

Bridging the Responsibility Gap: From Causalities to Ethical Decision-Making in Human-AI Collaboration

The integration of Artificial Intelligence (AI) into professional environments creates both opportunities and challenges. While AI systems can enhance decision-making, their growing autonomy and human-like features raise critical ethical questions: Who is responsible when decisions made with AI lead to harm? This project focuses on the "Responsibility Gap"—a phenomenon where responsibility for actions taken in human-AI collaborations becomes unclear. It aims to explore how responsibility is attributed between managers and AI systems, how this affects behavior, and what can be done to mitigate unethical decisions.

Our research tackles the urgent need to understand the behavioral and ethical consequences of working with AI. Managers often face situations where they must balance organizational goals, such as profit maximization, with ethical considerations, like consumer privacy and safety. The involvement of AI in these decisions can lead to responsibility shifting, where managers feel less accountable, potentially increasing harmful or unethical actions. By investigating these dynamics, we aim to uncover the factors that influence responsibility attribution and propose interventions to reduce negative behaviors.

The project will use innovative research methods, including experimental studies with real interactions between professionals and AI agents powered by advanced language models. These experiments will provide unprecedented insights into how managers collaborate with AI and how responsibility is shared in such partnerships. Our research will also test practical solutions to ensure that AI is used ethically in professional settings, safeguarding public trust and societal well-being.

The findings of this project will contribute to the responsible use of AI in business. By identifying the mechanisms that drive harmful behaviors and designing interventions to mitigate them, we aim to support managers in making ethical decisions while leveraging AI's potential. This work will also guide policymakers and organizations in developing frameworks for accountability and transparency, ensuring that AI technologies benefit society as a whole.