Introduction

As ethicists, our initial position was threefold.

Firstly, the rapid adoption of unregulated technological progress can drive transformative change in our markets but risks deepening inequality and can lead to unforeseeable catastrophic outcomes. Secondly, rapid AI development has major detrimental effects on our environment due to high energy consumption and creates water shortages through the reliance on increasingly ubiquitous data centers in developed countries. Lastly, we held the stance that accelerating AI weakens societal cohesion and humanistic capacities.

Through brainstorming and collaboration with Turing's Thinkers, representatives of policy think tanks, we established four initiatives: two short-term strategies that feed into a long-term direction. In the short term, we propose a company certification for research funding. This requires companies to demonstrate tangible commitment to displaced workers by providing training and opportunities to remain in the workforce. Second, an Academia-Industry Policy Initiative should focus on fostering constructive dialogue between key opinion leaders to ensure innovation aligns with societal needs. In the longer term, we propose a risk audit system requiring standardized evaluations to promote ethical and sustainable practices, offering financial incentives such as tax rebates for high scores. The idea is to have the short term policies help understand the accommodation required by stakeholders to build into longer term strategies that encompass everything related to AI adoption.

Short-Term Policy Recommendations

Every transformative change in human civilization is accompanied by humanity riding the technological wave while agreeing on certain shared principles—whether it pertains to aviation, climate change, or the invention of the seatbelt (Aeroplane Tech Editorial Team, 2024; Minhae Shim Roth, 2024). To address the immediate risks associated with artificial intelligence, we propose two short-term policies to be implemented globally.

The first is a *Certification for Research Funding* at a company or organizational level. This certification aims to ensure that both public and private research funds are allocated to organizations that are committed to ethical practices and those that portray a social responsibility to the society at large. This program will leverage governance frameworks, such as ethical review boards, and adopt risk management principles like those outlined in the NIST AI RMF and GDPR. To qualify, companies must demonstrate robust data governance practices and operational transparency. A key component of the certification is addressing workforce displacement resulting from widespread AI adoption.

Organizations will be required to invest in job retraining initiatives and adhere to equitable tax policies, showcasing a tangible commitment to supporting displaced workers

and reinforcing the social safety net. To ensure accountability and sustained compliance, periodic audits will assess adherence to these commitments over time.

Hand in hand, an *Academia-Industry Policy Initiative* would bridge the gap between advances in artificial intelligence, research, companies, and public policy. This initiative will create a collaborative organization that brings together subject matter experts and domain experts from academia, research institutions, and policymaking bodies to share insights, produce research, and develop evidence-based AI policies. Under a multi-stakeholder governance model, the organization will focus on addressing issues such as bias, transparency, and accountability, aligning its work with existing frameworks and allow for newer frameworks and guidelines to be framed.

In conclusion, these policies address immediate and short-term risks by emphasizing transparency, collaboration, and societal responsibility. The Company Certification safeguards the ethical allocation of research funds, while the Academia-Industry Policy Initiative fosters evidence-based decision-making. Together, they ensure that AI innovation aligns with public interest, balancing trade-offs and prioritizing stakeholder benefits.

Long-Term Policy Recommendations

In numerous private corporations that operate, expand, and govern their businesses in the United States and other economies, several auditing mechanisms are already in place. While it began with financial checks to protect public shareholders, the audit ecosystem now also includes environmental and sustainability standards. The essential benefit of ESG audits is to protect people—the stakeholders of the environment and society at large. Similarly, audits for AI integration, development, and propagation aims to protect users, clients, and all those affected by a corporation's interest in AI. We propose *Audit.AI* - AI auditing, expanding upon existing mechanisms, building further the best practices and functionalities of financial and ESG audits already in use, to ensure comprehensive coverage of a corporation's involvement in AI too.

The first necessity is to audit energy usage, specifically - increased carbon footprints resulting from workplace AI tools in search and chat, development processes, and computational upgrades. If AI tools are part of a company's regular offerings, their use by clients or users must also be measured. This is essential to assess a corporation's capacity to sustainably integrate, develop, and proliferate AI over the long term by incorporating cleaner energy sources into their operations. Since this builds on existing energy audits, it serves as an excellent starting point.

Additionally, audits must include "checks on the individual." These assess the economic impact—positive or negative—that a company's interest in AI has on all individuals associated with the company, including clients, employees, and contractors. Increased transparency is necessary for companies to engage with their stakeholders' interests. The

audit should measure how individuals have been economically impacted, including job security, performance, and benefits versus costs (such as data privacy loss or under-delivery).

A third aspect of the audit ecosystem should cover humane and inclusivity-outreach measurements. This includes auditing a company's efforts to upskill employees for an AI-enabled workplace and ensuring workers see capacity-building benefits from AI integration. This promotes responsible, safe, and privacy-protected use of AI while providing opportunities for creative engagements in the workplace or beyond. It is also essential to evaluate the preservation of an individuals' human capacities. This might be difficult due to the intangible nature of "humanity," and so metrics such as happiness indexes (World Happiness Report, 2024) and relational health audits (RHAs) (Holtzhausen, M., & Wielenga, C., 2023) can provide useful insights to inspire frameworks from. RHAs measure relationships between individuals and groups in organizations and can help determine the optimal scope of AI use to balance efficiency while maintaining human agency.

The long-term integration of *Audit.AI*, for corporations using AI, is to ensure compliance with certifications and governmental guidelines that would have matured as short-term measures. *Audit.AI* will be like a conclusive check for ensuring the benefit of both consumers of AI, and upstream stakeholders such as employees for and in corporations that adopt AI. The maturation of the *Academia-Industry Policy Initiative* over multiple iterations using feedback-loops will help establish standardized guidelines that will help streamline, evolve and fix standards for *Audit.AI*. At its core, it must continue prioritizing environmental sustainability, economic resilience, and human-centric considerations and as audits go, reflect the performance of a stakeholder via reports and dashboards that are transparent, interactive and generally conclusive. Better performing corporations can benefit from incentives like subsidized development costs - based on *Certifications of Research Funding* - and other accommodative measures that the *Policy Initiative* might find can be motivations for the larger adoption of Audit.AI by corporations as standard industry practice.

Collaboration and Compromise

We worked with the Policy Think Tank group to establish common objectives, find compromises, and propose balanced solutions addressing both the ethical and practical dimensions of responsible AI. Our shared goals included ensuring sustainable economic development, protecting the public interest, and fostering a regulatory environment that encourages responsible innovation without stifling growth.

The first effort is the promotion of a clean energy transition. By redirecting capital toward wind, solar, hydro, and small modular nuclear reactors, we can create diverse job opportunities, drawing lessons from historical infrastructure initiatives such as the New

Deal. These efforts expand employment possibilities for displaced workers and prepare the economy for AI-driven changes. In tandem, we recommended strengthening social safety nets—providing extensive job training, enhancing child tax credits and childcare, and even exploring *Universal Basic Income*.

Although these initiatives demand careful fiscal planning, they affirm our ethical duty to support individuals navigating rapid technological shifts. To address the complexities of AI-driven risks, we aim to develop a flexible governance framework. Instead of rigid regulation, this adaptable approach evolves alongside new technologies. Additionally, we proposed a certification process to ensure that companies benefiting from public research investments abide by ethical standards, support workforce transitions, and contribute their fair share of tax revenue. This balanced measure aligns the ethicists' emphasis on moral responsibility and social welfare with the Policy Thinkers' focus on practical incentives and sustainable economic growth. Finally, to maintain agility and insight, we urged the formation of an interdisciplinary policy forum connecting academia, industry, and policymakers. This collaborative network would provide real-time guidance as AI and related technologies evolve. By weaving together economic, ethical, and social considerations, our recommendations foster a resilient environment in which innovation thrives, ethical principles guide progress, and all stakeholders—from workers to corporate leaders—share in the benefits of an AI-driven future.

Conclusion

In the current paradigm, there is a need to contemplate and chart a course of action regarding the development of AI technologies. It is imperative to facilitate a healthy deliberation between the two schools of thought: the advocates of safety-first responsible AI and the effective accelerationists. With this in mind, we collaborated with Turing Thinkers and proposed bridging ethical imperatives with practical solutions through collaborative discussion. We advocate for initiatives that mitigate risks, promote sustainable practices, and ensure equitable benefits from AI technologies. In the short term, certification processes and academia-industry collaborations will guide responsible research and development, safeguarding public interests while fostering innovation. In the long term, risk audits and flexible governance frameworks will ensure that corporations act ethically and sustainably while balancing the human, environmental, and economic impacts of AI. These frameworks extend existing auditing practices, addressing environmental sustainability, economic resilience, and human-centric concerns such as job security and agency. Our recommendations provide a concrete pathway to balance the ethical and practical dimensions of AI development. They prioritize measurable actions, such as transitioning to clean energy sources, strengthening workforce support systems, and establishing interdisciplinary policy forums. In conclusion, responsible AI development requires deliberate action to align innovation with public interest. Our proposals not only address current challenges but also lay the groundwork for a resilient and inclusive future. By combining ethical oversight with practical incentives, we can

ensure that AI's transformative potential is harnessed responsibly, fostering sustainable progress and equitable benefits for all.

References

Aeroplane Tech Editorial Team. (2024, April 17). *The Evolution of Aviation Safety: A Historical Perspective - Aeroplane Tech*. The Insurance Universe. https://aeroplanetech.com/history-of-aviation-safety/

Holtzhausen, M., & Wielenga, C. (2023). A relational approach to human development in a crisis: perspectives from two case studies in South Africa during COVID-19. *Development Studies Research*, *10*(1). https://doi.org/10.1080/21665095.2023.2213844

Minhae Shim Roth. (2024, October 3). *When Were Seat Belts Invented?* HISTORY. https://www.history.com/news/who-invented-seat-belts

World Happiness Report. (2024). *World Happiness Report*. Worldhappiness.report. https://worldhappiness.report/

Interactions with LLMs:

Link to Doc