Ph.D. Thesis Proposal

April 2, 2025
Draft

Designing Evaluations of AI-Driven Systems for the Future of Work: A Risk Measurement Framework

Sara C. Kingsley
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

skingsle@cs.cmu.edu

April 2, 2025

Thesis Committee:

Beibei Li, Carnegie Mellon University, Co-chair
Jeffrey P. Bigham, Carnegie Mellon University, Co-chair
Jason I. Hong, Carnegie Mellon University
Saiph Savage, Northeastern University

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Copyright © 2025 Sara C. Kingsley

Keywords: Sensemaking, Decision Making, Embedded AI, Information Systems, AI Risk, AI Risk Mitigation, AI Risk Management, Future of Work, AI Evaluation, AI Risk Assessment, AI Audit, Sociology of Risk, Labor Economics, Human-Computer Interaction.

Abstract

This dissertation draws on the sociology of risk to explore how AI systems shape labor markets, especially through opaque and contested risks. It builds on sociological theories that frame risk as socially constructed, often hidden, and unevenly distributed, while advancing this framework by proposing methods to computationally detect and measure these risks.

- In **Chapter 3**, the dissertation examines how workers perceive, audit, and contest AI-driven decisions. It shows how notions of risk are negotiated between users, developers, and institutions—making visible the contested expertise in defining harm.
- **Chapter 4** continues this line by identifying hidden information risks in job advertising and developing a taxonomy of problematic events, with metrics showing how these disproportionately affect low-income workers.
- **Chapter 5** proposes novel computational critiques of techniques like Reinforcement Learning from Human Feedback (RLHF), arguing that without attention to diverse user risk perceptions, such methods can amplify rather than mitigate harm.

From these chapters, this Ph.D. thesis offers a framework for quantifying the risks of excluding user perceptions and disaggregated information choices or preferences from AI design. Together, the dissertation offers a unique contribution: It integrates sociological theories of risk with computational tools and empirical studies to surface, categorize, and begin to measure AI-related risks in labor systems-making the invisible visible.

These research threads are illustrated through case studies on user experiences with AI-enabled systems in the context of job advertising and job search. These systems, which include search engines and social media platforms, are vital for participating in labor markets. Individuals depend on these systems to earn income for essential living expenses, while employers invest billions in digital advertising to recruit qualified candidates. Inefficiencies arising from the design of these embedded AI systems can substantially increase the costs of hiring and job searching, affecting the economic stability of both workers and businesses.

Given these stakes, it is crucial to develop robust scientific approaches to identify and measure AI-enabled information system risks. Such approaches are necessary to design effective tools that mitigate unintended consequences. In addition, to prevent perpetuation or amplification of these risks in workplace environments, it is essential to assess the effectiveness of risk mitigation strategies used with these systems. However, the literature lacks a comprehensive framework and methods for computationally assessing the risks of embedded AI information systems in workplaces.

This Ph.D. thesis proposal aims to address the need to operationalize risk measurement by presenting a framework and measures for evaluating the risks associated with AI-enabled information systems. Through this work, our[1] dissertation objective is to contribute valuable guidance and tools for the design and application of risk assessment and management of information systems in workplaces and labor markets.

[1]Although this Ph.D. thesis proposal is based on research projects that I have personally led, this document predominantly contains the pronoun "we" and "our" out of respect to all of my collaborators who have contributed equally or invaluably to the research.

Dedication

This dissertation is dedicated to Clara Wang, Proteeti Sinha, Alex Mikhalenko, Rena Li, Jaimie Lee, Robert Lambeth, Will Nichols, Jiayin Zhi and the hundreds of Carnegie Mellon University students who have served as kind, generous, wickedly curious, intelligent, and dedicated teachers during my PhD journey. Your example and love of learning has changed my idea about what is possible for education and the future. Thank you.

Acknowledgements

My education and accomplishments reflect the deep, generous and unconditional support, encouragement, and advice from my parents, Susan and Charles Mitsis (the Bean), Edward D. Kingsley (my Tree) and Carol Pekar; my big brother and advocate, Thomas C. Kingsley, my awesome and hero sister-in-law, Sabrina Kingsley, my niece and nephew, Evelyn and Noah; my godparents Tom and Trisha Donovan, Joe Kenyon, Stephen Ross; George, Elizabeth, Stephen, Brooke, Matthew, Molly Kingsley, Sally Betts Goss; Chris, Pam, Liam and Hannah Kingsley. My dear, sweet dog, Moe ('Momo'). Equally, my advisors and PhD committee Beibei Li, Jeffrey P. Bigham, Jason I. Hong, and Saiph Savage; and faculty, collaborators, and mentors, including Alice Zhang, Motahhare Eslami, Hong Shen, Casey Fiesler, Araba Sey, Shamika Klassen, M. Six Silberman, Laura Forlano, Rediet Abebe, Elena Glassman, Emma Tosch, Abeba Birhane, George Obaido, Kehinde Arubela, Caitie Lustig, Andrea Gallardo, Rotem Guttman, Raewyn Duvall, Ian G. Williams, Caitlin McArthur, Liz Green, Mahmoud El-Sayed Shabana, Lauren Mcllvenny, Will Nichols, Clara Wang, Proteeti Sinha, Jaimie Lee, Rena Li, Robert Lambeth, Alex Mikhalenko, Rumman Chowdhury, Theo Skeadas, Rashi Rungta, Winter Mason, Siddharth Suri, Mary Gray, OpenAI's Red Teaming Network (RTN), Meta AI Foundation Lang/Speech Safety team, CMU's Software Engineering Institute, the Mayo Clinic workers, Kevin Epperson, Burhan Syed, Jeff R., and the rest of my global Network Engineering Crew, and many more. A special thanks to Dr. Ramayya Krishnan for your mentorship and for re-igniting my passion for researching at the intersection of AI and Labor Economics.

My deep thanks and appreciation to the U.S. government workers who work tirelessly for us, and who taught me the value of public service (and also computer programming!).

I owe my village here and beyond a debt of gratitude. Thank you.