Science-Policy Brief for the UN Multistakeholder Forum on Science, Technology and Innovation for the SDGs, 7 May 2025

Preemptive Governance for AI: Securing Health, Equality, Work, and Democracy for the SDGs

Nina M. Waals (n.m.waals@uva.nl), University of Amsterdam, Netherlands
Joyeeta Gupta (j.gupta@uva.nl), University of Amsterdam, Netherlands

Abstract

Artificial Intelligence (AI) is accelerating innovation—but at what cost? As its development intensifies, so do its unintended consequences, many of which jeopardize progress toward the Sustainable Development Goals. This brief focuses on four interlinked domains—health (SDG 3), gender equality (SDG 5), decent work (SDG 8), and global partnerships (SDG 17)—to illustrate how unregulated AI deployment is deepening structural inequalities, eroding democratic norms, and straining our planet's resources.

The environmental footprint of AI is rising sharply, with energy demands prompting major companies to quietly roll back net-zero commitments. This undermines climate action and, in turn, public health. In the labor market, AI is accelerating job polarization, hollowing out vital training grounds and forcing workers into either precarious gig work or elite technical roles. Gender disparities are similarly amplified, as biased training data reproduces and sometimes overcorrects harmful stereotypes, while algorithmic "solutions" fail to address deeper systemic injustices.

Most concerning is AI's role in the production and amplification of misinformation and disinformation. Deepfakes, hallucinated content, and weaponized narratives are now capable of undermining elections, destabilizing public health responses, and fragmenting social consensus. As information ecosystems become more automated, public trust in science and governance erodes. Meanwhile, creators and scientists face ethical dilemmas over contributing data to opaque AI systems.

Rather than reacting to harms after they occur, this brief argues for preemptive governance—the proactive application of democratic norms, sustainability principles, and global equity before AI becomes entrenched. Drawing on international frameworks and recent developments, it offers actionable recommendations to steer AI toward the public good and align its development with the SDGs.

Introduction

Artificial Intelligence (AI) has entered mainstream deployment, influencing everything from diagnostic systems to job recruitment to information flows. But as capabilities scale, so do the consequences of inaction. AI's rapid integration into daily life is outpacing the legal, ethical, and infrastructural systems meant to regulate it. This brief argues that we can no longer afford a reactive approach to governance. Instead, we must adopt preemptive governance: setting ethical and legal boundaries in advance of harm. This approach is essential if we are to meet the interconnected goals of health, equality, decent work, and global solidarity outlined in the 2030 Agenda.

AI's Environmental Health Costs (SDG 3)

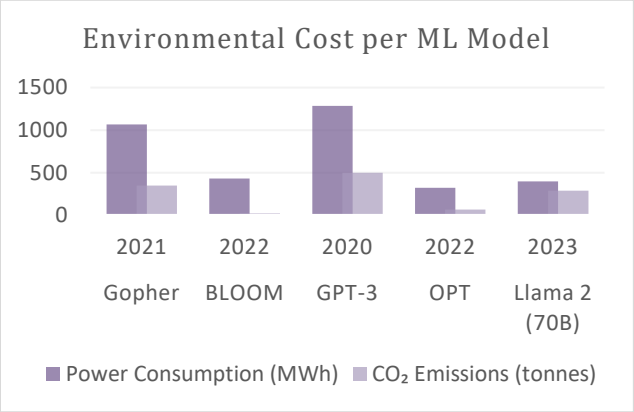

*Figure 1: This graph compares CO_2 emissions and energy use per machine learning model, highlighting the environmental impact of training large-scale AI systems. It includes estimates for leading models from 2020 to 2023.[**Error! Reference source not found.**,1,3]*

AI's energy requirements are rising exponentially, and much of that energy is sourced from fossil fuels[4]. Large-scale models require vast amounts of electricity to train and run, often in countries where grids still rely heavily on coal or natural gas. This contributes directly to global greenhouse gas emissions (see Figure 1). For example, the training of a single large language model

can emit over 250,000 kg of CO2—the equivalent of five round-trip flights between New York and London.[5] Consequently, tech giants like Microsoft and Google have seen dramatic increases in emissions—30% and nearly 50% respectively over recent years—largely due to AI-related expansion of data centers and hardware demand. These trends reflect a broader pattern: large multinationals are scaling AI without corresponding climate safeguards, often bypassing regulatory accountability in regions with lax environmental oversight.

These emissions aren't abstract. Climate-related health impacts—from respiratory illness caused by air pollution to increased vector-borne diseases—already affect millions. Rising temperatures and extreme weather events strain health systems, particularly in low-income regions.[5] The extraction of critical minerals for AI hardware further compounds environmental harms, generating toxic waste and fueling pollution in resource-rich but governance-poor regions. From an intergenerational lens, the consequences are even starker: the children of today will face escalating climate-induced health burdens unless AI systems are made more sustainable. The deployment of "green AI"—optimized for energy efficiency—and environmental auditing for AI projects should become standard policy to uphold SDG 3.

AI and the Erosion of Decent Work (SDG 8)

AI is disrupting labor markets at a scale unseen since the industrial revolution.[6] Yet unlike previous waves of automation, today's AI is targeting cognitive, white-collar work. Middle-tier jobs—legal assistants, data entry clerks, translators—are vanishing. These roles traditionally served as steppingstones to higher-skilled employment. Their disappearance risks trapping workers in either precarious, low-wage labor or requiring entry into elite, highly specialized fields with steep barriers to access.

As Bill Gates has noted, AI is creating a polarized labor market: "very rich people doing very well and everyone else scrambling."[7] Moreover, the removal of entry-level administrative roles removes crucial training grounds, making it harder for young workers to gain experience. Without intervention, AI will worsen structural unemployment, amplify class divides, and diminish pathways to social mobility.

To support SDG 8, policies should include protections for affected workers, such as transition assistance, skills retraining, and support for labor unions to negotiate fair AI integration. A fair future of work must be planned, not left to market forces.

Gender Bias in AI Systems (SDG 5)

AI systems trained on biased data often replicate and exacerbate existing gender inequalities.[8] Women are underrepresented in many datasets, and when included, are often mischaracterized. For instance, image recognition systems have misclassified female doctors as nurses and female engineers as "homemakers."[9]

Efforts to correct these biases can also backfire. A notable example was Google's image generator Gemini, which, when prompted with historical prompts like "U.S. senators in the 1800s," generated images of women of color—figures who were not historically in those roles (see Figure 2). This attempt to diversify visual outputs resulted in historically inaccurate portrayals, undermining credibility and fueling political backlash.[10] These superficial adjustments do little to address structural inequality and can undermine credibility in AI outputs.

Figure 2: The Verge (2024). "Google's Gemini AI continues to generate inaccurate historical images." Available from: https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical

Meanwhile, women are targets of AI-enabled harms, such as deep-fake pornography, that circulates without consent, often targeting journalists, activists, and politicians to intimidate or silence them.

The lack of gender-diverse teams in AI development exacerbates these issues. Regulations mandating gender bias audits, transparency in training datasets, and diverse design teams are necessary to make AI safe and equitable. SDG 5 cannot be achieved if AI continues to ignore or misrepresent half the population.

Misinformation, Disinformation, and Eroding Trust (SDG 17)

The proliferation of AI-generated content has dramatically lowered the barrier for producing convincing misinformation and disinformation. Tools like deepfakes and AI text generators can create fake videos, audio clips, and news articles that are difficult to distinguish from real, enabling malicious actors to spread false narratives at scale. This threatens SDG 17 (Partnerships for the Goals) by undermining the trust and democratic governance needed for international cooperation. In 2024, experts ranked AI-driven misinformation as a top global risk: the World Economic Forum listed "AI-generated misinformation and disinformation" as the second most likely source of global crisis in the near future[11]. Already, instances of AI-fabricated content sowing discord have been documented[12]. When citizens cannot trust the information ecosystem, it erodes public trust in institutions and between nations. Combating this will require collective action: platforms should label or curb AI-generated fake content, and governments may need to regulate AI use in political advertising and bolster digital literacy. Global partnerships are essential to establish norms and joint responses to the cross-border challenge of AI-fueled disinformation.

AI Training Data and Epistemic Justice

The rapid advancement of AI has been built on vast troves of data, much of it scraped from the internet without explicit permission from content creators. This raises ethical and equity concerns. Scientists, writers, and artists are seeing their publications, code, or artworks absorbed into AI training datasets without consent – a practice that challenges notions of intellectual property rights, consent, and epistemic justice (fair recognition of knowledge contributors). Many creators are uneasy that their work fuels commercial AI systems with no credit or compensation[14]. This problem is also highlighted by the current lawsuit by the New York Times against OpenAI. [13] Moreover, data extraction often reflects global power imbalances: AI firms in the Global North harvest content globally (sometimes termed "digital colonialism"), potentially exploiting knowledge from the Global South without due benefits sharing[15]. This dilemma undermines trust and fairness in the global research and creative ecosystem. Addressing it might involve new frameworks for data governance – such as requiring opt-in consent for using copyrighted works in AI training[14] or developing compensation mechanisms for creators. Ensuring inclusive and just AI development is crucial for the legitimacy and sustainability of AI innovations worldwide.

Policy Recommendations

To realign AI with sustainable development and preempt its harms, we recommend:

Adopt Preemptive AI Governance: Governments and the UN should establish proactive regulations and ethical guidelines for AI now (building on UNESCO's AI Ethics Recommendation) rather than reacting after crises. This includes risk-based oversight for high-impact AI applications and international coordination on AI standards.

Green AI Initiatives: Incentivize and mandate energy-efficient AI. AI companies must report and reduce the carbon footprint of training models. Promote the use of renewable energy in data centers and support research into low-power AI techniques to mitigate environmental impacts on health (SDG 3).

Workforce and Education Reforms: Prepare labor markets for AI. Invest in retraining programs to help workers transition into new roles alongside AI (supporting SDG 8) and update educational curricula to focus on skills complementing AI. Strengthen social safety nets for workers displaced by automation.

Diversity and Bias Audits: Embed gender and diversity considerations in AI design (supporting SDG 5). Companies and regulators should conduct regular AI bias audits for algorithms in hiring, lending, healthcare, etc., and take corrective action if disparities are found. Increase the participation of women and underrepresented groups in AI development via scholarships, hiring targets, and inclusive workplaces.

Combat AI-Driven Disinformation: Develop policies and tools to identify and limit AI-generated fake content. This could include watermarking requirements for AI outputs and stricter penalties for the malicious use of deepfakes. Nations should collaborate on best practices to safeguard elections and public discourse from AI-enhanced disinformation (supporting SDGs 16 and 17).

Fair Data and Knowledge Governance: Create mechanisms to protect creators' rights in AI development. Encourage transparency about AI training data sources and consider frameworks for content creators to opt out or receive compensation when their work is used. International intellectual property discussions (e.g. at WIPO) should address AI training data to ensure equitable outcomes and maintain global trust.

Conclusion

AI's promise can only be realized if its risks are anticipated and addressed. For too long, the tech sector has operated under the mantra of "move fast and break things." This culture of disruption prioritizes rapid deployment over long-term accountability, leaving governments and societies to clean up the fallout. A reactive approach to governance is no longer sufficient. The time has come for preemptive legislation—rooted in human rights, sustainability, and democratic accountability. By embedding these principles from the outset, we can ensure that AI strengthens, rather than undermines, our collective pursuit of the Sustainable Development Goals.

Acknowledgements

This project has received funding from the Netherlands' National Science Foundation Spinoza Price awarded to Prof. Joyeeta Gupta.

This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101020082).

References

1. **Luccioni AS, Jernite Y, Strubell E.** Power hungry processing: watts driving the cost of AI deployment? *arXiv*. 2023. Available from: **https://arxiv.org/abs/2311.16863**http://arxiv.org/abs/2311.16863
2. **Strubell E, Ganesh A, McCallum A.** Energy and Policy Considerations for Deep Learning in NLP. *Proc ACL*. 2019;3645–50. Available from: https://aclanthology.org/P19-1355
3. **AI Index Steering Committee**. *The AI Index Report 2024*. Stanford Institute for Human-Centered Artificial Intelligence; 2024. Available from: https://aiindex.stanford.edu/report/
4. **Luccioni A, Schmidt V, Lavoie A, et al**. Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. arXiv preprint arXiv:2211.02001. 2022.
5. **World Health Organization.** Climate change and health. WHO Fact Sheet; 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health
6. **International Labour Organization.** Generative AI and Jobs: A global analysis of potential effects on job quantity and quality. Geneva: ILO; 2023.
7. **Gates B**. The Road Ahead: AI and the Future of Work. Gates Notes; 2023. Available from: https://www.gatesnotes.com/
8. **UNESCO**. I'd Blush If I Could: Closing Gender Divides in Digital Skills Through Education. Paris: UNESCO; 2019.
9. **Dastin J.** Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. 2018 Oct 11. Available from: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/
10. **Robertson A.** Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis. The Verge. 2024 Feb 21. Available from: https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical
11. **World Economic Forum.** Global Risks Report 2024. Geneva: WEF; 2024.
12. **UNESCO**. Guidelines for the Governance of Digital Platforms. Paris: UNESCO; 2023.
13. **The New York Times Company v. Microsoft Corporation and OpenAI Inc.** Complaint. United States District Court for the Southern District of New York. Case No. 1:23-cv-11195. Filed 2023 Dec 27.
14. **Authors Guild.** *Generative AI & Copyright: The Authors Guild's Position*. 2023. Available from: https://authorsguild.org/advocacy/artificial-intelligence/faq/
15. **Birhane A, Prabhu V.** Large Image Datasets: A Pyrrhic Win for Computer Vision. *arXiv*. 2021. Available from: https://arxiv.org/abs/2006.16923