Feasibility Study: AI Intersectionality Project

Introduction

Today's AI risk assessments often overlook how systems can harm people with overlapping identities, like race, gender, and socioeconomic status. This blind spot puts organizations at risk of ethical failure, regulatory trouble, and public backlash.

Our solution fills this gap by combining simulated focus groups (via Plurals [3]) and structured discussions moderated by large language models (In Silico Sociology [1]). This helps surface hidden risks and supports the development of AI systems that are fairer and more inclusive.

We will validate our method by comparing it to tools like ExploreGen [2], using expert and AI-generated labels, and gathering structured feedback through Prolific studies.

With the EU AI Act [6] and growing pressure for ethical AI, there's a clear need for better tools. While challenges like model complexity remain, our approach is timely, innovative, and grounded in a strong interdisciplinary foundation.

Macro-Level Market Assessment

As AI becomes more widespread, concerns about its social, legal, and ethical effects, especially on marginalized groups, are growing. Traditional risk assessments aren't enough to handle these complex issues.

This growing awareness, combined with new regulations like the EU AI Act [6], is driving strong demand for better AI governance tools.

Intersectionality is now seen as key to fairness. Organizations want tools that can detect risks faced by people at the crossroads of multiple identities, like race, gender, and class. Our solution meets this need by simulating diverse perspectives and using LLMs to guide analysis.

Broader trends, such as the focus on algorithmic bias, explainability, and Ethics by Design, make inclusive tools like ours not just helpful, but essential.

Macro-Level Industry Assessment

The AI governance field is evolving quickly, and our project is well positioned within it. While there are more tools for AI risk assessment, few focus on intersectionality, leaving space for our socially aware and inclusive approach.

New players are entering the market, thanks to accessible technologies like LLMs and growing regulatory pressure. Still, our early start and unique method give us a competitive edge.

Large organizations and public institutions have strong buying power and expect effective, DEI-aligned solutions. This plays in our favor, as they're actively looking for tools like ours to meet ethical and legal obligations.

We depend on external LLMs and data, but we've designed our system to stay flexible, reducing supplier risk.

Alternatives like expert panels or traditional frameworks often miss the depth and scale we offer.

Team Assessment

We're a team of four master's students in Computer Engineering and Data Science, united by a shared drive to tackle the overlooked problem of intersectional harms in AI.

Though we're early in our careers, we've dived deep into intersectionality theory and explored tools like Plurals [3] and In Silico Sociology [1] to build a method that blends technical innovation with social insight.

We've already developed a working prototype and designed a clear validation plan, showing we can turn ideas into action.

As we grow our academic and research network, including collaborations through platforms like Prolific, we're laying the foundation for real-world impact in ethical AI.

Summary and Conclusions

As AI systems play a bigger role in shaping society, traditional risk assessments fall short, especially when it comes to intersectional harms at the crossroads of race, gender, age, and class. Our solution fills this gap by combining simulated focus groups, LLM moderation, and intersectionality theory.

What makes this project stand out is both the growing market demand, driven by regulations like the EU AI Act, and the originality of our method. Organizations are now expected to go beyond checklists and proactively detect hidden, systemic risks.

While we're still early in development, our approach offers a meaningful step toward more inclusive and responsible AI. Challenges like refining simulations or educating the market exist, but with solid planning and interdisciplinary skills, we're confident they can be managed.

References

[1] Kozlowski et al. In Silico Sociology: Forecasting COVID-19 Polarization with Large Language Models.
[2] Herdel et al. ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies.
[3] Ashkinaze et al. Plurals: A System for Guiding LLMs Via Simulated Social Ensembles.
[4] Rao et al. RiskRAG: A Data-Driven Solution for Improved AI Model Risk Reporting.
[5] Grand View Research. (2024). AI Governance Market Size, Share & Trends Report, 2030. Retrieved from https://www.grandviewresearch.com/industry-analysis/ai-governance-market-report

[6] European Parliament. (2024). Artificial Intelligence Act - Regulation (EU) 2024/1689. Retrieved from https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf