Governing AI Proactively: Cooperative Models of Anticipation and Accountability

Jared Katzman

University of Michigan
School of Information
apricity@umich.edu

1 Background & Motivation

The rapid advancements in artificial intelligence (AI) and other emerging technologies present significant social, economic, and ecological challenges. Current conversations on how society should effectively govern AI development to minimize potential harms have largely focused on risk-based governance frameworks, such as the use of risk and impact assessments (European Parliament 2024; Metcalf, Smith, and Moss 2022). These frameworks typically require corporate technology actors to conduct assessments of high-risk systems, balancing risk precaution with corporate innovation to allow rapid technological development within public guardrails (National Institute of Standards and Technology 2021). However, critical scholarship consistently argues that these expert-led risk assessment regimes are often inadequate (Stilgoe, Owen, and Macnaghten 2013). Corporate actors tend to limit their scope to foreseeable risks that are inexpensive to address and do not impact their profit margins (Selbst 2021). Moreover, these approaches frequently fails to account for the diverse impacts that new technological innovations have on affected individuals and communities, until after the harm has occurred.

My research directly responds to this gap by exploring how communities, particularly those frequently marginalized in the innovation ecosystem, envision and propose governance mechanisms that anticipate the impacts of new technologies. I build on theories of anticipatory governance and responsible research and innovation, asserting that effective governance must move beyond mere foresight to actively co-design institutions and infrastructures that expand community governance capacity (Guston 2013). My overarching research agenda aims to understand how to design sociotechnical systems that center post-capitalist imaginaries and how these models can be applied to new technological advancements like AI (Escobar 2018; Sharma, Kumar, and Nardi 2023). My work is situated within the Human-Computer Interaction (HCI) field, extending its existing calls for participatory and equitable design by pointing to economic democracy as a foundational design consideration.

2 Overview of Research Agenda

My dissertation includes three interrelated projects that build on the following research questions:

1. How can we better anticipate the possible impacts a technical innovation may have on society?
2. What governance models do marginalized communities propose for holding technology developers accountable for potential harms?
3. How can we design cooperative governance structures to anticipate the possible impacts of new AI systems?

The results of my first research project are currently under review at AIES, with early iterations presented in workshops at CHI (Katzman, Dillahunt, and Green 2025). For the second project, we published initial results as late breaking work at CHI (Katzman et al. 2025), and I am currently writing a full paper on our results. Then, I am currently in the scoping and planning phase for my third project, with plans to conduct data collection in the upcoming academic year. For the AIES student program, I would present on my current findings, the motivation for my third research project, and seek feedback on research design.

2.1 Conceptualizing Anticipatory Goverance within AI Development

While the rapid evolution of AI necessitates proactive governance, current development practices often fall short, narrowly focusing on predicting foreseeable risks and technical solutions, which fail to adequately address broader societal impacts or systemic power imbalances (Do et al. 2023; Liu et al. 2022; Metcalf and Moss 2019; Boyarskaya, Olteanu, and Crawford 2020). My work synthesizes insights from future studies, safety engineering, and anticipatory governance to define anticipation as an active practice that combines foresight with actionable governance (Bell 2003; Ericson 2011; Guston 2014)—implementing mitigations before deployment and establishing robust post-deployment response protocols. Recognizing the inherent unpredictability of future outcomes, a core conceptual contribution of this work is that the goal of anticipation isn't to perfectly predict and eliminate harm, but rather to build anticipatory capacity within development teams and institutions. This capacity empowers technologists with the methods, resources, and agency to foresee a wider range of scenarios, proactively

prevent harms, and effectively respond when challenges inevitably arise throughout the AI lifecycle. Ultimately, this conceptual framing provides practical recommendations to support technologists in better anticipating and governing the complex impacts of new innovations by fostering a continuous, proactive, and power-aware approach.

2.2 Designing Anticipatory Governance for Detroit's Food Sovereignty Community

As part of a larger research project on the role of technology in Detroit's food ecosystem, I explored how a community could implement anticipatory governance in practice. Through speculative design workshops, I investigated how the food sovereignty community (1) foresaw the impacts of new technologies and (2) what governance models they proposed to hold technology developers accountable for social impacts on their community. The workshops utilized technical scenarios to prompt reflection on potential impacts and accountability. These scenarios included a speculative digital platform for cooperative produce aggregation, a data dashboard for city tracking of food metrics, and real-world pilots like electric vehicles for mobile markets and autonomous robots for composting.

Throughout our workshops, participants leveraged their lived experiences to foresee a wide range of potential harms from technical innovations, from minimizing farmer autonomy to repressing cultural farming practices (Katzman et al. 2025). They consistently analyzed the power dynamics behind technical systems, asking: "Who gets to make decisions? For whom?" The core insight was that harms from technical systems are not inevitable or purely technical; they emerge from value misalignments and power asymmetries.

Thus, participants proposed the redistribution of power to avoid the set of foreseen harms (Chordia et al. 2024). They strongly advocated for community cooperatives as a governance model, viewing them as "ancestral technologies" for accountability, as legal regulation alone has often failed the Black community in Detroit. Cooperatives were seen as ensuring democratic decision-making (members, not capital investors, hold power), providing accountability through pre-agreed rules, and capable of furthering missions of social justice and solidarity. Our findings suggest for the HCI community that when designing governance structures, researchers should prioritize solutions that align with existing community accountability structures.

2.3 Proactively Governing AI within Cooperatives

Building on the insights from Section 2.2, my final dissertation project investigates how cooperatives can expand their capacities for anticipating and mitigating the social impacts of new technologies, such as AI. The Mondragón Corporation is one of the worlds largest worker-owned cooperatives, containing over 70,000 laborers and 35,000 member-owners across a network of 92 different cooperative businesses (Mondragón Corporation 2023). Each cooperative in Mondragón is committed to economic sustainability, providing good employment, and benefiting society. While some cooperatives want to implement more advanced AI systems into business practices, other cooperatives are hesitant because of concerns that AI may negatively impact worker welfare, cooperative culture, and the corporation's overall social mission. Working with cooperatives in Mondragón's network, this project aims to co-design proactive governance processes that can anticipate and respond to the possible impacts of AI. For instance, cooperatives may propose working with the university to update educational programs in anticipation of shifting labor demands. We aim to create a toolkit (e.g., resources and recommendations) to help cooperatives deploy and govern AI within their cooperative values.

Building on participatory design methods, I with my collaborator at Mondragón University have outlined a three staged research process:

1. **Interviews with Multi-Sector Co-ops**: Identifying 1-3 cooperative research partners to understand their AI strategy, tensions between AI and their social mission, and capacities for polycentric governance.
2. **Co-Design with Key Stakeholders**: Conducting participatory design workshops with stakeholders from partner cooperatives to co-design governance toolkits for AI implementation. This includes exploring AI's impact on governance structures and social missions and what capacities firms have to respond.
3. **Workshops with Co-op Network**: Conducting workshops with members from other cooperatives in the broader Mondragón network to validate and extend the toolkit, in order to build a larger community of practice.

3 Contributions and Significance

My research provides several significant contributions to the AIES community and the broader discourse on equitable technology development.

First, by centering the perspectives of marginalized communities, I expose the tensions between conventional technical imaginaries and the values of communities seeking to create radically alternative economies. My work materially demonstrates how developers can redistribute power within the innovation ecosystem to more proactively mitigate and respond to harms as they emerge in the innovation process. This requires interrogating questions of ownership, developing technology using democratic processes, and engaging in work that enables social justice. The emphasis on community cooperatives as a governance model embodies these forms of power redistribution, leveraging their historical significance in Black economic resiliency and their inherent structure around social missions.

Second, by analyzing AI governance through the lens of economic democracy, particularly within established cooperative structures like Mondragón, my research aims to provide concrete pathways for creating accountability mechanisms that challenge dominant capitalist models of technology development. It flips the traditional HCI focus from "designing tools for alternative economies" to "how can an alternative economy design our tools." My goal is to offer insights into how cooperative principles can inspire new approaches to algorithmic governance and AI ethics.

References

Bell, W. 2003. *Foundations of Futures Studies: Human Science for a New Era*, volume 1. New Brunswick, N.J: Transaction Publishers. ISBN 978-0-7658-0539-3 978-0-7658-0566-9.

Boyarskaya, M.; Olteanu, A.; and Crawford, K. 2020. Overcoming Failures of Imagination in AI Infused System Development and Deployment. arXiv:2011.13416.

Chordia, I.; Baltaxe-Admony, L. B.; Boone, A.; Sheehan, A.; Dombrowski, L.; Le Dantec, C. A.; Ringland, K. E.; and Smith, A. D. R. 2024. Social Justice in HCI: A Systematic Literature Review. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, 1–33. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-0330-0.

Do, K.; Pang, R. Y.; Jiang, J.; and Reinecke, K. 2023. "That's Important, but...": How Computer Science Researchers Anticipate Unintended Consequences of Their Research Innovations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16. Hamburg Germany: ACM. ISBN 978-1-4503-9421-5.

Ericson, C. A. 2011. *Concise Encyclopedia of System Safety: Definition of Terms and Concepts*. Hoboken, NJ: Wiley. ISBN 978-1-118-02866-7 978-1-118-02864-3 978-1-118-02865-0.

Escobar, A. 2018. *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press.

European Parliament. 2024. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).

Guston, D. H. 2013. "Daddy, Can I Have a Puddle Gator?": Creativity, Anticipation, and Responsible Innovation. In Owen, R.; Bessant, J.; and Heintz, M., eds., *Responsible Innovation*, 109–118. Chichester, UK: John Wiley & Sons, Ltd. ISBN 978-1-118-55142-4 978-1-119-96636-4.

Guston, D. H. 2014. Understanding 'Anticipatory Governance'. *Social Studies of Science*, 44(2): 218–242.

Katzman, J.; Dillahunt, T.; and Green, B. 2025. Anticipation Is More Than Prediction: Proactively Preventing Harms in AI Development. *Workshop on Sociotechnical AI Governance at CHI 25*.

Katzman, J. L.; Israni, A.; Okyne, S.; Winton, H.; and Dillahunt, T. R. 2025. Technological Tensions In Urban Food Sovereignty: Insights from Detroit's Farmers. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, 1–8. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-1395-8.

Liu, D.; Nanayakkara, P.; Sakha, S. A.; Abuhamad, G.; Blodgett, S. L.; Diakopoulos, N.; Hullman, J. R.; and Eliassi-Rad, T. 2022. Examining Responsibility and Deliberation in AI Impact Statements and Ethics Reviews. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '22, 424–435. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9247-1.

Metcalf, J.; and Moss, E. 2019. Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics.

Metcalf, J.; Smith, B.; and Moss, E. 2022. A New Proposed Law Could Actually Hold Big Tech Accountable for Its Algorithms. *Slate*, 5.

Mondragón Corporation. 2023. Humanity At Work. Annual Report.

National Institute of Standards and Technology. 2021. AI Risk Management Framework. *NIST*.

Selbst, A. D. 2021. An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology*, 35(1).

Sharma, V.; Kumar, N.; and Nardi, B. 2023. Post-Growth Human–Computer Interaction. *ACM Transactions on Computer-Human Interaction*, 31(1): 9:1–9:37.

Stilgoe, J.; Owen, R.; and Macnaghten, P. 2013. Developing a Framework for Responsible Innovation. *Research Policy*, 42(9): 1568–1580.