

Future of Institutions

DIALOGUE SUMMARY

Title/topic:

How to develop new institutional arrangements to help researchers, both in industry and academia, mitigate the ethical and societal risks that arise from their research

Organizers:

Center for Advanced Study in the Behavioral Sciences (CASBS) – Stanford University

Key takeaways:

This Dialogue centered around the question: “How can we effectively embed the consideration and mitigation of ethical and societal consequences of research into every aspect of the research process?” Current solutions to this question are limited in scope and focused on individuals’ intentions rather than collective mechanisms. For example, the purview of Institutional Review Boards (IRBs) is limited to risks facing human subjects involved in the research. The Common Rule does not consider risks to communities and societies. Other solutions rely on opt-in approaches or ethical reflection after the research is already complete. Particularly in the research fields of artificial intelligence (AI), machine learning, environment, and climate, investigators must consider how their discoveries could affect societies, well *before* the society is confronted with any deleterious consequences. From [discriminatory bail algorithms](#), to [inadequate charging stations for electric vehicles](#), to the [climate costs of cryptocurrency mining](#), our discoveries and developments affect individuals and societies.

Dialogue participants converged on three essential factors that are either nonexistent or ineffective in our current institutions:

- *Community-building and engagement.* Researchers must engage with communities to understand how new technologies and research might affect them; otherwise, researchers risk misinterpreting or overlooking relevant outcomes. Furthermore, community engagement must be brought in at the beginning of the research process when such engagement can still influence research design.
- *Information-sharing across institutions and between different organizational levels.* Many institutional solutions are erected within specific contexts, with little attempt at application of lessons learned in one context to others that could benefit from such lessons. Participants were particularly concerned with how to share information across institutions or organizations that have vastly different resources and privacy considerations, while prioritizing institutional transparency.
- *“Institutional change evangelism.”* As one participant noted, technological innovation can include institutional structures that inherently lead to inequity. For example, focusing on prototype development and initial market deployment of electric vehicles necessarily prioritizes some audiences over others (e.g., wealthy consumers). Unless researchers and developers consider how audiences may change as a technology scales, or how the act of scaling such technology could influence society, keeping institutional structures fixed throughout the life cycle of a technology or climate solution will hinder its development and scalability. Therefore,

there is a need for researchers and developers to consider the goals of and principles governing an institutional arrangement *before* seeking to scale.

For all these factors, Dialogue participants were less concerned with establishing a theoretical basis for why these factors matter and, instead, were concerned with establishing an evidence base for what approaches would best fit the context of research and development. To achieve this, participants strongly recommended institutional experimentation with rigorous evaluation, like the evaluations being done by the Ethics & Society Review and Actuate. While some cases exist, there is a need for institutional experimentation across a wide range of contexts.

The challenge(s):

This dialogue explored institutional solutions to the lack of required ethical and societal reflection, beyond human-subject based ethical issues, required by Institutional Review Boards of the research they oversee. Instead of focusing on all possible research, we focused on the fields of AI, computing, and climate change. Participants noted that institutional ethical issues these research fields face may even be shared by others.

Insights for the Future of Institutions:

Dialogue participants employed an institutional lens to consider how to bring ethics reflection further into the research process. Participants defined the general features of an institutional process that could achieve the ideal process for effective ethical reflection. This perspective ensures that all ethical and societal concerns are considered, not just those that are derived from specific research projects or areas.

Solutions:

The most actionable recommendation was to encourage and support institutional experimentation with rigorous evaluations. To understand how effective different institutional arrangements are across a wide range of contexts, there is a need for variations in institutional experimentation. What works in one setting will need to be tweaked or overhauled to fit in another. This cannot be done without robust literature and evidence that shows how arrangements might interact with each other, and how they would perform in different contexts.

Additionally, our conversation ended on the theme of uncertainty. To consider the future of institutions means addressing how these institutions can best adapt to and manage various domains of uncertainty, especially in relation to the development of technology, and maintaining the health and livability of our planet. We need better institutional arrangements in the public and private sectors — ones that crucially acknowledge that we must grapple not just with risk, but with fundamental uncertainties. Risk deals with known outcomes and probabilities that inform decision-making. Uncertainty, on the other hand, is present when we do not know the probabilities of outcomes and the possible permutations of events and outcomes. Our institutions lack the long-term, global, cross-disciplinary thought to best address the ongoing uncertainty.

Participants:

- Michael Bernstein, Stanford University
- Megan Finn, University of Washington
- Eric Horvitz, Microsoft Research
- Tom Kalil, Schmidt Foundation
- Zia Khan, The Rockefeller Foundation
- Margaret Levi, Stanford University
- Arun Majumdar, Stanford University

- Lara Pierpoint, Actuate Innovation
- Dustin Tingley, Harvard University
- Zack Ugolnik, CASBS at Stanford University
- Quinn Waeiss, CASBS at Stanford University