Steering AI for shared prosperity: practical lessons from 3 years of multistakeholder work at the Partnership on AI

Katya Klinova (Head of AI, Labor and the Economy at the Partnership on AI) and Stephanie Bell (Research Scientist at the Partnership on AI)

Rapid progress in the development and deployment of artificial intelligence (AI) promises to usher in the age of unprecedented breakthroughs in science, healthcare, education, and the fight with climate change. But aside from significant promise, AI also brings significant risks which are important to illuminate and address through new and existing science-policy-society interfaces. In particular, leading academic economists have made repeated warnings[1] [2] about the potential of AI to contribute to a dramatic rise in inequality around the world, as well as worsen the development prospects of low- and middle-income countries (LMICs). Not only is the advancement of AI poised to potentially undermine many LMIC's comparative advantage in global trade[3] and negatively affect employment in non-tradable sectors,[4] AI is also a prime example of an "inappropriate technology:"[5] one not well-suited to put LMIC's resource endowments to productive use.

While these warnings have resonated with policymakers and general audience media outlets, the response from the science and technology communities has been lacking, except for a limited investment in digital re-skilling programs. Large-scale efforts to examine and possibly redirect the overall trajectory of AI away from excessive automation and towards expansion of access to high quality jobs, especially in LMICs, are notably absent from the responsible AI discourse. Relevant suggestions coming from policy and technical communities have been limited to warranted but insufficient interventions on the labor supply side (such as calls for an expansion of retraining programs and social safety nets), which fail to explore

[1] Korinek, A. and Stiglitz, J.E., 2018. Artificial intelligence and its implications for income distribution and unemployment. In The economics of artificial intelligence: An agenda (pp. 349-390). University of Chicago Press. [2] Acemoglu, D. and Restrepo, P., 2019. The wrong kind of AI? Artificial intelligence and the future of labor demand (No. w25682). National Bureau of Economic Research. https://www.econstor.eu/bitstream/10419/196790/1/dp12292.pdf

[3] Rodrik, D., 2022. 4 Prospects for global economic convergence under new technologies. An inclusive future? Technology, new dynamics, and policy challenges, p.65.

[4] Pritchett, L., 2020. The future of jobs is facing one, maybe two, of the biggest price distortions ever. Middle East Development Journal, 12(1), pp.131-156.

[5] Acemoglu, D., 2020. How the Other Half Automates Project Syndicate. https://www.project-syndicate.org/magazine/artificial-intelligence-in-developing-countries-by-daron-acemoglu-2020-04

interventions that would address how technology alters the demand for human labor of various skill groups.

This brief presents interim lessons from the AI and Shared Prosperity Initiative—a multistakeholder experimental project founded three years ago to fill that gap, bringing together a small group of academic economists, technologists, civil society and labor leaders to form the Initiative's Steering Committee.[6] The Committee is charged with overseeing the development of policy and self-governance approaches for pragmatically guiding the development and deployment of AI in service of expansion of access to good jobs around the world.

We highlight four key lessons the AI and Shared Prosperity Initiative of the Partnership on AI can suggest for improving science-policy-society interfaces around AI governance.

1. **There is a need for multistakeholder groups (especially ones representative of affected communities) to form swiftly around frontier technologies and explore contours of potential problems as well as solutions.**

Technologists often lack the multidisciplinary expertise and lived experience needed to anticipate harms (and benefits) that others can see more clearly through their respective professional and personal expertise. Even when this differential expertise is intentionally sought out by technologists, as is often the case in "AI for good" efforts, some technologists can be prone to dismiss this information as non-expert or unimportant compared to their own technological skills.[7] These blind spots can lead to suboptimal technology development from the perspective of both society (where harmful impacts could have been prevented or benefits could have been added) and developing or implementing businesses (who lose out on value created by insights from broader stakeholder groups), meaning these collaborations hold promise for all stakeholders.[8]

In the case of the AI and Shared Prosperity Initiative, when the multistakeholder Steering Committee was formed, there had not yet been a comprehensive landscape analysis of the myriad ways AI could affect the labor market and workers. Drawing upon the diverse expertise and guidance of the group, the Partnership on AI created

[6] Please refer to the Initiative's webpage https://partnershiponai.org/shared-prosperity to see the full list of Steering Committee members, as well as the Initiative's latest updates and outputs.
[7] Sambasivan, N. and Veeraraghavan, R. (2022) 'The Deskilling of Domain Expertise in AI Development', in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '22), pp. 1–14. Available at: https://doi.org/10.1145/3491102.3517578.
[8] Partnership on AI (2021) Redesigning AI for Shared Prosperity: An Agenda. Partnership on AI. Available at: https://partnershiponai.org/paper/redesigning-ai-agenda/.

an agenda for research and action to steer AI towards better impacts on labor and the economy.[9] This agenda has since been cited by a leading AI company in its own efforts to understand and improve its labor market impacts,[10] and by academics highlighting practicable work on these issues.[11] [12]

Similarly, diverse and inclusive multistakeholder groups focused on the social impacts of frontier technologies enable better identification of technical as well as social solutions to problems posed. Technology creators are well-positioned to design and apply essential technical fixes and improved development processes to address problems, due to their technical expertise, proximity to the realities of AI development and deployment, and decision-making power within industry. However, they usually lack the ability as well as the expertise to implement social and governmental solutions. Multistakeholder groups enable a more comprehensive analysis of the diverse capabilities and power held by different actors in a given issue space, and thus a broader and more effective range of possible solutions to problems as they are identified. Early identification of potential impacts allows for the ability to steer frontier technologies in more socially beneficial directions; delayed formation of these groups wastes valuable time for action when harmful path dependencies could have been averted.

2. Reconciling frontier technologies' potential for harm across national borders and corporate structures requires creating new mechanisms for meaningful input from affected communities.

Frontier technologies have the potential to generate tremendous benefits for their creators and users, but the combination of their novelty with the structure of the global economy and regulatory systems around the world often means social harms can be dismissed as externalities to be dealt with later by other actors. Proactive harm identification enables problems to be addressed before they manifest in the real world, and become "someone else's problem." In the AI space, as with other frontier technologies, regulation lags innovation, and in many instances existing

---

[9] Partnership on AI (2021) Redesigning AI for Shared Prosperity: An Agenda. Partnership on AI. Available at: https://partnershiponai.org/paper/redesigning-ai-agenda/.
For a selection of endorsements from Steering Committee members occupying influential roles across industry, unions & worker organizations, and academia, see https://partnershiponai.org/pai-launches-new-phase-of-shared-prosperity-initiative/

[10] Manning, S. et al. (2022) A Research Agenda for Assessing the Economic Impacts of Code Generation Models. OpenAI. Available at: https://cdn.openai.com/papers/Economic_Impacts_Research_Agenda.pdf.

[11] Petersen, E. et al. (2022) 'Responsible and Regulatory Conform Machine Learning for Medicine: A Survey of Challenges and Solutions', IEEE Access, 10, pp. 58375–58418. Available at: https://doi.org/10.1109/ACCESS.2022.3178382.

[12] Korinek, A., Schindler, M. and Stiglitz, J.E. (2022) 'Technological Progress and Artificial Intelligence', in V. Cerra et al. (eds) How to Achieve Inclusive Growth. Oxford University Press, pp. 163–211. Available at: https://doi.org/10.1093/oso/9780192846938.003.0004.

laws are not sufficient to address social harms they may cause. For AI, this occurs both within countries (where companies can develop and implement new technologies, and externalize harms reduction and remediation to government) and across national borders (where automation technologies developed for countries with relatively scarce labor get exported to countries where labor is relatively abundant, leading to to an increase in unemployment).[13] International multistakeholder groups, thanks to the dynamics described in the first lesson, offer the potential to identify possible harms of frontier technologies before they arise, and in particular create opportunities to prevent cross-border harms that affected countries may not have sufficient resources or regulations to address.

3. Diverse participation in multistakeholder work helps advance its real-world application.

The AI and Shared Prosperity Initiative's Steering Committee consists of members representing a variety of fields and disciplines: technologists from the AI industry, academic economists, human rights experts, civil society leaders, and labor experts, including those representing workers advocacy organizations, traditional unions, and informal workers' collectives. Looking at the span of disciplines represented on the Committee, one could assume—correctly—that members of the Committee approach the question of how to govern the impact of AI on workers and the labor market from vastly different perspectives. In order to build cohesion within this group, before making any attempts at inviting the group to co-construct potential solutions together, we invested heavily over the first few months in developing a shared language and agreement around the core set of specific issues to be tackled within the scope of the Initiative. To our surprise, this turned out to be more attainable than initially anticipated, due, at least in part, to all members sharing a firm commitment to the goals of the Initiative. Many of the Committee members championed the goals aligned with the Initiative within their organizations or publicly since long before joining the Committee, while others became active in that capacity within their organizations due to their participation in the Committee. But in both cases, this championing of the Initiative's cause by its Steering Committee members has dramatically expanded the number of avenues open for advancing the adoption of the Initiative's recommendations around guiding AI development in service of shared prosperity. The diversity of fields represented by the Steering Committee members ensured the relevance of the Initiative's outputs to stakeholders in a variety of contexts: from engineers and researchers at technology companies looking to anticipate the societal impacts of their product decisions, to civil society and multilateral organizations building AI impact assessment frameworks, to worker

[13] Pritchett, L. (2020) 'The future of jobs is facing one, maybe two, of the biggest price distortions ever', Middle East Development Journal, 12(1), pp. 131–156. Available at: https://doi.org/10.1080/17938120.2020.1714347.

advocates and representatives preparing to negotiate over what technologies are brought into their workplaces and how they are introduced.

4. **Multistakeholder groups consisting of non-state actors can meaningfully complement the work driven by multilateral organizations and their member states.**

Non-governmental multistakeholder groups can flexibly bridge existing divides across national borders as well as corporate structures. While such structures lack regulatory or enforcement ability,[14] they allow for nimble experimentation and exploration of the solution space. Also, they often possess higher tolerance for risk-taking than what can usually be afforded by multilateral organizations comprised of state actors. International governmental organizations (IGOs) play a crucial role in technology governance and many of their functions cannot be replaced by non-governmental multistakeholder groups, but there is ample space for mutual complementarity between them. Non-governmental multistakeholder groups can generate, test, and create momentum behind promising interventions in the space of "soft" norms voluntarily adopted by private actors, offering proof-of-concept or other evidence in new or novel areas.[15] This in turn makes it easier for IGOs to generate support among their member states for similar interventions to be enshrined in shared frameworks and multilateral agreements subsequently referenced by national governments in their regulatory and legislative activities.

[14] MSI Integrity (2020) Not Fit-for-Purpose: The Grand Experiment of Multi-Stakeholder Initiatives in Corporate Accountability, Human Rights and Global Governance.

[15] Manning, S. et al. (2022) A Research Agenda for Assessing the Economic Impacts of Code Generation Models. OpenAI. Available at: https://cdn.openai.com/papers/Economic_Impacts_Research_Agenda.pdf.