
Call for Real-Time Transparency in AI and Society

Abigail Yohannes

With
Women in AI Safety Hackathon

Abstract

As AI and technology continue to rapidly reshape society, influencing job markets, economy, governance frameworks, social dynamics and more, tracking AI's societal impact remains fragmented with data representations scattered across multiple organizations with inconsistent public awareness that is both accessible to technical and non-technical audiences. To address this gap, I developed a proposal for a AI Societal Impact Tracker using Streamlit, with the aim of evolving it into a comprehensive real time reference, highlighting past and current influence of AI across multiple domains. This project will explore the current state of AI in society, investigating its effect on job markets, ethics, and policymaking. The goal is to leverage open datasets from organizations such as O*NET, Anthropic, TechPolicy.press as well as advocate for open source data from as many institutions and businesses willing to share their findings in these research areas. While I was unable to fully implement real-time data retrieval or visualizations within this iteration, the framework is structured to support continuous updates and interactive analytics in the future. By integrating diverse data sources and enabling public engagement through opinion submissions, this tracker seeks to provide policymakers, researchers, and the general public with a centralized, evolving resource for AI's societal impact. Future iterations will aim to maintain real-time monitoring, data visualizations and advanced analytics, striving to document and track AI's impact on society in a way that can be consistently engaged with.

Keywords: Societal impact, social norms, AI governance, transformative AI, economic impact, bias and misuse

1. Introduction

a. Problem Statement

With the proliferation of AI and its tentative positive and negative outcomes in relation to its impact on societal structures, there is a lack of real-time trackable resources. For the reports, articles, research, and products we can engage with, AI's influence on labor market shifts, automation risks, ethical concerns, and policy responses are often siloed across different organizations, disciplines, and datasets. This lack of centralized, continuously updated tracking creates gaps in knowledge, limiting policymakers, researchers, and the public from making informed decisions about AI safety and regulation. Additionally, the rapid development of new AI models and software causes discrepancies in societal impact studies. As the public is provided new updates on the developments of AI models or new products it seems the information is quickly rendered to be out of date for the current pace of model improvements. For example, even as some of the best experts in AI development and governance come together to design the International AI Safety Report (2025) the report states "Between the end of the writing period for this report (5 December 2024) and the publication of this report in January 2025, an important development took place. The AI company OpenAI shared early test results from a new AI model, o3" (The International Scientific Report on the Safety of Advanced AI, 2025). Just within a few weeks of the reports release, significant developments in AI are occurring while the rate of holistic reports come a few times a year depending on the agency. To keep up with the consistent development of new AI and its corresponding impact on markets and society, real time data and the continuation of centralized reports is essential. This current lack of centralized, continuously updated tracking creates gaps in knowledge, limiting policymakers, researchers, and the public from making informed decisions about AI safety and regulation. This project will aim to develop a framework for an AI Societal Impact Tracker, a web-based tool that aggregates, visualizes, and updates AI impact data across multiple domains. By combining open datasets from sources like Anthropic's Economic Index, O*NET's labor market trends, and AI policy databases, this tracker will provide a centralized, evolving reference for monitoring AI's societal effects.

While this version of the project has yet to achieve real-time data integration or visualizations, the underlying framework establishes a foundation for future development. The long-term goal is to enhance AI safety by ensuring that the societal risks and benefits of AI are continuously monitored, accessible, and interpretable for stakeholders. This work contributes to AI safety by democratizing impact data, fostering transparency, and equipping policymakers and the public with the tools needed to navigate AI's rapid evolution responsibly.

b. Background and Motivation

With AI increasingly being integrated into economic systems, governance structures, and daily life, its societal impact remains poorly understood and inconsistently tracked. Existing research highlights both positive and negative contributions, such as increased productivity and innovation, and significant risks, including job displacement, algorithmic bias, and the concentration of AI power among a few corporations (Krishna 2024). As a lot of AI risk is focused on underlying factors within LLM's, alignment, and AGI (powerful models), it's important to acknowledge how the development of AI is used by those who own and develop it. What are their value systems? How will AI help solve issues to problems that exist if decision makers and business owners don't have values that align with the general public. Additionally, as the general public may not agree on how society is governed and economic policy is designed, how would initiating more AI influence into these systems help society overall. Determining that AI is good or bad is a difficult question to answer. Another difficult question to answer is if people are inherently good or bad. As people are the builders of AI and aim to harness the use of AI, it's essential to track how it's used, especially by those who have the ability to make societal altering decisions. Those would be our government officials, tech leaders, big business leaders, and those who currently wield the ability to use AI for their goals. I'm posing a question that I am honestly not sure can be solved right now, maybe even ever. Just because AI has the ability to raise GDP's and potentially lift underserved areas out of poverty will those who currently hold those resources as well as maintain a hierarchy of political and economic influence enable AI to be democratized in this way. Will those with the ability to help others in the past without AI now use AI as a tool to uplift or use it to continue bridging the gap. These type of questions are addressed in Anthropic's CEO in his paper, *Machines of Loving Grace* where he states "I think that most people are underestimating just how radical the upside of AI could be, just as I think most people are underestimating how bad the risks could be" (Amodei 2024). I'm not sure my proposal has one answer, but rather it aims to act as a source for keeping the developments made in AI and its impacts accessible. By advocating for the release and deployment of as

much insightful real time data as possible, I believe we can work to promote **AI safety in terms of democratization. By making AI's societal impact more transparent and accessible, this initiative contributes to AI safety by ensuring both decision-makers and our constituents have real-time, data-driven insights into AI's risks and benefits.**

c. **Threat Model and Safety Implications**

The increasing integration of AI into economic, political, and social systems introduces **systemic risks** such as labor market instability, regulatory lag, algorithmic bias, and misinformation. AI-driven automation threatens job displacement and economic inequality, while AI governance often lags behind technological advancements, leading to **reactive policies** that fail to mitigate risks effectively. Additionally, AI systems can embed **racial, gender, and socioeconomic biases** into critical decision-making processes, exacerbating social inequalities. Furthermore, the rise of AI-generated content (e.g., deepfakes, synthetic media) poses challenges to **disinformation, public trust, and election security**. To address these threats, this project seeks to establish an **AI Societal Impact Tracker**, a centralized platform that consolidates fragmented data on **economic trends, AI policy updates, algorithmic fairness, and misinformation incidents**. By integrating diverse datasets in the future such as the **Anthropic Economic Index, O*NET labor market trends, and AI governance frameworks**, this tracker provides **real-time insights** for policymakers, researchers, and the public. It will enhance AI safety by supporting **proactive policy interventions, workforce adaptation strategies, and increased transparency in AI decision-making**. Ultimately, this initiative will work to make sure that AI's societal risks are **continuously assessed and mitigated before they escalate**, fostering greater accountability and responsible AI development.

2. **Discussion and Conclusion**

Unfortunately, due to development challenges and limited time, the project was unable to provide the level of insights originally envisioned. However, it successfully laid the groundwork for organizing multiple data sources and finding organizations releasing data, which can be leveraged to further develop and enhance this initiative to advocate for real time data on AI development in the future. Although overall impact can not be achieved in real time, we can still track and observe outcomes in real time.

3. **References**

United Kingdom Department for Science, Innovation & Technology. *International Scientific Report on the Safety of Advanced AI*. Gov.uk, Jan. 2025,

[https://assets.publishing.service.gov.uk/media/679a0c48a77d250007d313ee/International AI Safety Report 2025 accessible f.pdf](https://assets.publishing.service.gov.uk/media/679a0c48a77d250007d313ee/International_AI_Safety_Report_2025_accessible_f.pdf).([Accessed 3/9/2025])

Krishna, Venni V. "AI and Contemporary Challenges: The Good, Bad, and the Seary." *Journal of Open Innovation: Technology, Market, and Complexity*, vol. **10**, no. **1**, 2024, article 100178, <https://doi.org/10.1016/j.joitmc.2023.100178>. Accessed [insert date of access].

Amodei, Dario. *Machines of Loving Grace: The Ethics and Impact of AI*. <https://darioamodei.com/machines-of-loving-grace#3-economic-development-and-poverty>. (Accessed 3/9/2025)