G2O BRASIL 2024

T2O

Brasil 2024
Let's rethink the world

Task Force 05
INCLUSIVE DIGITAL TRANSFORMATION

Empowering Societal Engagement in AI: Aligning Ethics, Sustainability, and Development within the Digital Economy

Albert Sabater, Observatory for Ethics in Artificial Intelligence of Catalonia, University of Girona, (Spain)

Petra Ahrweiler, Technology & Innovation Sociology / Social Simulation Laboratory, Johannes Gutenberg University Mainz (Germany)

TF05

Abstract

In advancing the Sustainable Development Goals (SDGs) of the 2030 Agenda, our proposal for Task Force 5 (TF5) adopts a comprehensive participatory methodology that is pivotal in the realm of AI ethics and sustainable development, harnessing digital innovations for societal betterment while ensuring inclusivity and adherence to ethical standards. Our approach is strategically aligned with the objectives of the Digital Economy Working Group of the G20 Sherpa Track, and emphasise the imperative of embedding citizen engagement at the core of AI system design and application in a broad spectrum of AI for good strategies, including Decent Work and Economic Growth (SDG 8), Industry, Innovation, and Infrastructure (SDG 9) and Sustainable Cities and Communities (SDG 11). Building on this framework, we propose two principles through which AI technologists will be able to foster inclusive and equitable advancements:

(i) taking direct action and engaging in participatory relationships with diverse stakeholders, and (ii) recognizing the responsibilities and limitations of AI through a participatory realist culture. Since our project empowers stakeholders, particularly citizens, to actively engage in contesting and influencing algorithmic decisions and overarching AI system designs, not only augments the technological robustness of AI systems but also ensures their alignment with diverse societal needs, cultural sensitivities, and economic realities. The latter is seen as relevant because it enhances societal relevance, acceptability, and long-term sustainability of AI systems. Within this context, a task force will evaluate its challenges and opportunities while advocating for ethical and participatory governance frameworks to reflect a commitment to the multifaceted cultural, social, and economic contexts of various communities.

Keywords: Sustainable Development Goals, AI Ethics, Participatory AI Culture.

Diagnosis of the Issue

The integration of Artificial Intelligence (AI) in advancing Sustainable Development Goals (SDGs) requires taking into consideration a complex interplay of challenges, opportunities, and governance issues, each influencing the trajectory and impact of AI technologies on society. Generally, the interplay can be seen as a quest for a "better" AI that is fundamentally tied to ethical considerations that transcend traditional metrics like efficiency and operationality. This quest is rooted in a broader societal vision that aligns technological advancement with the core values of sustainability, justice, and peace. The emphasis on "ethical criteria" is not superficial and reflects a growing recognition in the literature that AI technologies must be developed and deployed in ways that support equitable and just outcomes. This perspective is informed by ethical theories and principles, such as those proposed by John Rawls (1991), Amartya Sen (2008) and Martha Nussbaum (2011), which advocate for equality, human dignity, and the development of capabilities as essential to social justice. These theoretical underpinnings provide a valuable framework for understanding the current ethical dimensions of AI, which require a fundamental shift "from rational to relational-in thinking about personhood, data, justice, and everything in between, and places ethics as something that goes above and beyond technical solutions" (Birhane 2021, 1).

Although in the context of the 2030 Agenda for Sustainable Development, AI technologies offer the potential to advance equitable outcomes, concerns have been raised regarding the emergence of a technocracy driven by automation, where societal systems become rigidly controlled by algorithms and technological rules (Newman et al. 2022).

This scenario, which prioritizes organizational efficiency, legitimacy, and profit, may hinder the adaptability and responsiveness of systems to societal needs, potentially

deviating from the equitable and people-centered approach advocated by the 2030 Agenda. Thus, the integration of AI into the digital infrastructures of states and the provision of public services, highlights the urgency of establishing an equity-oriented framework for AI systems. This is particularly critical in the absence of comprehensive sociotechnical understanding. Despite growing recognition among AI developers of the societal implications of their technologies, as observed by Joyce et al. (2021), there remains a tendency to view challenges such as data bias primarily through a technical lens, overlooking their roots in existing social disparities. Addressing these issues is vital, especially in the deployment of AI for the advancement of SDGs, where the potential for exacerbating inequalities is significant. Within this context, it is important to highlight that the discourse on algorithmic technologies, as described by Vicsek (2020), often reflects a sense of inevitability and a belief in technological determinism, which can overshadow critical discussions on the social and ethical dimensions of AI deployment. In this sense, the 2030 Agenda serves as a crucial framework for guiding the development and application of AI technologies, ensuring they contribute positively to societal goals and do not perpetuate or exacerbate social injustices, thus supporting the broader vision of sustainable and inclusive development. Of course, the challenges in achieving this "better" AI are multifaceted and also technical and cover several issues.

For instance, AI systems are prone to inherit biases present in their training data or algorithms, leading to discriminatory outcomes that can exacerbate existing social inequalities. In addition, the often "black box" nature of AI algorithms complicates understanding how decisions are made, challenging accountability and eroding public trust. Thus, ensuring transparency in AI operations is crucial for ethical scrutiny and maintaining societal trust. This is particularly relevant if we consider that AI systems might struggle with complex ethical decisions that require nuanced human judgment,

especially in contexts where cultural and individual values play a significant role. But while previous extensive reviews (Jobin et al. 2019) point to the ethical problems inherent in AI technologies and the plethora of ethical guidelines that have emerged in recent years, these general frameworks often fall short in addressing the implications of AI applications across different societal contexts, including governance challenges, economic interests, power imbalances, and the impact on democratic discourse. Thus, we believe that opportunities for "better" AI lie in harnessing participatory methodologies that engage diverse stakeholders in the design and governance of AI systems (Ahrweiler et al. 2024).

Recommendations

Our approach emphasises the role of dialogue with policymakers and the integration of empirical research with vulnerable populations. Such participatory approaches ensure that AI technologies are developed in close consultation with those most affected by their deployment, fostering inclusivity and ensuring that AI systems are attuned to the varied needs, cultural sensitivities, and economic realities of different communities.

However, as indicated by the ongoing discussions and expected agreements on ethical standards, governance is a critical aspect of ensuring that AI technologies contribute positively to society. This requires the need for governance frameworks that are flexible, context-aware, and capable of accommodating the diverse ethical landscapes in which AI operates. Such frameworks should advocate for ethical and participatory governance, reflecting a commitment to the multifaceted cultural, social, and economic contexts of various communities.

By fostering an environment where AI is developed and governed in line with ethical principles and societal values, we can move closer to realizing the vision of "better" AI

that supports social justice and contributes to the broader goals of sustainable development. For this purpose, we propose the following 5 recommendations:

1. *To develop and continuously update ethical frameworks for AI* that are sensitive to the diverse contexts in which AI operates. These frameworks should be dynamic to accommodate evolving societal norms, technological advancements, and regulatory changes. They should also be informed by broad stakeholder engagement, including input from marginalized and vulnerable communities, to ensure that the frameworks are inclusive and representative of diverse societal needs and values.

2. *To implement participatory governance models* that actively involve a wide range of stakeholders in the decision-making processes related to AI development and deployment. This should include policymakers, technologists, civil society organizations, and, crucially, the end-users of AI technologies, especially those from underrepresented groups. Such engagement will help ensure that AI systems are developed and governed in a manner that is transparent, accountable, and aligned with the public interest, thereby enhancing trust and acceptance among all societal sectors.

3. *To encourage and support interdisciplinary collaboration* between technologists, social scientists, ethicists, and other relevant stakeholders to explore the socio-technical complexities of AI. This approach should aim to integrate qualitative and quantitative research methods to gain a holistic understanding of the impacts of AI on society. Interdisciplinary collaboration can foster the development of AI technologies that are not only technically advanced but also ethically sound, culturally sensitive, and socially beneficial, ultimately contributing to the achievement of sustainable development goals.

4. *To implement comprehensive ethical and impact assessments* throughout the lifecycle of AI systems, from design to deployment and beyond. These assessments should evaluate the potential social, economic, and cultural impacts of AI technologies, with a particular focus on identifying and mitigating risks related to bias, discrimination, and inequality. By making these assessments an integral part of the AI development process, organizations can proactively address ethical concerns and ensure that AI systems contribute positively to society and do not inadvertently harm vulnerable populations.

5. *To develop educational programs and public awareness campaigns* to enhance AI literacy among all segments of society, including policymakers, developers, and the general public. These initiatives should cover the ethical dimensions of AI, the potential benefits and risks of AI technologies, and the rights of individuals in the context of AI decision-making. By fostering a well-informed public discourse on AI, these efforts can empower individuals to engage in meaningful discussions about AI ethics, advocate for their rights, and participate actively in shaping the development and governance of AI technologies in a way that aligns with democratic values and social justice principles.

Scenario of Outcomes

Our strategy is in harmony with the structural ethos of SDGs, underlining the significance of engaging in meaningful dialogue with policymakers and melding empirical investigations with insights from communities at the margins. Such collaborative practices would ensure that AI innovations are co-created with those who stand to be most impacted by them, thereby promoting inclusiveness and tailoring AI solutions to meet the diverse cultural, social, and economic needs of various groups. This collaborative modality not only would bolster the technical integrity of AI systems but would also elevate their social pertinence and acceptance, paving the way for AI interventions that are both sustainable and equitable in addressing issues of social justice in the 2030 Agenda.

More specifically, for SDG 8 (Decent Work and Economic Growth), our strategy would be in line with ethical AI systems facilitate the emergence of new job sectors by driving innovation in fields such as green technology, sustainable agriculture, and digital services. These sectors prioritize inclusive hiring practices and workforce development programs, ensuring that the benefits of economic growth are widely shared and contribute to the reduction of income inequalities. In addition, with AI-driven educational platforms, it is possible to co-design with labor organizations and educational institutions, accessible upskilling and reskilling opportunities to enable workers to adapt to the evolving job market and to ensure that automation and digital transformation lead to job enhancement rather than displacement.

The use of participatory governance models in AI development would be paramount to ensure that labor rights are embedded in the design of AI systems used in workplace management and monitoring, promoting fair labor practices and safeguarding against exploitative uses of technology.

Regarding SDG 9 (Industry, Innovation, and Infrastructure), AI can accelerate the development of sustainable industrial processes that reduce environmental impact, enhance energy efficiency, and minimize waste. But ethical AI development needs to prioritise eco-friendly materials and processes, contributing to the creation of circular economies that support sustainable growth. Also, AI technologies play a crucial role in planning and developing smart, resilient infrastructure that serves the needs of all community members, including the most vulnerable, thus ethical frameworks that ensure that AI-driven projects are accessible, equitable, and aligned with community-led sustainability initiatives are key. Of course, ethical AI fosters a culture of open innovation and collaboration, breaking down barriers to technology transfer and knowledge sharing between countries. Although this may lead to a more equitable global innovation ecosystem where developing nations can harness AI for sustainable development and bridge the digital divide while promoting inclusive growth, it is imperative for governance models that are adaptable, contextually informed, and capable of embracing the wide array of ethical terrains AI traverses.

Finally, leveraging ethical AI in alignment with SDG 11 can transform urban landscapes into more sustainable, resilient, and inclusive communities. For instance, by incorporating AI into public service delivery, urban planning, disaster response, community engagement, and resource management, cities can optimize infrastructure and services to meet diverse needs efficiently. This holistic approach can facilitate that urban development is participatory, data-driven, and responsive to environmental and social

challenges, thus paving the way for livable cities that prioritize the well-being of all inhabitants and the planet.

By cultivating the proposed scenario of outcomes where AI is conceived and managed in adherence to ethical standards and societal values, we edge closer to the aspiration of a "better" AI that champions social justice and aligns with the expansive objectives of sustainable development. Incorporating a participatory approach to the development and deployment of ethical AI, adds a valuable grassroots dimension to technological initiatives. As previously stated, this perspective emphasizes the importance of deriving ethical guidelines and principles from the lived experiences, needs, and values of diverse urban and rural populations, especially those who are often marginalized or underrepresented in decision-making processes.

By engaging directly with community members, our participatory or "ethics from below" approach ensures that AI solutions are not only technically sound and efficient but also deeply rooted in the social fabric and cultural contexts of urban communities.

Thus, it promotes a participatory model of governance where residents have a significant say in how AI is implemented in their environments, from traffic management to waste reduction systems. This bottom-up ethics methodology fosters trust and collaboration between technology developers, city planners, and the public, leading to AI interventions that are more attuned to the nuanced dynamics of urban living and capable of addressing specific local challenges. Ultimately, integrating a participatory or an "ethics from below" approach enriches the ethical landscape of AI in SDGs, ensuring that technology serves as a tool for empowerment and inclusivity, enhancing the quality of life for all people while respecting their rights, dignity, and cultural diversity.

References

Ahrweiler, Petra, Nigel Gilbert, George Kampis, Zsolt Jurány, Albert Sabater, Martha Bicket, Blanca Luque, and David Wurster. "Using ABM and Serious Games to Create 'Better AI'." In Proceedings of the 2024 Annual Simulation Conference (ANNSIM'24), edited by Philippe J. Giabbanelli, Istvan David, Cristina Ruiz-Martín, Bentley Oakes, and Raul Cardenas. American University, Washington, DC, USA, May 20-23, 2024.

Birhane, Abeba. "Algorithmic injustice: a relational ethics approach." *Patterns* 2, no. 2 (2021).

Jobin, Anna, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 1, no. 9 (2019): 389-399.

Joyce, Kelly, Laurel Smith-Doerr, Sharla Alegria, Susan Bell, Taylor Cruz, Steve G. Hoffman, Safiya Umoja Noble, and Benjamin Shestakofsky. "Toward a sociology of artificial intelligence: A call for research on inequalities and structural change." *Socius* 7 (2021): 2378023121999581.

Newman, Joshua, Michael Mintrom, and Deirdre O'Neill. "Digital technologies, artificial intelligence, and bureaucratic transformation." *Futures* 136 (2022): 102886.

Nussbaum, Martha C. *Creating capabilities: The human development approach*. Harvard University Press, 2011.

Rawls, John. "Justice as fairness: Political not metaphysical." In *Equality and Liberty: Analyzing Rawls and Nozick*, pp. 145-173. London: Palgrave Macmillan UK, 1991.

Sen, Amartya. "The idea of justice." *Journal of human development* 9, no. 3 (2008): 331-342.

Vicsek, Lilla. "Artificial intelligence and the future of work–lessons from the sociology of expectations." *International Journal of Sociology and Social Policy* 41, no. 7/8 (2021): 842-861.

G2O
BRASIL 2024
BUILDING A JUST WORLD
AND A SUSTAINABLE PLANET

T2O Brasil 2024
Let's rethink the world

Let's
rethink
the world

CEBRI
Rethink Tank

FUNDAÇÃO
ALEXANDRE
DE GUSMÃO

ipea
Instituto de Pesquisa
Econômica Aplicada