

Redesigning AI for Shared Prosperity: an Agenda

Executive summary



PARTNERSHIP ON AI

Executive Summary

Artificial intelligence is expected to contribute trillions of dollars to the global GDP over the coming decade,¹ but these gains may not occur equitably or be shared widely. Today, many communities around the world face persistent underemployment, driven in part by technological advances that have divided workers into cohorts of haves and have nots. If advances in workplace AI continue on their current trajectory, they could accelerate this economic exclusion. Alternatively, as this Agenda outlines, advances in AI could expand access to good jobs for all workers, regardless of their formal education, geographic location, race, ethnicity, gender, or ability.

To put AI on this alternate path, *Redesigning AI for Shared Prosperity: An Agenda* proposes the creation of **shared prosperity targets**: verifiable criteria the AI industry must meet to support the future of workers. Crucial to the success of these targets will be thoughtful design and proper alignment with the needs of key stakeholders. In service of that, this Agenda compiles critical questions that must be answered for these targets to be implementable and effective. This living document is not a final set of questions, nor a comprehensive literature review of the issue areas covered. Instead, it is an open invitation to contribute answers as well as new questions that will help make AI advancement more likely to support inclusive economic progress and respect human rights.

The primary way that people around the world support themselves is through waged work. Consequently, the creation of AI systems that lower labor demand will have a seismic effect, reducing employment and/or wages while weakening the worker bargaining power that contributes to job quality and security. Recent research by Acemoglu and Restrepo (2019) has shown that job task displacement by technology has outpaced job task reinstatement since the mid-1980s; AI deliberately directed to match or exceed human performance on basic tasks is poised to accelerate this trend. Moreover, current government policies – including tax structures that advantage capital investments over labor expenditures – further encourage the development of automation technologies at the expense of workers. Combined, these trends could lead to a further “hollowing out” of middle-class jobs, reducing opportunities for economic mobility and leaving more workers with few options beyond low-wage, low-quality, precarious jobs. When viewed through a geographic lens, a similar removal of rungs on the economic ladder could occur, with the gains of AI largely going to the wealthy countries where its developers live while low- and middle-income countries lose service and manufacturing work.

The outcomes described above are not inevitable. Companies, governments, international organizations, and workers have opportunities to steer the trajectory of AI in a different direction: one where AI genuinely expands and complements the productivity of human workers instead of replacing them with “so-so automation”² – technologies that eliminate jobs but fail to deliver meaningful productivity boosts that lead to creation of better jobs elsewhere in the economy. To this end, this Agenda proposes that the AI industry take up shared prosperity targets, either voluntarily or with regulatory encouragement. These targets would consist of commitments by AI companies to create (and not destroy) good jobs – well-paying, stable, honored, and empowered ones – across the globe. Broad adoption of shared prosperity targets would help lead AI away from the path of exacerbating inequality and proliferating underemployment.

- 1 Bughin, J., Seong, J., Manyika, J., Chui, M., Joshi, R., 2018. Notes from the AI Frontier: Modeling the Impact of AI on the World Economy (Discussion Paper). McKinsey Global Institute.
- 2 Term coined in Acemoglu, D. and Restrepo, P., 2019. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), pp.3-30.

To date, no metrics have been developed to assess the impacts of AI on job availability, wages, and quality. Additionally, no targets have been set to ensure new products do not harm workers, either in aggregate or by category of potential vulnerability. Without clear metrics and commitments, efforts to steer AI in directions that benefit workers and society are susceptible to unbacked claims of human complementarity or human augmentation. Currently, such claims are frequently made by organizations that, in reality, produce job-displacing technology or employ worker-exploiting tactics (such as invasive surveillance) to produce productivity gains.

The Agenda proposes a set of ideal attributes for metrics to assess whether AI technologies are “good jobs positive,” modeled after carbon neutrality/negativity targets that have helped change institutional behavior in recent years. It additionally proposes domains for assessment (impacts on job availability, remuneration, and quality) and considers a scope of assessment (the direct impacts of an implemented technology, as well as select indirect impacts).

The success of the targets to be developed relies on their support by critical stakeholders in the AI development and implementation ecosystem: workers, private sector stakeholders, governments, and international organizations. Support within and across multiple stakeholder categories is particularly important given the diffuse nature of AI’s development and deployment. (Technologies are often created in separate companies and separate geographies than where they are implemented.) Directing AI in service of expanding access to good jobs offers opportunities as well as complex challenges for each set of stakeholders. The Agenda outlines questions that need to be resolved in order to align the incentives, interests and relative powers of key stakeholders in pursuit of a shared prosperity-advancing path for AI.

- *Workers* offer substantial expertise on what could be improved to increase their productivity and satisfaction, but are rarely given opportunities for genuinely empowered participation in the AI development and deployment process. Strengthened worker organizing would increase their ability to avert exploitative workplace AI, and support the creation of worker-complementing AI, though the diffuse nature of AI development and deployment may limit company-based unions’ ability to shape technology. Alternative corporate governance structures, as well as new practices, norms and institutions are likely needed to provide workers with substantive, empowered opportunities for input.
- *Private sector stakeholders* (AI developing and deploying companies, investors) face real and perceived pressure to focus on profits and returns on investments. However, expansion of AI systems focused on top-line growth underpinned by worker-complementing, productivity-enhancing technologies rather than cost-cutting could be more lucrative in the short-run as well as the long-run. This alternative trajectory could offer investors an additional benefit of supporting a labor market robust enough to encourage continued economic growth through household consumption.
- *Governments* can direct AI in service of good jobs creation through their R&D funding (which builds foundations for private sector growth) and procurement policies, as well as by refining labor market policies to prevent abusive practices, support workers navigating job transitions, as well as address intellectual property and privacy challenges created by the increasing use of workers data.

- *Civil society, international organizations and educational institutions* working across borders will be critical for identifying and mitigating the adverse impacts of AI on workers' rights, and encouraging the development and deployment of technology that fosters shared prosperity.

While this summary has focused on what is known of the subjects outlined above, the objective of the Agenda is in many ways to focus on the critical questions that remain unanswered; these questions anchor each section of the full report. It is our hope that this Agenda will catalyze the research and debates around automation, the future of work, and the equitable distribution of the economic gains of AI, and specifically on steering AI's progress to reduce inequality and support sustainable economic and social development. We enthusiastically invite collaboration on the design of shared prosperity targets. For more information on how to get involved, please visit partnershiponai.org/shared-prosperity.

The full Agenda is available at partnershiponai.org/shared-prosperity.