An Expert Panel Discussion

Embedding Ethics and Equity in Artificial Intelligence and Machine Learning Infrastructure

Moderator

Malaika Simmons, MSHE

National Alliance against Disparities in Patient Health

AIM-AHEAD, sponsored by NIH

Expert Panel

Malaika Simmons, MSHE (moderator), is the Chief Operating Officer for the National Alliance against Disparities in Patient Health (NADPH), a data-driven nonprofit health research organization. She uses her background in research, psychology, and design thinking to promote empathy-based leadership with over 20 years of experience in corporate policy, training, and program management, operating at the convergence of health research, IT, and consumer issues. Malaika has a strong commitment to improving the health of underserved populations and more broadly supports these efforts through her organization Momentology Media. Through Momentology, Malaika fulfills her passion for the elimination of economic and health disparities in minority communities by using her proprietary framework to empower women to own businesses, champion health causes, and enter the corporate and political landscapes to affect change in their corner of the world. The NADPH is a nonprofit health research organization applying scientific research and technology, community education and outreach, and advocacy to reduce health disparities and improve public health outcomes for our nation. NADPH operates through a national partnership ecosystem that spans government, academic, nonprofit, and industry stakeholders. One of NADPH's major objectives is to bridge the gap in access to advanced precision-based health care in underserved and under-resourced communities, NADPH seeks to do this by providing access to information, subject matter experts, and technology that help community members: (1) raise their awareness and education about the array of factors impacting their quality of health and health care; (2) capture, manage, and understand their health information across its various domains; and (3) use this increased knowledge and access more actionably to make better decisions about their health and health care.

Rachele Hendricks-Sturrup, DHSc, MSc, MA, is a health scientist, engagement expert, and bioethicist serving as Chief Data Governance Officer at the NADPH. Her research and practice centers on examining and addressing ethical, legal, social, and implementation issues at the forefront of policy and innovation. Dr. Hendricks-Sturrup's research and work at NADPH involve engaging persons with lived experience (PWLE) in inequity to provide them with access to information, subject matter experts, and technology that can help PWLE understand the broad array of factors that impact their health, healthcare, and well-being.

Gabriella Waters is an artificial intelligence and machine learning researcher and the Director of Research and Operations at the Center for Equitable AI & Machine Learning Systems at Morgan State University in Baltimore, MD. She is also the director of the CoNA (Cognitive & Neurodiversity AI) Lab. She is passionate about increasing the diversity of thought around technology and focuses on interdisciplinary collaborations to drive innovation, equity, and ethics in the development and application of AI tools. In her research, Gabriella is interested in studying the intersection between human neurobiology/learning and intelligent systems that make use of that foundation for improved human–computer synergy. She develops technology innovations, with an emphasis on support for neurodiverse populations, as well as the creation of more robust systems that function in the absence of human interaction.

Laurie Novak, PhD, MHSA, FAMIA, is associate professor and director of the Center of Excellence in Applied Artificial Intelligence in the Department of Biomedical Informatics (DBMI) at Vanderbilt University Medical Center. Her work is focused on the design and implementation of artificial intelligence (AI) and other informatics tools in a variety of clinical environments. She earned a PhD in anthropology from Wayne State University in 2005, and a master's degree in health services administration from the University of Michigan School of Public Health in 1994. Dr. Novak has collaborated successfully with data scientists, computer scientists, engineers, clinicians, ethicists, and humanities scholars. She has led research funded by the National Library of Medicine, the Robert Wood Johnson Foundation, the Baptist Hospital Foundation, and the Stead Foundation, and co-led or participated in studies funded by NSF, NIH, PCORI, AHRQ, CMS, and IBM Corporation. She is an active mentor for graduate students in biomedical informatics and directs courses on technology and society, and workflow, user centered design, and implementation.

Martin Were, MD, MS, FIAHSI, FAMIA, FACMI, is a professor of biomedical informatics and professor of medicine, vice chair for diversity, equity, and inclusion within the Department of Biomedical Informatics at Vanderbilt University Medical Center (VUMC). He has over 15 years of experience in developing, implementing, and evaluating a range of health information systems that are relevant to resource-limited settings, including the ethical and equity implications of such technologies. Current projects include use of artificial intelligence/machine learning (AI/ML) and bioengineering approaches to develop bloodless smartphone-based hemoglobin and bilirubin analyzers. Dr. Were is a co-investigator within the Ethics sub-core of the NIH-funded Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity, and in the Ethics core of the Bridge to Artificial Intelligence (Bridge2AI) program.

Sajid Hussain, BSc, MS, PhD, is associate provost for research and the director of the Office of Sponsored Research and Programs (OSP). In 2009, he joined Fisk University as an associate professor in the Department of Mathematics and Computer Science. Before Fisk, he worked as associate professor and assistant professor at Acadia University, Canada, 2005–2009. He received a PhD in electrical engineering from the University of Manitoba, Canada, in 2004. Dr. Hussain is interested in applying machine learning techniques for interdisciplinary research projects related to health care, bioinformatics, digital humanities, and criminal justice. He has published more than 80 refereed peer-reviewed articles. His research is financially supported by several grants such as $1.17M—NSF NDSA, $1.28M Department of Energy—Environment Management, and $2.25M—NSF Implementation Award. He has co-organized several journal special issues, conferences, and workshops. He is a senior member of IEEE.

Open camera or QR reader and
scan code to access this article
and other resources online.

Embedding Ethics and Equity in Artificial Intelligence and Machine Learning Infrastructure
National Alliance against Disparities in Patient Health

Moderator: Malaika Simmons, MSHE[1]

Participants: Rachele Hendricks-Sturrup, DHSc, MSc, MA,[2]
Gabriella Waters,[3,4] Laurie Novak, PhD, MHSA, FAMIA,[5,6]
Martin Were, MD, MS, FIAHSI, FAMIA, FACMI,[7,8] and Sajid Hussain, BSc, MS, PhD[9]

[1]*Chief Operating Officer and Human-Centered Design Expert, National Alliance against Disparities in Patient Health, Woodbridge, Virginia, USA.*
[2]*Chief Data Governance Officer and Project Director, National Alliance against Disparities in Patient Health, Woodbridge, Virginia, USA.*
[3]*Director of Research and Operations, Center for Equitable AI and Machine Learning Systems, Baltimore, Maryland, USA.*
[4]*Director of the Cognitive and Neurodiversity AI Lab (CoNA), Morgan State University, Baltimore, Maryland, USA.*
[5]*Associate Professor, Biomedical Informatics,* [6]*Director of the Center of Excellence in Applied AI, Vanderbilt University Medical Center, Nashville, Tennessee, USA.*
[7]*Professor of Biomedical Informatics and Medicine,* [8]*Vice Chair of Diversity, Equity, and Inclusion, Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA.*
[9]*Associate Provost for Research, Fisk University, Nashville, Tennessee, USA.*

The National Institutes of Health's (NIH) Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD) Consortium was established in 2021 with a mission to address factors that undermine achieving health equity through the design, use, and application of Artificial Intelligence/Machine Learning (AI/ML), including the lack of the following:

- An adequately diverse workforce
- Adequate data and data infrastructure
- Adequate community engagement
- Adequate oversight, governance, and accountability
- Consensus that ethics can strengthen innovation.

The National Alliance against Disparities in Patient Health (NADPH) leads the Infrastructure Core within

EXPERT PANEL DISCUSSION

AIM-AHEAD, which functions to assess AIM-AHEAD AI/ML system user ''needs and constraints and pilot and test different data and computing infrastructure, tools, and governance models including data policy and organizational models.''[1]

NADPH also co-leads the AIM-AHEAD Ethics and Equity Workgroup in partnership with faculty at institutions within AIM-AHEAD's Infrastructure Core. This collaboration importantly serves to advance AIM-AHEAD's ethics and equity initiatives through faculty, student, and researcher engagement on the topic of embedding ethics and equity into AI and ML infrastructure. This roundtable is a continuation of these discussions with thought leaders and experts at Fisk University, Vanderbilt University, and Morgan State University.

In the spirit of AI discovery and exploration, this roundtable includes responses generated by ChatGPT (enclosed in the addendum), an AI chatbot, using questions provided by the AIM-AHEAD Infrastructure Core with comments provided by the roundtable moderator and participants. Our purpose for including ChatGPT in this roundtable is to give readers an opportunity to compare responses that emerged from interactive discussion among people to AI-generated responses to complex questions on embedding ethics and equity in AI/ML infrastructure. We note that the ChatGPT responses do not offer sources for the information provided; therefore, we are unable to identify plagiarized text (see Supplementary Material online at www.liebertpub.com/doi/10.1089/big.2023.29061.rtd).

Mrs. Simmons: **My name is Malaika Simmons. I am the Chief Operating Officer for the National Alliance against Disparities in Patient Health (NADPH) and co-lead on our qualitative research work; I am also a Program Manager for the Infrastructure Core at the National Institute of Health-supported AIM-AHEAD program. I am extremely proud and grateful to be a part of this forum today with some of my illustrious colleagues.**

Dr. Hendricks-Sturrup: My name is Dr. Rachele Hendricks-Sturrup. I am the chief data governance officer and a project director at the NADPH. As a bioethicist and data governance expert, I direct our mixed methods research projects and engagement initiatives and oversee our data governance policies and procedures. I am also the cochair of the National Institute of Health-supported AIM-AHEAD Ethics and Equity Workgroup.

Mrs. Waters: My name is Gabriella Waters. I am the director of research and operations at the Center for Equitable AI and Machine Learning Systems (CEAMLS) at Morgan State University in Baltimore, Maryland, where I am also the director of the Cognitive and Neurodiversity AI Lab (CoNA). The research being conducted at CEAMLS touches on every discipline, every industry, and anything in which AI can possibly have a role. I have the very fortunate luck of being the person who is the filter for all of these different topical areas and helping to guide the researchers and support them in the work that they are doing. I am very excited to be a part of this discussion. My own research focus is more on cognitive and neuro-symbolic systems.

Dr. Novak: My name is Laurie Novak. I am an associate professor of Biomedical Informatics in the School of Medicine and Director of the Center of Excellence in Applied AI at Vanderbilt University Medical Center (VUMC) in the Department of Biomedical Informatics (DBMI) in Nashville, Tennessee. I am an anthropologist by training, and I have a master's degree in health services administration as well. I conduct ethnographic research on the implementation of technology in organizational settings and have also done a lot of research on the experience of chronic illness in everyday life. All of my research and experiences are coming together now to focus on implementation of AI. I work on the ethics cores in the NIH initiatives *Bridge2AI and AIM AHEAD*.

Dr. Were: My name is Martin Were, professor of biomedical informatics and medicine at VUMC in Nashville, Tennessee, and vice chair of diversity, equity, and inclusion within the DBMI. My work largely revolves on clinical informatics and global health informatics, and increasingly working more in the equity space.

Dr. Hussain: My name is Sajid Hussain. I am the associate provost for research at Fisk University in Nashville, Tennessee. My background is in data science. I am also collaborating with Dr. Talitha Washington (PI) at Clark Atlanta University, Atlanta, Georgia and Dr. LaTanya Robertson (Co-PI), Howard University, Washington, DC, for the National Data Science Alliance (NDSA). It is an NSF INCLUDES project.[2] The goal is to train 20,000 African American data scientists by 2027. We are working with all 100 Historically Black Colleges and Universities (HBCUs) for capacity building. I am excited to be part of this amazing initiative. I am from the computer science department, working in AI ML algorithms, and collaborating with other colleagues in social sciences and life sciences to apply these AI/ML (ML techniques). In the past, I had more of an engineering role working on communication protocols and sensor networks. But I am currently more interested in applying AI/ML—essentially a complete protocol stack from top application level to hardware level—and really enjoy working with colleagues from different areas.

Mrs. Simmons: **We definitely have a very diverse and eclectic group here, which is fantastic. For the first question, let's consider engineering infrastructure, which can be both structural and digital. With AI and ML infrastructure examples in use range from automated robot surgeries to advanced data analytics and health research. What comes to mind in terms of applying ethical practices across these examples based on your subject matter expertise?**

Dr. Hendricks-Sturrup: As a bioethicist, I give a great deal of consideration to questions like these. And as a member of the AIM-AHEAD Infrastructure Core, I have had to think even more about how to apply bioethics to the practice of developing AI in ML infrastructure. I think ethics starts at the infrastructure development level, not only at the data analysis level, or even as many do today, considering ethics as an afterthought to both of those processes in AI ML development and use.

So, what does that mean? Ethics and equity, within the scope of AI/ML infrastructure, means that we are engaging powerful and less powerful stakeholders, as well as neutral third parties, very early in the process of AI/ML development and system implementation, offering and protecting their seats at the table. That is, creating and protecting safe spaces for them to engage in transparent and collaborative discussions, address power imbalances and/or information asymmetries, and contemplate potential downstream impacts or effects across the range in which AI/ML tools would be used.

Mrs. Simmons: **Neutral third parties are particularly important to help mitigate power imbalances through effective moderation and translation of differing views, and also ensure multistakeholder discussions do not drift away from critical questions: Are we being intentional about designing AI/ML technology such that it can reach and serve the most vulnerable stakeholder in the room? Will our actions leave our most vulnerable stakeholders behind or place them at a direct or indirect risk of harm if their mere existence or lived experiences are devalued or not considered in the process?**

Dr. Were: To answer this question, it is important to define and agree about what we mean by AI/ML infrastructure because people might have different definitions of it. AI/ML infrastructure is relevant to all the stages of AI/ML lifecycle—from data acquisition, data preparation, algorithm development, and deployment. The infrastructure, then, involves data storage and management, the computer resources needed, with considerations on security and fault tolerance. Infra-

structure also touches on the networks and the ML operations platforms that are needed to run the AI/ML models once they are implemented. The infrastructure must be secure and scalable, the right governance structures must also be in place, and the right team must be assembled to manage each element of the outlined AI infrastructure. And then when you talk about things like automated robotic devices, the concept of AI intelligence of things, where you combine AI technologies with the internet of things also becomes relevant in this discussion.

When you think about the life cycle approach to AI/ML infrastructure, it is important to recognize that ethical issues can arise at each stage in the life cycle. Second, ethical issues have to be considered at multiple levels. Dr. Hendricks-Sturrup and her colleagues have done work in this area, examining the ethical issues at an individual level, at an organizational level, and also at a societal level. Each level brings up slightly different considerations around ethics, which I am sure we will discuss in detail. A different way to think about the ethical issues is to consider issues that are cross-cutting across the life cycle such as security; support of the infrastructure for consent and provenance of the data; governance mechanisms at each of those stages; the cost-benefit of the infrastructure, especially when compared with other proven things in place; the risk distribution for the infrastructure; or the equity issues that arise.

Dr. Hussain: What comes to my mind is that all stakeholders and all aspects should be covered. If you take robotic surgery as an example, we cannot think about the patient in the same way that we think about the doctor. So, we need to consider the concerns of all stakeholders, including their privacy concerns, as mentioned earlier, are at different levels.

We should be conscious about bias and it should be explainable. It is not a black box where we do not know what is happening inside. It should be traceable and transparent, so we know how these models are developed. That is a challenging problem; we must be conscious about privacy. For example, as we are simulating the movements of the surgeon with these devices, we need to collect lots of doctors' data, and at the same time, the patients' data. We want our device to act like a surgeon, but at the same time, preserve the privacy of all the surgeons from whom we collected data. I believe that becomes much more important when we have autonomous systems. Ownership, reproducibility, and liability are big issues as well in these AI-driven automated systems, as are licensing and patient involvement accountability. We need to apply our AI ethic principles at all levels, stages, and to all elements. If I go back to our critical thinking model, just like we have the Paul-Elder model,[3] and

it says, to apply intellectual standards to all elements of reasoning (questions, points of view, assumptions, etc.) to develop intellectual traits (humility, fair-mindedness, empathy, perseverance, etc.); similarly, we need to apply critical thinking to elements of your thought, of your concept, not only as a whole, but also on each individual piece. And coming back to our software engineering life cycle model, I will say, it should be an iterative approach. It is not like a waterfall model, where the requirements and design are not reconsidered, but is an iterative process for continuous improvement. You analyze all these different stages, and then you go back and revisit it again.

Dr. Novak: I am going to point us to some of the scholarly research that has been done on infrastructures. There was some work by Leigh Star and Karen Ruhleder on infrastructures that was funded by the NSF. This has been documented in the information systems literature[4,5] and in other places. There is a classic article called *Steps Toward an Ecology of Infrastructure Design and Access for Large Information Spaces* by Star and Ruhleder,[6] which talks about characteristics of infrastructure. This is a nice layer to add onto the framework that Dr. Were laid out already. If we think about roads as infrastructure, for example, you need to know how to use a road. You need to know not to get on the freeway on your bicycle. You need to know which side of the road to drive on. There is a lot about driving that comes with experience, not that comes with just learning from a textbook. We all have had those experiences. A couple of important insights from Star and Ruhleder are how infrastructure is *learned as a part of membership* and infrastructure is *built on an installed base*. I described that with roads. All of us who are drivers are part of a community of practice. We all understand how to interact with each other in cars, and that we are all driving on roads that have been built up over time from previous roads. I think these concepts apply to AI infrastructure as well.

If you are part of an organization that already has a big research infrastructure and you have already been doing analogous research, such as genomics, for example, you can build on that installed base. You have a set of researchers who really understand how to interact with each other, and how to access services, use those resources to generate grant proposals, and really expand your role in AI. If you do not have that installed base, if you are not a member of that community of practice, you must find your way to enable your own access to this infrastructure without accidentally getting on the freeway on your bicycle.

I think that problematizing these characteristics of infrastructure and thinking through how those char-

acteristics apply to AI can help us identify ethical issues that may not surface in other ways. These concepts can surface inequities in design, inequities in implementation, and inequities in use.

Mrs. Waters: I believe part of the framework that we need to look at has to do with the end user, the patients, and how they are impacted by these systems. While I respect the entire process of the AI/ML infrastructure, we still need to keep pushing the idea forward that there are real people at the end of these systems. The framework must do several things. It must respect autonomy and provide informed consent, it needs to ensure fairness and equity, it needs to be transparent and accountable, there must be some measure of data privacy and security, and there must be some environmental ethics. So, if I put each of these into the framework on the cognitive science side of things with the user in mind, when you are looking at autonomy and informed consent, there is decision fatigue and cognitive overload. As an example, you may have been going from form to form and input after input after input. After a while, you are tuning out, and that is impacting saliency and attention. Are people paying attention to what is relevant at that point? So give them what is relevant upfront in your architecture. Highlight the key aspects of the AI's involvement. Make sure patients really understand their choices. Instead of burying things in the terms of service or use, put the patients' needs first to help them to make the decision early on. Don't saddle them with a lot of extraneous information.

We must also ensure fairness and equity—we already know about bias in AI, but there is cognitive bias as well. ML algorithms are trained on biased data in many cases. These intelligent systems are in environments with people who have certain cognitive biases in mind already. Your infrastructure and your architecture must take these biases into account so that patients receive the best care possible.

There is blind trust versus critical trust. We need to make sure that the trust that we attempt to cultivate in AI and technologies that use it is not blind. We cannot promise that these systems are a panacea for all ills. Users must be educated enough to question how the AI comes to a decision so that they can make better choices. We must also consider our cognitive footprint. This is in the environmental ethics area, and I want to touch on the concept quickly. Because there is a cognitive footprint—or mental strain—that a tool can bring; an AI system can be designed to make it easier or make it harder for a user. If you are ethically deciding on how to design your systems, your AI should reduce the cognitive load, streamline the task, and present the information in a digestible way. That needs to be considered in ethics as well.

Mrs. Simmons: **Thank you. What kind of evidence or use cases might support the development of ethical principles or standards to guide AI/ML infrastructure development and implementation in health research settings?**

Mrs. Waters: There is nothing analogous in health care to an AI institutional review board (IRB). We have an IRB to protect human subjects for a reason in research, but we do not have that in health care, where human subjects' lives are essentially in people's hands. How do we have a diagnostic for bias and misrepresentation? What are we doing in those areas with predictive health analytics, with these kinds of things in mind? All of our AI-driven predictions influence choices made by doctors and patients. The predictive analytics side is typically on the provider's side and not on the patient's, which means we are not looking at a full picture. We are not getting or providing enough information for patients to decide whether or not this intervention is the best option, if it is recommended by the AI, if it is an AI-supported decision, or things of that nature. People are susceptible to the availability heuristic. AI systems can exploit that tendency or leverage to place relevant information in concise chunks in a prominent location.

Privacy in mental health predictions—there is a sanctity to the individual thought process. What is happening with AI is, as we continue to train it on individuals' thoughts and patterns of behavior, that becomes a commodity. We are not ethically considering that every patient interaction that informs the model and trains it further is infringing on that sanctity of thought. We are not giving permission to allow this to happen or to restrict some of this model training. We are just moving forward in the realm of AI assisting. ''*Assisting*'' is the key term. We want these systems to be an *extension of the human*. So, in the case of something like robotic surgery, you are not replacing the surgeon, but the technology should seamlessly integrate with that surgeon's expertise. We still need the human in the loop.

This segues into personalized medicine, where we are big fans of adaptive treatment recommendations—all of these require cognitive insight that we must consider as we ethically derive these models.

Mrs. Simmons: **Is there any use case, or what sort of evidence, supports the development of ethical principles in health research settings?**

Dr. Novak: I am partial to implementation research that uncovers tensions that we might not have anticipated. When you have a tool that is being implemented, one of the first questions I like to ask is who is this making more work for? And—spoiler alert—it is almost always the nurse!

So, how do these tools impact workflow? That is one thing we need to understand. Another tension is, as Mrs. Waters mentioned, related to autonomy. For example, in clinical decision support, we may be asking a provider to rely on a potentially highly accurate prediction tool that is nontransparent, that is, the provider cannot see how the algorithm works? We need to understand that tension. Giving providers more autonomy through creating new ways of visualizing explainability that helps them understand why the algorithm is saying what it is saying—maybe that's where we need to go.

I think we need evidence and findings from rural cases. We need information from under-resourced settings as well as highly resourced settings so we can understand the impact of the technological sophistication of the environment on the safety and use of the tools. That will tell us something about what management infrastructures we need to put in place in all settings to keep patients safe. We need evidence with stakeholder perspectives. We need to start working more closely with community members, patients, family members, to fully understand—what do you want to know about these tools that are being used in your care?

Dr. Hendricks-Sturrup: I should say that I have published some work already[7] on the topic of what should be required or contemplated to build a system of health. I approached this question through a structural engineering systems design lens. I do think this same lens can be applied to AI/ML development as another area of engineering system design within the broader system of health.

In my estimation, and perhaps to Dr. Were's earlier point, there are three layers to consider to identify where ethics plays a role across the entire system of health: the micro level, meso level, and macro level. At the micro level, there is the patient–doctor interaction or the health researcher and health research participant interaction. Then we think about the meso level—navigating an encapsulated system within the broader system. That is, how the patient and/or provider navigate the buildings in which they interact and the structural systems they are required to use for those interactions. This includes but is not limited to electronic health record systems (EHRs) and platforms that contain information used to drive AI/ML development, AI/ML analytics tools embedded within EHRs, AI-driven ambient charting at the bedside, and computable phenotypes used to screen patients for inclusion in a clinical trial. Lastly, at the macro level, you think about the broader system or environment that encapsulates activities, functions, resources, hazards, etc., at the meso and micro levels. It is at the macro level where answers to specific questions can be

EXPERT PANEL DISCUSSION

found: How is the system being paid for? How is it being maintained? How is it being monitored as far as the practices and procedures that are implemented within that broader system? How are those policies and procedures informed by government, or business operations, or even business interests?

I think we need to look at those three layers to contextualize system boundaries, the bounded rationality that exists naturally within and across those three levels, and where AI/ML might address or reinforce desirable or undesirable characteristics within and across those three layers within the system of health. From there, we are able to better identify opportunities to integrate ethics and equity at every single level. For instance, as we examine patients' and providers' lived experiences across those three layers—what system-level boundaries that they are up against in the pursuit of ethics and equity in AI/ML development and implementation? How might they perceive AI/ML-augmented system design criteria across those three levels within the system? What are acceptable versus unacceptable trade-offs among key stakeholders with conflicting views?

Educating AI/ML system designers and developers is critically important as well. What are their codes of standards and ethics? What are policies to which they are expected to adhere or moral values that they might personally carry both at home and at work? These questions and considerations are what I think about as far as ethics implementation within a broader system of health.

Dr. Were: I will respond to this question from a narrow lens—meaning, not thinking about the ethical guidance around AI/ML in general, but specifically on ethical issues around AI/ML infrastructure development. When I think about the types of evidence and use cases that would be required to support the development of those principles and standards, I almost think of myself as being in the shoes of other policymakers, or academia, or decision-makers to answer this question. Some examples of use cases that might support development of the ethical principles and guidance include the following:

- Demonstration of inequities (disparities) in affordability and access to various types of AI/ML infrastructure and associated patient outcomes
- Security breaches to AI/ML infrastructure (e.g., data storage) with compromised protected health information
- Inability of AI/ML infrastructure to scale appropriately
- Difficulty in demonstrating and enforcing provenance of data—such as queries to a large language model sending identifiable patient data to developers of the large language model (LLM)

- Impact of nonstandardized AI/ML infrastructure on interoperability between systems
- Demonstration of negative impact when clear governance structures of the AI/ML infrastructure are lacking.
- Failed and suboptimal deployment of AI/ML models resulting from ML Ops platform issues
- Projects evaluating robustness of implemented AI/ML infrastructure
- Evidence of disproportionate distribution of risks and benefits caused by the implemented AI/ML infrastructure.

Dr. Hussain: We need to ensure that the protected groups are indeed protected. When we are developing infrastructure and these tools, make sure that AI is conscious about the protected communities and make sure our algorithms are not biased against them. We need to have metrics to assess how much of this is done, which will help companies that are developing these tools or making these products. Amazon and other companies launched products that were not sensitive for facial recognition equally across ethnic groups[8] and it was an embarrassment for the companies. It should be transparent, whatever that model is, so people can verify it, reproduce it, and validate it. Robust not just in terms of fault tolerance, but robust in terms of AI and ethics principles.

Mrs. Simmons: **The next question is the engineering field has, as a whole, shifted from ethical awareness of the individual engineer to systems-level ethical awareness. How might this awareness extend into the practice of AI/ML engineering and systems design?**

Dr. Hendricks-Sturrup: I think a few questions are important to explore around design, engineering, and ethical awareness. A key one is, if every single engineer or developer had no unethical intent or even approached their work according to every ethical principle there is, how might that not translate into ethical outcomes in the real world? Another one is, what are, in fact, the ethical values and principles of the developer or the engineer versus the ethical principles of the system in which the engineer is operating? Lastly, what ethical imperatives live within the broader system of operations in which the technology is intended to be used? These questions are very important to consider, first and foremost. I think from there, that leads to another question of whether one could experience certain vulnerabilities in the process of integrating or implementing what we might consider an ethical tool into an unethical system that preexists or predates the ethical tool's development. I think, as far as awareness goes, one interesting thing

about AI is that some might argue that it is technology that can become artificially self aware. Overall, I think this lends to a broader existential question around what is artificial versus human awareness of self in in an environment where human-derived constructs, whether ethical or unethical, control the broader system in which an AI/ML is implemented or operated.

Dr. Novak: A systems-level perspective allows us to look at interaction of elements in the system, and the impact of those elements, and their interactions on what the system produces. In my experience, what we often do with these systems or ecological analyses is focus on less powerful actors. We are going to study how this software engineer interacts with a nurse when they are trying to develop this tool and how this nurse interacts with this doctor while they are trying to use this tool. Those are actors who have no power in the larger scheme of things. I would love to see more work that follows the money, essentially.

In anthropology, we call this ''studying up.'' Who is paying for the system, and what is their goal? That needs to be detailed and presented. Is there a way to attach that information to this algorithm as it moves throughout the world? If you are a doctor using an algorithm taking care of a patient, you should know if the algorithm was funded by an insurance company. Lay out the incentives; democratize the knowledge of those incentives. I think all of us are beginning to understand now that *we* are the product, for instance, in social media. We need to understand our relationship with these AI tools and where we stand in relation to these tools in terms of our power. Help us all understand the consequences of different funding models, how we can influence the management and control of these systems, and their dissemination and use.

Dr. Were: Ethical awareness at the individual engineer level (''microethics'') implies that the individual is following long-standing and well-accepted professional codes of conduct and ethical standards. Systems-level ethical awareness, in this context, presumed to be ''macroethics,'' has been defined as ethics concerned with the collective social responsibility of the engineering profession and societal decisions around technology.

You are moving from what ethical expectations there are of the individual engineer to what we expect as the body of engineers and professionals and as a society around the ethics around AI and ML. This systems-level ethical awareness brings into light pressing issues of the unintended physical and nonphysical consequences of the technologies. As Dr. Hendricks-Sturrup was saying, if everybody is operating at an individual level very ethically, how can we think of things that are unintended as related to these

technologies? Because if we think along those lines, then it can stimulate policies and guidance around how these systems should be designed and used. For the engineering profession and society, it brings up the issue of assuming a collective social responsibility to these systems and for increasing our scrutiny of the impacts of these systems on society.

Dr. Hussain: Since we are dealing with systems and these system boundaries that are unclear with data going from one component, one interface into the other, we should clarify who owns the data. Where are the data? Where data are stored, who can access it? Data ownership becomes very important, especially when we are dealing with the system level: different parts working together, cooperating together, and processing. They need each other's information. We must be careful with what is considered my data and what is value-added information added on top of it. Think of it as an industrial plant with fluid going from one processing unit to another. The same fluid transforms into different phases. It is the same idea—data are moving from different stages and being transformed. So, who owns that data? Also, we have to consider compliance issues. It is different to have compliance at individual level or for one particular model, but when it is integrated—part of the whole system—we need to be more cautious, especially around the boundaries. Who is using my model and where is it being used? If I have information in the pipeline—who is feeding me and who am I feeding to? We need to consider aspects of compliance both for my input as well as my output. Those system boundaries compliance issues are essential. And very importantly, user consent is essential. We ask a user to sign the consent form but are we really getting their fully informed consent? Or do they just see a black box full of jargon and they just say, ''Ok, fine with me?''

Systems need to be reliable, robust, and efficient. Sometimes people say that convenience and privacy are opposite—you cannot have both. But the thing is we can try to design a system somewhere in between. We can get the benefits of AI in the system, at the same time, not compromising on our basic rights.

Mrs. Waters: There are five principles of gestalt that say that humans perceive the whole structure or the patterns of the sum of its parts, but we do not apply that to AI and ML. The system has combined ethical implications that are different than its individual components. We only want to apply the ethics to the individual components—how the image is being read by a convolutional neural network and what the output is for the AI, or how the insurance company's AI is deciding on who should be covered for what within the personalized medicine AI. But they are all integrated

into one system. If we perceive the whole as being greater than the sum of the parts, then why aren't we addressing the whole? We are trying to ethically apply band-aids to different parts of the AI that have been deployed. It is a part of the cognitive process in humans to have these feedback loops that are constant and learning through an iterative process. Why not apply that to AI? We can design AI systems to learn ethically and refine their processes based on real-world feedback. There are a lot of analogs to human behavior in the macro and micro that we simply do not apply to the ML algorithms in which we could be doing a better job. And I wouldn't be true to myself if I did not say that part of the macro-level solution to this is diverse and interdisciplinary collaboration. I will say that every day for the rest of my life because if you want ethical design, you must have diverse input.

We should have teams of ethicists, sociologists, and social workers—all of these domain-specific experts—to ensure the development of a systems-wide ethical scope and not just a narrowly focused one. We have cognitive intelligence that tells us that groups of problem solvers can outperform a single high-ability problem solver. You need different lenses and viewpoints so that you can do this effectively. Then you must have ethical guardrails and constraints. AI can very easily iterate and amplify, often to a fault, so you need to put some boundaries around this sandbox so that it plays well within those constraints or within a reasonably expected performance rate. Systemic ethical education needs to be a part of the experience when you are in school so that you are thinking ethically, even before you are in the work world, so that these things are, as one of the members of Morgan's Board of Regents Dr. Shirley Malcom, says, "… *baked in and not bolted on.*"

Mrs. Simmons: **Prior work shows there are six ethical themes present in current engineering systems design literature: integrating ethics and equity center perspectives into design; recognizing system boundaries; developing augmented system design criteria; managing trade-offs and conflicting values; educating systems designers; and applicability to engineering systems of health. In your perspective, how might these themes apply—or not apply—to the scope and practice of AI/ML engineering and system design?**

Mrs. Waters: The diverse applications of health care AI means that we must have a great deal of ethical considerations that go from the ground up, regarding the data sources—which are the end all—be all, start to finish of the conversation. The algorithms, the implementations of them, they all must be scrutinized for their potential bias and inequities. AI needs can benefit

from more grounding for improved contextual understanding and solving the symbol grounding problem. We are shaped through filters of past experiences. AI can be trained to actively integrate ethics and equity into their filters to produce a more equitable outcome. It just requires a different way of looking at these systems. If you are developing this technology, again, where it is baked in instead of tacked on at the end? You can have better optimization of the output and recognition of the boundaries with this in mind.

AI is not the solution to everything. It is not going to be perfect in every single-use case. We need to admit that and be clear about what the system can and cannot do. It is analogous to bounded rationality in people. Humans are making decisions within the constraints that are available to them with the information at hand. We must do the same thing with AI, which is to develop some augmented system design criteria. You need to know, as we have all said repeatedly, things like fairness, transparency, resilience, robustness. We cannot keep saying, "It's a black box, secret sauce kind of thing." We must be clear about how these systems are arriving at the decisions that they are because, otherwise, we cannot address anything. If you are just going to hide behind it saying that it is a trade secret, we cannot address it. Unless you have a brand new algorithm that no one has ever heard of, it is not that big of a secret. Let us help to understand how these systems are operating so we can fix them.

There will be trade-offs in this technology. Sometimes, what your institution values may not be what is valuable for your patients. Being able to articulate that and to provide the right weights in your labeling and in your data so that the AI can understand that and iterate on that is important. But you must make a decision: What might be good for one situation is not necessarily going to be good for another. This is where we must consider where we are going to meet in the middle to solve this particular conflict ethically.

Dr. Hussain: We must cover all aspects in the engineering basic design. Our issues are not for the end user alone. Of course, end user is the most important, as well as those who benefit from it, too, who are making the products—if it is not benefiting the overall system, if it is not increasing the bottom line, if the whole system is not integrated together, it will not work; it will stay siloed in small pieces. All stakeholders must see value in it for an integrated system to work. We want to benefit the weakest link, which is the end user. But at the same time, we cannot ignore the money. People will see that applying this AI principle not just as helping the needy or those who are deserving, but the businesses' bottom line will benefit, as well. It can be tracked. Because when you see that providers (doctors or businesses) have trust in your

products, they will promote your tools. For these AI systems to work, we cannot address the compliance issue unless we get the common stakeholders in place. Companies will benefit too. I believe the United States and other countries that apply rigorous ethical standards will have an advantage over those countries whose standards are different and perhaps less stringent, so patients' health is not well followed or documented. Those areas may lose people, because they will not be trusted. Those who will respect the data, those who will respect the end user, will prosper. We have the tools now to provide precise medicine, customized to each and every individual based on the individual. We can have very refined models tailored to each individual user, which I believe is also good for businesses. I strongly believe that these integrated engineering applications will help the end user. They will help the companies as well. Take, for example, glucose monitoring. A few years ago, it was difficult for a patient to monitor their own blood glucose level, but now they can use their continuous glucose monitoring system, a Libre reader, or any Dexcom device that continuously monitors glucose levels.

AI is scalable and can be applied in many applications, but of course, there are privacy and other issues, which must be addressed. Although currently we face challenges with ethics issues, hopefully, the ethics issues will be addressed and these integrated AI models will win eventually.

Dr. Were: I found the themes very self-explanatory in many ways and all the themes have direct relevance to AI/ML engineering. I would like to add a recommendation that AI/ML engineers and teams should walk through each of these themes at the beginning of their processes. This will clearly and deliberately identify the issues that arise, and where potential harms exist, teams should implement strategies to eliminate or mitigate these harms, while maximizing benefit. These themes should also be applied systematically across all stages of the AI/ML development lifecycle, as opposed to simply looking at them holistically.

Dr. Novak: I think these five ethical themes and how they are applied is interesting. Integrating ethics and equity-centered perspectives into design is critical. These efforts need to be funded. Funding organizations, whether they be corporations or government institutes, need to be clear that they want this. And they need to pay for it appropriately.

There is this thing called algorithmovigilance, which is an organizational capability. Recognizing system boundaries means you are able to understand when the system is biased, make sure it is not drifting, understand if it is drifting or becoming inaccurate because of changes in the underlying data or whatever the case may be, or being used inappropriately. That is a capability that organizations need to have in place that will enable them to manage these boundaries. In terms of systems being used inappropriately, this is similar to the evolving way we think about quality and safety in health care in general, which is not to focus on the ''bad actor.'' Let's figure out how the system created this situation where this algorithm is being used inappropriately. For example, are providers so busy and overworked that they reflexively accept AI-generated defaults built into electronic health records? Why are they so busy? That is a result of the system that they are working in. Understanding that system, the incentives, and what pressures people are experiencing in the system is a part of all of this.

We urgently need more guidance on developing augmented system design criteria. Everyone is improvising, trying to develop visualizations for explainability. People are truly trying to do well, but we do need more guidance based on empirical research. Managing trade-offs and conflicting values—what *are* our values? We often do not even articulate what our values are. I think that when we have done research, when we bring a predictive algorithm or the prototype of a predictive algorithm to a provider and say, ''OK, what does this information *mean* to you when it comes to predicting your patient's risk for a heart failure complication?'' Perhaps they have not thought about this before. They do not have a preconceived set of values related to this type of predictive technology. I think we must be very careful, related to that, to the metaphors that we use. If we say, ''Well, it's kind of like a lab test. You just bake it in with all of your other decisions that you use to decide whether to discharge the patient,'' or whatever the case may be. Well, maybe it is *not* exactly like a lab test. Maybe we need to think more critically and carefully about the metaphors and analogies that we are using to implement these tools. Educating system designers is great but I think we also need to educate the general public and those applying the tools in their work. In terms of applicability to engineering systems of health, what is the role of patients in everyday people in this whole system? How can they have a stake in the resulting consequences or products? Maybe we need to be thinking about new financial and ownership models.

Mrs. Simmons: **Thank you. It is interesting that you mention educating people, as did Mrs. Waters and Dr. Hussain, in terms of we do not know what we are signing, and that we sign our rights away. I will say that there is streaming content right now on our favorite streaming stations about AI use. The takeaway was to always check the terms and conditions, but how many of us really do that?**

EXPERT PANEL DISCUSSION

Dr. Hendricks-Sturrup: I might take a little bit of a different approach to going through each of these ethical themes that are present in the literature. I think integrating ethics and equity-centered perspectives into design is very much intertwined with recognizing system boundaries. Often times, we do not know what is possible as far as integrating ethics and equity perspectives. We do not know what is possible or what will be appreciated without understanding the system boundaries that are at play, whether they are structural, or cognitive, or even part of the collective persona of those operating within the system. So again, looking at the individual versus the collective operating within the system, observing complex outcomes that are sometimes unpredictable yet might cause unexpected friction within the system—all of this could render an ML tool suboptimal in its performance within that system.

Developing augmented system design criteria is certainly very important and something that should be done on the front end—but again, something that should not be done by itself. It should be done with a collective of informed and diverse stakeholders— those who hold greater power versus lesser power—to manage any trade-offs, power imbalances, intractable disagreement, and/or conflicting values among themselves. Protected space and time for those stakeholders to engage in potential conflict resolution to reach consensus around system design criteria and thus function and operate with one another on a day-to-day basis is critical to build and maintain systems of health.

We can think about all of the themes you mentioned, Malaika, as functions within the broader system of health, which is very different from other types of systems, like transportation, or the food industry, or what have you. Heath systems, not to be confused with systems of health, operate very differently. Health systems function very differently depending on geography—such as rural versus urban health systems. Many would reasonably expect these two types of systems to function very differently from one another within the broader system of health. Therefore, AI/ML developers should take special care to observe and appreciate those differences by engaging stakeholders within and across those geographic settings, for example, early in the AI/ML development process to augment engineering designs though both an ethics and equity lens and in a culturally sensitive manner that take into account desired and undesired system boundaries based on multiple stakeholder perspectives.

Mrs. Simmons: **We just talked about the systems design themes. Let's turn to ethics, fairness, and trustworthiness as ethical imperatives for critical infrastructure within systems of health, such as medical ethics, principles of beneficence,** **nonmaleficence, respect, and justice. In your perspective, what are the ethical imperatives to which AI/ML infrastructure engineers and system designers should adhere?**

Dr. Hussain: I will group them in five categories. The first thing is to make sure, when we are designing the system, be conscious about the bias. That is extremely important when we are making these models. The second thing is that the data, privacy, confidentiality, must be addressed. Third, ensure that the transparency and reproducibility can be verified and validated. Fourth, ensure that those who are performing it are accountable and that they are following the compliance guidelines. And lastly, the fifth one—we should not forget diversity, equity, and inclusivity, of course. Ensure that when we are crafting this ethics model, we set out to treat others the way we want to be treated. If everyone plays by these rules, it will be an ideal system. But we need to make sure we have the metrics evaluation assessment on the five areas I summarized.

Dr. Were: At the end of the day, the same imperatives of ethics, fairness, and trustworthiness should be adhered to by AI/ML teams—aligning with well-known under ethics, teams must address several things, namely how the infrastructure support consent mechanisms, approaches to secondary use of collected data, and protections of these data with robust governance mechanisms in place. Teams need to critically assess risks and benefits of the infrastructure across the AI/ ML lifecycle and implement mitigation mechanisms where risks exist and adjust this.

Issues arising from bias in collection and use of data, or access to the infrastructure and access to AI/ML products should also be addressed as well as the distribution of risks and benefits.

Under fairness and equity, these teams must ensure that developed and implemented AI/ML infrastructure do not exacerbate inequities and, where possible, that they contribute to narrowing existing societal gaps. There should be a strong focus on systems that serve areas and people most at need, and not just those with ability to pay. As such, implemented infrastructure should not disproportionately benefit particular groups, entities, or geographies. Finally, systems should ideally be accessible to all those who need them, and approaches to assure equitable access implemented.

Trustworthy AI/ML infrastructure is robust and assures security at all stages of the lifecycle and privacy of data within it. It should also be explainable and transparent to key stakeholders and users.

Mrs. Waters: Benevolent design under beneficence means that the AI has to be designed with a primary

intent to benefit patients and health care providers to enhance care quality and accessibility—point blank, end of story. If there is nothing present that speaks to that, then your system is never going to perform in an optimal way. Do no harm, your digital nonmaleficence—your system should be designed to avoid causing harm, either through misdiagnosis, data breaches, algorithmic biases, whatever it may be. It must be designed with that in mind. If you have not checked that box at the very minimum, your system is not optimized to provide the best outcomes. Inclusive and respectful design is needed to adapt the principles of respect into these AI systems. The system must be able to acknowledge the diversity of patients and providers and understand their unique needs, contexts, and values. That is not impossible but it requires forethought. It means you must design with this in mind.

Fairness, equity, and justice—AI systems must ensure that benefits and risks are equally distributed to avoid disparities in care—period, end of story. Governance accountability and responsibility—you must have clear accountability in the design and operation of these systems. If errors or ethical breaches occur, there needs to be a mechanism to address that and provide some means of preventing the recurrence of that kind of activity from happening in the future. We all are on board with transparency and explainability. You must be able to understand how the system makes these decisions. and that explanation must make sense to humans.

And then there is continuous ethical learning—because the health contexts are going to evolve as we understand more and more, and as we learn more and more, the systems must be able to learn and adapt ethically alongside these evolutionary processes throughout the industry, throughout the caregiving process. The AI system needs to be able to adapt to a patient's health changes. It cannot be rigid and inflexible. There must be a pathway to continuous learning for the systems, just like there are for the humans.

Dr. Novak: Is it an ethical imperative or is it just being a good person to be a team player on interdisciplinary teams—working and respecting innovative colleagues of multiple disciplines, whether they be clinical, technical, social science, or community members? I do not know if that is an ethical imperative, but I think we need to do that. Also, we need a process to document all the stakeholders in every project at the outset of the project. Document who is funding it. Document who the user of the tool will be, but also who the impacted people will be. The user might be a doctor. Impacted people might be patients or they might be all the people whose data were used to create the

technology. So, understanding all of the affected groups and their interests, I think, should be a part of every single project.

I believe that an ethical imperative is to have standardized communication mechanisms for communicating with each other about AI tools at all levels. These mechanisms are sometimes referred to as model facts labels. Technical people might have one version, and physicians or nurses or respiratory techs might have a different translated version of it. Community members might have yet a different translated version of it. But it summarizes what an algorithm is, how it was created and its performance, and carries some standardized labeling.

Dr. Hendricks-Sturrup: When we consider AI/ML development, we think about the end user being the primary stakeholder within the system. I am not sure if I completely agree with that because it is likely the case that the end user of the technology is not necessarily who bears the burden or benefit of the technology. We should consider, however, the broader societal impacts of that tool's implementation. Take, for example, algorithms used in clinical care and research that involve race correction. The main end users of those tools are health care providers. However, in this case, the provider as the end user is not the primary stakeholder that bears the burden or experiences the broad societal impact to that tool's implementation—it is the patients who experience both the personal and societal benefits and burdens following the use of that tool in this case.

For this reason, there should be a distributive justice lens applied to ascertaining the personal and societal benefits to AI/ML implementation in health care and research. This specifically requires contemplation around how benefits and burdens are naturally shared across the health system, through health insurance, health care distribution, health service distribution, etc.[9]

Lastly, I think it is important to educate developers about the importance of not substituting data, or data infrastructure, or data elements for strong community engagement. What do I mean by that? Limiting the use of statistical methods and tools that can be deployed across the system to impute data, where rather, it is better to actually go out to communities and engage them appropriately in the collection of data about them, versus using data that developers might assume represent those communities.

Mrs. Simmons: **Thank you. And now for our last question. Many of us in this forum have had these discussions in and outside of the *Aim Ahead* rooms and work groups, but prior work shows that persons with lived experiences in inequity are often excluded from and/or not effectively involved in**

data generation, including selecting and acquiring data sources, sharing, and use processes that involve their data. So, I want to expand a little bit on what Sajid mentioned about the golden rule. Instead of treat others as you would like to be treated, Christopher Voss in *Never Split the Difference*[10] says, you should treat others as they want to be treated. So how do we effect that in AI and ML is really the question. What comes to mind as far as strategies to meaningfully engage these persons more in AI/ML infrastructure development and use?

Dr. Hendricks-Sturrup: The approach we take at NADPH is embedding persons with lived experience at all stages of the data life cycle. That means from the time of acquiring data to the time of analyzing it to the time of disseminating evidence generated from that data collection process. Persons with lived experience, as the data subjects, must be involved at every step of the way to gain context behind quantitative data, correct biases in data driving the performance or performance drift of AI/ML tools, and comment on or react to outputs generated from the AI/ML tools. As data professionals, bioethicists, engagement experts, and researchers, it is incumbent upon us to ensure there is protected time that can be spent exclusively on this form of engagement. Without this engagement, we will run into issues that stem from suboptimal data collection and interpretation and, ultimately, dismal outcomes and effects within the health system, especially for patient stakeholders that are systemically excluded from AI/ML health research.

Mrs. Waters: You must design a deeply inclusive system. If you are not trying to do that, you are going to miss a lot of things through your sieve. You may find that it has too many large holes in it to actually capture the information that you need. To remedy this, you have to actively and cognitively engage those with lived experiences, because otherwise, you are just doing more tokenism. So, you need to bring people to the table. As was mentioned earlier, how do we do that? Because for some of us, this is embedded in the systematic approach that we have when we are designing systems, but for those who are not sure where to start, begin in the participatory data collection phase. Make sure that communities are engaged in that process, and allow them to have a say in how, when, and what data are collected about them. Solicit their help with codesigning workshops. Organize sessions where community members can cocreate and assist in how these AIs even get the objective set for them in the first place. Feedback loops—how can people from these communities actually give feedback that is meaningful and that can be acted upon? Cultural competency training—it is critical to have AI/ML

engineers trained in cultural competency so that they can understand nuances in the different communities that their systems may be deployed in. Transparent data governance—as I stated before, we want to know how the data are collected, used, stored, protected, how the rights of those people's data points, and all of their information, are safeguarded. Tailored data literacy programs—it is not just enough to bring people to the table, but you need to help them understand what it is they are going to be seeing, what they are going to be asked. Then again, ensuring that marginalized groups or all groups in the communities are represented in committees or bodies that oversee the development and deployment of these systems should be a priority. You need to guarantee their perspectives are influencing decision-making.

Dr. Hussain: Engage our current infrastructure entities like nongovernmental organizations (e.g., United Negro College Fund), community members, institutions, our churches, scouts, etc. Try all other entities that people trust so that we can be sure to reach them. Ensure our budgets reflect that we care about all stakeholders. We need to spend money for outreach so those affected are not just ignored.

Dr. Novak: We need to provide incentives for and support projects that implement community engagement. Pay for it. Create a widespread initiative to educate the general public on the basics of AI. If we are trying to engage people to help us design tools better, they really need to understand the tools. We cannot just ask questions and then interpret those. We really need meaningful participants, which means they need to understand something about what we are doing. Create these motifs and interpretations that people can actually use, like the model facts label that we talked about earlier. Think about the complex things that we all interact with and we understand pretty well. We are fairly successful at interacting with our cars. We understand when it is low on gas. We know how to fix that problem in the car. We have a dashboard that gives us information that we can interpret. It tells us when there is an emergency happening versus when something just needs to eventually be checked at the dealership. So, we can understand complex things if they are visualized for us effectively. Also, I would just throw in that things are made easier to use with the input of the average user. These are basic design research concepts; I look forward to the day when we are doing this more frequently.

Dr. Were: I wanted to extend the conversation around comprehensive community engagement and collaborative partnership to say that that engagement also has to identify the trusted decision-makers for communities that are vulnerable. Sometimes you can try to engage individuals, but they might not be able to

comprehend the issues you are discussing. Identifying the trusted decision-makers thus becomes very important. At what stage should they be involved? My colleagues have mentioned that they should be involved in all the stages. I wanted to suggest that this involvement should be extended even further upstream—to also ensure that engagement also occurs at the ideation, planning, and resource allocation stages. This is because oftentimes, people might come to communities with solutions in hand and resources already allocated. It's been mentioned to some degree, but the aspect of listening to the communities through multiple channels, and doing repeated engagements becomes important. Again, you want to understand the needs, constraints, and priorities of these communities as opposed to simply telling them to provide inputs on a solution that you have already predetermined. These engagements should inform design and development of the solutions that directly address and respond to the priority needs of these communities. Obviously, the infrastructure that you develop as part of your "engaging them" must be affordable and accessible to them and should also respond to their key challenges. As has been mentioned by colleagues, the communication approach has to assure understanding and comprehension of the trustworthiness, fairness, and explainability of the proposed solutions.

Meaningful engagement also involves deliberate capacity building of persons from populations that have experienced inequity. The capacity building has to occur at each stage of the AI/ML lifecycle, enabling members of these communities to be embedded within teams working on AI/ML solutions and the associated infrastructure. Team members who have lived experiences of inequities will not only serve as champions to represent the needs of their communities but will also bring new perspectives and insights that can enrich the products developed.

Finally, where possible, mechanisms are needed that allow communities that have experienced inequities to independently evaluate the deployed AI/ML infrastructure and systems, and to assess the performance and suitability of these systems for their needs.

Mrs. Simmons: **This has been enormously enlightening. To briefly summarize, the call is to embed ethics and equity, and to promote what Mrs. Waters calls "digital nonmaleficence." There are some valuable things that arose from this discussion, such as visible value—both monetary and nonmonetary—ownership, governance, autonomy, and including trusted decision-makers through education and integration that are engaged early as possible and throughout the process.**

Thank you to all of our participants and discussants today. It has been my pleasure to host this event.

Supplementary Material

ChatGPT supplemental material

Author Disclosure Statement

References

1. AIM-AHEAD. Infrastructure Core. Available from: https://www.aim-ahead.net/infrastructure-core/ [Last accessed: September 5, 2023].
2. National Science Foundation. NSF INCLUDES. Available from: https://new.nsf.gov/funding/opportunities/inclusion-across-nation-communities-learners [Last accessed: August 21, 2023].
3. The Foundation for Critical Thinking. Available from: https://www.criticalthinking.org/pages/certification-in-our-approach/1308 [Last accessed: September 5, 2023].
4. Hanseth O, Lyytinen K. Design theory for dynamic complexity in information infrastructures: The case of building internet. J Inf Technol 2010;25:1–19; doi: 10.1057/jit.2009.19
5. Monteiro E, Pollock N, Williams R. Innovation in information infrastructures: Introduction to the special issue. J Assoc Inf Syst 2014;15(4):4.
6. Star SL, Ruhleder K. Steps toward an ecology of infrastructure design and access for large information spaces. Inf Syst Res 1996;7(1):111–134; doi: 10.1287/isre.7.1.111
7. Maier A, Oehmen J, Vermaas PE (eds). Handbook of Engineering Design. Springer Link. 2020.
8. Raji ID, Buolamwini J. AIES '19:. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. Proceedings of AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society; ACM Digital Library. 2019; https://dl.acm.org/doi/10.1145/3306618.331424.4 [Last accessed: September 20, 2023].
9. Distributive Justice (Various Abstract Listing). Available from: https://www.sciencedirect.com/topics/social-sciences/distributive-justice [Last accessed: September 5, 2023].
10. Voss C. Never Split the Difference. Harper Collins: New York, NY; 2016.