Carnegie Mellon University

Sijia Xiao
Postdoctoral Researcher
Human Computer Interaction Institute
Pittsburgh, USA

xiaosijia@cmu.edu
applexiao.com

Research Statement

I am a researcher in Human–Computer Interaction (HCI), Social Computing, and Responsible AI. Drawing on theories of social justice, sensemaking, and decision-making, my work develops frameworks and systems that make information infrastructures not only technically robust but also socially accountable. My research has two goals:

- **Complement technological harm governance's focus on prevention with collective repair.** Much existing work on technological harm emphasizes preventing future incidents through evaluation, correction, and deterrence, or punishing those responsible. Far less attention has been paid to what happens after harm occurs—how to repair its impacts, rebuild trust, and support those affected. My research shifts this focus toward **reparation**, which asks how individuals and organizations can make amends and restore relationships, and **deliberation**, which examines how people collectively interpret harm, negotiate responsibilities, and decide on paths toward redress. I explore these processes in contexts such as online harassment, misinformation, and AI bias and risk, generating insights for governance, information policy, and accountability in organizational systems.
- **Broaden Responsible AI's focus beyond harm to also consider benefits.** Responsible AI efforts often center on identifying and mitigating risks after systems are built, with limited attention to how benefits are defined or weighed against those risks. This post-hoc orientation contrasts with industry practice, where AI development typically begins by articulating business or user value and only later considers potential downsides. As a result, Responsible AI considerations often arrive too late to influence foundational design decisions. My research examines AI innovation practices and develops tools and frameworks that integrate Responsible AI and innovation goals, helping practitioners maximize social benefits while minimizing risks through more balanced and anticipatory approaches to AI development.

The broader goal of my agenda is to design sociotechnical systems that foster trust, support collective sensemaking, and embed societal values. **Methodologically, I employ qualitative interviews, participatory design, experiments, large-scale data analysis, and system building.** My work is informed by social and justice theories of decision-making, sensemaking, and restorative and transformative justice. My research has been published at **CHI, CSCW, TOCHI, and AIES**, and has been cited by **The Guardian, CNN, BBC, and The New York Times**. I have received a **CHI Honorable Mention** and funding from the **CMU–NIST Cooperative Research Center**, the **Berkeley AI Policy Hub**, and the **Center for Technology, Society and Policy**.

1 Lines of Research

1.1 Understanding Technology's Impacts through Social and Justice Theories

To address technological harm, we must first understand how people experience it**. My research examines how design choices—such as platform features, recommendation systems, and moderation tools—shape people's sense of justice, influencing what they view as a fair response to harm and what values become embedded in infrastructures.** These choices define whose voices are heard, whose suffering is recognized, and what forms of response are possible. Drawing on sensemaking theories, I showed through interviews with conspiracy believers and ex-believers how online communities reinforce spirals of ambiguity and how leaving such belief systems often depends on trusted relationships [3]. **This socio-technical perspective shifts conspiracy research from focusing on**

Carnegie Mellon University
Sijia Xiao
Postdoctoral Researcher
Human Computer Interaction Institute
Pittsburgh, USA
xiaosijia@cmu.edu
applexiao.com

false information to how people make sense of ambiguity. The work has been cited by **The Guardian, CNN, BBC, and The New York Times**.

Extending this lens, **I studied how restorative justice principles apply to online governance.** With adolescents who experienced harassment, I co-designed activities **that surfaced unmet needs—such as sensemaking, emotional support, and transformation—that current moderation practices overlook, and I identified both the opportunities and challenges of restorative justice as an alternative to punitive governance** [4,5]. I also showed how young people reconfigure Instagram through "finstas" to create feminist counter-publics that enable vulnerability, reciprocity, and collective sensemaking [2]. Together, these studies demonstrate how sociotechnical design shapes whether people can interpret harm, seek support, and pursue reparation, **underscoring the importance of justice-oriented, value-embedded technology development and governance.**

1.2 Designing Sociotechnical Systems for Deliberation and Repair

Recognizing harm is not enough; I build systems that scaffold collective deliberation and create pathways toward repair. I developed **SnuggleSense**, a system that helps online harm survivors build peer support by constructing action plans for addressing harm. The system combines interactive prompts with algorithm-recommended peer suggestions, which proved effective in restoring survivors' agency and confidence [6]. **SnuggleSense demonstrates that sensemaking itself can be empowering: by helping survivors connect their reflections to concrete actions and shared experiences, it transforms isolation into collective healing.**

I also led an interdisciplinary team to create **Vipera**, a system for professional AI auditors that identifies blind spots in evaluating generative AI models through visual analytics and LLM-driven guidance. Vipera introduces a new model of human–AI collaboration in auditing, using visual scene graphs to organize patterns across large image spaces while prompting auditors to explore overlooked risks. **Our evaluation showed that this approach improved both coverage and efficiency, providing a scalable and transparent approach to AI accountability.** In addition, I developed **ConsensUs**, which visualizes disagreement in group deliberation and shows how transparency can reshape how groups negotiate consensus [1]. By revealing where and why people disagree, ConsensUs helps groups move from implicit conflict to constructive deliberation, encouraging shared understanding rather than conformity. Together, these systems demonstrate how information infrastructures can be designed to embed justice-oriented values, **moving beyond documenting and understanding harm to actively structuring deliberation, enabling repair, and fostering accountability.**

1.3 Balancing Benefits and Risks in Responsible AI

I am expanding my work on social justice into the field of Responsible AI. I developed the **AI Harm Reparation Taxonomy**, the first systematic framework for classifying how organizations and communities respond to AI harms. Analyzing over 1,000 incidents, I found that most responses end at symbolic acknowledgment, and only a few extend to tangible remedies or systemic reforms [8]. **This work reframes accountability as the pursuit of remedy and reform rather than mere recognition, highlighting structural gaps in how AI harms are addressed and repaired.**

This insight led me to examine **why responsible practices rarely shape AI systems before harm occurs.** Through interviews with AI decision-makers [9], I found that early-stage innovation is driven by technical feasibility and business opportunity, with little structured reflection on trade-offs, alternatives, or long-term societal consequences. Many leaders pursue AI amid hype and fear of missing out, often overestimating benefits and overlooking risks. **Current Responsible AI efforts, such as auditing and impact assessments, typically occur after**

Carnegie Mellon University

Sijia Xiao
Postdoctoral Researcher
Human Computer Interaction Institute
Pittsburgh, USA

xiaosijia@cmu.edu
applexiao.com

systems are built, treating responsibility as an evaluation step rather than a design principle. My research shows why Responsible AI struggles to influence early design practices and highlights the need for frameworks that weigh risks and benefits together**, ensuring that questions of who gains, who bears the cost, and which values are prioritized become integral to decision-making rather than afterthoughts once key choices are set.**

2 Research Agenda

2.1 Risk–Benefit Frameworks for AI Innovation

My interviews with AI decision-makers revealed a misalignment between Responsible AI research and industry innovation practices. To address this gap, my future research will develop **risk–benefit frameworks and decision-support tools** that translate Responsible AI principles into early design contexts—helping organizations maximize benefits while anticipating and mitigating risks starting from the earliest phases of AI innovation, and throughout the AI life cycle. **This line of research is currently supported by funding from the National Institute of Standards and Technology–Carnegie Mellon University Cooperative Research Center (AIMSEC) and the Society of Actuaries.**

This agenda will advance through **synthesizing prior frameworks, co-designing with practitioners, prototyping systems that operationalize deliberation, and evaluating their effectiveness in shaping early AI decisions**. Beyond organizational practice, this work will also **inform policy and standards for AI risk management and information governance, offering models that regulators and industry bodies can adapt to guide responsible innovation.**

2.2 Organizational Pathways to Fairness and Repair

While organizations often acknowledge harm, their responses are typically symbolic and short-lived. **What is missing are durable processes that can be resourced, sustained, and institutionalized, such as complaint pipelines, survivor support, and follow-through on remedies.** My prior work has provided insights into the opportunities and challenges of embedding alternative justice and governance models into existing organizational practices. One effective approach is adapting existing practices toward alternative goals in justice, for example, transforming punitive moderation explanations into reflections on the impact of actions, or redirecting time spent debating proper punishment toward supporting survivors and fostering accountability and learning among perpetrators.

My future work will build these pathways with institutions that carry responsibility and processes for governance, including nonprofits, compliance offices, companies, and public agencies. I will combine **surveys, participatory collaborations, and system building** to design and evaluate tools that support communal healing for affected individuals and long-term accountability for those who cause harm. These pilots will generate evidence on cost, staffing, and long-term adoption, guiding partners and regulators in developing infrastructures and policies that scale repair across large platforms, information management systems, and organizational governance contexts. **By grounding reparation in existing workflows and budgets, this research will demonstrate how repair can evolve from one-off responses into sustainable institutional practice.**

2.3 Generative AI and the Future of Scientific Knowledge

Across my research, I have examined how people deliberate about technology's risks and benefits and how they repair its impacts. Generative AI raises these same questions in the domain of scientific knowledge production. **In fields such as computational social science, large language models are increasingly used to simulate behavior, generate synthetic data, and prototype theories. These tools can accelerate research and broaden access, but may also homogenize ideas, flatten complexity, and reshape how scientists experience agency and critique.**

Carnegie Mellon University

Sijia Xiao
Postdoctoral Researcher
Human Computer Interaction Institute
Pittsburgh, USA

xiaosijia@cmu.edu
applexiao.com

I will study how researchers deliberate about these risks and benefits, and how they envision repair when core scientific values are compromised. Using **surveys, experiments, and longitudinal studies**, I will analyze how generative AI changes the questions scientists ask, the perspectives they consider, and their sense of agency. **The findings will guide funders, oversight bodies, and institutions in evaluating AI-supported research and building sustainable governance and information policy practices that deliberate on and repair the impacts of generative tools on scientific knowledge production.** Overall, my future work integrates Human–Computer Interaction, Social Computing, and Responsible AI policy and practice to design sociotechnical systems and frameworks that not only advance technology but also foster reflection, deliberation, and repair around their impacts.

REFERENCES

[1] Weichen Liu, **Sijia Xiao**, Jacob T. Browne, Ming Yang, and Steven P. Dow. *ConsensUs: Supporting Multi-Criteria Group Decisions by Visualizing Points of Disagreement. ACM Transactions on Social Computing (TSC)*, 2018.
[2] **Sijia Xiao**, Danaë Metaxa, Joon Sung Park, Karrie Karahalios, and Niloufar Salehi. *Random, Messy, Funny, Raw: Finstas as Intimate Reconfigurations of Social Media. Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2020.
[3] **Sijia Xiao**, Coye Cheshire, and Amy Bruckman. *Sensemaking and the Chemtrail Conspiracy on the Internet: Insights from Believers and Ex-Believers. Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2021.
[4] **Sijia Xiao**, Coye Cheshire, and Niloufar Salehi. *Sensemaking, Support, Safety, Retribution, Transformation: A Restorative Justice Approach to Understanding Adolescents' Needs for Addressing Online Harm. Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2022.
[5] **Sijia Xiao**, Shagun Jhaver, and Niloufar Salehi. *Addressing Interpersonal Harm in Online Gaming Communities: The Opportunities and Challenges for a Restorative Justice Approach. ACM Transactions on Computer-Human Interaction (TOCHI)*, 2023.
[6] **Sijia Xiao**, Haodi Zou, Amy Mathews, Jingshu Rui, Coye Cheshire, and Niloufar Salehi. *SnuggleSense: Empowering Online Harm Survivors Through a Structured Sensemaking Process. Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2025.
[7] Yanwei Huang, Wesley Hanwen Deng, **Sijia Xiao**, Motahhare Eslami, Jason I. Hong, and Adam Perer. *Vipera: Towards Systematic Auditing of Generative Text-to-Image Models at Scale. Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI LBW)*, 2025.
[8] **Sijia Xiao**, Haodi Zou, Alice Qian Zhang, Deepak Kumar, Hong Shen, Jason Hong, and Motahhare Eslami. *What Comes After Harm? Mapping Reparative Actions in AI Through Justice Frameworks. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, 2025.
[9] Shixian Xie, **Sijia Xiao**, Cindy Peng, Ganesh Mani, Motahhare Eslami, and John Zimmerman. *Investigating How Leaders Decide on AI Innovations: Opportunities for HCI. Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2026, under review.
[10] Yanwei Huang, Wesley Hanwen Deng, **Sijia Xiao**, Motahhare Eslami, Jason I. Hong, and Adam Perer. *Vipera: Blending Visual and LLM-Driven Guidance for Systematic Auditing of Text-to-Image Generative AI. Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2026, under review.