

## Topic 1 - Accountability and Responsibility

This essay examines the responsibilities of both governments and corporations in regulating intelligent technology, arguing that both entities must collaborate to establish ethical, transparent, and accountable AI systems. While governments create regulatory frameworks, corporations developing AI technologies also bear significant responsibility for ensuring ethical practices.

Governments must strike a delicate balance between promoting innovation and mitigating risks, ensuring that regulations are adaptable to the diverse applications of AI. Thoughtful regulation can promote reliable, robust, and trustworthy AI applications. However, governments should not take a one-size-fits-all approach as there is no single definition or use case for AI.<sup>4</sup>

Governments play a vital role in determining and regulating the role of AI in law, commerce, and society. The UK government, for example, has taken a significant step forward with its pro-innovation approach to AI regulation. This comprehensive method for governing the development and use of artificial intelligence aims to provide clarity and coherence to the currently fragmented AI regulatory landscape. The proposed framework establishes five guiding principles: safety, security, and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress. Moreover, the framework recognises the importance of supporting tools, such as assurance techniques, voluntary guidance, and technical standards, in promoting responsible AI innovation. By adopting this pragmatic and proportionate approach, the UK seeks to strengthen its position as a global leader in AI, drive economic growth and prosperity, and increase public trust in the technology.<sup>12</sup>

As AI continues to reshape the workforce, governments must also grapple with the complex socio-economic implications of automation. The increasing adoption of AI technologies has the potential to displace workers across a wide range of industries, exacerbating existing inequalities and creating new challenges for social cohesion and economic stability.

To mitigate these risks, governments must invest in reskilling and upskilling initiatives that prepare workers for the jobs of the future, as well as exploring innovative solutions such as universal basic income (UBI) to provide a safety net for those displaced by automation. However, if a UBI is implemented in only one or a few nations, it may raise justice concerns, as some workers who contribute to the effectiveness of the UBI through their labor may not receive its benefits. A UBI implemented in only a few nations could raise concerns about fairness, as some workers may contribute to its effectiveness without receiving its benefits, potentially exacerbating existing inequalities. To address these socio-economic challenges, governments need to prioritise policies that support workforce transitions and economic equity, ensuring that the benefits of AI are widely shared and do not disproportionately disadvantage any particular group.<sup>10</sup>

A further demonstration of how governments have proactively taken on their responsibility in regulating AI in law, commerce, and society is the European Commission's Artificial Intelligence Act. This comprehensive regulatory framework aims to ensure trustworthy AI by

imposing harmonised rules for high-risk AI systems, banning those with unacceptable risks, and requiring compliance measures such as risk management and human oversight. Additionally, it sets transparency obligations for other AI systems and restricts the 'real-time' use of biometric identification in public spaces for law enforcement. The act also promotes innovation through regulatory sandboxes and support for Small Medium Enterprises (SMEs), with penalties for non-compliance reaching up to 6% of global turnover. The European Commission's AI Act demonstrates the government's responsibility in creating a safe, human-centric AI environment that balances the need for innovation with the imperative to protect public safety and trust. By providing clear guidelines and robust oversight, the European Commission aims to balance the need for innovation with the imperative to protect public safety and trust, ensuring that AI systems are developed and deployed responsibly. This approach to AI regulation sets a strong foundation for the responsible development and deployment of AI technologies across the European Union.<sup>2</sup>

Building upon the European Commission's efforts, the OECD Principles on Artificial Intelligence, adopted in May 2019, represent a significant milestone in this regard, providing a set of international standards for the responsible development and deployment of AI systems. The OECD is dedicated to ensuring that AI benefits society as a whole. Their efforts include measuring and analysing the economic and social impacts of AI, engaging with stakeholders to identify good practices for public policy, and promoting trustworthy AI through the OECD AI Policy Observatory. The organisation also provides a framework for classifying AI systems to help policymakers assess opportunities and risks, and an AI Incidents Monitor to document and understand AI-related hazards. International cooperation, through bodies like the OECD, is crucial for setting global standards and sharing best practices.<sup>7</sup>

In addition to the OECD, multi-stakeholder initiatives play a vital role in fostering collaboration and knowledge-sharing among governments and other stakeholders. For example, the Partnership on AI's AI Incident Database project aims to create a global repository of AI incidents and failures, enabling developers, researchers, and policymakers to learn from past mistakes and develop more robust and reliable AI systems. By bringing together diverse perspectives and expertise, these initiatives contribute to the development of comprehensive and effective AI governance frameworks.<sup>8</sup>

Although governments establish the framework for responsible AI governance, corporations developing AI technologies bear the ultimate responsibility for ensuring ethical practices in the development and deployment of AI systems. As the architects of intelligent systems, AI corporations hold a great deal of power and responsibility in shaping the future of AI. Embedding ethical considerations into the design process from the outset is crucial, as the choices made by AI corporations can have far-reaching consequences for individuals and society as a whole. The importance of corporate responsibility in AI development is further underscored by the Brookings Institution, which highlights that technology companies must prioritise ethical values, transparency, and accountability to ensure that AI systems align with human values and promote fairness.<sup>1</sup>

Recognising the need for corporations to adhere to ethical guidelines, governments are taking steps to enforce responsible AI practices through legislation. For example, the

European Union's proposed Artificial Intelligence Act (AIA) seeks to establish a comprehensive legal framework for AI, with a risk-based approach to regulation. The AIA would require high-risk AI systems to undergo conformity assessments, including independent audits, before being placed on the market. Additionally, the establishment of national supervisory authorities and a European Artificial Intelligence Board would provide ongoing oversight and enforcement of AI regulations.<sup>2</sup>

Fostering public trust in AI systems requires transparency in how these systems make decisions. However, the increasing complexity of AI often results in 'black box' models, making it difficult for users to understand the underlying processes. To address this challenge, AI practitioners should prioritise explainability and consider users' needs from the start of the development process, rather than relying on post-hoc interpretations that may be ambiguous and fail to increase trust. Companies must be transparent about how data supports the development of AI solutions to build public confidence.<sup>11</sup>

To effectively mitigate the impact of AI, corporations need to engage in partnerships with various stakeholders to ensure a balanced and inclusive approach. For example, companies like KPMG have established frameworks to ensure AI is developed and deployed responsibly, emphasising transparency, fairness, and accountability.<sup>5</sup> McKinsey emphasises the need for corporations to invest in training programs aligned with future job demands, especially in high-tech and STEM fields, to help workers adapt to the changing job market shaped by AI adoption.

Corporations must also address the potential impact of AI on the workforce. As AI adoption reshapes the job market, corporations need to invest in training programs aligned with future job demands, especially in high-tech and STEM fields, to help workers adapt to the changing landscape. By proactively upskilling and reskilling employees, companies can mitigate the disruptive effects of AI on employment while fostering a more resilient and adaptable workforce.<sup>6</sup>

Beyond regulation, corporations have a vital role to play in building public trust and understanding in AI. Initiatives like "Partnership on AI," a coalition of major tech companies, aim to promote the responsible use of AI and foster public engagement through open dialogue and educational efforts. These initiatives highlight the role of corporations in promoting digital literacy and engaging the public in meaningful conversations about AI's impact on society.<sup>9</sup>

In contrast to the previous examples that highlight the benefits of responsible AI development, the use of AI in military contexts, specifically in the development of lethal autonomous weapons systems (LAWS), raises grave concerns. LAWS are designed to autonomously select and engage targets without direct human intervention, leading to critical questions about accountability, transparency, and the potential for unintended consequences. Opponents of LAWS argue that the absence of human judgment in the decision-making process could result in unintended casualties and potentially escalate conflicts. To address these risks, it is imperative that developers and corporations prioritise transparency, conduct rigorous testing, and ensure human oversight in the development of autonomous systems. Moreover, regular public reporting on AI deployment practices,

including discussions of ethical implications and decision-making processes, can foster trust and accountability. This approach aligns with the growing calls for international regulations to govern the use of AI in military applications, with the aim of preventing misuse and safeguarding against potential harms.<sup>13 3</sup>

While the previous section highlights the potential risks associated with AI in military contexts, it is equally important to recognise the proactive steps taken by some corporations to address ethical concerns in AI development and deployment. A notable example of this is IBM, which has demonstrated leadership in the industry by proactively integrating ethical considerations into its AI development and deployment processes. Under the guidance of Francesca Rossi, IBM established an internal AI ethics board to oversee the company's global technology ethics efforts and ensure that ethical principles are embedded at every stage of AI development. This board is responsible for overseeing IBM's global technology ethics efforts. IBM also trains all its employees in 'ethics by design' methodology, ensuring that ethical principles are integrated into every stage of AI development. To operationalise these ethical commitments, IBM has developed several open-source toolkits, such as AI Explainability 360, AI Fairness 360, and the Adversarial Robustness Toolbox. These tools help make AI models more transparent, fair, and robust, addressing issues like bias and explainability. By making these tools freely available, IBM aims to foster a culture of ethical AI development across the industry. IBM's efforts highlight the importance of internal governance, employee empowerment, and the development of technical tools to ensure that AI systems are developed and used responsibly. This approach not only helps build public trust but also sets a benchmark for other corporations in the industry.<sup>14</sup>

Navigating the landscape of responsible AI governance is complex and requires a multifaceted approach. Governments play a crucial role in establishing robust regulatory frameworks that balance innovation with risk mitigation, as exemplified by the UK's pro-innovation AI regulation approach and the European Commission's Artificial Intelligence Act. These initiatives aim to provide clarity, promote public trust, and ensure that AI systems are safe, transparent, and fair. Moreover, international standards, such as the OECD Principles on Artificial Intelligence, emphasise the importance of measuring economic and social impacts and promoting trustworthy AI through collaborative efforts.

As the primary developers of AI, corporations hold significant power and responsibility in shaping the technology's future. IBM's establishment of an internal AI ethics board and the development of open-sourcing toolkits demonstrate the importance of internal governance and ethical training in fostering a culture of ethical AI development across the industry. Engagement in partnerships with various stakeholders, like the Partnership on AI, is crucial for promoting public engagement and digital literacy. By embracing transparency, accountability, and fairness, and involving multiple stakeholders, we can create a resilient and responsive governance framework that ensures AI benefits everyone and unlocks its potential for global positive change.

## Bibliography

1. **Brookings.** How to address AI ethical dilemmas. Available at: <https://www.brookings.edu/articles/how-to-address-ai-ethical-dilemmas/> (Accessed: 19 June 2024).
2. **European Commission. (2021)** Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM(2021) 206 final. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206> (Accessed: 12 May 2024).
3. **Floridi, L. et al. (2018)** 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations', *Minds and Machines*, 28(4), pp. 689-707. Available at: <https://doi.org/10.1007/s11023-018-9482-5> (Accessed: 27 May 2024).
4. **Google AI.** Building a responsible regulatory framework for AI. Available at: <https://ai.google/static/documents/building-a-responsible-regulatory-framework-for-ai.pdf#:~:text=URL%3A%20https%3A%2F%2Fai.google%2Fstatic%2Fdocuments%2Fbuilding> (Accessed: 19 June 2024).
5. **KPMG. (2024)** Trusted AI framework. Available at: <https://kpmg.com/xx/en/home/services/kpmg-trusted-ai/trusted-ai-framework.html> (Accessed: 5 May 2024).
6. **McKinsey. (2023)** Generative AI and the future of work in America. Available at: <https://www.mckinsey.com/mgi/our-research/generative-ai-and-the-future-of-work-in-america> (Accessed: 17 May 2024).
7. **OECD. (2019)** Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. Available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (Accessed: 12 May 2024).
8. **Partnership on AI. (2021)** AI Incidents Database. Available at: <https://partnershiponai.org/workstream/ai-incidents-database/> (Accessed: 8 May 2024).
9. **Springer.** AI powered narrative building. Available at: <https://link.springer.com/article/10.1007/s44163-022-00023-7> (Accessed: 19 June 2024).
10. **Sticker, M. (2013)** 'A merely national 'universal' basic income and global justice'. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1111/jopp.12289> (Accessed: 15 May 2024).

11. **Thomson Reuters. (2023)** 'The road to building trust with AI systems'. Available at: <https://www.thomsonreuters.com/en-us/posts/our-purpose/the-road-to-building-trust-with-ai-systems/> (Accessed: 17 May 2024).
12. **UK Government. (2023)** 'A pro-innovation approach to AI regulation: government response'. Available at: <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper> (Accessed: 5 May 2024).
13. **Umbrello, S., Torres, P. and De Bellis, A. F. (2020)** 'The future of war: could lethal autonomous weapons make conflict more ethical?' *AI & Society*, 35(1), pp. 273-282. Available at: <https://doi.org/10.1007/s00146-019-00879-x> (Accessed: 27 May 2024).
14. **Weforum.** Case study on IBM ethical use of artificial intelligence technology. Available at: <https://www.weforum.org/agenda/2021/09/case-study-on-ibm-ethical-use-of-artificial-intelligence-technology/> (Accessed: 19 June 2024).