THE EVOLVING LANDSCAPE OF HUMAN RESEARCH WITH AI – PUTTING ETHICS TO PRACTICE

2024 EXPLORATORY WORKSHOP

Office for Human Research Protections

Program Book

INTRODUCTION

INTRODUCTION

The use of artificial intelligence (AI) in biomedical and social and behavioral research raises fundamental questions about what impact these tools have on humans in the present and may have on us in the near future. We need to think about what AI to create, how to use it, what risks AI can pose, how we can prepare for and control these risks, and how this might inform future policy. This workshop explored the ethical and practical considerations for the use of AI in research involving humans.

OBJECTIVES

The purpose of OHRP's Exploratory Workshops is to provide a platform for open dialogue and exchange of ideas between stakeholders in the regulated community. This workshop on the use of AI in research:

- Explored the ethical considerations and principles for governance pertaining to the use of AI in biomedical and social and behavioral research.
- Considered some of the practical challenges of applying the Belmont Principles and the Common Rule to research involving AI.
- Discussed challenges with maintaining public trust and aligning AI with human values.

AGENDA

Time	Sessions
	Opening
9:45 a.m.–10:00 a.m.	**Welcome from OHRP** Speaker: Jeffery Smith, M.P.P., *Deputy Director, Certification & Testing Division, Office of the National Coordinator for Health IT, U.S. Department of Health and Human Services*
Session 1	**Exploring the Ethics and Governance Principles for Human Research with AI**
10:00 a.m.–11:20 a.m.	**Moderator:** Jessica Vitak, Ph.D. *Full Professor, College of Information, University of Maryland* 1. **Defining AI** Speaker: Kevin McKee, A.B., *Staff Research Scientist, Google DeepMind* 2. **How Is AI Currently Used in Research Involving Human Subjects?** Speaker: Craig Lipset, *Co-Chair, Decentralized Trials & Research Alliance* 3. **How to Create a Responsible and Responsive Research Program for AI** Speaker: Reid Blackman, Ph.D., *Founder and CEO, Virtue Consultants* 4. **Governance Across the AI Spectrum: Key Considerations from Data to Deployment** Speaker: Stephanie Batalis, Ph.D. *Research Fellow, Center for Security and Emerging Technology, Georgetown University* 5. **Creating a Trustworthy Health AI Ecosystem—Establishing Best Practices Based on Principles of Equity, Fairness, Safety, Transparency, and Reliability from the Outset** Speaker: Michael Pencina, Ph.D., *Chief Data Scientist, Duke Health; Vice Dean for Data Science, Director, Duke AI Health; Professor of Biostatistics and Bioinformatics, Duke University School of Medicine*
11:20 a.m.–12:00 p.m.	**Panel Discussion**
12:00 p.m.–1:00 p.m.	**Lunch**

AGENDA

Time	Sessions
Session 2	**Examining the Challenges of Applying the Belmont Principles and the Common Rule to Research Involving AI**
1:00 p.m.–1:50 p.m.	**Moderator:** Eric Mah, Ed.D., M.H.S., *Associate Dean, Clinical and Translational Research, University of California, San Diego* **1. Key Regulatory Terms and AI Tools in Human Subjects Research (HSR)** Speaker: Iris Jenkins, Ph.D., *Director of Research Integrity and Consultation, Virginia Tech* **2. How Does the Use of AI in Research Test the Notions of Personal Privacy and Identifiability of Data or Biospecimens?** Speaker: Benjamin C. Silverman, M.D., *Senior IRB Chair, Human Research Affairs, Mass General Brigham; Director of Ethics, McLean Institute for Technology in Psychiatry, McLean Hospital; Chair, Mass General Brigham Embryonic Stem Cell Research Oversight (ESCRO) Committee; Assistant Professor of Psychiatry and Faculty Member, Center for Bioethics, Harvard Medical School* **3. Disclosing the Role and Use of AI in HSR** Speaker: Sara Gerke, Dipl.-Jur. Univ., *Associate Professor of Law and Richard W. & Marie L. Corman Scholar, University of Illinois Urbana-Champaign College of Law*
1:50 p.m.–2:30 p.m.	**Panel Discussion**
2:30 p.m.–2:45 p.m.	**Break**
Session 3	**Exploring the Challenges with Maintaining Public Trust and Aligning AI with Human Values**
2:45 p.m.–3:35 p.m.	**Moderator:** Jeffery Smith, M.P.P., *Deputy Director, Certification & Testing Division, Office of the National Coordinator for Health IT, U.S. Department of Health and Human Services* **1. Data Privacy Isn't as Important as You Think** Speaker: Reid Blackman, Ph.D., *Founder and CEO, Virtue Consultants* **2. AI and Values: Alignment, Privacy, and Autonomy** Speaker: Karina Vold, Ph.D., *Assistant Professor, Institute for the History and Philosophy of Science and Technology, University of Toronto* **3. The Role of AI in Biomedical and Health Research - A Research Participant's Perspective** Speaker: Hugo Campos, *Participant Ambassador, NIH* All of Us *Research Program*
3:35 p.m.–4:15 p.m.	**Panel Discussion**

SPEAKER BIOGRAPHIES

SESSION 1

Jessica Vitak, *Ph.D.*
Full Professor, College of Information, University of Maryland

Jessica Vitak is a full professor in the College of Information and director of the Human-Computer Interaction Lab (HCIL) at the University of Maryland. Her research evaluates the privacy and ethical implications of new technologies that collect data in our homes, schools, and workplaces. She seeks to understand how privacy concerns play a role in technology adoption and use, and she develops tools and resources to help children and adults make more informed decisions when using technology and sharing sensitive data.

Kevin McKee, *A.B.*
Staff Research Scientist, Google DeepMind

Kevin McKee conducts research spanning machine learning, social psychology, and human-agent interaction. His projects focus in particular on the design of inclusive and cooperative artificial intelligence (AI) systems, touching on topics such as leveraging machine learning to promote group cooperation, applying Rawls' concept of the veil of ignorance to align AI systems with human values, and exploring the ethical and epistemological risks of replacing human participants with AI surrogates. Currently, Kevin is working to develop sociotechnical methods for evaluating large language models, in partnership with external experts and community stakeholders. Kevin also frequently collaborates on diversity, equity, and inclusion efforts, including co-founding and leading Google DeepMind's queer employee group.

Craig Lipset
Co-Chair, Decentralized Trials & Research Alliance

Craig Lipset is Co-Chair of the Decentralized Trials & Research Alliance, a global nonprofit organization dedicated to improving trial access. He is founder of Clinical Innovation Partners, providing advisory and board leadership with pharmaceutical and technology companies, health systems and research networks, advocacy, and investors. Craig is Vice President of the Foundation for Sarcoidosis, Adjunct Assistant Professor in Health Informatics at Rutgers University, and Fellow at DIA. He serves on the Advisory Council for HL7 Project Vulcan and External Stakeholder Board for IMI Trials at Home, as well as the Advisory Boards for EveryCure and the i-Cubed innovation center at the Duke Clinical Research Institute. Craig was previously the Head of Clinical Innovation and Venture Partner at Pfizer and on the founding management teams for two successful startup ventures.

Reid Blackman, *Ph.D.*
Founder and CEO, Virtue Consultants

Reid Blackman, Ph.D., is the author of *Ethical Machines* (Harvard Business Review Press, 2022), creator and host of the podcast *Ethical Machines*, and Founder and CEO of Virtue, a digital ethical risk consultancy. He is also an advisor to the Canadian government on their federal AI regulations, was a founding member of EY's AI Advisory Board, and a Senior Advisor to the Deloitte AI Institute. His work, which includes advising and speaking to organizations including AWS, US Bank, the FBI, NASA, and the World Economic Forum, has been profiled by the *Wall Street Journal*, the BBC, and *Forbes*. His written work appears in *Harvard Business Review* and the *New York Times*. Prior to founding Virtue, Reid was a professor of philosophy at Colgate University and UNC-Chapel Hill. Learn more at reidblackman.com.

Stephanie Batalis, *Ph.D.*
Research Fellow at Center for Security and Emerging Technology, Georgetown University

Stephanie Batalis is a Research Fellow at Georgetown's Center for Security and Emerging Technology (CSET). Her research examines a number of issues at the intersection of AI and the life sciences, including how emerging technologies will impact both biomedical innovation and U.S. biosecurity. Before joining CSET, Dr. Bataliz was the STEM Policy Fellow at the North Carolina Biotechnology Center where she focused on economic and workforce development initiatives in North Carolina's life sciences ecosystem. She earned her Ph.D. in biochemistry and molecular biology with a focus in structural and computational biophysics from Wake Forest University's School of Medicine and a B.A. in molecular and cellular biology from Vanderbilt University.

Michael J. Pencina, *Ph.D.*
Chief Data Scientist, Duke Health; Vice Dean for Data Science, Director, Duke AI Health; Professor of Biostatistics and Bioinformatics, Duke University School of Medicine

Michael J. Pencina is Duke Health's chief data scientist and serves as vice dean for data science, director of Duke AI Health, and professor of biostatistics and bioinformatics at the Duke University School of Medicine. His work bridges the fields of data science, health care, and AI, and builds upon Duke's national leadership in trustworthy AI. Dr. Pencina co-founded and co-chairs Duke Health's Algorithm-Based Clinical Decision Support Oversight Committee and serves as co-director of Duke's Collaborative to Advance Clinical Health Equity. He spearheads Duke's role as a founding partner of the Coalition for Health AI whose mission is to increase the trustworthiness of AI by developing guidelines to drive high-quality health care through the adoption of credible, fair, and transparent health AI systems. Dr. Pencina is an internationally recognized authority in the evaluation of AI tools and algorithms. Guideline groups rely on his work to advance best practices for the application of algorithms in clinical medicine. With over 100,000 citations, he has been recognized by Thomson Reuters/Clarivate Analytics among the world's most "highly cited researchers" in clinical medicine and social sciences.

SESSION 2

Eric Mah, *Ed.D., M.H.S.*
Associate Dean, Clinical and Translational Research, University of California, San Diego

Eric Mah is Associate Dean at the University of California San Diego, where he leads the administrative operations of its central clinical research program. Prior to UC San Diego, he led the Ethics & Compliance Office for UC San Francisco and held leadership positions at the Institutional Review Board (IRB) offices for UC Davis and UCLA. He has served on conference planning committees for Public Responsibility in Medicine and Research and as past-chair of its Diversity Advisory Group. More recently, he helped co-found the Consortium for Applied Research Ethics – Quality (CARE-Q.org), a joint effort between the University of California and Stanford University, to support quality improvement efforts across IRB programs nationally. He is currently serving a 4-year appointment on the Department of Health and Human Services Secretary's Advisory Committee on Human Research Protections.

Iris Jenkins, *Ph.D.*
Director of Research Integrity and Consultation, Virginia Tech

Iris L. Jenkins is the Director of Research Integrity and Consultation at Virginia Tech. In this role, she serves as the institution's Research Integrity Officer, directs and administers the research integrity/ responsible conduct of research training program, and directs and administers the institution's research ethics consultation service. Prior to this role, Dr. Jenkins led the Human Research Protection Office at the University of Massachusetts Amherst overseeing daily operations, serving as a voting member of the Institutional Review Board, and advising the campus research community on matters relevant to human subjects research. Dr. Jenkins has over 14 years of research ethics and compliance experience having also served in roles supporting Institutional Animal Care and Use and Institutional Biosafety Committees.

Dr. Jenkins earned a B.A. in biological sciences from Mount Holyoke College and an M.S. in plant pathology from the University of Arizona. She conducted human subjects research while earning her Ph.D. from the University of Massachusetts Amherst in neuroscience and behavior. Her own research experience influences her work as an ethics administrator. She is enthusiastic about helping researchers accomplish their goals while maintaining the highest ethical standards and remaining in compliance with applicable regulations and policies.

Benjamin C. Silverman, *M.D.*
Senior IRB Chair, Human Research Affairs, Mass General Brigham; Director of Ethics, McLean Institute for Technology in Psychiatry, McLean Hospital; Chair, Mass General Brigham Embryonic Stem Cell Research Oversight (ESCRO) Committee; Assistant Professor of Psychiatry and Faculty Member, Center for Bioethics, Harvard Medical School

Benjamin C. Silverman is the Senior IRB Chair at Mass General Brigham, Human Research Affairs. Additionally, Dr. Silverman is currently the Chair of the Mass General Brigham Embryonic Stem Cell Research Oversight Committee, Director of Ethics for the Institute for Technology in Psychiatry at McLean Hospital, and serves as an instructor in psychiatry and a faculty member in the Center for Bioethics at Harvard Medical School. Dr. Silverman received his medical degree from the Johns Hopkins University School of Medicine, completed his psychiatry residency at the MGH McLean Adult Psychiatry Residency Training program, and completed sub-specialty fellowship training in addiction psychiatry through Mass General Brigham.

Sara Gerke, *Dipl.-Jur. Univ.*
Associate Professor of Law and Richard W. & Marie L. Corman Scholar, University of Illinois Urbana-Champaign College of Law

Sara Gerke is an associate professor of law and Richard W. & Marie L. Corman Scholar at the University of Illinois Urbana-Champaign College of Law. Her current research focuses on the ethical and legal challenges of artificial intelligence and big data for health care and health law in the United States and Europe.

Professor Gerke has more than 60 publications in health law and bioethics, especially AI and digital health. Her work has appeared in leading law, medical, scientific, and bioethics journals, including *JAMA*, *Science*, and *Nature Medicine*.

Professor Gerke previously served as an assistant professor of law at Penn State Dickinson Law. Prior to that, she held the position of Research Fellow in Medicine, Artificial Intelligence, and Law at the Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School.

SESSION 3

Jeffery Smith, *M.P.P.*
Deputy Director, Certification & Testing Division, Office of the National Coordinator for Health Information Technology

Jeffery Smith is the Deputy Director in the Certification & Testing Division at the Office of the National Coordinator for Health Information Technology (ONC), where he oversees and implements policies related to the ONC Health IT Certification Program. He has served in this capacity since 2020. Previously, Mr. Smith served as Vice President of Public Policy at the American Medical Informatics Association and at the College of Healthcare Information Management Executives (CHIME), where he served as lead government affairs liaison to federal agency and congressional staff on matters related to health IT and health informatics.

Mr. Smith holds a bachelor of arts in political science from Kansas State University and a master's degree in public policy from the University of Maryland, where he specialized in health care and technology policy. Mr. Smith has published in *Health Affairs, the Journal of the American Medical Informatics Association,* and *Applied Clinical Informatics*. He also authored a chapter on public policy for *Clinical Research Informatics 3rd Edition*.

Karina Vold, *Ph.D.*
Assistant Professor, Institute for the History and Philosophy of Science and Technology, University of Toronto

Karina Vold is an assistant professor at the Institute for the History and Philosophy of Science and Technology at the University of Toronto (U of T). She is also a Research Lead at the U of T Schwartz Reisman Institute for Technology and Society, an AI2050 Early Career Fellow with the Schmidt Sciences Foundation, a Faculty Associate at the U of T Centre for Ethics, and an Associate Fellow at the University of Cambridge's Leverhulme Centre for the Future of Intelligence. Dr. Vold specializes in philosophy of cognitive science and philosophy of artificial intelligence, and her recent research has focused on human autonomy, cognitive enhancement, extended cognition, and the risks and ethics of AI.

Hugo Campos
Participant Ambassador, NIH All of Us *Research Program*

Hugo Campos lives with hypertrophic cardiomyopathy, a genetic heart condition that raises his risk of sudden cardiac arrest. This has driven him to become a passionate advocate for patient autonomy, empowerment, and access to medical data. For more than a decade, he has served in various roles that align closely with this mission, including Participant Ambassador for the NIH's All of Us Research Program, Co-Chair of the Community Advisory Committee for the California Partnership for Precision Nutrition (CAPPN), and Co-Lead of the Patient Engagement Working Group for PCORI's THRIVE Trial. These experiences have deepened his understanding of the vital partnership between patients, clinicians, and researchers in achieving better health outcomes.

In 2023, Campos embraced generative AI to empower himself, his family, and other patients. Using large language models for personalized health insights, decision support, clinic visit preparation, and various other use cases, he witnessed the immense potential of generative AI for patient empowerment. This experience deepened his commitment to revolutionize patient engagement, ensuring AI solutions meet real-world needs, enhance equity, and give individuals control over their health, potentially untethering them from a profit-driven, inefficient, and inequitable health care system.

CONTENTS

WELCOME AND INTRODUCTION

- *Julie Kaneshiro, M.A., Acting Director, Office for Human Research Protections (OHRP), U.S. Department of Health and Human Services (HHS)*
- *Jeffery Smith, M.P.P., Deputy Director, Certification and Testing Division, Assistant Secretary for Technology Policy/Office of the National Coordinator for Health Information Technology (ASTP/ONC)*
- *Yvonne Lau, M.B.B.S., M.B.H.L., Ph.D.; Director, Division of Education and Development (DED), OHRP*

Remarks by the Acting Director of OHRP

Ms. Kaneshiro welcomed everyone to OHRP's 7th Exploratory Workshop. She expressed particular appreciation for the speakers and moderators for sharing their insights on the increasingly pervasive use of artificial intelligence (AI) in research. OHRP's focus is on protecting the rights, welfare, and well-being of human subjects, and the aim of the panelists is to help the office explore the important ethical, legal, and practical implications of this technology for human subject protection. Questions for discussion include: How can the research community think about the challenge of integrating ethical considerations into research that uses AI? How are these tools being used? How might they be used in the future? What concerns arise from the use of these tools?

Ms. Kaneshiro introduced the new OHRP Director, Dr. Mary (Molly) Klote, who will officially begin work in mid-October.

Ms. Kaneshiro introduced Jeffery Smith, Deputy Director of the Certification and Testing Division in the Office of the National Coordinator for Health Information Technology (ONC), to offer opening remarks. This office is responsible for overseeing strategy and policy related to AI in HHS.

Opening Remarks by Jeffery Smith

- Jeffery Smith, M.P.P., Deputy Director, Certification and Testing Division, Assistant Secretary for Technology Policy/ Office of the National Coordinator for Health Information Technology (ASTP/ONC)

Mr. Smith said the topics to be discussed at this meeting arise at various phases of the AI lifecycle. The lifecycle includes planning and design, data collection and management, model building and tuning, verification and validation, model deployment, operation and monitoring, and real-world performance evaluations. The lifecycle serves as a framework for product development, a blueprint for institutional strategy development, and a way of identifying or developing public policy related to AI. When viewed as a framework for public policy, the AI lifecycle can help identify different agencies and actors that may be relevant and may have roles to play. For example, ONC primarily focuses its energies on the predevelopment phase. Through its certification process, ONC regulates electronic health records (EHRs) that are used by the majority of hospitals and office-based physicians in the United States, and so has a nationwide footprint that provides a fair amount of leverage to influence the nation's health data infrastructure.

In 2023, HHS finalized the HTI-1 Final Rule to advance health information technology (IT) interoperability and algorithm transparency. When a certified health IT developer supplies an AI tool that offers predictive decision support, it must provide users with information about how that algorithm was designed, developed, tested, and evaluated, as well as how it should be implemented. The new rule, "Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing" does the following:

- Requires that risk analysis and risk mitigation strategies be identified and applied.
- Establishes policies and controls for governance, including how data are acquired, managed, and used.

Requirements pertain to ONC-certified health IT, which supports the care delivered by more than 96% of hospitals and 78% of office-based physicians in the United States.

Various authorities regulate aspects of AI. For example, the U.S. Food and Drug Administration (FDA) regulates machine-learning-driven device software functions; ONC regulates certified developers of EHRs; and the Office of Civil Rights (OCR) regulates covered entities against discrimination, including using AI tools. Mr. Smith said he envisions a "regulatory mosaic" in which these authorities coordinate. As AI develops rapidly, Mr. Smith sees the need not only to consider additional policies, but also to reconsider numerous existing policies (e.g., the Common Rule) to determine which are still appropriate and applicable. Some remain fit-for-purpose, while others will require modification to be relevant and effective. While modifications are inevitable, Mr. Smith observed that some existing policies have more "stretch" than we may at first think.

Mr. Smith suggested three questions he hoped panelists would address:

- What areas of ethical concern at the intersection of AI and health services research (HSR) are the priority? He observed that "when everything is a priority, nothing is a priority."
- Are there priorities that are more tractable than others? Which can be tackled in the near future?
- What steps can be taken to establish practical guidance and policy?

Mr. Smith noted that translating the principles of the *Belmont Report* into the Common Rule was a "towering feat." Existing policies and procedures must be leveraged whenever possible.

Remarks by the DED Director

Dr. Lau shared that, given the fast pace at which issues related to human research with AI are being considered, DED found it challenging to create a relevant agenda for the workshop. She noted that a webinar or symposium on the subject seems to be held almost every week. DED decided to focus the symposium on issues related to ethics and government. She welcomed the distinguished panel of speakers, thanked them for their time, and introduced Dr. Jessica Vitak, the moderator for the first session.

SESSION I: EXPLORING THE ETHICS AND GOVERNANCE PRINCIPLES FOR HUMAN RESEARCH WITH AI

- *Moderator:* Jessica Vitak, Ph.D., Full Professor, College of Information, University of Maryland

Session I Introduction

- Jessica Vitak

Dr. Vitak introduced each of the five speakers presenting remarks in the session, which focused on principles of ethics and governance that apply to human research with AI.

Defining AI

- Kevin McKee, A.B., Staff Research Scientist, Google DeepMind

Mr. McKee focused his remarks on "demystifying" AI and exploring how traditional research ethics might apply to AI. Often, conversations about AI can be difficult because of speculative ideas from the world of science fiction that make it seem hard to understand. A clear shared understanding of what AI is and how it operates is an essential foundation for discussion. He offered an "operational definition" of AI: AI is any human-made process or system that makes decisions or solves problems.

Noting that AI comes in many different forms, Mr. McKee explained that rule-based AI and learning-based AI systems are very different. "Eliza," an AI psychotherapist chatbot, is an example of a rule-based system. Eliza looks for patterns in communication and responds according to set rules. In contrast, other AI systems may be capable of learning through trial and error as they are used. A second distinction is between narrow AI (for example, a system that can only play chess) and

general systems (for example, ChatGPT, which can address a range of tasks and respond to a wide range of questions). It is also important to distinguish between predictive systems and generative AI (genAI). A predictive system might identify what video a user wants to watch based on what they have chosen to watch in the past. GenAI systems go beyond this to generate novel content, such as video and text, based on the context and the data fed to it. Mr. McKee cautioned that genAI systems will incorporate and even amplify biases found in their sources.

While modern AI systems demonstrate strong, sometimes superhuman capabilities in specific domains, their capabilities typically come with unintuitive limitations and boundaries. People can be blind to these unexpected failures. For example, generative systems train on large amounts of human-generated text, allowing them to learn many underlying patterns. The flip side of this is that they can also learn social patterns of stereotyping and bias that are encoded in the data the systems are fed.

Scientists are currently exploring many diverse ways to integrate AI systems into different stages of the research process. Mr. McKee reviewed several emerging use cases (adapted from Messeri & Crockett, 2024):

- As studies are designed, researchers may use AI to search for, evaluate, and summarize scientific literature and generate new hypotheses.
- At the data collection phase, researchers may use AI to simulate data points from natural complex systems, including those involving human participants.
- As data are analyzed, researchers may use AI to curate and analyze datasets to produce new insights.
- At the peer review phase, researchers may use AI to evaluate scientific results, papers, and other artifacts.

What are the implications for AI and research ethics? Mr. McKee stressed the following points:

- AI is a broad, diverse, and expanding class of technologies.
- Ethics for research with human participants and AI will need to be flexible and contextual.
- Navigating the landscape of ethics and AI will require ongoing dialogue among scientists, policymakers, ethicists, and technology developers.

How Is AI Currently Used in Research Involving Human Subjects?

- Craig Lipset, Co-Chair, Decentralized Trials and Research Alliance

Mr. Lipset described the ways in which AI is currently used in clinical trials.

During the study design phase, automation using genAI helps organize and author protocols. GenAI can also gather historical and contemporary data to improve the design and assess the feasibility of possible approaches to implementation—for example, by simulating trials. AI may also be used to envision how the data a study yields can drive downstream processes.

AI is used to select appropriate sites, investigators, and participants. It can assess the feasibility of possible sites for a specific study, as well as predict the performance and associated risks for possible principal investigators. AI can help manage contracting and budgeting processes. Investigators can use AI to help them select the right participants and help achieve the desired diversity among them. Streamlining the recruitment process, AI can match EHRs with real-world data. It also offers automated chart review, saving time and facilitating analysis. Once potential participants are chosen, AI can further assist by personalizing the informed consent process to align with the specific participant's culture, language, preferences, and education level. Bots can help with screening, consent, and support. Mr. Lipset noted that while people previously resented having to interact with bots, they are increasingly comfortable with them.

AI is playing a variety of roles in study conduct. It can support participants by identifying safety signals based on historical

data. Risk-based monitoring algorithms can be used to flag a variety of associations among relevant medical data. Other useful functions include quality oversight and signal detection, safety follow-ups and narratives, and fraud detection.

Once data from the study are available, AI can be used to structure data and, if desired, create digital twins and synthetic data to further the analysis process. It can help with statistical programming and automate data flow. Other applications include automated quality control and chart review, safety review, and narrative authoring, which may be facilitated by eSource data (data initially recorded in electronic format) and natural language programming (NLP), which uses machine learning to enable computers to understand and communicate with human language. AI can also be used to identify novel endpoints or sensors.

In regard to reporting and submission of study findings, AI can help with authoring, automating quality reviews, and managing regulatory questions.

Turning to the subject of future uses of AI in the research context, Mr. Lipset highlighted groundbreaking work in the field of AI-supported data capture. For example, in 2023, FDA issued a Letter of Intent Determination regarding an "automated depression and anxiety severity measurement product using multiple behavioral and physiological indices of depression in a machine learning (ML) model" (DDT-IST-000014-LOI-1). Mr. Lipset also pointed to products said to be capable of analyzing facial expressions, speech behavior, and vital signs, among other observable features.

In an analysis of clinical trials by 16 companies, Mr. Lipset found that the studies lacked "digital maturity" in the areas of analytics, clinical trial transparency and data sharing, and automation (Properzi & van Tongeren, 2024). Significant challenges to achieving digital excellence in trials include a lack of internal capabilities, institutional culture, lack of talent, and a need to find the right partners to move forward.

How to Create a Responsible and Responsive Research Program for AI

- Reid Blackman, Ph.D., Founder and CEO, Virtue Consultants

Dr. Blackman distinguished between "AI for good," in which people are using AI in pursuit of ethical goals such as reducing poverty, and "AI for not bad," in which research goals are ethically neutral but users do want to identify possible sources of ethical risks and avoid them. He expressed interest in the challenges posed by both of these objectives, and said people need a clear sense of the risks in order to mitigate them; often the understanding of both risks and mitigation strategies is too thin.

A significant ethical risk can arise in the use of software that learns by example, Dr. Blackman said. Objectives might be to identify an impressionist painting, to distinguish between good and bad resumes, or to examine medical profiles of people who develop diabetes in order to draw conclusions related to risk factors. Conclusions may be erroneous if investigators have an insufficient number of examples, the wrong examples, or if the examples are not representative. For example, a facial recognition algorithm used only examples of white men and could not recognize the features of a black woman. This type of error can be mitigated through the use of clear definitions.

Dr. Blackman said some ethical risks are probable simply because of the way AI works. In the effort to get as many examples as possible to develop a reliable algorithm, investigators are likely to violate individuals' privacy. The examples are effectively within a "black box" invisible to algorithm users, who may be unaware of ways in which the sample on which the algorithm is based is biased.

Additional ethical risks of generative AI include "hallucinations" based on false information and the "sleazy salesperson" syndrome in which data are misused. Adding complexity to the challenges is the fact that responsibility for errors may be difficult to assign.

Dr. Blackman stressed that AI developers are largely unprepared to address substantive, qualitative ethical questions.

Examples include:

- What is the appropriate metric for bias? Which of the various incompatible metrics is appropriate for the specific sample in question?
- Who should have control of the data?
- Is our chatbot objectionably manipulative?
- What is the appropriate benchmark for safe deployment?
- Should we defer to the AI's "judgement"?
- Does this place an undue burden on the user? Our employees?
- Is this our responsibility or the user's or the government's?
- Is this explanation good enough? Is it suitable for the participants?
- Should we defer to our client's/that culture's ethical standards?

It is important to ensure that the right people are involved in answering these questions for each application, Dr. Blackman said.

Governance Across the AI Spectrum: Key Considerations from Data to Deployment

- Stephanie Batalis, Ph.D., Research Fellow, Center for Research and Emerging Technology, Georgetown University

Dr. Batalis focused her remarks on possible roles for governance—defined as actions that shape the direction of the future—across the spectrum of AI applications. These interventions might require compliance (laws or rules) or might be voluntary (incentives). They might be enabling (intended to encourage a particular outcome) or restrictive (an attempt to prevent something from happening). Finally, governance might be public (coming from government) or private (for example, from industry).

Dr. Batalis briefly discussed a variety of "nodes" of present possibilities for governance. These include identification of needs or problems that AI might be able to address, or "use cases"; accessing the resources needed to address a need or problem; model development and deployment; the feedback cycle; and real-world outcomes.

Before AI models are built, governance can influence research priorities by offering incentives to indicate high-impact research areas, articulating norms, or making funding available to encourage promising work. Dr. Batalis suggested it may help establish norms for responsible AI through outreach to researchers and by engaging them in norm development. Governance can also influence future directions by making resources available to support certain directions but restricting their availability for others. Various government initiatives and public-private partnerships may also be influential.

Governance may also affect who develops AI through the choices it makes on access to resources and incentives, Dr. Batalis noted. By promoting best practices, standards, and specifications, it can shape how models are developed. Regulations, incentives, and governance initiatives can also encourage preferred target outcomes. Where AI models already exist, governance can use outreach, regulations, guidance, best practices, and incentives to promote or limit access to particular models.

Dr. Batalis suggested that by streamlining the regulatory process, governance can promote the adoption of AI and facilitate its use by providing guidance. As AI is implemented, governance can ensure that information needed to assess outcomes is collected and that research is conducted to understand these outcomes. Models can be improved through an iterative process as feedback is employed. Governance can also provide incentives and funding to encourage further work on promising models. With a view toward supporting desirable real-world outcomes, governance can promote certain interventions in a variety of ways.

Dr. Batalis cited biodata as an area in which future uses of AI may call for governance. AI can be useful in identifying underlying patterns in biodata, such as the relationships among certain characteristics. However, this requires huge amounts of data, potentially including medical histories, genome sequences, test results, medical images, and lifestyle factors. These data are obviously personal in nature and may pose privacy or security risks to participants. Consequently, governance may aim to mitigate these risks by taking steps to protect security and privacy, regulating or influencing who has access to what type of data, limiting data bias by incentivizing research to better understand what it means for a database to be complete, engaging with database subjects to support an informed choice regarding inclusion of their data, and encouraging or discouraging research choices in future AI development in the field.

In conclusion, Dr. Batalis stressed the following points:

- Governance is more than regulation. Laws and rules are not the only tools available for shaping choices.
- Governance can influence the trajectory of AI through actions such as collecting information, promoting or restricting certain options, providing information, or encouraging promising lines of effort.
- In order to choose appropriate tools, governance goals must be clearly identified. As tools are identified, it is crucial to realize that influence is "not a zero-sum game": a particular mechanism will not achieve every goal, so there is opportunity for effective "layering" of initiatives to accomplish government's aims.

Creating a Trustworthy Health AI Ecosystem—Establishing Best Practices Based on Principles of Equity, Fairness, Safety, Transparency, and Reliability from the Outset

- Michael Pencina, Ph.D., Chief Data Scientist, Duke Health; Vice Dean for Data Science, Director, Duke AI Health; Professor of Biostatistics and Bioinformatics, Duke University School of Medicine

Dr. Pencina said that we have a "Wild West" of algorithms at present, with so much focus on development and technological progress that insufficient attention is paid to an algorithm's potential value, quality, implications for health equity, or adherence to ethical principles.

There is plenty of opportunity for mistakes in algorithm development and application, Dr. Pencina said. Some developers do not follow the rigor required of research, and algorithms have been applied in inappropriate settings. Quality studies are needed to uncover serious errors. For example, he cited a study of a widely implemented model for predicting sepsis that performed poorly (Wong et al., 2021). An algorithm used to make decisions about health care was shown to be racially biased. In this case, authors observed that "bias occurs because the algorithm uses health costs as a proxy for health needs. Less money is spent on Black patients who have the same level of need, and the algorithm thus falsely concludes that Black patients are healthier than equally sick White patients" (Obermeyer et al., 2019). These are human errors, not computer errors, he stressed. They stem from the way systems are set up.

The field needs to do better, Dr. Pencina urged. Clearly, no single regulatory agency can review all the emerging prediction models. However, he said, this does not absolve model developers and users from the obligation to demonstrate their effectiveness and safety.

The Coalition of Health AI (CHAI) has articulated the principles of trustworthy health AI. These include:

- Ensure that AI technology serves the human person.
- Define the task we want AI to accomplish.
- Describe what the successful use of an AI tool looks like.
- Create transparent systems for continuously testing and monitoring AI tools.

Duke University's School of Medicine seeks to put these principles into action. Dr. Pencina quoted the following mission statement:

Out of our primary focus on patient safety and high-quality care, our mission is to guide Algorithm-Based Clinical Decision Support (ABCDS) tools through their lifecycle by providing governance, evaluation, and monitoring.

In support of this mission, all electronic algorithms that could impact patient care at Duke Health fall within the scope of the ABCDS Oversight Committee, which conducts "checkpoint reviews" at various stages. Each algorithm must be registered in the Federation of Local Health AI Registries (described below). In addition, Duke AI researchers are encouraged to reach out to the Institutional Review Board (IRB) for review. Duke University's School of Medicine actively seeks ways to monitor the use of algorithms and protect them from human error.

Dr. Pencina said that because budget constraints make it impossible to pay equal attention to all algorithms, the ABCDS Oversight Committee's review process is based on risk. Algorithms based on standard of care are considered low risk. Those that are based on clinical consensus that has not reached the level of standard of care are considered medium risk, while those derived from data are flagged as high risk. Those that are considered high risk receive the greatest amount of attention.

Dr. Pencina described the Federation of Local Health AI Registries (Pencina, McCall, & Economou-Zavlanos, 2024), which encourages every organization that applies algorithms to health decisions to inventory these algorithms. The registry records and tracks all health AI technologies deployed in clinical care and operations within a given health system. Dr. Pencina stressed that hidden algorithms that are applied without awareness can be dangerous. It is important to talk to vendors about how they have populated the data used in algorithm development and to create transparent communication about algorithms with patients. To this end, Duke has a website that lists all the AI used in patient care at Duke.

Panel I Discussion

What's the biggest priority?

Dr. Vitak began by asking panelists what they saw as the biggest priority from an ethical standpoint looking forward.

Dr. Blackman responded that it was "getting the political will to do something about it." AI is new and complex. He argued that "we more or less know what to do." Organizations need the alignment and leadership to address emerging issues. It is necessary to educate leaders about why this is important and motivate them to devote resources to addressing it.

Dr. Jenkins saw the biggest issue as public education. People need to understand the ethical implications of AI. She said her colleagues have high expectations for AI, but most are unaware of its limitations and accept AI without question. She stressed raising awareness of the need for "guard rails." Dr. Batalis agreed and said people tend to see AI as either wonderful or scary. It is essential to be frank about both the possibilities and the challenges.

Mr. Lipset highlighted the importance of transparency and trust. The data that are fuel for AI models come from human beings, and people need to be conscious of that lineage.

Ms. Sara Gerke saw the priority as educating IRBs on the complex issues involved and making sure there is guidance. IRBs have different levels of expertise. Already, increasing numbers of AI applications and a range of privacy regulations—at the state as well as federal level—make decision-making in this new area especially challenging. Guidance is needed on what should go through IRB review.

The role of the IRB

Dr. Vitak asked what the role of the IRB should be, given that AI applications generally fall outside the narrow definition of human subjects research. Many of the cases described by speakers are not human subjects research, although there is potential for harm to human subjects. IRBs are not well prepared to understand those harms or know how to prevent them.

Dr. Pencina said that, taking the broad view, the Helsinki Principles apply quite well to AI. He believes that encouraging IRBs to go beyond their normal role to partner with AI developers and engage in a two-way dialogue is worthwhile. However, he asked, "How do we define the scope?" The scope of the IRB's authority must be conceptualized in terms of risks and benefits for individuals.

Dr. Silverman said the issue of subject identifiability is particularly important in determining what is and is not human research. While he said there should be "no question" about whether the IRB has a role, there is also no question that it cannot do it alone. The dialogue with AI developers should be part of a larger and well-integrated governance process within the institution. He added that most people who work on AI in medical centers do recognize the need for IRB involvement. However, there are also many people working alone in this type of research who have no idea that oversight or infrastructure might be important.

While Mr. McKee agreed with the need for boundaries around what the IRB reviews, he noted that a surprisingly low number of AI research projects involving participants are currently reviewed by IRBs (McKee, 2024). Researchers working with AI need to be educated that there are expectations and ethical responsibilities when human subjects are involved.

Dr. Mah expressed concern about the need to avoid "mission creep" for overburdened IRBs, particularly given the likely lack of understanding of the issues involved. Dr. Batalis agreed that it was important to see what elements of responsibility for oversight can be distributed. However, she said, many of the concepts involved are familiar, such as the potential identifiability of human data in datasets and the need for thoughtful interactions during the development process to identify ethical concerns.

Dr. Vitak noted that a listening member of the public, citing the fact that investigators can "push people around" with their unfamiliar technical vocabulary, had asked: "How do we keep our seat at the table? How do we make sure our voices are heard?"

Educating and involving the public, including subjects

Dr. Pencina highlighted the valuable role that algorithm users can play in their development. He gave the example of an algorithm that was intended to help nurses in their duties, but nurses refused to use it because it was not consistent with their experience. Subsequently, nurses were engaged in helping to rebuild the algorithm; ultimately, they liked and used the product. Engagement is critical.

Dr. Blackman said that educating the general public about AI is not a major priority. The public should not be shut out, but their input is generally poorly informed and they are just not that interested. Engagement may be appropriate for a single project, but there are thousands of models. Even if members of the public were educated enough to participate, the job is just too overwhelming in terms of scale. Instead, the IRB needs to know how to ask the right questions: "Prove you did this." "Did you check this?"

Dr. Pencina, however, pointed to specific areas in which input from the population of patients contributing data is important. One is in the design of informed consent forms, which are often designed by lawyers without taking stakeholders into account. Dr. Batalis added that people who bear the risks should understand those risks.

Mr. Campos also underlined the importance of including participants and communities in the process. Without communication and participation, there will be no trust. Admittedly, however, people currently have varying levels of ability to contribute. He cited the *All of Us* study as a good example of how to honor the voices of the general public.

Ms. Gerke suggested that the standard for effective informed consent in regard to AI should be: "What would a reasonable patient like to know?"

Dr. Silverman stressed that there is a limit to the level of detail a person actually needs or would want to know. If you need a ventilator, that ventilator uses software, and we do not take the time to educate people about it. What we do need them to know is that in any U.S. medical center, their medical records will be used for research.

Mr. Lipset noted that while stakeholder participation in design may be helpful, it should never be conflated with oversight. The issue of responsibility for oversight of the study is entirely separate.

Researcher training

Dr. Vitak turned the discussion to the question of how researchers who will be using AI are prepared to assess risk. How do we encourage them to engage more deeply with ethical issues?

Dr. Blackman argued that young people entering the workforce do not have the power to analyze and respond to ethical concerns. Their attempts to do so will only slow the workflow. Teach them to be capable of raising issues; they need to be able to "smell the smoke" and alert people to potential problems in a way that does not put their jobs on the line.

Dr. Mah said that a lot of the training in the field of ethics in HSR is not meaningful and misses key issues. Instead, people simply click through required online training with little thought. If there is any new training, it should be effective and practical. It is important not to reinforce the perception of IRB review as merely a "checkbox system." He suggested that the field is confused about what quality means in a regulatory environment. Dr. Vitak agreed and noted that, encouragingly, some machine learning conferences have required authors to describe potential downstream consequences of their research. Still, many authors do not know how to engage in that kind of reflection. Mr. McKee also said that expanded training would be helpful and that it should offer developers increased exposure to ethical ideas and concepts.

An online listener said there is a lot of pushback from industry about how ethical review slows development. How can governmental influence be leveraged to ensure that ethicists have a seat at the table? Dr. Blackman responded that identifying and addressing such issues simply does take time, and there is no point trying to convince people it does not. It is important to stress that the purpose is to avoid hurting people. Dr. Batalis further suggested that government can help by providing a framework and guidance for this type of assessment. In addition, it should consider appropriate incentives for doing so.

SESSION II: EXAMINING THE CHALLENGES OF APPLYING THE BELMONT PRINCIPLES AND THE COMMON RULE TO RESEARCH INVOLVING AI

- *Moderator:* Eric Mah, Ed.D., M.H.S., Associate Dean, Clinical and Translational Research, University of California, San Diego

Session II Introduction

- Eric Mah

Dr. Mah observed that the opportunity presented by this workshop demonstrates OHRP's commitment to the research community. He urged panelists to explore the risks and challenges associated with AI and determine how the ethical framework of the Belmont Principles and the regulatory framework of the Common Rule could be applied to research that incorporates these evolving tools.

Key Regulatory Terms and AI Tools in Human Subjects Research

- Iris Jenkins, Ph.D., Director of Research Integrity and Consultation, Virginia Tech

Dr. Jenkins focused her remarks on how key terms within the Common Rule apply to AI tools that might be used in human subjects research. She began by reviewing regulatory definitions of "research" and "human subject" according to the Common Rule (45 CFR 46):

- Research means a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.
- Human subject means a living individual about whom an investigator (whether professional or student) conducting research: (i) Obtains information or biospecimens through intervention or interaction with the Individual, and uses, studies, or analyzes the information or biospecimens; or (ii) Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens.

OHRP has developed a decision tree used to determine whether or not an activity is human subjects research. First, it must be determined whether the activity is a systematic investigation designed to develop or contribute to generalizable knowledge (45 CFR 46.102[l]). If so, it is still necessary to determine whether human subjects are involved:

- About whom? Are people involved?
- Information or biospecimens obtained through intervention or interaction?
- Identifiable private information?
- Identifiable biospecimens?

It is possible to have an activity that is not research but does involve human subjects (for example, an evaluation of a specific activity that will not yield generalizable knowledge).

Dr. Jenkins offered four scenarios as examples of activities that might or might not be research. The first scenario, which does not involve AI, seeks to evaluate the effects of television viewing on food intake by providing participants a meal while they view television and another meal while they are not viewing television. This is research (whether or not it is published). A second scenario assesses the impact of an educational intervention by a robot on food intake; one set of subjects receives this intervention and another does not. This still meets the definition of research.

The third scenario is "a bit more complicated," Dr. Jenkins said. Investigators propose to train an algorithm to predict the amount of food a person will eat while viewing television based on various demographic information. They will then feed previously collected, deidentified data into an AI tool to reach their conclusions. While this is a systematic investigation, researchers would say they are simply trying to train their algorithm. Since the information used is deidentified secondary information, it may not be seen as involving human subjects.

In the fourth and final scenario, identifiable information will be deidentified and fed into an AI tool to generate findings. Although they used identifiable private data, investigators might argue that their study is not "about" any person. Even if the scenario is considered human subjects research, it may be exempt if findings are recorded in such a way that subject identity cannot be readily ascertained. Dr. Jenkins suggested that IRBs probably do not interpret this type of scenario in a consistent way.

Dr. Jenkins observed that many AI projects will not readily fit the definition of "human subjects research." Further, many of those that do meet this definition may fall into the exempt category and receive limited oversight.

Concerns related to privacy and confidentiality remain, however. People who contributed data may not have understood it would be used in this way. Determining what constitutes "publicly available" data is also not a straightforward process. Further, even when data are deidentified, datasets can be combined in such a way that an individual's identity can be pinpointed. This means that individuals may face some risks to which they never consented.

How Does the Use of AI in Research Test the Notions of Personal Privacy and Identifiability of Data or Biospecimens?

- Benjamin C. Silverman, M.D., Senior IRB Chair, Human Research Affairs, Mass General Brigham; Director of Ethics, McLean Institute for Technology in Psychiatry, McLean Hospital; Chair, Mass General Brigham Embryonic Stem Cell Research Oversight (ESCRO) Committee; Assistant Professor of Psychiatry and Faculty Member, Center for Bioethics, Harvard Medical School

Dr. Silverman observed that almost all AI research projects start with the use of large datasets (e.g., of medical records data) to build, train, and validate an AI model. This process often requires data sharing and combining datasets. The guiding principle is "the more data the better." By incorporating massive amounts of data, investigators hope to ensure representative datasets.

Dr. Silverman cautioned that AI model development in research poses unique challenges related to privacy and confidentiality and transparency about data use. IRBs care about the potential for subjects to be individually identified when their data are used, stored, and shared because of regulatory requirements. The *Belmont Report's* principle, "respect for

persons," is especially relevant. The Declaration of Helsinki specifically states that there is a duty to protect the privacy of research participants and the confidentiality of their personal information. Under the Common Rule, identifiable data for an individual can be considered as a human subject in itself, and adequate provisions to protect the privacy of subjects and maintain the confidentiality of data may be required (45 CFR 46.11[a][7]). In general, IRBs must ensure that risks to subjects are minimized, which includes the risk of violations of privacy and confidentiality. IRBs strive for minimum necessary data use, deidentifying and anonymizing data when possible (45 CFR 46.111[a][1]).

AI poses unique challenges to the methods and safeguards IRBs typically rely on to meet these ethical and regulatory requirements. With AI, as noted above, the more data the better. How can this goal be reconciled with the aim of minimizing data use? With AI, deidentification may no longer be possible. Also, informed consent for large datasets is difficult if not impossible; if required, it could lead to less representative datasets and more biased algorithms.

Many methods of deidentification exist. It is possible to deidentify data by deleting key data points, obfuscating or transforming data, coding or linking the data (which would make them indirectly identifiable) or anonymizing the data, a process in which identifiers are irreversibly stripped. However, the science of reidentification is increasingly used to overcome typical methods of deidentification. It is possible to use auxiliary information and datasets to link information (data mosaic effect). The power of big data and generative AI increases the probability of reidentification, essentially amplifying the power of the data mosaic effect. Even further, enhanced pattern recognition makes it possible for AI to overlap seemingly disconnected datasets to pinpoint an individual. Information that has been anonymized and scrubbed of all identifiers can now be reidentified with emerging AI strategies. Consequently, given the new reality of identifiability using generative AI, deidentifying and anonymizing data as a method or safeguard to minimize risks may provide a false sense of security.

The Common Rule allows for judgment about whether or not the identity of the subject may readily be ascertained, and people may come to different conclusions. Dr. Silverman argued that in reality, there is a continuum between identifiable and not identifiable data, and studies that use AI are likely to fall somewhere between extremes.

Dr. Silverman asked, if we take the stance that data cannot be deidentified, what does this mean for consent requirements? Should consent be required? And if so, should we work toward creation of large consented datasets for AI model development? He proposed that if the data are identifiable, then consent may be required. Requiring consent would, he argued, improve transparency about data use, show respect for persons, and honor the autonomy and agency of research participants. The approach could mirror consent considerations for other potentially identifiable data, such as whole-genome sequencing data.

However, Dr. Silverman admitted that requiring consent is likely to be impracticable for the large datasets required for AI model development. Further, it is likely to result in reduced data quality and less diverse datasets, in part because people who enroll in research studies are generally not a representative population. The approach also has the potential to cause more data bias and worsen algorithmic bias from the subsequently developed AI models, ultimately creating bigger justice concerns. Additionally, it can be argued that despite the evident risk of reidentification, use of the large and representative datasets required for AI model development may be permissible under both Common Rule and Health Insurance Portability and Accountability Act (HIPAA) criteria for waiver of consent, though features of AI pose challenges to those criteria.

In conclusion, Dr. Silverman stressed the following:

- AI model development in research requires large and representative datasets. Most medical AI models are trained and validated on large existing datasets such as medical records, which are typically accessed under a waiver of consent.
- Generative AI and emerging AI strategies challenge our traditional understandings about personal privacy and identifiability of data. In the context of emerging AI technologies, it may no longer be possible to truly deidentify data.
- Researchers, Human Research Protection Programs (HRPPs), and IRBs may need to find new ways to minimize risks when using data for AI model development and may need to adjust review pathways accordingly.
- For large datasets required to develop AI algorithms, there is a tension between the possible benefits of requiring consent and the potential for less representation and more bias and discrimination.

- Beyond consent, public notification and education about the use of personal and medical data for research are critical to enhancing transparency and trust. This is not exclusive to AI model development research.

Disclosing the Role and Use of AI in HSR

- Sara Gerke, Dipl.-Jur. Univ., M.A., Associate Professor of Law and Richard W. & Marie L. Corman Scholar, College of Law, University of Illinois Urbana-Champaign

Ms. Gerke highlighted a few common uses of AI in human subjects research. AI can be used to improve trial design by identifying biases in datasets (such as gender imbalance) and simulating trials to show flaws in design. It can help recruit and engage appropriate subjects. Some forms of AI can make the informed consent process more understandable by personalizing it to better communicate with specific subjects. AI can help ensure compliance with regulatory requirements. It can also collect and analyze data from human subjects, including qualitative and real-time data. As trials proceed, AI can be useful in risk monitoring and assuring safety. For example, it can monitor adherence to drug regimens. Finally, Ms. Gerke noted that AI can be itself a study subject.

AI is rapidly transforming the drug development landscape, challenging regulators to keep up with the pace of change. Ms. Gerke noted the rapid increase in the number of submissions to the FDA referencing AI (approximately 300 submissions from 2016 to the present). The use of AI ranges from drug discovery to clinical research.

While AI creates opportunities for improved health care, it may also raise safety considerations. She noted that although the FDA has reviewed and authorized marketing for 950 AI-based medical devices, these devices typically do not need to show clinical evidence.

Informed consent

Ms. Gerke discussed issues related to informed consent. After reviewing regulatory definitions of "research" and "human subject," she observed that some AI projects may not fall under these definitions. The definition is met in some instances, such as when an investigator conducts a clinical trial to test the AI model for its safety and effectiveness. Informed consent follows from one of the *Belmont Report*'s principles, "respect for persons." In general, the Common Rule requires that consent be voluntary, that the subject be given sufficient information to make an informed decision, that the language be understandable to the subject, and that no exculpatory language be included.

Under the Common Rule, informed consent must include the purpose of the research and procedures, reasonably foreseeable risks and discomforts, benefits, appropriate alternative procedures or courses of treatment, confidentiality, compensation and medical treatment, contact information, voluntary participation, and future use of data and biospecimens (if applicable). Other elements that might be included are unforeseeable risks, termination of participation, additional costs, consequences of withdrawal, new findings, the number of subjects, commercial profit from biospecimens, disclosure of clinically relevant results, and the use of whole-genome sequencing.

Ms. Gerke posed the question of whether investigators are required to disclose the use of AI in research and whether, if it is not required, they should do so anyway. She argued that although it is unclear in some cases whether the use of AI in human subjects research must be disclosed to participants under the Common Rule, doing so promotes transparency and trust. Although the diverse uses of AI in human subjects research make disclosure challenging, nothing prevents researchers from doing so.

What must or should be disclosed depends on the context in which AI is used. For example, Ms. Gerke identified the key types of information that she believes should be included on the label of AI- or machine-learning-based medical devices: model identifiers, model type, model characteristics, indications for use, validation and model performance, details on the datasets, preparation before use and application, model limitations, warnings and precautions, alternative choices, and privacy and security implications.

Panel II Discussion

Disclosure of AI components

Dr. Mah asked speakers to consider when it was appropriate to disclose AI study components to subjects. Mr. Campos said AI components definitely should be disclosed. Dr. Vitak added that she "couldn't think of a reason not to disclose." There are no drawbacks to disclosure, but the likely result of failure to disclose is lost trust. Data science has a "huge distrust issue" with the public. In her experience, people who use social media do not think of their data as public. If it is being scooped up and used, they want some notification that this is happening. Dr. Batalis said it is particularly important to disclose breaches that may make subjects vulnerable and to do this in a timely way.

Mr. Lipset agreed that respect for persons requires disclosure of how their data are being used, but he stressed that this is different from having to disclose all use of AI in a research context. Some uses of AI, such as to recruit subjects or staff, are irrelevant to subjects and need not be disclosed.

Dr. Jenkins observed that the issue of disclosure of the use of AI is distinctly different from informed consent. In most cases, the subject is not being asked to consent to the use of AI, but rather is being informed that this is the case.

Dr. Silverman said the answer must be, "It depends." The more complex question would be, "disclose what?" Doctors do not typically disclose how they are making decisions or where they studied. Are we justified in holding AI to a higher standard?

Dr. Mah asked whether disclosure might lessen participation. For example, today almost all food is genetically modified in some way; however, this is not a public message that is stressed, though in the past genetically modified food was considered controversial. Mr. Campos said that no one needs to be protected from the truth, and investigators should take whatever time it takes to communicate clearly. Mr. Lipset suggested that requiring innovators to disclose their use of human data should not have a chilling effect on their work. An innovator should be proud to tell research participants how their data were used. If that is not the case, there may be a problem.

Dr. Blackman opined that if the goal is to empower a person to make the right decision about participation, they should be given the right amount of information to make this possible. The way to accomplish this is not to tell them everything that is true. People can be overwhelmed by too many facts that are not relevant.

Dr. Vitak pointed to food labels divulging information about nutrition as an example of how to facilitate informed decision-making. This idea has been used by Apple and Google in providing information about data collection by mobile apps. She said employing "layered labels" is a useful way to provide more data to people who actually want it. For example, a QR code could provide additional information that goes beyond what the label provides. This provides transparency without overload. Similarly, we have a baseline for consent, but perhaps we need a way to provide more in-depth information to those who want it. Mr. Smith suggested that the notification of nutritional content makes a difference whether anyone looks at it or not, in that it holds the entity putting forth those claims accountable.

Confidentiality: Advice for IRBs

Dr. Mah noted that IRBs sometimes think they are the last defense of people's privacy, but increasingly, people have different attitudes toward what is important to share. Younger people, for example, may be less concerned about privacy. What advice would panelists give to overloaded IRBs as they consider the issues of privacy and confidentiality in regard to AI?

Dr. Silverman's advice was succinct: "Never worry alone." IRBs play a critical role, but they need to look around for people who have the expertise they lack. In 5 or 10 years, he maintained, every single protocol will have some AI involvement. It is problematic for IRB review to be the last stop in moving the protocol forward, because concerns put forward at that point make the IRB appear to be an obstacle. People writing applications need to be encouraged to collaborate with the IRB as the protocol is developed so the review process is not perceived as an obstacle.

Mitigating the risk of bias

Mr. Smith invited panelists to further consider the challenge of achieving representative data and mitigating risks. One challenge is that underrepresented groups may want their data withdrawn, undermining the applicability of an AI tool.

Dr. Mah relayed a question from a member of the audience, who also specifically asked the panelists' opinions on whether individuals should be allowed to remove their data from large datasets. Dr. Mah commented that this might be impractical; further, if enough samples are removed from underrepresented populations, it could introduce bias. Dr. Vitak agreed that that there was a potential for bias but argued that this could be avoided by engaging underrepresented populations and hearing their voices before and during data collection ("data antisubordination"). Ms. Gerke observed that samples might be biased in any case because some minority groups face reduced access to the health care system.

Dr. Silverman said that secondary use of existing data should draw concern from the IRB in some cases. For example, reidentification of samples that may contain data about illicit substance use could have serious consequences for individuals. Nevertheless, IRBs are told in the regulations that they should not consider future risks, and the effects of bias in a data sample are generally downstream. The fact that data are biased does not directly affect the subject from whom data were taken; it affects future subjects.

Dr. Mah observed that there tend to be two schools of thought among IRBs. Some believe it is their role to ensure the study is the best possible study; others think this is overreaching. IRBs are often paternalistic about privacy, but subjects may not be equally concerned. IRBs also tend to believe that questions of bias are resting on their shoulders, but these have not yet been solved in the larger cultural context.

Noting that there is an important difference between "output" and "outcome," Dr. Blackman said it is possible to have a biased algorithm that is used constructively by a well-informed human. Output is narrow, a sliver of what the outcome of algorithm use might look like. If a particular AI product benefits some groups more than others, should the groups that do benefit be denied a life-saving advance?

Mr. Smith suggested that it might not be the IRB's role to make this type of judgement. Perhaps a flawed algorithm will be refined later. Dr. Mah said that the IRB system is not designed to solve problems the AI community itself has not solved. Mr. McKee noted, however, that while it is not an IRB's job to sort through the complexity of such issues, it would still be in the interests of transparency and accountability to have such limitations on the record.

Ms. Gerke said that there are clearly new challenges associated with AI and we cannot expect IRBs to face them alone. They need support, and there is a lot of work to be done. FDA needs to figure out how to evaluate AI tools, such as generative AI, and bring them into the clinic. If the regulator does not yet know where it needs to go, how can we expect IRBs to figure it out for themselves?

Dr. Vitak said that there should be some entity that performs this type of meaningful analysis. If an algorithm prescribing care is not appropriate for the population at a particular hospital, this could cause harm. Affected communities need to be informed and given a voice. Dr. Mah closed the discussion by reiterating the view that the IRB must partner with others at the institution who have clearly defined roles in reviewing AI projects. Informed partners can help the IRB analyze ethical issues and ensure that community interests and concerns are clearly understood.

SESSION III: EXPLORING THE CHALLENGES WITH MAINTAINING PUBLIC TRUST AND ALIGNING AI WITH HUMAN VALUES

- *Moderator:* Jeffery Smith, M.P.P., Deputy Director, Certification and Testing Division, Office of the National Coordinator for Health IT, HHS

Session III Introduction

- Jeffery Smith

Mr. Smith invited panelists to a discussion on how to ensure that future uses of AI are aligned with human values so that public trust can be retained.

Data Privacy Isn't as Important as You Think

- Reid Blackman, Ph.D., Founder and CEO, Virtue Consultants

Dr. Blackman said that "we should care far less about 'our data' being collected than we should about what it's used for." He stressed that he does believe privacy is important and that his presentation, while he hoped it would give food for thought, is "only part of the meal."

Dr. Blackman highlighted three "confusions." First, he said that the outcry against companies "stealing my data" is not always justified. The slide from "these data are about me" to "these data are mine" is questionable. More premises are needed to justify control of particular personal data.

The second confusion relates to harms attributed to violations of privacy. Dr. Blackman provided examples that if someone observes that Achilles is sensitive about protecting his heel and uses that knowledge to kill him, it is not the observation of his vulnerability that is wrong but how that knowledge is used; people own kitchen knives that could be used to commit a crime, but owning them is not a crime.

The third and final confusion Dr. Blackman cited is that all data collection is surveillance. He felt the term was used "sloppily." For example, he said, if someone who knows a subject coincidentally happens to notice their name in a set of aggregated data that have not been anonymized, that is not surveillance. However, if an employee intentionally seeks highly sensitive information about someone, that does constitute surveillance. The level of the violation depends on its nature.

"Am I letting companies off the hook?" Dr. Blackman asked. He stressed that surveillance is serious and people should be deeply concerned about it. Certain uses of data should certainly be illegal. We need to keep asking how the data are being used. This is important because if all data collection by AI were somehow halted, wrongdoing might be avoided, but would also stop a lot of "good-doing." Health care data, properly used, may save millions of lives.

AI and Values: Alignment, Privacy, and Autonomy

- Karina Vold, Ph.D., Assistant Professor, Institute for the History and Philosophy of Science and Technology, University of Toronto

What is value alignment and why is it challenging? Dr. Vold focused her remarks on issues that arise in the effort to build an advanced AI system that is aligned with values. She quoted the aspiration expressed in the Asilomar AI Principles: (2017): "Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation."

Would a very smart system be more benevolent than humans? Dr. Vold suggested that an AI system might in fact be intelligent but not ethical. She quoted Bostrom (2014, p. 107): "Intelligence and final goals are orthogonal: more or less any level of intelligence could in principle be combined with more or less any final goal."

Dr. Vold cited four challenges in aligning AI with human values.

1. Identifying human values. Humans are often confused and conflicted. Also, cultures differ in their understanding of values, opening the door to "moral pluralism." Is it possible to identify and implement a complete set of possible values? Or is this, as many pluralists believe, an example of an unsolvable moral dilemma?

2. The King Midas problem. Encoding values in a machine is not easy. Dr. Vold cited the story of Midas to illustrate the importance of stating goals clearly. Midas' wish was not articulated carefully enough and his own food, and his beloved daughter, were turned to gold at his touch. Similarly, a robot told only to clean up as much dirt as possible from a room might focus on a specific spot and not cover the entire room.

3. Moral imperfection. We actually do not want a machine to be just like us, Dr. Vold cautioned. This would only amplify our own imperfections.

4. Moral progress. Even if we could build a machine to match only the best of us, we would not want to stop there; we would want to allow it to progress.

Dr. Vold then turned to the areas of privacy and autonomy, where she stressed the distinction between instrumental and intrinsic value. Do personal data have intrinsic value, or does their value stem only from their utility? If nothing is done with

data collected from an individual, has that person's privacy been violated?

Privacy issues relate to the values of autonomy, reputation, intimacy, bodily integrity, self-development, dignity, and fairness. Autonomy, more specifically, implies the ability of an individual to reflect on and decide freely about values, actions, and behavior and act on those choices. If others have access to information about an individual, they can use it to influence that person, threatening their ability to make independent decisions. In itself, attempting to influence people does not raise a moral problem. If one appeals to a person's rationality, the person is still in control of the outcome. In contrast, personalized targeting risks being manipulative:

- It is more pervasive and covert. As choices are made through the medium of digital technology, we lose the ability to detect the ways in which our choices are influenced.
- Targeted ads are more likely to be deceptive. When it is only necessary to speak to one individual at a time, it is much easier to warp the truth and avoid being challenged.

Dr. Vold highlighted three key technological advances around personal data that have important implications for privacy and autonomy and their associated values:

1. Data collection. The type and amount of data that can be collected has increased. Smart phones carry a tremendous amount of information about their owners. The human brain has imperfect memory, but a smart phone can recall perfectly where you shop, who you called, what you searched for on the internet, and much more.
2. Data storage and access. A few powerful companies are able to buy and sell the tremendous amount of data that is now available on each of us. This asymmetric power is concerning.
3. Data use. As discussed above, the ways in which personal data can be used have changed dramatically. Companies and organizations now have the means to personalize all sorts of interventions and influence our decision-making.

The Role of AI in Biomedical and Health Research—A Research Participant's Perspective

- Hugo Campos, Participant Ambassador, NIH *All of Us* Research Program

Mr. Campos identified himself as a person living with a genetic heart condition who has participated in several studies and research projects. He has been an active user of generative AI since 2023. He observed that as AI becomes more integrated with health care and medical research, it raises questions related to trust, power, and ethics. These questions must be taken seriously, for history shows us that unchecked power and lack of transparency can lead to exploitation.

While predictive AI has been used for years without much public awareness or concern, generative AI will require stronger ethical safeguards because of its ability to mimic human communication and its potential to interact directly with both research participants and researchers, Mr. Campos said. This has also been called "relational AI." Relational AI can create a false sense of closeness between AI and humans, increasing the risk of manipulation, misinformation, and the blurring of boundaries in its interactions with humans. He gave the example of a friend who had what she thought was a conversation with a helpful staff member who resolved an error in her medication shipment; later, she learned that she was actually talking to a chatbot.

AI interactions are blurring the lines between human and machine, and this may undermine trust. Mr. Campos asked, with this in mind, should the consent process mention AI explicitly? Should predictive AI be treated differently from relational AI?

Mr. Campos highlighted three top concerns: transparency, reliability, and accountability.

1. Transparency. As a participant, he said, he does not care to know all the details about how AI works. But in order to trust it, he must trust the people and the institutions conducting the work. He needs to know that he will not be harmed.
2. Reliability. He expects that AI used in human subjects research will provide accurate, useful, and helpful responses. Mr. Campos stressed that hallucinations generated by AI do undermine trust.

3. Accountability. It is important to identify mechanisms to remedy harms caused by wrong decisions made by the AI. If the AI makes a mistake and harms a participant, who is responsible for the algorithm? Who is monitoring the process and will know if this happens?

Mr. Campos noted that moral principles often differ across cultures, communities, and populations. Training an AI system on the ethical norms of one group and applying them to another group is a form of oppression. In medical research, this might mean imposing our views on issues like gender roles, mental health, autonomy, death, and decision-making on somebody else. Also, using AI to generate synthetic data may risk amplifying or perpetuating existing biases. He also suggested that the use of generative AI for participant recruitment, retention, and engagement strategies is concerning because of its potential for influencing behavior, coercion, and manipulation.

Risks associated with AI can be mitigated if researchers proactively design AI with respect for human dignity, fairness, and transparency in mind, Mr. Campos said. It is important to engage diverse stakeholders in the cocreation of research, especially if subjects belong to communities that are underrepresented in biomedical research. Risks associated with generative AI can be further reduced by communicating results and risks clearly, as well as by educating people about the use of AI in research.

Panel III Discussion

Mr. Smith valued the broad exploration of the implications of AI for public trust and human values. He stressed the importance of taking a pragmatic view of the topic and appreciated Dr. Blackman's "somewhat provocative" presentation on what does and does not matter. He underlined Dr. Vold's clarity on the subject of privacy and autonomy, as well as Mr. Campos' focus on transparency and accountability.

Mr. Smith noted, in the United States, laws exist that focus not on the possession of certain data, such as genetic material, but on how they are used. It is recognized that no single agency can be saddled with the whole responsibility to ensure ethical use of such sensitive data. He invited general reflections from the panel before moving on to specific questions.

Targeted marketing

Dr. Blackman said he does think privacy and data stewardship are important, but he was not convinced that we need to be worked up about targeted marketing. Are ads generally bad? Targeted or not, ads are not generally rational; most marketing is not trying to engage our reasoning, and he was inclined to think this is not in itself unethical. What makes this situation a particular source of concern with AI?

Dr. Vold responded that when she was a child, toy commercials always came on when children were likely to be watching. That is targeted advertising. However, there was usually only one television in a household. Now, each individual may have a computer recording his or her individual preferences. Findings are not only used to influence what jeans are marketed to an individual (not a major concern) but also to frame the news they receive and the health options they are offered. It would be possible for different individuals to visit the same website and for each of them to find different information. It is as if people were sitting across from each other in the same restaurant and seeing different menus. Depending on how personal data are used, the stakes can be high enough to justify concern, she noted.

Surveillance

Dr. Vold asked Dr. Blackman how he defines surveillance and observed that in terms of his illustrative scenario, companies may have a lot of information about Achilles, but Achilles typically knows little or nothing about these powerful companies. The power dynamic is very different. Is this relevant? Dr. Blackman responded that he did not have a tight definition of surveillance to offer, but in his view, it is systematic and has something personal about it. People talk about companies "knowing" certain things about them, he said, "but I don't think Walmart knows things." Dr. Blackman said he was not sure the power dynamic mattered very much. A lot of organizations have a lot of power and may know a lot about us, but in an impersonal way. He was not convinced the power dynamic matters.

Access to personal data

Mr. Campos said he is not so concerned about the fact that data are collected about him, but he does care about what kind of data are collected and whether or not they are shared. For example, he has an implantable defibrillator that provides

information directly to the clinic, but it is difficult for him to access the information it provides. He finds this objectionable. If data are collected by a medical device he is wearing, he would like someone to review the data and share them with him so that he can better manage his health.

Privacy and autonomy

Dr. Batalis was interested in Dr. Vold's comments on privacy and autonomy. She wondered if Dr. Vold could comment on whether some types of information were private and should not be collected and used at all. For example, some companies use AI to provide therapy sessions. Should those highly personal data be collected?

Dr. Vold responded that using apps to facilitate cognitive behavioral therapy with a human therapist offers a useful service to people in remote places. However, one app she knows of has some serious privacy issues. This is the type of data that occurs in a private discussion and should not be collected.

Dr. Vold said when data are collected, people need to be concerned about how they are shared. People know that companies have their data, but may not object. However, they do not know how those companies are trading and selling their data behind the scenes to entities they know nothing about. Finally, there may be data that should not be used to target individuals. For example, it may be fine to target people on the basis of chosen behaviors such as a love of hiking, but not on the basis of their racial or gender identity.

Changing views of privacy

Dr. Mah said that ideas of what privacy means have shifted, but the IRB world has not kept up. People no longer read end user license agreements and privacy statements and there is less concern about the issue. We have given up vast amounts of information to companies, and we have had little or no choice in the matter. He himself is unconcerned about using a mapping app that knows where he goes as long as it guides him there successfully. Patients' data are being shared as soon as they enter a hospital. Where are the real risks to individuals in this context? If it helps to improve notes for physicians and improves health care by giving up some privacy, should that be a concern?

Ms. Gerke said that the level of concern one should have about privacy is highly dependent on the context. Using Google maps is a choice. If an individual wants the convenience, they use the app. However, the implant that Mr. Campos has did not offer him a real choice. He could give up his data or risk dying. This is very different. Ms. Gerke wrote a paper with colleagues in 2020 on the subject of access to patient data from pacemakers; the cardiologist can only access a PDF summary, while the manufacturer retains the raw data (Cohen, Gerke, & Kramer, 2020).

Individual risk

Dr. Mah said he was still not hearing of cases in which there are privacy risks that affect the individual. Are there cases in which there are significant risks to the individual associated with AI that merit concern?

Mr. Lipset was concerned that AI in the context of social media and AI in the context of health are being conflated. He noted that the cost of medical care is the leading cause of bankruptcy in this country, so the stakes are high. There is no comparable cost associated with our use of social media. He added that the fact that data are used for commercial purposes makes the situation different. Dr. Blackman rejoined that he was not sure that it was a bad thing that a company was making money and noted, "That's usual." Mr. Lipset said that if someone is using his data to make money, he should be able to participate.

Dr. Blackman said there might be a legitimate concern in a situation in which a chatbot is used to forge a deep emotional connection with a subject, then the connection ends and the app pulls the plug. Emotional harm could result. People do in fact form such relationships, he noted, and this is a plausible source of harm to an individual. Dr. Mah pointed out, however, that the individual may have consented to engage in the conversation and may have been told it would end. Most IRBs would require the terms of the agreement to be disclosed. However, they might not approve a study in which people are not told they will be talking with a chatbot. The informed consent process should make clear that when the study ends, they may feel lonely or depressed.

There has been a major shift in the last decade, Dr. Vitak reported, in how researchers think about privacy risks. There is much more concern about collective and networked privacy over individual privacy. Helen Nissenbaum has been an important thinker in this area (via her work on contextual integrity) and has stressed the importance of social context and community norms in assessing privacy violations (Nissenbaum, 2009; 2019).

Two types of data should be clearly distinguished in this discussion, Dr. Batalis cautioned: data that are observational and data generated for the purpose of completing a study. "Some data would not exist if I had not participated in a study. When I walk into a café, I know that my presence may be observed, but when I walk into a doctor's office my expectations are different," she said.

Dr. Silverman said the example of the chatbot that stops having intimate conversations with a vulnerable individual is not unique to the AI world. A participant in a randomized controlled trial may receive a drug that works well for them but never have access to it again. He suggested that the only potential for individual harm unique to the AI world would be if the AI is autonomously making medical decisions, which does not typically happen now. In some instances, people will need to know something about the algorithms that are being used to make decisions about their care.

Mr. McKee said the chatbot scenario does represent a heightened risk for vulnerable individuals, especially given the increased capacity of language models for personalized response. The social risks associated with this type of AI are increasingly recognized and discussed by AI ethicists.

Dr. Blackman asked whether there are studies a researcher might propose to which an individual might consent, but the IRB would say "No, you cannot participate?" Dr. Mah gave the example of a biomedical study in which people were being asked to forgo a certain established standard of care. The IRB might say there is an alternative, and we will not allow you to be exposed to this harm. Dr. Silverman said we routinely run Phase I trials with drugs that have never been given to humans, but in such cases we rely on preclinical trial data and a transparent consent process to approve such studies. There are limits to risks from preclinical data (e.g., fatality) that would be acceptable to most IRBs.

Dr. Vold said that there are potential individual harms associated with the application of nonrepresentative datasets. Individuals may agree to a study but not know how data for an algorithm used in the study were developed and whether they represent people like themselves. X-rays have been known to miss certain conditions such as pneumonia in people with darker skin. Another risk cited by Mr. Smith is race-based medicine, in which race has improperly been used as a variable, resulting in harmful bias.

REFERENCES

Asilomar AI Principles. 2017. Future of Life Institute. Retrieved from https://futureoflife.org/open-letter/ai-principles/

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies.* University of Oxford.

Cleveland, M. E. C., Aikens, B., De Maio, F., et al. (2024). Efforts in organized medicine to eliminate harmful race-based clinical algorithms. *JAMA Netw Open*, *7*(3). e241121. https://doi.org/10.1001/jama.2024.1121

Cohen, G., Gerke, S., Kramer, D. B. (2020). Ethical and legal implications of remote monitoring of medical devices. *Milbank Quarterly*, *98*(4), 1257–1289. https://doi.org/10.1111/1468-0009.12481

Gerke, S. (2023). "Nutrition facts labels" for artificial intelligence/machine learning-based medical devices—The urgent need for labeling standards. *The George Washington Law Review, 91*(79). https://www.gwlr.org/wp-content/uploads/2023/03/91-Geo.-Wash.-L.-Rev.-79-2023.pdf

McKee, K. R. (2024). Human participants in AI research: Ethics and transparency in practice. *IEEE Transactions on Technology and Society*. https://ieeexplore.ieee.org/document/10664609

Messeri, L., Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627, 49–58. https://doi.org/10.1038/s41586-024-07146-0

Nichol, J. and Tazbaz, T. (2024). Blog: A Lifecycle Management Approach toward Delivering Safe, Effective AI-enabled Health Care.

Nissenbaum, H. (2004). Symposium, privacy as contextual integrity. *Washington Law Review*, 119. https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford University Press.

Nissenbaum, H. (2019). Contextual integrity up and down the data food chain. *Theoretical Inquiries in Law*, *20*(1), 221-256. https://doi.org/10.1515/til-2019-0008

Obermeyer, Z. et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464) 447-453. https://doi.org/10.1126/science.aax2342

Patel V., Shah, M. (2021). Artificial intelligence and machine learning in drug discovery and development. *Intelligent Medicine*, *1026*(2), 134-140. https://doi.org/10.1016/j.imed.2021.10.001

Pencina, M. J., McCall, J., Economou-Zavlanos, N. J. (2024). A federated registration system for artificial intelligence in health. *JAMA*, *332*(10), 789–790. https://doi.org/10.1001/jama.2024.14026

Properzi F., van Tongeren, T. (2024). The state of digital excellence in the global pharmaceutical industry, 2023: Clinical operations. CT Consulting.

U.S. Food and Drug Administration (2023). Using artificial intelligence and machine learning in the development of drug and biological products. Retrieved from https://www.fda.gov/media/167973/download

Volk, K., Whittlestone, J. (2020). Privacy, autonomy, and personalized targeting: Rethinking how personal data is used. In Veliz, C. (ed.) *Report on data, privacy, and the individual.* Center for the Governance of Change.

Wong, A., Otles, E., Donnelly, J. P., et al. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med.*, *181*(8), 1065-1070. https://doi.org/10.1001/jamainternmed.2021.2626