Selected Papers of #AoIR2024:
The 25th Annual Conference of the
Association of Internet Researchers
Sheffield, UK / 30 Oct - 2 Nov 2024

RETHINKING AI FOR GOOD: CRITIQUE, REFRAMING AND ALTERNATIVES

Faranak Hardcastle
Australian National University

Sujatha Raman
Australian National University

Christer de Silva
Australian National University

Jenny L. Davis
Vanderbilt University

Ehsan Tavakoli-Nabavi
Australian National University

The pursuit of technological systems for global well-being has given rise to AI for social good (AI4SG) initiatives across academia, governments, private corporations, and civil society. However, the definition of “the good" within this context remains ambiguous, raising concerns that the term may become an empty signifier, lacking conceptual or practical meaning. This ambiguity poses a challenge, potentially allowing AI actors to invoke “the good” to justify their interventions without having to be accountable for the promises they make about the benefits of AI. In turn, such ubiquitous claims to be doing good undermine collective societal capacities to work towards ways of ensuring that future AI developments do indeed align with a normatively viable sense of public good. This paper is motivated by the premise that the extensive social science scholarship on AI and other emerging technologies contain several critical insights that can help inform a public good framework that can be the basis for dialogue with those who are the imagined beneficiaries of AI4SG or overly abstract principles in this regard. For example, social science work helps us unveil implicit as well as overlooked values, identify boundaries where caution or avoidance of AI applications may be as warranted

Suggested Citation (APA): Hardcastle, F., Raman, S., de Silva, C., Davis, JL., Tavakoli-Nabavi, E. (2024, October). Rethinking AI for Good: Critique, Reframing and Alternatives. Paper presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from http://spir.aoir.org.

as adoption in the pursuit of public good and recognise the conditions that need to be in place for this pursuit to be plausible (Davis et al, 2021; Davis, 2023).

We therefore propose to develop a socially situated public good framework, to scrutinize existing AI4SG initiatives, their embedded meanings of "the good" and their potential alignment or misalignment with values associated with two key areas of scholarship. The first is the broad field of technology and society studies; our way into this is via a public good test developed for quantum technologies that was informed by this literature (Roberson et al, 2021). Our aim is to refine, adapt and elaborate this 'test' into a public good framework that is more closely situated in and reflective of insights from the second area of scholarship, namely: critical AI studies (used as an umbrella term to encompass the fields of critical data/algorithm studies (Kitchin & Lauriault, 2014; Gillespie & Seaver, 2015; Dalton et al, 2016), design justice (Costanza-Chock et al, 2018), and algorithmic reparation (Davis et al, 2021).

In developing our public good framework, we first examine a body of literature that seeks to define "the good" in AI4SG in narrowly prescriptive ways *or* that sidesteps this question altogether by associating AI with broad expressions of global good, notably, the United Nations Sustainable Development Goals (SDGs) (Chui et al, 2018; Cowls et al, 2021).[1] For example, "the good" may be construed in some of this work through universal ethical principles drawn from bioethics (Taddeo & Floridi, 2018) to identify the potential for harm and maleficence (Floridi et al, 2020). In other cases, authors try to connect their definitions of "the good" to design values and data processes (Tomašev et al, 2020) or practical applications and eventual outcomes (Cowls et al, 2021).

However, much of this literature adopts a vision of AI as a neutral tool, rather than one that is shaped by specific framings, narratives, and interests as highlighted by critical social science work. We argue that even when principles and definitions are offered, they tend to be abstracted from the complex social realities and contexts that shape both the design and implementation of AI technologies and their ability (or inability) to create social good, especially in the global South. Promises of social progress are made but the question remains who defines progress and what technologies are needed to achieve it. We therefore delve into how dominant framings of AI4SG stabilize and normalize, probing whether these initiatives fulfil their promises, asking what Northern-led AI for Social Good means in a global context and to what effects does it operate particularly in the global South? In sum, we highlight several limitations in how AI4SG is operationalised, impacting their practical applicability in specific contexts, and raising questions about their global relevance.

In the second part of the paper, we synthesise insights from another body of literature that delves into the social, material, and historical context surrounding AI4SG initiatives and algorithmic developments more broadly. We build on these insights to extend and adapt Roberson et al's (2021) public good test which requires assessing diversity in research agendas, social orders, and research society networks. In particular, we

[1] For more detail, please see https://sdgs.un.org/goals

highlight the need to understand why and how general frameworks or public good tests might fail, despite the best intentions or recognition of social realities.

Critical social science literature highlights the importance of considering the sociomaterial dimensions of AI4SG that might be in tension with its normative objectives: persistent power imbalances often prioritize technological imaginaries while perpetuating existing structural inequalities, including colonial legacies and global disparities (Magalhães and Couldry, 2021; Madianou, 2021). Additional concerns include exploitative labour practices (Fejerskov, 2022) and research dynamics (Stahl et al., 2023), overexploitation of resources (Stahl et al., 2023), and long-term harm (Green, 2019). Ongoing critical research in this space grapples with rethinking ways forward (Aula & bowles, 2023; Moorosi et al, 2023).

We argue that these critical insights can be integrated into a public good that better reflects structural disparities and social relations in shaping research agendas, social orders, and research-society networks to define “the good” in and around AI. Additionally, focusing on practice enables challenging assumptions that: 1. "the good" is intrinsic to goals such as achieving the SDGs, which can obscure how these goals are pursued and their practical implications for intended beneficiaries, and 2. "the good" is intrinsic to universal principles, often neglecting the systemic and structural nature of issues and their specific contexts.

A practice-based approach would require mapping the system around the problems at hand within their wider social, cultural, and historical context. We could start by developing an understanding of how people are affected by them, and how existing local practices for addressing them can be supported and enhanced. Then we could explore if and how developing AI could be helpful within this system: Whether design strategies less common in AI such as downscaling, or interventions that do not centre on AI might be more appropriate in specific contexts (Hanna & Park, 2020). Assessing the sociomaterial dimensions of “the good” as well as its normative viability, enables moving beyond one-size-fits-all solutions to create interventions that are more responsive and adaptable to the specific needs of the communities they aim to serve.

Our effort to co-produce such a framework will undoubtedly demand numerous iterations and persistent adjustments for this challenging undertaking. We conclude with reflections on the potential use of this framework to foster dialogue with diverse stakeholders, test and co-produce a revised framework, and inform the ongoing development of AI initiatives, particularly in the global South.

References

Aula, V., & Bowles, J. (2023). Stepping back from data and AI for good – Current trends and ways forward. *Big Data & Society, 10*(1), 20539517231173901. https://doi.org/10.1177/20539517231173901

Chui, M., Harryson, M., Manyika, J., Roberts, R., Chung, R., van Heteren, A., & Nel, P. (2018, December). Applying AI for social good. *Discussion Paper. Notes from the AI*

Frontier. McKinsey Global Institute. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.pdf. Last accessed on 1/03/2024

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark and database of AI for social good initiatives. *Nature Machine Intelligence, 3*(2), 111–115. https://doi.org/10.1038/s42256-021-00296-0

Costanza-Chock, S. (2018, July 16). Design justice, A.I., and escape from the matrix of domination. *Journal of Design and Science*. https://doi.org/10.21428/96c8d426

Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical data studies: A dialog on data and space. *Big Data & Society, 3*(1). https://doi.org/10.1177/2053951716648346

Davis, J. L., Williams, A., & Yang, M. W. (2021). Algorithmic reparation. *Big Data & Society, 8*(2), 20539517211044808. https://doi.org/10.1177/20539517211044808

Davis, J. L. (2023). 'Affordances' for machine learning. In *2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 324–332). ACM. https://doi.org/10.1145/3593013.3594000

Fejerskov, A. (2022). *The global lab: Inequality, technology, and the experimental movement*. Oxford University Press.

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics, 26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Gillespie, T., & Seaver, N. (2015, November 5). Critical algorithm studies: A reading list. *Social Media Collective (blog)*. https://socialmediacollective.org/reading-lists/critical-algorithm-studies/. Last accessed on 29/02/2024

Green, B. (2019). 'Good' isn't good enough. In *NeurIPS Joint Workshop on AI for Social Good*. Vancouver, Canada: Track 3, Public Policy. https://aiforsocialgood.github.io/neurips2019/acceptedpapers_track3.htm

Kitchin, R., & Lauriault, T. (2014, July 30). Towards critical data studies: Charting and unpacking data assemblages and their work. *SSRN Scholarly Paper*. Rochester, NY. https://papers.ssrn.com/abstract=2474112

Madianou, M. (2021). Nonhuman humanitarianism: When 'AI for good' can be harmful. *Information, Communication & Society, 24*(6), 850–868. https://doi.org/10.1080/1369118X.2021.1909100

Magalhães, J. C., & Couldry, N. (2021). Giving by taking away: Big Tech, data colonialism, and the reconfiguration of social good. *International Journal of Communication, 15*(0), 20.

Moorosi, N., Sefala, R., & Luccioni, S. (2023). AI for whom? Shedding critical light on AI for social good. https://openreview.net/forum?id=vjwYYlA8Pj

Hanna, A., & Park, T. M. (2020, November 20). Against scale: Provocations and resistances to scale thinking. *arXiv*. https://doi.org/10.48550/arXiv.2010.08850

Roberson, T., Leach, J., & Raman, S. (2021). Talking about public good for the second quantum revolution: Analysing quantum technology narratives in the context of national strategies. *Quantum Science and Technology, 6*(2), 025001. https://doi.org/10.1088/2058-9565/abc5ab

Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). AI for good and the SDGs. In *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges* (pp. 95–106). SpringerBriefs in Research and Innovation Governance. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-17040-9_8

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752. https://doi.org/10.1126/science.aat5991

Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., et al. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications, 11*(1), 2468. https://doi.org/10.1038/s41467-020-15871-z