, for example, or even vaccines. In a way that the existing kind of one‑and‑done existing FDA approval process is not a great fit.

Third cut: Most algorithms in the United States used in health care fall entirely outside of FDA's jurisdiction. That IS: Typically, there's no federal regulator looking at ANY of this. At best, you have a process like the one Professor McCradden was talking about, and Professor Spector‑Bagdady has talked about as well, regarding internal review. But that's largely an internal undertaking.

And fourthly, I'll throw out there something I feel ambivalent about: There's many ways in which artificial intelligence is actually like other things in the health care system. But patients have a very different feeling about it, at a kind of ground, normative narrative level. And there's a question for me here about how much to take that as a reason to structure our ethical analysis around it, versus saying, well, here's an analogy to this and to that, and it's not really that different so we can live with what that is.

KAYTE SPECTOR-BAGDADY: Great. Thank you. That was really helpful. So Professor McCradden, what do we do? No, I'm just kidding! (laughing) I won't do that to you. But you know, so you just sort of headlined the most recent issue of AJOB about artificial intelligence, and there are a lot of great articles in there... thinking about all these good ideas that you put forward, that I know there are a lot of links to in the chat. So can you talk to us more about ‑‑ so, Glenn highlights the limitations of our current governance structure. The kinds of things your team has been trying to do to sort of plug those leaks, so to speak.

MELISSA McCRADDEN: Yeah, thank you for that question. And I think, um, especially what Professor Cohen has highlighted is really that we may be seeing a much greater shift toward institutional accountability as the primary mechanism that tries to address these issues of harms? And in some ways, that's actually very much in fitting with AI, because we also know that there's, there's... so many challenges related to the generalizability ‑‑ whether that's even a desirable property in many cases? And HOW much local context is valuable for optimizing the performance of AI systems? So I think it's actually... I think it actually fits really well? Um, with the technical approach... that ‑‑ with the technical considerations specific to the machine learning space. And really, the way that we've tried to frame it is consistent with the, the movement within health equity for generally, which is looking at issues ‑‑ like bias, like different treatments ‑‑ as an issue of patient safety. And so, if we can use algorithms to help us increase the documentation of how different patients are treated, to try and more provide that empirical foundation for addressing the gaps, I think is really important?

And then with respect to the, the clinical research piece. So, we have a bit of an advantage in the fact that... pretty much all of our systems go through the same process. And one of the things that we're currently trying to establish is a, a standard set of... recommendations across different stages. So, in the exploratory stage, some of the things that we ensure all applicants who are going forward with machine learning models do... IS explore the model's performance with respect to different patient groups. We then replicate that through the silent trial process. So when we're running a model, at the point when it's not reaching clinical decision‑making space but it is being assessed against real‑world patient outcomes, that can help us better tailor the implementation period. So for example, we know if we find during the silent trial that there are certain patients on whom the model is just not performing well, but it might be performing well on others, as we move to translation you might give guidance to providers to say these sorts of patients, don't use the predictors for these sorts of patients because we have this. And we continue to work on that problem in the development space.

That being said, it's a very... it's a very BIG issue? And I think that... you know, the conversation about how to really... not just mitigate bias, but actually effectuate equity? Within machine learning models? Is something that, um... I think is possible! And, it just requires us to have a little bit of a different conversation with respect to how we are using some of these models.

KAYTE SPECTOR-BAGDADY: Yeah. Those are great ideas. Um. Although... you know. So, Professor Cohen. Between the two of you, I've heard ‑‑ so, it's not ‑‑ it's not up to the individual physician. Right? It can't be institutionally regulated. It's not FDA; it's not CMS. Then who? Who does the enforcement? Who ensures that those ethical ideas that you're putting forth, Professor McCradden, actually come into play and are actually applied?

GLENN COHEN: So, I tend to think ‑‑ so, again, it's ‑‑ yes, yes, yes, and yes. Right? But it's not that I don't want to absolve anyone of responsibility. But I DO think that if you ask me who is the ‑‑ so I am, again, very law school. Right? Least cost of order, who has the most say in this, who has the least control? To me, when institutions are adopting algorithms, they are kind of the locus of control. And I put a lot more emphasis, responsibility, and I would probably put more liability on them. Because they have the sophistication to evaluate, if they're large enough to purchase and to develop. And they also, I think, have enough control over the various pieces. But! Asterisk here, which has to do with exacerbating equity. Imagine you've got great algorithms that really do a lot of good work. If the idea is that every institution has to go through a significant process before adoption, resource pore‑setting ‑‑ which are the places sometimes where AI can do the most work ‑‑ may be the least well‑equipped to do this kind of analysis. So it would be great to kind of have clearinghouses, reliance agreement, share agreements between institutions to enable this kind of stuff so the work isn't done over and over again. But second asterisk, little cross‑cross here, work on bias and translation suggests to us even if you do a great RCT, great analysis at your institution, that doesn't always predict the model's performance at another institution. Because of both physician characteristics and patient characteristics. So it's just very hard to know if these things work in advance, and expensive and difficult? And in some ways that may box out certain institutions from adopting them. And that's not great either.

KAYTE SPECTOR-BAGDADY: Mm‑hmm. So just to come back at you one more time, 'cause I know you can take it. So Professor McCradden, in her most recent article, suggested that perhaps institutional IRBs might be the appropriate group to assess those algorithms. Who do you think at institutions should be actually doing the work? Is this a new committee? Is it in an IRB? Is it call research ethics and see if Kayte answers and knows what she's talking about? What should we do?

MELISSA McCRADDEN: I ‑‑ oh, sorry.

GLENN COHEN: Another ‑‑ oh, after YOU, Professor McCradden! Sorry, please!

MELISSA McCRADDEN: I didn't know if that was coming to me, or back ‑‑

KAYTE SPECTOR-BAGDADY: Both of you guys! Keep talking!

MELISSA McCRADDEN: Okay, so one of the principles that I think has been... WELL embedded into IRBs is the perspective that all of us bring a different lens to assess the scientific and ethical validity of a particular activity. So, research ethics boards, for instance, have embedded members of the community, can seek out local experts and do community engagement as a part of a review process. And so, I... it's not... I think that IRBs are well situated to do that, because that is currently within their mandate. And for the people who have historically, um, you know, kind of motivated IRBs to better include community‑based and social harms? I think that IS well situated? I do also accept the critiques that many things ‑‑ many AI systems may not come through an IRB pathway, and that's certainly important. But I wonder if we can bring the elements of IRB review that HAVE made it, you know, a, a foundation of good clinical research...! Into that, into that arena. If that makes sense.

KAYTE SPECTOR-BAGDADY: Mm‑hmm!

GLENN COHEN: I don't think we're that far apart on this. So I think, you know, the last thing Professor McCradden said, and particularly in the U.S. context where the QA/QI loophole is big enough to drive a truck through when it comes to artificial intelligence. Right. I think as a formal matter, much of this does not currently get reviewed by IRBs. Now, whether the expertise of IRBs are a good match for algorithms, I think that some portions of the questions that we want to ask are? But there are other portions that are not. And this has to do with ‑‑ so, let me put the question the following way: I've sat on a lot of IRBs in my life. When I think about the kinds of expertise I want on a typical drug trial or a community behavioral trial, right, sitting around the table, there is SOME overlap in the expertise I want in understanding a very complex algorithm ‑‑ and particularly something about, that's non‑explainable or non‑interpretable? But there's a lot of divergence as well. And part of me wonders whether there might be some division of responsibility, and some ways of training certain members of the IRB to sit on a committee that's also staffed with many more technical people. But I think a lot of it is going to be up to the institutions, because right now, unlike the IRB system? If you have something reviewing these algorithms, that's good on you, because you have someone like a Kayte Spector‑Bagdady at your institution, who is smart about this and knows that this is important! But it's entirely voluntary, and I think the same movements we've had towards getting IRBs institutionalized, it's worthwhile to think about what institutionalized ethics by design work looks like for algorithms.

KAYTE SPECTOR-BAGDADY: That's a great place for us to stop our mini discussion, because I'm getting so ‑‑ I want to be greedy and just ask all my questions. But I will allow the audience members to ask a few as well, 'cause I think we've gotten some great ones.

All right. So, to kick us off. So, I think that this is directed at you, Professor Cohen. How does the deviation use case for liability apply to "precision/personalized" medicine currently with the increase in Precision Medicine groups, particularly in oncology? Some AI/ML approaches contribute to the decision‑making already. What are we seeing? What can we predict?

GLENN COHEN: Yeah, so I should say that oncology is not a specialty ‑‑ so, I'm not a clinician by training, and oncology is not a space I've particularly studied closely. But what I will tell you is in some places, we are moving towards a future with the standard of care itself involves the kind of algorithmic analysis or machine learning analysis. One way to disrupt this tort law towards the standard of care would be a court world where, in fact, it becomes the standard of care to consider artificial intelligence in medicine. And just to give you a tangible example ‑‑ that might seem what, really? But there was a time in cardiology ‑‑ or EKGs, I should say ‑‑ where we didn't have the kind of algorithms that translate the raw data to the outputs we see now, and where people were training to read the raw data. Now, that's crazy. Nobody would ever ‑‑ maybe there's a particular use case where they would. But for the most part, we've had this move. And one could easily see this becoming the standard of care in other places. If there's proof that it adds a lot of value, and it can be demonstrated in a way that is transparent and that is measurable for patients.

KAYTE SPECTOR-BAGDADY: Great. Thank you. And I think this next one is a good one for Professor McCradden. So, you were talking about focus on accuracy as a potential harm for thinking about model validation. Couldn't the same be said as well for how "bias" is being measured? Are we, you know, are we sort of implanting that back in? Seems discussions around "justice" focus on the nature of the political context of medicine, and less about machine learning. What say you?

MELISSA McCRADDEN: Yeah, that's absolutely true. And, and to be honest, like, sometimes I do... question the extent to which, um, machine learning, and particularly continuous learning algorithms, are the right solution for a particular problem. And, I do think that it depends very much... on the particular use case that you're looking at? And, the context in which you are applying... a technical solution to that. So you can imagine, for example. Let's say we have a particular, um ‑‑ let's say we have a difference in referral rates, to surgical ‑‑ to, um, people who are candidates for surgeries. And say we have a difference between two demographic groups of patients. We might consider whether the use of a algorithm, ML‑based or not, can first push us toward equity. Once we have equity, THEN we might be able to apply more continuous learning approaches to continually be optimizing. But the issue is that when we have this sort of, when we have a status quo for certain kinds of things where there IS such a HUGE inequity... it's really challenging! To think about how... like, we could be further optimizing for some people and leaving other people behind. And I think that that really ‑‑ it takes a conscious decision, that needs to be made? And sometimes there are ways to compensate. So, as an example, if you have... let's say you have a model that works really well at detecting skin cancer on fair‑skinned people, but less well on darker‑skinned people. Is there a way that we can use that algorithm to streamline access for those for whom the model works the best, while also prioritizing people who ‑‑ on whom it works the least, for perhaps in‑person consultation? And to improve the resources of clinicians for accurately detecting skin cancer and other malignancies on people with darker skin.

So I think that it is ‑‑ and I totally accept that sometimes machine learning may NOT be the right answer to certain kinds of problems, and we have to think really carefully to what is the best approach to optimize outcomes in a particular case.

KAYTE SPECTOR-BAGDADY: Got it. But ‑‑

GLENN COHEN: Dr. ‑‑

KAYTE SPECTOR-BAGDADY: Oh, sorry, go, Glenn.

GLENN COHEN: Sorry, can I jump in for one second on this topic? So I want to pull out three things that I think are implicit in Professor McCradden ‑‑ maybe she doesn't agree they're implicit, but I think are worth surfacing. One is, again, what question. Which is just to say, the existing practice of medicine has huge amounts of racial and other kinds of bias in it. Right? And the perfect shouldn't be the enemy of the good. So for me, when I evaluate what artificial intelligence in medicine does, I ask myself how is it doing against existing practice, and how is it doing against an ideal of where I want to get?

Second thing is just to say, and I think this was in her slides more than this last answer, is that I think people ‑‑ so, if you're going to walk away from this talk? One thing not to walk away is the problem is just the datasets. Right. That's like bias 1.0? It turns out ‑‑ and Ziad Obermeyer's paper, I think, does a great job of getting at this: Problem's not the dataset, it's parameterization to cost. There's a huge number of things designed in these algorithms, intentionally or unintentionally, that produce different kinds of impacts, disparate impacts. Right? And that's really challenging, because when I say look at the dataset, okay, I can look at the dataset. But when I have to think about all of these things, it becomes much more of an ethical practice than it is merely a one‑step solution.

And the third thing to surface, and I think this is to me the thing I feel most ambivalent about. We focus a lot on race; a little bit on gender, which makes sense. It tracks us by constitutional law and classifications and the like. But when it comes to medicine, the fact that a model performs worse for some patients than other patients? That need not track race, gender, or even any other categorization that preexists us. It could be men between the ages of 46 and 49 who are above BMI X. Right? And one of the interesting questions, I think, is we have a lot of moral and also legal history of thinking about race and gender as special, and anti‑subordination, and reasons to focus upon it. Do those reasons carry over in the AI space? Or, we need to think about all these kinds of biases similarly or differently? I think that's an interesting, challenging question, of which I haven't myself come to a conclusion of what I think.

KAYTE SPECTOR-BAGDADY: Yeah, and then I was also going to add... Professor McCradden, when you made those points, the other thing that occurred to me is that if we DO offer additional information about the demographics of the communities who have validated our AI algorithms to help this guide clinical decision‑making, then we know who these advances are gonna be applicable to. Right? They're gonna continue to be applicable ‑‑ and I'm looking at our list of participants! All the good research many of our colleagues have done, to establish these incredible differences on the basis of different demographics... in these AI databanks. And so I worry it will just be sort of self‑affirming of, oh, we've only validated this on uneducated white people, and therefore that's only where we can help people. Which of course isn't your point, but makes it just more complicated.

One question I have for you, Glenn, that I think can be two combined. So, sort of going back to your first bucket, and. Any thoughts on how health systems trying to implement AI models should engage and involve patients on whom those models are being applied, without falling prey to tokenism from having a single early adopter patient as an advocate? What lessons can AI implementation... whoops, sorry ‑‑ learn from other areas ‑‑ IRBs, et cetera? I think is a good question for both of you.

GLENN COHEN: Great, so maybe I'll start, which is to say I'm a big believer that consent is the wrong way to go and governance is the right way to go. Which is just to say whether it's data collection or the actual deployment of artificial intelligence on patients. The more you put it on an individual to make a decision in their case, the more you're basically asking someone to make a decision which they are poorly equipped to do, and they are likely not to make a decision ‑‑ I'm not ‑‑ I'm very well‑aware this literature? And I'm not equipped to make these decisions! Instead what I think is the right way to go is governance, and I think it's what the question suggests. When you do governance, it's easy to see how you do well patient governance when you're talking about a rare disease group, for example. You partner with members of that disease group, you have representatives on it. You have to worry about bias and conflict of interest, but. It's easy to see how to go about doing it. When you have a artificial intelligence that's going to be run on the general population... here I think it's a little bit more challenging to know who the right representatives are. The IRBs are one model. A different model would be a kind of priortarian model, where you focus on those who are the worst off or the most likely to be negatively, adversely affected by this. So you think about the people who have the most skin in the game if this goes badly, and you oversample that. So that's an idea. Professor McCradden may have other ideas.

MELISSA McCRADDEN: Yeah, I love Dr. Singh's question, because this is something I'm quite passionate about as someone who works in a pediatric institution. Some of my work currently is engaging children and youth regarding their views on AI. And I think one of the really important things that sometimes we miss out on is that we always ask adults about their opinions and their thoughts. And actually, it's young people who are the ones who are not only engaging with technology in a way that we're not, but they're also contributing to shaping the evolving norms! Surrounding different ideas about privacy, and ideas about social value of data, and things like that. And so, one of the things that, um, a model that I really like that I know you, you use a fair bit in the United States, is like a ‑‑ I believe, a community accountability board, or a CAB? I think is the correct acronym. So, we've thought about using something like that, not in terms of kind of the tokenistic method that I think Dr. Singh is concerned about, but in considering things that we prioritize for translation. And in considering some of the best practices surrounding translation, in terms of what Professor Cohen was saying about governance. So, can our young patients, whose knowledge of being health care patients themselves is so ‑‑ is so rich? And provides such a valuable, um... epistemic advantage. That they can contribute to saying how procedurally things should go, and considerations for all of us. Rather than sort of the, here we made this AI model; do you like it? Um, sort of approach.

KAYTE SPECTOR-BAGDADY: Yeah. Great. Um, so talking practicality. So how we put, um, the rubber to the road, so to speak. I'm gonna skip down a couple questions, 'cause I think that this is important. And also because I've heard this debated a lot in other circles, and I'd really like to hear YOUR answer. But it's a question about sort of how collecting more granular data ‑‑ particularly on the basis of things for which people have been discriminated against ‑‑ is a good ‑‑ whether that's an appropriate answer to combat AI bias. For instance, are we really in better shape in the U.S. because we've been collecting "race" for a long time? The collection of "race/ethnicity" categories is also deeply entangled with systems of racial oppression and inequity. Is seeking more granularized categorization really the best way to approach this? I think is a good question, and let's do that. But also! With the past few weeks in mind... sure! Right now, we might be collecting these data in order to help people, and to do people good! But we can think back, um, on data that we have generated in the past, and in the future might be used to harm people! I'm thinking of ‑‑ you know, for example, in Michigan, we have a trigger law on abortion that has a statute of limitations that goes back six years! You know? So something that somebody might have done under one set of circumstances, then, might be prosecuted on another. And that was a big hard question? But you are big smart people, so. Love to hear your answers.

MELISSA McCRADDEN: Yeah. That is an excellent question, and I will be totally frank in saying that... the past few weeks, seeing what's been going on, has actually very much... you know, called pause to some of the things that I thought would BE the best practices going forward. And I really, I honestly don't have a clear answer for that at this point. I think that, you know, as a woman, as a member of the LGBTQ community, like, this is something that I'm deeply concerned about? I... the notion of ‑‑ you know. Most of my life, I've been like, sure, I'll volunteer my data! I'm happy to have it go toward good. Um, and there's sort of a different ‑‑ you know, when you realize that you might be on the receiving end... um, of some concerning practices, it really DOES change the stakes. And I think that that's an incredibly important point. And so, I think that goes back to, you know, what Professor was ‑‑ Professor Cohen was talking about a little bit, with respect to do patients need to be informed? About the use of AI. If it's involving this sort of data collection. Um, I wonder! I wonder if this is ‑‑ if we might kind of think about particular... um, situations wherein stigma, or harms and concerns about discrimination, might be certain kinds of use cases where we MIGHT need to think about whether patients... have to choose! Whether or not their data is used in these algorithms. And I realize that's kind of a, a big bucket of worms? And would be very difficult to implement? Um. But I do think that, you know, especially ‑‑ a lot of the releases about, you know, using... menstrual tracking apps, and things like that, and the kinds of information that's tracked in there. It really has changed the stakes. And I'm sorry that I don't have a better answer for that at this point.

GLENN COHEN: (no audio)

KAYTE SPECTOR-BAGDADY: Oh, we can't hear you ‑‑ oh, there you go, you're back.

GLENN COHEN: Oh, good. I'll jump down. Also being a man talking about these issues, right? So it's, ah ‑‑ I want to be careful about, you know, kind of my position on this and not kind of speaking for people. One thing I'll say, more generally, maybe, is the following. That, at a philosophical level? Some people are worried about privacy; other people are worried about inferences equally. If you're worried about inferences about you, right, then I think it's actually just a much harder problem. Right. If you say, I don't ‑‑ and I use the example ‑‑ pregnancy's actually a great example, right. Someone wants to know you're pregnant. They break into your physician's office and steal your test results. Versus, they make inferences based on changes in your body weight, changes in your flush pattern, changes in your alcohol consumption, and the like. That doesn't involve a privacy violation, per se, on a health level. Instead, it's a lot of inferences about you from other sources of data. I know how to protect against, ah, privacy? To some extent? Whether we'll do it is another matter. It's much harder to have a legal regime that protects against inferences ‑‑ and the ones that do it the best! And I know we're talking to a lot of geneticists in the room. Is kind of genetic discrimination law. But, it has a very ambit. Right. So for me, when I think about a lot of these issues, if your main concern is about adverse treatment on the basis of inferences about you, I would put more of my eggs in the bucket of antidiscrimination law. But of course, that requires either legislative or judicial common law development, at a time when those are also very difficult to do in this space we're talking about.

KAYTE SPECTOR-BAGDADY: Yeah. Thank you guys so much. Not to derail us, but I think that these issues are all interrelated and important.

Um, okay. So we have four minutes left. So I might bring us back to where you started us, Professor Cohen, in talking about the patient and the informed consent conversation, and what that should look like. Because I know we're in a context where we keep piling more and more and more onto the informed consent conversation ‑‑ particularly about secondary uses. You want to commercialize? Well, you know, the patient going in for their cancer surgery, they should know about commercialization! Right? What about like the AI algorithm's bias? Can you... talk to us more specifically about what these informed consent forms should include, what this conversation should be? Is it okay that at the University of Michigan, we say this is, um... we are a learning center? There's learning here? Does that, does that cover it, if patients sign it?

GLENN COHEN: Yeah, so first let's talk ‑‑ there was a question in the Q&A about the current practice. The current practice is to say very little about the subject. And the place that it deviates is where they want to advertise it. Like the da Vinci. I've never looked, but I guess that's explained to a greater extent.

Now, about what the world SHOULD look like that in respect. This is the way I think about it, at least. I think there are many forms of artificial intelligence that are not that different from things physicians are already doing in their minds, or decision aids. Where we deviate from that to patients where I think the patient should be worried is where I'd want the most informed consent. So for me, when we're talking about things that have not been well validated? Right, they don't necessarily have to be interpretable or explainable, but they have to be validated and well tested. And the applications in a particular instance have high risk to patients? Those would be the instances where I'd like to see more disclosure and more informed consent, going along with what we do for other more high risk or experimental forms of therapy. So I would probably bisect the world of AI to less validated and treat them more like other medicines like that.

KAYTE SPECTOR-BAGDADY: Great, thank you. And Professor McCradden, last word. Or a few.

MELISSA McCRADDEN: I think ‑‑ (chuckles) No, I think that was a great point. I also think there's an element to which we should be asking patients and families what their preferences are. And not to say that, you know, whatever they say goes. But to think about where they're coming from, where they're meeting us with respect to this conversation about AI? Because it's also a possibility that, to address some of the concerns that maybe aren't, um, manageable within the context of a single clinical encounter, we could consider repositories of information available on hospital websites that can inform about AI, inform about issues like bias. What do we do to protect your privacy? How do we de‑identify? And things like that. And so I think really connecting with our patients and families to better understand what their needs and wants ARE can help us figure out a way to enhance, ah, their ability to consent, or assent, or give permission. Or, simply, accept and endorse, the activities that go on at learning health care systems.

KAYTE SPECTOR-BAGDADY: Great. Thank you guys so much. I'm going to turn again to Professor Cho to clear ‑‑ to... bring us out.

MILDRED CHO: All right! Wow! You all really hit it out of the park today. Ah, so thank you to, ah, the panelists for their presentations. And to Kayte for her wonderful moderating. And I do wanna remind everyone that the recording of this, and all the resources posted in the chat, will be available at ELSIhub.org. I think I'm gonna wanna go back and look at this again. It was amazing. So, for those who can, please join us in our post‑event discussion. The link will be posted in the chat. There it is. And, see you on June 17th. So this is gonna be the third Friday of the month, not the second. June 17th, next time. For our ELSI Friday Forum, for The Impact of the Dobbs Decision on ELSI Research and Practice. So you'll also receive a survey on how to improve this Forum. Please do fill it out, if you can. This is how we know to bring new topics and speakers to you. Hope to see you in the post‑event discussion, and I wish you a wonderful weekend. Thank you.