

Projects

fair assignment: the team-size rules (2 or, preferably, 3 members), while ensuring every student is assigned to either their first or second choice.

Project 1: Intersectional Harms in Workplace Incidents

Guilherme Turina De Melo

Birk Halseth

Marc Font Ruana

Project 2: Identifying AI Risks for Non-Human Life in Urban Spaces

Baris Tan Unal

Kaan Riza Akarcay

Mustafa Burak Erkocak

Project 3: From Chats to Work: Which Roles ChatGPT Plays at Work

Xiangxi Li

Gabriele Agosta

Sarra Lajnef

Project 5: Mapping AGI Risks in the Workplace

Jonas Behnisch

Alberto Mutti

Project #1

Guilherme Turina De Melo, Birk Halseth Marc Font Ruana

Intersectional Harms in Workplace Incidents

Harms from AI used in workplaces can be worse for people with intersecting identities, such as young women and lower-income people of color.

Prior work has not closely studied how intersecting identities within one person shape the causes and effects of workplace AI incidents.

We address this gap by:

- collecting workplace AI incident reports;
- coding how intersecting identity factors relate to each incident's context, mechanism, and harm, following the procedure in [1]; and
- releasing a dataset and analysis code so others can reproduce our results.
- References

[1] *Why AI Harms Can't Be Fixed One Identity at a Time: What 5,300 Incident Reports Reveal About Intersectionality.*

Project #2

Baris Tan Unal, Kaan Riza Akarcay, Mustafa Burak Erkocak

Identifying AI Risks for Non-Human Life in Urban Spaces

AI systems used in streets, parks, and other urban spaces can change how non-human life moves, feeds, reproduces, and survives in cities. Yet existing work does not provide a clear, systematic map of AI-related harms across the full range of non-human life in urban environments.

We contribute the following:

- Compile a structured list of concrete AI uses in urban spaces.
- Develop a categorization of non-human life in urban spaces.
- Design a method that uses an LLM to elicit risks by asking it to reason from the perspective of a selected non-human being (for example, a cat) for a given AI use.
- Release a dataset and analysis code so others can reproduce our results.
- References

[1] Why AI Harms Can't Be Fixed One Identity at a Time: What 5,300 Incident Reports Reveal About Intersectionality

Project #3

Xiangxi Li, Gabriele Agosta, Sarra Lajnef

From Chats to Work: Which Roles ChatGPT Plays at Work

To study how artificial intelligence may reshape work, Anthropic mapped Claude queries to occupational tasks to estimate which tasks are most exposed to automation by LLMs. It is not clear whether the same patterns hold for ChatGPT.

We test this by mapping millions of publicly available ChatGPT interaction logs [1] to tasks and skills in the Occupational Information Network (O*NET), including detailed work activities (DWAs). We:

- map ChatGPT interactions to O*NET tasks and DWAs;
- map the same interactions to the roles a chatbot can take (for example, friend, colleague, or trainer), using the role set in [2];
- evaluate both mappings with a human-in-the-loop protocol, with a focus on separating plausible automation targets from spurious matches; and
- release a dataset and analysis code so others can reproduce our results.

[1] *WildChat Dataset*

[2] *Frictionless Love: Associations Between AI Companion Roles and Behavioral Addiction*

Project #5

Jonas Behnisch, Alberto Mutti

Mapping AGI Risks in the Workplace

Mapping the risks of artificial general intelligence (AGI) in the workplace matters because AGI systems could change how work is assigned, checked, and controlled.

Prior work proposes a method for mapping risks from workplace AI agents, but it does not cover AGI systems [1].

We will map workplace risks of AGI by:

- Defining workplace AGI. Review the AGI literature (well beyond [3,4,5,6]) and define the main components of workplace AGI, using a multi-layer framework like Figure 1A in [1]. This includes cases where an AGI hires humans to do tasks on its behalf [2].
- Generating and testing scenarios. Build and use a large language model (LLM) pipeline to generate and validate workplace-grounded AGI risk scenarios, similar to Figure 1B in [1].
- Releasing a dataset and analysis code so others can reproduce our results.

[1] Unaccountable Delegation, Fading Skills: Mapping the Risks of Workplace AI Agents

[2] <https://www.wired.com/story/ai-agent-rentahuman-bots-hire-humans/>

[3] <https://arxiv.org/pdf/2309.10371>

[4] <https://sciendo.com/2/v2/download/article/10.2478/jagi-2014-0001.pdf>

[5] <https://arxiv.org/abs/2501.03151>

[6] <https://ieeexplore.ieee.org/abstract/document/11096544>