Redesigning AI for Shared Prosperity: an Agenda

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 by Acemoglu and Restrepo (2019).

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, partnershiponai.org/shared-prosperity.

Table Of Contents

Foreword

The AI and Shared Prosperity Initiative (AISPI) is a multi-stakeholder effort that explores ways to proactively guide the development and deployment of Artificial Intelligence (AI) in service of globally shared prosperity. Advancement towards shared prosperity here is understood as expansion of access to good jobs for all workers, independently of educational level or background, geographic location, race, ethnicity, gender and ability.

This document is a living agenda of the Initiative. It compiles the questions that stakeholders still need answers to in order to successfully steer AI in the service of the above goals. The document is not meant to be thought of as a final set of questions, nor a comprehensive literature review of the issue areas covered, but instead as an invitation to everyone interested to contribute research and answers as well as new and refined questions that would make AI advancement more likely to support globally shared prosperity.

The Agenda is structured as follows:

- Chapter 1 reviews the arguments for why the current trajectory of AI poses a risk of dramatically exacerbating inequality and creating large pockets of entrenched poverty, as well summarizes the key determinants of the direction of AI progress;

- Chapter 2 describes open questions on setting verifiable targets for shared prosperity and defining metrics that could help evaluate whether a given AI investment, development or deployment-related decision serves to expand access to good jobs;

- Chapters 3-6 examine the interests and concerns of key stakeholders who could play a critical role in directing AI in support of shared prosperity:

    - Chapter 3: Workers, unions, and worker organizations
    - Chapter 4: Private sector stakeholders (AI developing and deploying companies, institutional and venture capital investors)
    - Chapter 5: Governments
    - Chapter 6: Civil society, international organizations and educational institutions

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, respect for human rights. We invite people and organizations to collaborate with the Initiative on the advancement of this research or contribute to it through their own independent efforts. For more information on how to get involved, please visit partnershiponai.org/shared-prosperity.

Acknowledgements

The Agenda was written under the close guidance of **the AI and Shared Prosperity Initiative's Steering Committee.** Critical foundations for this work were laid during nine Steering Committee deliberation sessions, which were held between September 2020 and April 2021 to debate the arguments, ideas, opinions, and evidence relevant to the Initiative. PAI is deeply grateful to the Steering Committee members for the generosity, thought leadership, and commitment they invested into guiding the development of the Agenda.

We are very grateful to **Anton Korinek** for serving as a Senior Academic Advisor to the Initiative in addition to his Steering Committee role, and for directly contributing to the Agenda drafting.

We thank **Strea Sanchez** and **Mario Crippen** of United for Respect for participating in the Steering Committee deliberations as guest experts. We thank the Future of Work(ers) program at the Ford Foundation for funding support, and especially **Rachel Korberg, Ritse Erumi** and **Rachel Wasser** for their advice and championship.

The Agenda has benefited greatly from the contributions of the current and former PAI staff members, especially **Samir Goswami** who contributed to the Agenda drafting, **Mark Latonero**, **McKane Andrus, Tina M. Park, Marlena Wisniak, B Cavello, Nicholas Anway, Hudson Hongo, Joshua Pikus, Rosie Campbell** and **Madhulika Srikumar** who provided valuable ideas and advice.

Though this document reflects the inputs of many PAI Partner organizations, it should not be read as representing the views of any particular organization or individual within the Steering Committee or any specific PAI Partner.

Please cite as follows:
Partnership on AI, 2021. *Redesigning AI for Shared Prosperity: an Agenda.* Available at: partnershiponai.org/shared-prosperity

The AI and Shared Prosperity Initiative's Research Group

Avital Balwit, Research Scholar at the Future of Humanity Institute, Oxford University. Avital wrote multiple chapters of the Agenda.

Stephanie Bell, Research Fellow at the Partnership on AI. Stephanie wrote multiple chapters of the Agenda and served as its lead editor.

Katya Klinova, Head of AI, Labor and Economy Programs at the Partnership on AI. Katya has led the work of the AI and Shared Prosperity Initiative since inception and wrote multiple chapters of the Agenda.

Cullen O'Keefe, Associate Counsel for Policy & Governance at OpenAI. Cullen contributed to multiple chapters of the Agenda.

Rafael Proença, PhD Student at the John F. Kennedy School of Government, Harvard University. Rafael contributed to multiple chapters of the Agenda.

The AI and Shared Prosperity Initiative's Steering Committee

Daron Acemoglu, Institute Professor of Economics at the Massachusetts Institute of Technology

Dunstan Allison-Hope, Vice President at Business for Social Responsibility

Michael Chui, Partner at the McKinsey Global Institute

Jack Clark, co-chair of the AI Index and former Policy Director at OpenAI

Andrea Dehlendorf, Executive Director at United for Respect

Arturo Franco, Vice President for Insight at the Mastercard Center for Inclusive Growth

Jessica Fulton, Vice President at the Joint Center for Political and Economic Studies

Ryan Gerety, Senior Advisor at United for Respect

Stacey MacGrath Gifford, Research Staff Member - Future of Climate, Governance of Science and Technology at IBM Research

Deborah Greenfield, Senior Advisor at the US Department of Labor and former Deputy Director-General for Policy, International Labour Organization

Anton Korinek, Professor of Economics at the University of Virginia

Andrew Kortina, Co-founder of Fin Analytics and Venmo

Jody Medich, Principal Design Researcher, Office of the CTO at Microsoft and Founder of Superhuman-X

Sabelo Mhlambi, Founder at Bantucracy

Pamela Mishkin, Policy Staff Member at OpenAI

Shakir Mohamed, Senior Staff Scientist at DeepMind

Paloma Muñoz Quick, Advisor at the UN Working Group on Business and Human Rights and OHCHR B-Tech

Grace Mutung'u, Research Fellow at the Centre for Intellectual Property and Information Technology at Strathmore University

Lama Nachman, Intel Fellow & Director of Anticipatory Computing Lab

Reema Nanavaty, Executive Director of the Self Employed Women's Association

Aiha Nguyen, Program Director, Labor Futures Initiative at Data & Society

Rahul Panicker, Chief Research and Innovation Officer at the Wadhwani Institute for AI

Damon Silvers, Policy Director and Special Counsel at AFL-CIO

Sarah Treuhaft, Vice President of Research at PolicyLink

Part I. What kind of AI would serve shared prosperity and advance the creation of good jobs?

1: Introduction

Will the advancement of artificial intelligence expand access to good jobs[3] or further widen the gap between the haves and have nots? Leading academics have warned about the current trajectory of AI leading to greater inequality (see, for example, Korinek and Stiglitz 2019, Acemoglu and Restrepo 2019, 2020). This Agenda proposes the pursuit of a different trajectory that supports the development of an inclusive global economy. The Agenda outlines what a better trajectory of AI should look like and how to implement it in practice.

The Agenda does not call to slow down AI advancement or categorically avoid automation, but instead proposes that some types of advances in AI should be minimized, while others should be invested in more heavily. AI presents powerful opportunities to revitalize productivity growth and improve living standards and quality of life for present and future generations. But for these outcomes to occur equitably and be shared widely, intentional steps must be taken. The Agenda asks what these steps should be, which stakeholders should be taking them, and how they should be implemented, while considering the incentives and constraints those stakeholders face.

1.1. Operating theories and guiding assumptions

1.1.1. Job availability, remuneration and quality

It is crucial to keep job availability, remuneration and quality in mind when examining how to steer AI in service of expanding access to good jobs. The Agenda emphasizes AI's impact on labor demand as one of the key determinants of the changes AI brings to availability, remuneration and quality of jobs. A reduction in labor demand directly leads to a decrease in employment and/or a decrease in wages.

Moreover, lower labor demand also frequently leads to lower job quality and proliferation of precarity, as it reduces the options workers have available as well as their bargaining power. Lower job quality can manifest itself not only in the form of lower wages but also lower benefits, unpredictable hours and wages, worse safety conditions, reduced autonomy and job control, increased surveillance and privacy violations, and neglect of basic rights, among other impacts. Many workers who lose their jobs remain persistently under-employed and become more willing to accept lower quality jobs or enter into informal employment or non-standard (including "gig") employment (Benanav 2020).

3 Section 2.3 discusses factors constituting a good job for the purposes of the Shared Prosperity Initiative

AI-induced negative shocks to labor demand can, therefore, degrade all aspects of workers' well-being, not only wages or job availability. Because of that, the Agenda focuses on the question of AI's impact on jobs availability and remuneration, while also explicitly keeping track of job quality. It would not advance shared prosperity if AI helped to create more jobs than it eliminated, but those jobs were more precarious, more prone to worker exploitation, or less decent in other ways.

AI also impacts the distribution of income through automated decision-making in economically relevant contexts, like credit access, housing allocation, hiring and more. The Fairness, Accountability and Transparency research community produces scholarship examining this important category of mechanisms, and as such they are not included in this Agenda. For an overview of this work, see Mehrabi et al. 2019, Mitchell et al. 2021.

1.1.2. Recognizing unevenly distributed benefits and losses

Ensuring the total number of good jobs created as a result of an AI advancement exceed the number of good jobs destroyed is necessary but insufficient to meet the goals of the Shared Prosperity Initiative. The gains might fail to compensate for the losses, for instance, if those accrue to workers of different educational attainment levels, in different locations, or in different employment arrangements (such as contract status). Disparate impacts often correlate with race, gender, age, immigration status, income level and other factors, all of which merit dedicated attention.

Workers are often told to prepare and "upskill" themselves in order to move from the losing group to the benefitting group, without an acknowledgement of a highly unequal playing field. The burden of instability and the need to continuously adapt to new skill requirements can be overwhelming for many workers, in part due to structural inequalities owing to entrenched racial, ethnic, gender, colonial, digital, geographical, language, able-ist and other divides, including those created by the technology itself. The economic system AI development happens within allows for a "privatization of gains, and socialization of costs" (Mazzucato 2011), with individual workers, their families and communities forced to bear much of the burden.

1.1.3. Types of AI advancement depending on labor impact

Economists refer to technologies that increase the demand for labor as labor-using. Not all technologies are labor-using; the opposite are labor-saving technologies, which reduce overall demand for labor in the economy and lead to falling wages and employment.

Among innovations that increase the demand for labor and are thus labor-using, some increase the demand for capital or other factors of production by even more than the demand for labor, leading to a higher percentage increase of capital earnings than labor earnings. Technological change that benefits labor and capital equally is called labor-neutral. If all changes induced by AI were neutral in this sense, it would guarantee that labor earnings grew at the same rate as the earnings of capital. If labor benefits relatively more than capital, the technological change is called biased in favor of labor, and this type of progress would actively contribute to reducing inequality. These concepts could also be applied in a more nuanced way: technological change can be biased in favor of or against specific categories of labor (for example, workers without college degrees).

Chapter 2 discusses how technologies that are labor-neutral or biased in favor of labor can be identified by organizations that develop or deploy AI, and how to develop an operational framework for steering AI progress to serve greater economic inclusion.

1.1.4. Insufficiency of a redistribution-only solution

The Agenda does not discourage redistributive policies, given their large potential to support the sharing of AI-produced economic gains. It does not, however, assume redistribution will be sufficient for sustaining an inclusive, rewarding, and broadly prosperous economy. Relying on redistribution alone is a fragile strategy, especially in a world where most people's livelihoods presently rely on waged work, robust international redistribution mechanisms remain politically elusive, and a hyper-concentration of wealth gives the super-rich outsized political powers. The Agenda argues that at least for the foreseeable future, expanding access to good jobs is a necessary and critically important component of steering AI towards shared prosperity, which should be complemented but not replaced by redistributive policies.

In particular, a drive to "liberate" people from the need to work using automation technology does not account for the world-spanning consequences of job automation. Redistributive policies are far more likely to facilitate transfers of wealth within countries, not between them. As will be discussed in section 1.2.2, workforces in low- and middle-income countries stand to be impacted by the rise of AI, but might not be able to count on income transfers. Even if the advancement of AI could generate sufficient gains to fully support the world's unemployed population and cross-border redistribution could be arranged, the political stability of a world with even more concentrated economic power and reduced or eliminated bargaining power of the unemployed remains to be seen.

Counting on the benevolent or socially inclined behavior of actors who have historically not acted in such a fashion effectively converts the basic needs of economically vulnerable populations into a gamble. Additionally, the feasibility of AI creating a post-scarcity world remains unproven and unsupported. Excusing the entrenchment of inequality produced by AI deployments today in the name of an uncertain, AI-backed post-scarcity future unnecessarily sacrifices the needs of present-day workers, their families, and their communities in the name of highly speculative benefits for future generations.

None of this implies that technological opportunities to automate dangerous, physically or mentally taxing work should be foregone. But "freeing" people from their existing work can only be thought of as beneficial if alternative sources of income or better work opportunities are made available to them. As long as the short- and long-term feasibility (and even the desirability) of supporting the world's population through a global basic income remain uncertain, the only responsible option is making sure that technological change creates more good jobs than it eliminates.

1.2. Trajectory of AI progress and its determinants

This section gives a very brief overview of the major arguments that have unfolded recently in the literature on AI and labor demand. At the time of writing, the long term impact of COVID-19 pandemic on the employment and automation patterns remains to be seen. The pandemic has imposed a "shadow cost" on in-person labor, boosting incentives to accelerate automation (Korinek, Stiglitz 2021b). Year 2020 has also clearly demonstrated the depth of technological transformations that many organizations were able to undertake with unprecedented speed. Recent reports point to an uptick in AI-powered automation, the use of robotics, and robotics process automation (Lund et al. 2021).

1.2.1. AI's impact on productivity and labor demand

AI will boost productivity, but there is no agreement on the size of the increase. Even less certain is whether workers will share in the gains associated with increased productivity through higher wages or employment, with many warning that technological change brought about by the current trajectory of AI will be biased against labor. (Lane, Saint-Martin 2021). Concerns have been raised about overall productivity growth in the age of AI remaining slow due to frontier growth failing to be shared across the economy. Unlike the frontier firms of the past, today's "superstar firms" employ only a tiny fraction of the workforce and buy very little from the rest of the economy. (Gutiérrez, Philippon 2020).

Many academics warn that AI advancement might exacerbate recent negative economic trends observed in the US and multiple OECD countries, including the overall decline in labor's share of national income, the widening gap between average and median labor compensation, and the stagnation and decline in real wages of workers without college degrees, as well as labor-market polarization and the hollowing out of the middle-paying jobs (Autor, Mindell, Reynolds 2020). Those trends, which have been unfolding over the past three decades, are often partially attributed to the bias of technological change in favor of people with advanced degrees, who have seen their wages grow much faster that everyone else's in the US. AI, on its current path, is expected to be plagued by skill bias as well.

Acemoglu and Restrepo (2019) show that, unlike in the four decades following WWII, in the recent three decades since 1987 technological change has been labor-saving, in other words the pace at which waged tasks were replaced by machines has been greater than the the pace of creation of new waged tasks for humans. They warn that AI is likely to continue this trend which would lead to further downward pressure on labor demand, and caution against AI proliferating "so-so technologies": those that generate task displacement, but do not bring step-shift productivity gains, such as self-checkout kiosks, robotized customer support agents, etc. (Acemoglu, Restrepo 2020).

1.2.2. AI's impact on low- and middle-income economies

AI-powered automation is likely to accelerate premature deindustrialization across the developing world. That would force the export-oriented growth model to retire even faster as a viable path to development, since an expansion in the range of human tasks that can be automated decreases the relevance of low-income countries' main comparative advantage: a competitively paid labor force (Rodrik 2018, Korinek and Stiglitz 2021a).

Recent evidence from Africa suggests that global trends in technology have already led to the adoption of excessively capital-intensive techniques in countries where much more labor-intensive techniques would be expected and appropriate given the labor market conditions: shortage of formal sector jobs, young and growing labor force, and the proliferation of a learning crisis (Diao et al. 2021; World Bank 2017).

An aging workforce and the relative price of labor in the economies dominating the production of automation technologies, coupled with the borderless nature of technological deployments, leads to excessive levels of automation "spilling over" to countries facing a dire need of jobs (Pritchett 2020).

AI advancement is at present concentrated in a handful of high income countries and China. Their governments might be able to capture AI-generated gains and redistribute those to their impacted workforces. Countries not currently in possession of significant AI talent, R&D and investment capacity might end up dealing with job-displacement occurring due to cross-border technology deployments, but without any upside to capture and redistribute. The concentrated nature of AI development might also lead to an increase in the income gap between the AI development-hosting countries and the others: the former will get richer while the latter will be disadvantaged, potentially leading to an overall increase in cross-country inequality.

1.2.3. AI's impact on workers rights, well-being and quality of jobs

As has been noted above, a decline in job quality is one of the most common ways in which a negative shock to labor demand can manifest itself, along with employment and wage decreases. A few concerns have rightfully attracted the most attention in the conversation about AI's impact on job quality. Those include AI enabling increases in worker monitoring and surveillance, exploitative and excessively punitive practices, and unpredictability of algorithmic shift scheduling. Concerns have also been mounting around growing work intensity and injury rates as a result of introduction of robots or algorithmic work oversight, the proliferation of API-enabled per-task payment, and increasingly narrow definitions of work eligible for compensation, among others (Moradi, Levy 2020; OECD 2019).

For an in-depth overview of documented as well as anticipated impacts of AI on all aspects of workers' well-being (human rights, physical safety, financial, intellectual and emotional well-being, sense of meaning and purpose), please see Partnership on AI's "Framework for Promoting Workforce Well-being in the AI-Integrated Workplace" (2020).

1.2.4. What factors determine the current trajectory of AI progress?

Government policies directly and indirectly shape economic incentives of innovators and affect whether AI gets channelled into automation beyond socially optimal levels and/or into other harmful practices. These include:

- Tax policy: the current effective tax rate on labor exceeds the effective tax rate on capital in the US, incentivizing development of labor-saving technologies (Acemoglu, Manera, Restrepo 2020);
- Focus of government R&D investment: presence of publicly-funded research in a particular application area makes commercial investment in that area less risky (Mazzucato 2011);
- Interest rate policy: low interest rates make investments in capital and automation more attractive;
- Government procurement practices: large government contracts for development and deployment of labor-saving technology can spur its proliferation (Korinek, Stiglitz 2020);
- Immigration policy: excessive labor mobility restrictions lead to an over-investment in automation technologies, which get deployed not only domestically, but also in low- and middle-income countries (Pritchett 2020);
- Labor laws and regulations: restrictions prohibiting or limiting practices that undermine workers' well-being, such as fair scheduling laws[4] or workplace surveillance restrictions, reduce incentives to create, invest in, and proliferate exploitative technologies.

4 See, for instance, New York City's Fair Workweek Law, which took effect in November 2017

These and other incentive-shaping policies are discussed in chapters 5 and 6. It is also relevant to note that at present AI development is taking place against the backdrop of historically low levels of union membership and corresponding worker power. The share of the workforce covered by collective agreement is at its lowest in Germany (under 60%), Canada (under 30%), and across the OECD (just above 30%), while the UK and the US have seen the largest proportional drops in union membership, to just under 30% and 12% respectively (Autor, Mindell, Reynolds 2020). Also notable is the major role that big technology companies play in setting the direction of AI advancement, spending tens of billions US dollars on AI R&D annually (Bughin et al. 2017).

Given the prominence of big tech companies in funding AI research, the vision and ideas embedded in the culture of those companies play a meaningful role in determining its focus (Acemoglu and Restrepo 2020). Similarly, practices that gain popularity in the AI research and development communities have significant influence on the trajectory of AI. For example, it is very common to chase state-of-the-art performance on benchmark datasets (such as ImageNet or SuperGLUE). Availability or unavailability of benchmark datasets therefore heavily shapes which domains AI progress happens in, and what counts as progress (Raji et al. 2020). In particular, the propensity to compare the performance of an AI system to human performance, fueled by both media headlines and benchmark practices, further increases the focus of AI research on "beating" humans instead of seeking to complement them and team up with them (Siddarth et al. forthcoming).

2: Shared prosperity targets

Multiple influential organizations have published AI principles that include laudable statements like "The economic prosperity created by AI should be shared broadly, to benefit all of humanity"[5] or "AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being".[6] But to the best of our knowledge, no organization has translated these principles into verifiable targets pursued in practice.

This section discusses the challenges around setting such targets and identifying metrics that would help evaluate whether a given AI development or deployment decision advances shared prosperity. Overcoming these challenges is necessary to practicably guide AI development towards non-destruction and co-creation of good jobs.

The section employs an analogy to environmental impact metrics. Specifically, there are valuable lessons in the efforts to change the direction of technological progress in the energy sector, steering it away from fossil fuels and towards renewable sources via advancing a broad adoption of carbon emission reduction targets.[7]

The current trajectory of AI risks depressing labor demand and subjecting millions of people around the world to a life of under-employment and poor-quality jobs. Can a better trajectory for AI be encouraged by the adoption of shared prosperity targets that are (i) sufficiently precise to be expected to lead to a better outcome, but (ii) sufficiently simple to be practicably measurable and intuitive (like goals around zero carbon emissions)? What measurable criteria should the targets cover? What regulatory steps are necessary to stimulate the adoption of the shared prosperity targets?

2.1. Components and requirements

2.1.1. An actionable framework

Targets.

What concrete targets around AI impacts would make AI advancement more likely to support shared prosperity?

Here, targets can be both quantitative and qualitative, and could be adopted by organizations developing and deploying AI either voluntarily or as a result of external imperatives (regulation, pressure from the public or employees, etc.)

Requirements to be satisfied by the shared prosperity targets include:

- Transparency: the definition of being "good jobs-friendly" should be public knowledge
- Verifiability: it should be possible to unambiguously determine whether an organization is making meaningful progress towards its declared shared prosperity targets.

5 Principle #15 of Asilomar AI Principles: https://futureoflife.org/ai-principles
6 Principle #1 of the OECD AI Principles: https://www.oecd.org/going-digital/ai/principles/
7 To the best of our knowledge, the analogy was first described on October 18, 2019 by Daron Acemoglu during his talk at the PIIE conference "Combating Inequality: Rethinking Policies to Reduce Inequality in Advanced Economies"

- Supportability: shared prosperity targets should be grounded in the best available theoretical and empirical knowledge around fostering a good jobs economy, and be updated as new evidence emerges.

- Coherence with existing standards and requirements.

As discussed in the Introduction, if progress towards the shared prosperity targets is evaluated only in aggregate, that might disguise disparate racial, gender, caste, and other group-based impacts. This warrants a dedicated examination of deltas by different identity categories (race, gender, age, educational attainment level, income distribution percentile, etc).

What requirements are missing or should be refined?

Implementation blueprints.

What processes and governance mechanisms can be incorporated into the AI development process to help avoid imposing excessive economic burdens on society?

What "off-the-shelf" steps can organizations take to progress towards shared prosperity targets?

What stakeholders should be involved in evaluating the progress?

What audit trails should be generated to document efforts, learn and improve?

What public disclosures need to be made, and how often?

Today, organizations interested in reducing their environmental impact have access to a range of helpful blueprints, action plans, and sector-specific pathways towards reducing greenhouse gas emissions.[8] These include best practices in climate risk assessment and governance, energy management, waste management, emission reduction target-setting and reporting practices, even if the implementation of each of these might be unique for every organization.

Organizations will be more likely to pursue shared prosperity targets if there are blueprints for advancing towards them, as well as clear and measurable progress criteria. A growing body of literature develops insights on how to advance towards AI transparency, accountability, safety, privacy and more. Practical guidance for avoiding negative labor market impacts produced by AI development is currently lacking.

Organizational decisions unrelated to AI can also be guided by the concern for shared prosperity, for example decisions about investment in worker training, building innovation ecosystems, fostering a talent pipeline, etc. Simultaneously, implementation blueprints need to guard against incongruencies between efforts to develop and deploy AI in service of shared prosperity and practices that may counter that, such as tax avoidance, union busting, lobbying for legislation that weakens worker rights and protections.

8 See, for example https://sciencebasedtargets.org/sectors or https://www.climateaction100.org/approach/the-three-asks/

2.1.2. A need for precise definitions

How can shared prosperity targets be defined precisely enough to avoid labor impact-washing and risks associated with vague goals like "augment humans instead of displacing them"?

Investing in AI applications that augment the productivity of human workers and create new tasks for them is often discussed as a recipe for countering the negative economic trends described in section 1.2. This has led to a proliferation of AI companies that describe their products as complementing humans instead of displacing them. A number of risks are inherent in the vagueness of such claims, prompting the need for a more precise specification of a desirable path for AI.

For example, what conditions are needed to ensure a productivity enhancing technology does not contribute to harmful work intensification, and increased physical or cognitive strain? How can productivity monitoring technology be ensured not to enable exploitative and punitive work environments? Finally, what does it take to ensure workers share in the benefits of their increased productivity through higher compensation or improved working conditions?

"Complement humans instead of augmenting them" goals unaccompanied by measurable commitments are too ambiguous and can be used as cover-ups by organizations that develop worker-exploiting AI systems marketed as "worker-augmenting". Organizations genuinely conscious of their economic and labor impact should be wary of unbacked claims and welcome the adoption of measurable targets, which would allow them to credibly differentiate themselves.

2.2. Accounting for the impact on labor income and employment

What observable metrics capture AI's impact on labor demand, i.e., on both labor income and employment?

Which of those metrics should be incorporated into the shared prosperity targets?

Technological change affects labor income and employment through a number of direct and indirect ways. This section examines each of those effects. Direct impacts are reviewed first for the ease of the discussion progression, though the bulk of a given AI system's economic and labor impacts might arise through the indirect and downstream effects propagating throughout the economy. Thus, the sequence of this section should not be assumed to reflect the relative size or importance of impacts: those discussed last in this section can often be the largest.

2.2.1. Direct impacts

What observable metrics reflecting direct impacts of AI on labor demand should be incorporated into the shared prosperity targets?

The most direct effect that an organization has on labor demand is via its own hiring and compensation practices. An AI system that leads the organization to hire new workers and/or increase the pay of existing workers would contribute to greater overall labor demand in a given labor market category. Conversely, rolling out a technology that induces the organization to shed certain types of workers lowers labor demand in that category.

In addition to looking at absolute impacts on workers, it is necessary to examine relative impacts as well, namely whether each category of labor benefits at least proportionately or more than proportionately compared to other factors of production. An AI system roll-out is neutral or biased in favor of a group of workers if it induces the organization to increase the compensation of this group proportionately or more than its spending on other factors. Relative impacts are captured by the change in labor's share (i.e. the ratio of what the organization pays to workers in the form of wages divided by the total value added that is created by the organization).

In these calculations, it is crucial for the change in labor's share and other metrics to be assessed by worker category, not just in total. Otherwise, the outsized wages of C-level executives and other high-power, high-wage workers may mask the reality faced by other workers. Worker categories of greatest concern can be identified by examining the changes in labor's share by wage distribution percentiles, as well as race, gender, age, educational attainment level and other characteristics.

2.2.2. Impact along an organization's value chain

How can the shared prosperity targets be designed to guard against gaming or circumvention?

If an organization sells AI solutions to other organizations, what restrictions can it impose on the downstream uses, and through what mechanisms?

What actions of customer organizations are to be considered outside of any mechanism of influence of the selling organization?

Tracking the change in the labor share for a given organization might present a biased picture if the organization outsources, for example, its lowest-paid work to sub-contractors. In such a case, it might look like all categories of its own full-time workers are well-compensated, while inadequate compensation or working conditions of the sub-contractor are hidden away from the spotlight. For instance, AI companies frequently outsource labeling of datasets, which is completed by subcontracted or crowdsourced workers whose income may be below a living wage (Gray, Suri 2019). To account for that, shared prosperity metrics must track not only the organization's own labor's share over time, but the labor's share of its upstream suppliers.

Similarly, the bulk of an organization's impact on labor demand may occur from selling AI solutions used by customers in ways that increase or undermine their customers' labor demand—rather than labor demand within the AI-developing organization's own walls. This underscores the importance of incorporating the downstream impacts of new technologies on the labor share in the targets for shared prosperity. AI-developing organizations can influence or restrict the use of their technology by customers via licensing agreements, by refusing sales to bad actors, or via design features that prevent certain uses. Further work on mechanisms that would allow developers to prevent downstream uses of their AI systems that undermine labor demand and/or reduce labor's share of income is urgently needed.

2.2.3. Indirect impacts

What observable metrics reflecting indirect impacts of AI on labor demand should be incorporated into the shared prosperity targets?

To analyze the overall effects of the introduction of a new technology on the economy, it is not sufficient to focus on a static snapshot of a single organization. There are important equilibrium effects in the economy whenever the prices and quantities of the products created by the organization change. The rest of this section discusses those effects in brief; for a more detailed description and examples see the framework in Klinova and Korinek (2021).

First, there is a direct demand effect. If a new process allows an organization to increase sales, either by offering new products and services or by offering existing ones at better prices, quantity demanded will increase. By implication, the organization will require more input of all the variable factors required to produce more, including different categories of labor.

There are also vertical effects along the organization's value chain – for example, if an organization expands production, its suppliers in turn need to increase production, which will generally increase labor demand in the economy in absolute terms. Depending on how much labor the suppliers rely on, this effect may be neutral, biased in favor of labor or against labor.

Moreover, there are horizontal knock-on effects on the competitors of the organization as well as the producers of other goods and services that are complements or substitutes to what the organization produces. For instance, a new service by an e-commerce company might impact workers employed by brick-and-mortar stores that are neither suppliers nor customers of that e-commerce business, leading to a reduction in the labor's share of income in the economy overall even if the labor's share of the e-commerce company remains the same. A new product or service might also allow consumers to save time or money, or lead to a creation of other forms of consumer benefit, freeing up resources to boost demand in other sectors of the economy, which can lead to a corresponding increase in employment in those.

Finally, a productivity effect arises when a new technology makes factors redundant, for example when workers are laid off. This effect captures that the workers can now be employed elsewhere, producing more output and thus making the economy more productive. The magnitude of this productivity effect depends on how many workers are made redundant as well as how quickly and productively they can actually be reemployed.

As the analysis moves from the direct effects of an organization's choices to more indirect demand and equilibrium effects, it becomes more difficult to make precise predictions about the effects on the economy's labor share. One important question is if there are impacts on labor demand that are sufficiently indirect to be considered outside of an organization's area of influence, analogous to "Scope 3 emissions"[9] in the case of environmental impact.

2.3. Accounting for job quality impacts

How can the shared prosperity targets incorporate organizations' impact on quality of jobs?

What aspects of AI's impacts on workers are most important to include?

9 Description of Scope 3 emissions can be found at: https://www.epa.gov/climateleadership/scope-3-inventory-guidance

In addition to the income and employment impacts discussed above, companies must consider impact on job quality.[10] As mentioned in the Introduction, these categories of impacts are closely related, and all have substantial impacts on workers' quality of life. While debate exists around the margins of what constitutes a high-quality job and how to best categorize the elements therein, there is general consensus on the importance of five non-financial aspects of worker well-being: human rights, physical safety, intellectual well-being, emotional well-being, and sense of meaning and purpose (Partnership on AI 2020). These aspects of well-being consistently appear in respected measures of job quality, from countrywide measurements like the ILO's Decent Work indicators,[11] the OECD's quality of working environment inventory, and the UK RSA's Job Quality Metrics, to company-level approaches like MIT's Good Jobs Institute Scorecard and The Aspen Institute's Working Metrics collaboration.

The ILO indicators and OECD inventory are particularly useful guides for the shared prosperity targets, as they not only offer clear and measurable indicators, but also have been endorsed by member states with a wide range of business contexts and cultural values (187 member states, including worker and employer representatives in the case of the ILO, and a selection of mostly high-income economies in the case of the OECD). Work must be done to convert these measures—designed for countrywide assessments—into useful, rigorous metrics and tools to be used at a company level. All of the challenges of measuring direct and indirect impacts discussed above apply here as well, with additional obstacles created by the fact that these aspects of job quality are often measured using surveys, potentially requiring substantial collaboration from customers or contractors to understand the full scope of impact. Best practices from supply chain auditing organizations are likely to prove helpful in identifying practical and effective ways to measure downstream effects.

2.4. Formats for adoption of the shared prosperity targets

What existing formats of company assessment and reporting are most promising for the shared prosperity targets?

Are there opportunities to integrate the shared prosperity targets into existing measurements, assessments, or reporting (including disclosures mandated by regulations or laws)?

What existing institutions could offer rigorous and objective assessment or evaluation of adherence to the shared prosperity targets?

10 A complete accounting of labor demand impact would include attributes of job quality as a part of true compensation. Without explicit measurement of these attributes, however, traditional measurements like work-hours and wages can obscure elements like working conditions. Care must be taken to ensure double-counting does not occur across metrics.

11 The ILO describes decent work as involving "opportunities for work that is productive and delivers a fair income, security in the workplace and social protection for families, better prospects for personal development and social integration, freedom for people to express their concerns, organize and participate in the decisions that affect their lives and equality of opportunity and treatment for all women and men." (For additional information, see the ILO website on decent work: https://www.ilo.org/global/topics/decent-work/lang--en/index.htm)

Ease of adoption for companies will be critical to the success of the shared prosperity targets. Beyond the target attributes and implementation blueprints discussed in 2.1.1, the format of the targets themselves will play a major role in determining ease. Two potential ways of ensuring ease include either integrating with existing assessments and reporting, or modeling the targets after them. In the former category, AI companies are exploring a variety of responsible/ethical AI assessments that could be constructed to encompass the targets, including documentation of impacts in model cards, inclusion in AI ethics audits, or inclusion in independent digital ethics rankings (e.g., The New America Foundation's Ranking Digital Rights). In the latter category, the Sustainable Accounting Standards Board offers an illustration of how to structure new assessments to resemble reporting requirements that have become second nature for businesses—in this example, financial accounting standards.

Another approach to creating ease of adoption for companies would be to rely on an external body for assessment or adherence verification. Numerous certification schemes (e.g., FairTrade or organic goods, LEED standards for sustainable architecture, Human Rights Campaign's Best Places to Work for LGBTQ Equality) have followed this approach. The external body could solely act as an assessor or certifier, or it could also advise companies on how to meet the shared prosperity targets. In some analogous examples, new organizations were created for the explicit purpose of assessment or certification; in others, existing bodies expanded their responsibilities. No universally agreed upon AI ethics certifier presently exists, but existing organizations with related mandates might be well suited for role expansion. Governments could also choose to require companies to report their impacts on labor demand, in the same way that they require standardized financial reporting and are exploring the same for environmental sustainability. Alternatively, they could institute labor demand and job quality impact assessment requirements similar to environmental impact assessments required for policies, programs, or projects.

Companies' business models determine which of their AI development and deployment decisions have the most relevant impacts on shared prosperity; returning to the climate change analogy, an oil and gas company would have a different set of relevant factors for achieving carbon negative targets than an agribusiness. Similarly, relevant factors measured and tracked under the shared prosperity targets will likely differ between companies and business models, meaning companies themselves may be best positioned to do the measurements, and generate a documentation trail that can later be audited by a certifier or another independent organization.

Part II. Actors critical for steering AI towards shared prosperity and creation of good jobs, their interests and concerns

Part II of the Agenda reviews the interests, concerns, and levers of influence of the relevant groups of stakeholders: workers, industry, governments, civil society, and international organizations.

3: Empowering vulnerable workers throughout the AI development and deployment process

Employers have authority over their workers and their working conditions that outstrips even governments' authority over their citizens in some regards (Anderson 2017). With relatively few constraints, employers decide the conditions under which their workers spend a third of their waking hours, including expectations for the work they complete and the environment and culture in which they complete it. Throughout much of the world, workers can be dismissed and thus stripped of their livelihood without cause, offering employers substantial coercive authority over their employees. For vulnerable workers,[12] understandable fears of job loss further exaggerate this authority. As discussed in the Introduction, the ability of businesses to meet their needs for labor while offering little to no job security, low wages, and poor working conditions is a function of multiple factors: labor demand, worker bargaining power, and legal or regulatory requirements (among others). AI has the potential to transform the scale and scope of each.

In the context of these workplace power dynamics, new technologies like AI have the potential to exacerbate harmful practices. However, when created with higher regard for workers, AI can bring many advantages: the elimination of dangerous work, the reduction of repetitive or unfulfilling tasks, new paths for career advancement, and the opportunity to accomplish previously impossible tasks or levels of productivity, among others. While managerial culture largely determines introductions and applications of technologies—not the other way around—the availability of technology still affects whether workers are subject to tools of exploitation at an individual or systemic level, or whether the tools work to their assistance and benefit.

The degree to which workers have agency and autonomy in their jobs greatly impacts their well-being (Pfeffer 2018). Finding opportunities for workers, especially vulnerable workers, to participate in shaping the role of AI in their workplaces is therefore critical for worker well-being, as well as beneficial for employers (discussed further in chapter 4). Creating substantive and meaningful opportunities for participation (i.e., with workers able to speak freely without fear of repercussions or retaliation, and with the expectation that their participation will affect outcomes) requires workers to have some level of authority or influence (e.g., through unions, sectoral bargaining, or worker technology committees in government labor departments).

12 For example low-wage workers, workers without in-demand credentials, undocumented workers, workers from demographic categories subjected to higher rates of workplace abuses, and workers in geographies with high formal unemployment

3.1. AI's impact on worker power and autonomy

How is AI changing worker power, especially workers' abilities to advocate for themselves (as individuals or groups), or their autonomy over their work?

AI products have already been deployed in workplaces throughout the world, from robotic assemblers, to schedulers, to surveillers, with more applications to come. Though the systems vary in terms of underlying technologies and tasks performed, when considering their impact on workers, they can largely be categorized by the roles they fulfill in the workplace: managers (with decision-making authority or control over workers), peers (completing the same or related tasks as workers), or assistants (completing supporting tasks for workers). In each role, use cases exist that can either benefit or harm workers. As with the power dynamics of their human equivalents, workers have more concerns about AI bosses and peers, while assistants are largely viewed as helpful.[13]

Given its comparative level of authority over workers, manager AI has the highest potential to shape job quality, worker autonomy, and worker power. Companies use manager AI to hire, schedule, train/mentor, and surveil workers. Manager AI systems often deepen existing workplace interactions and power relationships, for better and for worse (Kellogg et al. 2020). For instance, in workplaces with high regard for workers, training systems could provide workers more personalized feedback, offering better guidance on what they do well and how they can improve their performance. In other workplaces with less regard for workers, the same systems could be used to punish workers for falling below expectations set by the company. Assigning managerial tasks to AI systems introduces the possibility of what Gray and Suri (2019) describe as "inadvertent algorithmic cruelty," with decisions made by technologies designed to follow employer preferences. Such systems lack empathy for challenges faced by workers in their very human personal lives, or acknowledgement of workers' abilities to navigate these out-of-work challenges when afforded flexibility and job control.

Some manager AI products, like surveillance tools, can transform existing workplace dynamics so thoroughly that they are experienced as different in kind by the workers (Athreya 2020). Surveillance tools are created and sold for a wide variety of work settings and industries, from cameras and motion sensors in warehouse and retail work to desktop monitoring software for anyone whose work primarily entails sitting in front of a computer (e.g., piecework data labelers, human resources professionals, coders, investment bankers). Previously, the most surveilled workplaces could at most create the impression of a Benthamite panopticon, with workers self-regulating and self-disciplining because they are unaware of whether they are being monitored in that moment—but without the ability to actually monitor each worker at every moment.

AI's ability to work with new and large-scale categories of data have led to the creation of systems in which workers are genuinely under perpetual surveillance, leaving no space on the shop floor out of the boss's sight (Delfanti, Frey 2021). This comprehensive surveillance harms workers' individual privacy and autonomy on the job as well as their ability to organize. Employers are provided tools that enable them to set productivity targets that sacrifice workers' health and well-being, and workers are required to navigate

13 There is a distinction between technology that workers view as assistants and technology that their managers or company leaders see as their assistants. Some surveillance technology is marketed as an assistant to workers, as it gives them clearer insight into whether they are spending their time on tasks that are of high value to the company and can help them prioritize their time, for instance. When such surveillance technologies are used to set more aggressive output targets, or to penalize lower performance, these technologies are better classified as manager AI from a worker's perspective. Similarly, workers may knowingly or unknowingly train assistants in how to perform their own tasks; where this is the case, and particularly when these technological assistants go on to assume those tasks, these technologies are better understood as peer AI.

environments some of them describe as being more like prisons than workplaces. Similarly, employers often orient AI scheduling software entirely around company and customer desires, not worker needs. This practice disrupts workers' out-of-work lives when they're called in at the last minute to meet customer surges (or leads to punitive measures when they cannot work last-minute shifts) and systems are frequently set to cut worker hours just shy of benefits requirements.

Peer AI, or automating and displacing AI, has clear negative impacts on job quantity and quality, as laid out in the Introduction. Higher unemployment and underemployment undermines worker power in the long-term. In countries without strong job guarantees or employment support for most workers (e.g., outside places like Sweden and Japan), workers view peer-like AI and automation with some wariness. One survey found that a majority of US retail workers anticipate new technology will negatively impact the quality of their job (including wages, hours, and benefits) within the next 10 years (Corser 2019). In areas with stronger protections for worker retention, job transitions, or income replacement, this wariness is less prevalent, with workers embracing some automation technologies. Taking one such example, in Japan, automation is more common due to current and anticipated impacts of workers aging out of the workforce and low immigration. In instances where existing workers' jobs are automated, societal expectations of companies to provide lifelong employment for their workers means many workers are retained even after their tasks have been automated (Kageyama 2017). 68% of the Japanese and in a recent multi-country Pew survey supported robots doing many jobs humans have done in the past, versus only 41% of US respondents (Pew Research Center, 2020).

AI serving as workers' assistants offers the highest potential to improve workers' lives, by taking on tasks that are dangerous (e.g., performing the deadliest and most injurious tasks in mining and offshore drilling, under the supervision of workers who previously had to put their bodies on the line to), or time-consuming and unfulfilling (e.g., rote paperwork at desk jobs), as well as by expanding the frontier of what's possible through AI/human collaborations. It also has the potential to protect workers by monitoring health and safety conditions, and assist labor inspectors in their jobs. Still, depending on the use cases and deployment of these AI worker assistants, their implementation can threaten worker well-being. Examples include deskilling work and leaving lower wages for what remains, or claiming all of the "recharge time" created by these more routine/simple tasks in a role—leaving only tasks requiring more strenuous emotional or intellectual engagement without any corresponding increase in pay per hour.

Regardless of the roles assumed by workplace AI technologies, they share a common trait of operating in ways not described to the workers who interact with them. With AI worker assistants, the outcomes produced by the systems may be more important than this trait. To a bakery cashier using an AI-driven visual product identification system, it normally matters more that the system is swift and accurate in its identification and accounting of a customer's lunch than it does that the cashier could explain the underlying technology to the customer. In instances where the systems exercise authority over workers, however, that lack of explainability and transparency is directly relevant to workers' interests and well-being. AI decision-making systems can reduce transparency and accountability of managers and leaders, with the supposed greater wisdom of the model overriding negative outcomes it produces for workers (Doshi-Velez, Kim 2017; Kellogg et al. 2020). Moreover, tools that appear to treat workers neutrally on their surface in application cause impacts disproportionately born by specific subgroups of workers (e.g., demand-driven, last-minute scheduling software applies equally to all workers but is particularly hard for workers who are also caretakers—and this out-of-work duty disproportionately falls to women). Across the board, these systems may leave workers feeling less autonomous and empowered; decisions are made for them by systems for which they have not been offered functional explanations, and which leave no avenue of appeal.

3.2. Challenges incorporating worker voices in AI development and deployment

How do AI development and deployment processes need to be modified to foster substantive input from affected workers, including workers from different companies and geographies?

What steps should be taken to ensure workers are empowered participants?

Workers possess unique insight and expertise into their roles and tasks through the accumulated wisdom of doing their jobs, a form of applied knowledge that the anthropologist James Scott (1998) argues can only be created through experiences themselves. Yet, they are normally only engaged in AI development at the level of user research studies (if that), and AI deployment when they are being trained on new systems that they must use (Power 2012a). New technologies (including invasive systems like constant surveillance) can be introduced into their workplaces without their freely given agreement. In some jurisdictions, employers are merely required to make workers aware of monitoring, not to ask for their consent; where consent is requested or required, it may be accompanied by the threat of job loss for not signing, undermining any pragmatic or ethical understanding of the concept of consent. With few exceptions (such as Germany's works councils), these opportunities for worker engagement with development or deployment do not constitute meaningful or substantive participation in shaping the direction of AI creation and implementation.

Despite having unrivaled insight into their own roles, low-wage, frontline, and vulnerable workers are rarely treated as experts in their work (Power 2012b; Robinson, Schroeder 2009; Yorke, Bodek 2005). Instead, their employers and the companies that sell AI products to their employers often understand them as costs to be reduced, or "human capital" that could and should be delivering higher levels of productivity.[14] This attitude shapes if, when, and how AI developers choose to engage workers in product creation, limiting their creative influence as well as their veto power over anti-worker product features or use cases. Improving this engagement is particularly important given the knowledge and empathy gap between developers and the vulnerable workers their technology affects. The fast-moving, specialized field of AI creates another assumed constraint on worker participation—developing and deploying companies rarely keep workers or their representatives informed about new and emerging technologies and their potential applications, though managers and executives in the company may be wholly tasked with this responsibility. This process of corporate decision-making, alongside a lack of worker bargaining power, leaves workers multiply disadvantaged in their ability to provide input into the development and deployment process.

14 Seeking ways for workers to be more productive is not in itself in direct conflict with treating workers as a source of expertise on their work. Seeking workers' input and empowering them in AI development and deployment is one possible path to higher productivity. However, when it comes to low wage workers, unfounded beliefs that they do not work hard enough are prevalent. Increased surveillance, control, and punitive measures—which may incorporate race, gender, age, and other biases, and are in direct conflict with worker well-being—are frequently offered as solutions to this perceived problem, at workers' expense (Baldissarri, Andrighetto 2021; Delfanti, Frey 2021; Shih 2002).

First, the dynamic above means workers do not receive information about emerging technologies until decisions are already made and technologies appear in their workplaces, meaning workers lack sufficient information to proactively advocate for the adoption of beneficial AI, the prohibition of harmful AI, or specific conditions for the implementation of AI systems. Second, because they are not understood to be relevant decision-makers or advisers, technology companies rarely create transparency or explainability documentation for frontline workers as an audience, further limiting their ability to participate in an informed way. Third, explainability issues with AI technologies mean implementing managers and leaders may themselves be unable to offer sufficient explanations for workers about how a given technology functions, further circumscribing workers' ability to engage. Finally, the fact that AI deploying companies are frequently purchasers of others' technologies—not developers themselves—adds additional layers of complication in the ability of workers to engage developers and vice versa, requiring the establishment of cross-company and often cross-geography modes of participation. New AI products are often launched or introduced in workplaces without robust, prospective research into their impact on workers, meaning the executives or managers who implement them also lack complete information about the tradeoffs involved. Substantive, informed worker participation in AI development and deployment is valuable for workers as well as their employers and the companies that sell workplace AI products (see Chapter 4 for more detail), but these obstacles must be overcome to obtain it.

3.3. Mechanisms for change

What institutions and mechanisms could empower workers throughout the AI development and deployment process?

How can existing and emerging worker organizations best contribute to promoting the shared prosperity targets?

How can informal workers be empowered to have a say in these processes?

Institutions and mechanisms exist for empowered worker participation (e.g., unions and worker organizations, user research) though they may need additional capability-building or coordination to make them effective in the context of AI development and deployment. Reimagined corporate governance and ownership structures, mechanisms, and principles (e.g., worker co-determination and board seats, cooperative ownership), may also offer promise for worker participation on technology decisions. Workers could also play a role in regulatory agencies and law-making; this is discussed further in Chapter 5.

3.3.1. Unions and worker organizations

Technology does not inherently dilute worker rights or wages, as discussed in the Introduction. Its impacts on workers depend on the type of technology and the conditions of its deployment, which include workers' relative bargaining power. Much as occurred with technological advances in the postwar era, it is possible for AI to generate meaningful value for workers as well as their employers.

Labor unions have a particular interest in ensuring that AI does not further erode workers' rights, as well as that AI augments workers' overall productivity and thereby provides further opportunities for income generation, professional advancement, or a better work-life balance. A central recommendation of the AFL-CIO's Future of Work report is for companies to ensure that workers have a say in technologies before they are deployed, so that they benefit workers, and build upon their expertise rather than replace it (AFL-CIO 2019). This is already practiced in many German companies, where works councils offer institutional mechanisms for soliciting and capturing worker input on technology introductions. In establishments with physically demanding work, works councils appear to encourage higher levels of implementation of digital technologies (Genz et al. 2019).

As mentioned in section 1.2.4, union membership is at a historic low across the OECD, resulting in the erosion of worker power and voice. This erosion has not happened organically, but is partially a result of persistent legislative attempts to dilute the right to collectively bargain and corporate strategies that discourage unionization in some countries. Globalization and the liberalization of trade have also played a role in the decline of unions in some wealthy countries, as companies moved work overseas to decrease labor costs; stronger worker protections and rights throughout the world as well as more liberal immigration policies could act as countervailing forces to this phenomenon. Better protecting and enforcing workers' rights to unionize may require additional government intervention in some countries. This is particularly relevant in light of the fast growth of highly surveilled gig work and its biggest companies' classification of their workforce as independent contractors; this is discussed in more detail in chapter 5. Other important strategies to increase union power, union membership, and union support for worker needs include increasing leadership roles held by individuals from underrepresented groups, and improving communication and outreach to workers who do not currently enjoy the right to collectively bargain.

Unions are not the only way that workers represent themselves, particularly in home-based or subcontracted work in both formal and informal enterprises. There are rich and effective forms of worker-led, non-union organizing throughout the world. For example, worker organizations representing the interests of subcontracted and home-based garment and textile workers in India, China, Brazil and South Africa have led to increased labor protections (Tilly et al. 2013). In the US, Coworker.org offers a slate of digital tools, trainings, and other resources for workers looking to organize on workplace issues inside and outside of the technology industry.

3.3.2. User research and testing of AI products

Inside AI-developing organizations, user research and testing is a crucial part of the development cycle. This research rarely offers opportunities for empowered worker participation (e.g., the ability to eliminate exploitative features, or end development or implementation of products whose outcomes would create negative impacts for workers. Engaging workers earlier in the process (e.g., creating products based on needs and wants they identify) and empowering their participation (e.g., ensuring they can freely offer their perspectives without any threat or perceived threat of retaliation; providing them with sufficient information to engage broadly in the scope of the work, rather than narrowly on aspects like ease of use; and giving them decision-making power or influence) could provide opportunities for AI developers and workers to bridge the empathy gap and steer AI towards shared prosperity.

Some principles of "design justice" developed by the eponymous collaborative network are particularly applicable to thinking through approaches to empowered worker participation in product design and testing. In particular, the list of design justice principles highlights the importance of centering "the voices of those who are directly impacted by the outcomes of the design process," prioritizing "design's impact on the community over the intentions of the designer," and believing that "everyone is an expert based on their own lived experience" (Costanza-Chock 2020, Design Justice Network 2021). Such modifications to the development cycle would also enable AI products to benefit from reverse tutelage and solidaristic forms of working, establishing a dialogue with workers as creative partners co-creating technological future, rather than merely subjecting them to it (Mohamed, Png, Isaac 2020).

3.3.3. Corporate governance and ownership structures

Corporate governance structures also create opportunities for worker participation. In the US—one of a handful of countries dominating AI development—the corporate model is presently anchored on a concept of shareholder primacy, defined by an overarching duty of corporations to deliver a return on investment for shareholders, agnostic of social impacts (provided they are within the bounds of the law). This concept has seen challenges from multiple angles in recent years, from the creation of the B-Corporation structure that explicitly accounts for social impacts in governance mandates; to the rise of investment strategies focused on environmental, social, and governance (ESG) factors; to the growing interest in "stakeholder" capitalism, a corporate governance philosophy that encourages companies to operate in service of customers, suppliers, workers, and communities as well as shareholders. These alternative structures and philosophies explicitly account for the needs of workers, and more widespread adoption could provide incentives for companies to create and adopt AI that encourages shared prosperity.

Other existing corporate models are structured to encourage substantive worker participation. German corporate structures, for instance, have codetermination provisions that enable workers to elect between one third and one half of a company's board seats, depending on the size of the company. In these board roles, non-managerial workers have the opportunity to contribute to major corporate decisions, including large investments and outsourcing decisions. Cooperative ownership structures or labor-managed firms similarly enable workers to influence corporate decision-making, with worker-owners in charge of the company's direction and business decisions, including its technological choices (Scholz 2016). Finally, some companies have launched "shadow boards" as a way to involve workers outside the traditional leadership hierarchy in vetting ideas and providing different perspectives based on their own expertise (Jordan, Sorrell 2019). Participants in shadow boards can be drawn from anywhere in the company—and companies have in fact seen higher levels of success and identified new rising stars—by issuing open calls for interest rather than having managers and leaders handpick members. While shadow boards are non-binding and there has not yet been a documented instance of their use in technology decisions, such an approach would be in line with other previous use cases.

4: The benefits and challenges of adopting shared prosperity targets for private sector stakeholders

An ecosystem of private sector actors holds tremendous influence over the technological trajectory of AI and its applications in workplaces. This group of actors sets corporate R&D agendas, identifies use cases, value propositions, and development processes for converting AI research into products, invests in bringing those products to market, and decides on their implementation in workplaces. This chapter examines the interests and constraints of AI-developing and AI-deploying businesses separately, but it should be noted that some businesses combine both of these functions.

Influential private sector stakeholders include:

- AI-developing business leaders
- AI-deploying business leaders and managers
- AI developers and product managers
- AI investors

Each stakeholder category influences different aspects of AI development and deployment, and wields that influence through different mechanisms. AI-developing businesses serve as the interface between more fundamental AI research and its real world applications (or, put differently, between research supply and market demand). AI-deploying businesses provide some market demand signals for developing businesses through their decisions to adopt new technologies (or not), as well as set the conditions for their use in the workplace. Technically specialized AI developers often possess the clearest insight into what is feasible given the latest research, and their talents are necessary resources to compete in the AI market. Finally, AI investors (e.g., venture capitalists, strategic corporate investors, and large institutional investors) make bets on technologies and firms—building firms' financial, intellectual, and social capital—and in so doing fundamentally alter the odds of success. Each stakeholder's position holds different opportunities, incentives, and challenges in steering AI towards shared prosperity; the rest of this chapter addresses these by stakeholder category.

4.1. AI-developing businesses

What is the business case for creating AI applications that boost labor demand?

What incentives do businesses have to empower worker participation in AI development?

How can they solicit worker needs as demand signals for new AI products?

How can AI-developing companies design commercialization approaches in support of shared prosperity?

AI-developing businesses can shape AI's effect on shared prosperity in three major ways: defining tasks for AI to execute, charting commercialization approaches, and identifying their target customer markets. In a perfectly functioning AI development and deployment flywheel, AI-developing businesses understand the cutting edge of what's technologically possible, and how to match it with business needs of workplace AI deployers. Those deployers ideally provide clear, high-value demand signals for their technological needs, which developing businesses then translate into research demands and later into products, keeping the flywheel spinning.

The present flywheel, unfortunately, neither matches this ideal nor supports shared prosperity. First, as discussed in section 1.2.3, the dominant measure of AI research success is performance at a task compared to a human—a standard that biases research in favor of developing AI that could replace workers in their jobs. Obvious use cases and value propositions for converting research aimed at these goals into commercial products involve worker automation, leaving non-developing, AI-deploying businesses with a set of automation-focused products to choose from. Demand shown by the purchase of these automation technologies may thus be a false signal arising from a constrained option set, not a demonstration of genuine preference if a broader range of AI product types were available. Second, the demand signals from AI-deploying companies—to the extent that they influence the research and development process—emanate from a specific segment of AI-deploying company employees holding particular views of company needs: leadership and management. While it makes sense for AI developers to focus on product needs identified by those with decision rights over corporate funds, an alternative, similarly sensible approach could focus on the needs of those directly executing the work that builds those funds: frontline workers. While frontline workers are sometimes consulted in user testing of workplace technologies (as discussed in more detail in section 3.3), treating the needs and frictions they identify in their jobs as critical demand signals for product development would open a new frontier for AI, channeling research to better complement workers.

Paying closer attention to the experiences and needs of frontline workers and creating AI systems that better complement their existing skills and make work safer would not only improve workers' well-being, but create the potential for companies deploying such systems to increase productivity and become more competitive. Fixation on cost-cutting through workplace automation can in the long term erode the customer base and lead to a disproportionate shrinkage of labor demand: as has been documented by Autor, Dorn and Hanson (2018), Case and Deaton (2020), and Goldstein (2017), disappearance of good jobs triggers a chain of domino effects that devastate entire communities. It is thus in the best interest of AI developing companies and their clients to ensure they are contributing to an expansion, or, at a minimum, non-subtraction of good jobs.

Moreover, the present AI development flywheel fails to explore immense opportunities to profit and expand human flourishing by treating frontline workers as creative partners and collaborators, rather than one-off sources of training data. Tapping workers' expertise and creativity on an ongoing basis does require sharing with them the full context of their participation in creating a new product, properly compensating them for it and giving them a seat at the table and negotiation power. This strategy opens up opportunities for infinite reinvention and improvement of AI systems.[15]

Commercialization approaches of AI B2B companies afford them varying degrees of control over how their products are deployed. Products could be sold to customers who control deployment; accessed via APIs; or wholly owned, controlled, and deployed through consulting services. Target customer base and business model decisions meaningfully affect the ability of AI developing businesses to steer their technologies towards shared prosperity.

15 According to Silicon Valley pioneer Jaron Lanier, treating workers with dignity and valuing their expertise would be no less transformational to software development than the introduction of "Deming Philosophy" was for Toyota in the twentieth century, which connected customer feedback directly with the frontline workers whose creativity was essential in improving the products.

AI developing businesses do face real and perceived challenges in their pursuit of labor-demand creating innovations. Without sufficient public R&D (discussed further in chapter 5), all but the biggest players may lack the research capabilities to create new foundational technologies, or find it difficult to justify the upfront costs needed. This is exacerbated by the fact that creating useful and non-extractive human-complementing technologies is an inherently more difficult research problem than creating technologies to automate atomized human tasks. From a business strategy perspective, moving into white space product domains comes with more uncertainty than competing in established fields with more defined paths to success. This uncertainty manifests in the R&D process, where more work is required to create products for market as well as in market creation and shaping for new and novel technologies. Company leaders and boards of directors concerned with their fiduciary responsibilities may see these moves as too risky.[16] Moreover, investors and public markets value firms viewed as tech companies more highly than firms viewed as service companies, putting pressure on companies to demonstrate that their business model is not highly reliant on human employees.[17]

4.2. AI-deploying businesses

How do AI-deployers business incentives currently influence the AI product market?

What incentives could steer products and their deployment towards shared prosperity?

What role do consumer preferences play in shaping business incentives for pursuing shared prosperity targets (e.g., conscious consumption vs. consumer cost-cutting)?

What incentives do businesses have to empower worker participation in AI deployment?

The current state of the market for workplace AI technologies places many AI-deploying businesses in the position of technology "takers" rather than genuine drivers of the shape of the product market. While AI-developing businesses are necessarily in conversation with their AI-deploying customers (especially if these two activities are taking place within the same company), AI products' paths are mostly set upstream of the deployers. There are, however, ways in which deployers have the potential to more effectively influence the workplace AI product market, and directly control the deployment environment, both of which are critical to steering AI towards shared prosperity.

4.2.1. Product market influence

AI-deploying businesses exert influence on the development and deployment cycle through their choice of AI systems to purchase. Though media coverage of AI advances often focuses on automation (which largely falls under cost-cutting measures), businesses using AI to increase top-line growth are more frequently realizing gains than those using it to cut costs (Balakrishnan et al. 2020). Automation can offer gains for deployers, especially in countries whose tax structure de facto incentivizes headcount reduction (see section 1.2.4). However, these gains tend to be realized in the short-term without adding to long-term company growth. In the case of "so-so" technologies (see section 1.2.1), automation does not even meaningfully contribute to productivity gains: it simply trades out paid labor (which has broad social benefits, as good jobs come with positive externalities) for a combination of unpaid or precarious labor and cost-saving, the economic value of which is most often captured by a very narrow group.

16 A parallel set of challenges exists for AI deploying companies, though in their case the risks and uncertainties are more tied to use case and implementation uncertainty, rather than the research and development itself.

17 Some companies (including many gig platform companies) respond to this pressure by subcontracting human work or asserting that the human workers their model relies on are in fact independent contractors, or belong to other non-employee statuses. National and local governments around the world have had divergent responses to these maneuvers; their impact is discussed in more detail in Chapter 5.

AI-deploying businesses could also influence a new AI product trajectory by involving their workers in shaping AI product demand, converting underexplored potential for technological innovations into concrete value. Workers actually performing tasks are best-versed in what it takes to do the work well (see section 3.2) and thus have clear insight into a set of underexplored, potentially rewarding use cases. As noted in section 4.1, to the extent that AI-developing businesses take demand signals from their customers, those signals tend to be generated by leadership and management, not frontline workers.

These workers presently lack dedicated time from their employers to consider these questions, monetary incentives to compensate them for taking on this high value work, and the respect or standing in many companies to believe that their opinions and ideas will be taken seriously (Power 2012a, 2012b). The International Labour Organization and others promote the approach of social dialogue, in which employers and workers share information and collaborate with each other (and sometimes government actors) to democratically build consensus on workplace issues. The model offers useful guidelines for these interactions as well. Workers and employers both must have independent, empowered representatives, sufficient technical knowledge of the subject at hand, vested interest in the dialogue from all parties; the dialogue must also work to understand differentiated impacts on different populations (e.g., by gender, race, immigration status, and age), and work to protect vulnerable groups (Ishikawa 2003). Though little has been written so far about social dialogue and workplace AI, recent work has suggested it as a promising approach to ensuring shared benefits from technological changes in workplaces (Global Deal 2020, International Labor Office 2018, Vogel 2017).

4.2.2. Deployment environment control

AI-deploying companies generally have a higher degree of control over deployment environments than developing companies. The terminal deployer controls which technological features to implement, the managerial and organizational cultural context in which they are used, targets and metrics for human tasks (with or without AI collaborators), and the implications for human skills required, hours worked, and compensation paid. This constellation of factors (and more) set the conditions relevant to the shared prosperity targets, and demonstrate that making sure that an AI system serves shared prosperity is not a matter of a single decision, but a web of decisions that need to be made in concert with each other.

A body of research exists across industries demonstrating the benefits of workers' job satisfaction to their overall productivity (see, for instance, Harter et al. 2002, Böckerman, P. and Ilmakunnas, P., 2012). Workplace AI intersects with many aspects of worker satisfaction (e.g., autonomy, empowerment, trust, and achievement), and can impact them positively or negatively. It may have unprecedented and unpredicted effects even when compared to prior technological introductions, due to the nature of AI products (e.g., limited explainability or transparency, use in meaningful decisions previously made by humans, deployment as "colleague" as well as "manager"). Research on worker perceptions of AI is nascent; research on workplace AI's effects on worker satisfaction is more nascent still. As AI becomes more widespread in different industries, it will be possible for companies to more thoroughly account for its anticipated effects on worker satisfaction and engagement, as well as to more carefully set the conditions for its incorporation into the workplace.

4.2.3. Deployment risks and opportunities

By prioritizing different financial goals for AI deployment (e.g., long-term vs. short-term gains, revenue growth vs. cost reduction, real economic gains vs. accounting gains), companies can collectively steer AI development towards products more supportive of shared prosperity.

Current narratives of incentives for AI-deploying companies lean in favor of taking advantage of cost-cutting opportunities through automation as they arise, in order to stay competitive and reinvest in other areas of growth. As the scope for automation continues to expand, and the impact is felt directly in communities, becoming an ever-steadier drumbeat in media coverage, companies who choose to automate jobs may face public backlash. The very term "AI" is increasingly associated with the notion of job destruction, which might impact consumer perceptions of AI-deploying companies. While evidence is mixed on the degree to which consumers' personal ethics affect their purchasing decisions, a recent US consumer survey indicated 4 in 10 consumers are willing to walk away from brands that don't align with their values, and nearly two-thirds of consumers surveyed wanted companies to take a stand on cross-cutting social issues like fair employment practices (Barton et al. 2018). Additionally, mature multinational corporations have staked future growth opportunities on the rise of new consumer markets in emerging economies. Introduction of automation technologies into these labor markets has the potential to stall growth and hollow out the anticipated consumer class, adding additional pressure on B2C companies to carefully consider the systemic effects of their automation decisions. The flipside of this effect, however, is that AI which builds real economic value and productivity by working collaboratively with humans could rapidly accelerate growth in these regions.

4.3. AI researchers, developers and product managers

What opportunities do specialized workers like AI developers have to steer AI towards shared prosperity?

What challenges must they overcome to do so?

How do the shared prosperity targets align with common values of specialized Silicon Valley workers, or Valley ideology about technological innovation?

AI developers hold substantial influence as the scarcest "factor" in AI: R&D. Compute resources, while expensive, have a supply function determined by availability of capital. Cutting edge research and development skills, however, cannot be generated so easily, forcing companies to compete for scarce talent. The race for talent is manifested in the high salaries paid to this set of employees, but researchers do not make their employment decisions on compensation alone. They pay attention to companies' values and the impact of their products in the world when choosing between various opportunities in the private sector and academia.

The influence of specialized technical employees on product trajectory does not stop at their employment decisions; they also have the opportunity to exert influence on the design of their research and products, from setting the initial research questions and optimization goals through to deciding on the model design and data sets to use. While shared prosperity is not presently a major subfield of AI ethics, the principles behind it are strongly aligned with foundational AI ethics principles (e.g., the Asilomar principles).

Silicon Valley frequently sees itself as working at the frontier of human progress, and trying to make the world a better place. As recent scrutiny of the tech giants has shown, some ideas of how to accomplish that have unintended consequences, and their creators may wish they had known in advance how their technologies interacted with broader social and economic systems. For developers of this mindset, the shared prosperity targets offer a framework to understand the economic impacts of their work, and guidance for designing products for a more inclusive economy.

Despite their relatively influential positions within tech companies, developers still have numerous constraints on their ability to direct product development. They are often employees, not owners, and thus subject to leadership decisions (even if they are more trusted voices of influence for those leaders). Their skillset may not be well-suited to running economic impact assessments, making company-level commitments all the more important.

4.4. AI investors

What incentives and opportunities do investors have to pursue the shared prosperity targets?

What relevant constraints do they face, and how can those be addressed in the design of the shared prosperity targets?

How do these differ between investor types (e.g., VCs, institutional investors)?

AI investors[18] encounter different opportunities for influence and incentives regarding steering AI in service of shared prosperity depending where along the value chain they invest. Early stage funders (e.g., angel investors, VCs) invest directly in startup AI companies, products, and their researchers, assuming substantial risk in order to receive a higher payout upon success. Corporate investors are also active in the AI startup space (making up 16% of investors in top US AI startups, vs. 9% of investors in top US startups overall) (Kagan et al. 2021). These corporate investors include the investment arms of other technology companies, who can expand their R&D footprint through strategic investment, and whose incentives largely follow those discussed for AI-developing companies in section 4.1.

All of these early-stage investors can influence AI development trajectories by deciding in whom to invest, providing access to networks and contacts, and offering business and industry advice. Opportunities and constraints for this group are closely tied to assessments of risks and returns; identifying new technologies or new markets for them at the right time can mean astronomical returns, but misidentifying the promise of products or markets, missing the timing by being too early or too late converts investments to losses. The dynamics underpinning these risks are similar to those referenced in section 4.1 for AI-developing companies, though investors normally assess and balance them across a portfolio of companies. Investment in this space (and thus overall influence), however, is dwarfed by internal corporate R&D; one estimate holds internal corporate AI R&D investment to be 3-5 times the size of AI startup investment (Bughin et al. 2017).

18 "Investors" is used in this section to refer to external investors, rather than internal company investments. Internal company investments (e.g., corporate investment in their own AI R&D) are discussed in section 4.2

Later in the value chain, large institutional investors are shareholders in both blue chip technology companies with large AI footprints as well as many of the customers of AI products. These large institutional investors also have substantial investments in consumer-facing, publicly traded, multinational companies, whose performance is tied to labor market performance and the existence of disposable income for lower- and middle-class customer segments. For these investors, long-term performance of the overall global economy is more important than the short-term or even long-term performance of a given company; their risk exposure (and potential for upside) is systemic, and their shareholders are seeking returns across decades-long time horizons. This has led some such investors, including BlackRock and the Canada Pension Plan Investment Board to explicitly focus on factors that could contribute to the deterioration of global economic growth over the long-term, including factors frequently associated with ESG investment strategies. This set of institutional investors faces higher exposure to risks introduced by labor market disruptions than early-stage investors do—particularly long-term disruptions that reduce consumer spending and retirement savings. For this group, the potential gains produced by AI-driven automation must be considered in the context of potential negative impacts like the elimination of paths to the global middle class in emerging economies, and mass unemployment and underemployment.

5: The role of government in encouraging AI that enables shared prosperity

As institutions given unique authority by citizens to safeguard their collective well-being, governments have both a stake in the outcomes of shared prosperity and a role to play in bringing them about. Good jobs and a high level of prosperity first and foremost bring well-being for citizens, along with political and social stability, a strong tax base, and lower reliance on government services, among other advantages. Governments also benefit from technological progress, which supports economic growth, the ability to export higher value-added goods and services, and undergirds increased state capability. These incentives offer strong reasons for governments to shape the trajectory of AI towards continued innovation and shared prosperity.

Governments possess unique power, financial resources, and visibility to steer AI development and deployment towards shared prosperity and support vulnerable workers along the way. They can encourage development paths through approaches like funding or conducting certain types of research, setting procurement and contracting standards, joining international associations or agreements, and restructuring taxes, as well as through more directive actions like legislation, regulation, and enforcement litigation. The actions of democratic governments additionally carry a deep legitimacy and have the capacity to set widespread norms. Governments have a mandate to ensure citizens can lead flourishing and meaningful lives. Steering the paths of world-shaping technologies like AI to people's benefit and supporting society through periods of major technological change are natural extensions of this mandate. However, not all governments are benevolent and willing to invest in ensuring that AI supports the prosperity of their citizens; many are actively using AI to suppress dissent, surveil and subjugate populations (Feldstein 2019). But even with the issues of corruption and malevolence set aside, not all governments have the same capacity and levers of influence at their disposal to steer technology or support their workers, and innovators often show greater deference to regulators in certain jurisdictions. Low- and middle-income countries in particular face a multitude of constraints on their ability to influence the trajectory of AI development and manage its impact on their economies.

The remainder of this chapter discusses the policy levers governments have in steering technological change in service of shared prosperity and in supporting workers affected by technological transitions; the chapter concludes with a look at the specific challenges faced by many low- and middle-income countries.

5.1. Steering AI development and deployment towards shared prosperity

Outside of directly encouraging the adoption of shared prosperity targets by AI developing organizations, governments can steer AI towards shared prosperity through industrial policy, funding and conducting worker-complementing AI research, and restructuring taxes to avoid incentivizing excessive automation. Such policy measures are not, however, without potential critics: namely, that such actions are "picking winners and losers," introducing unfairness (through potential capture by special interests) and inefficiency (through potential bureaucratic red tape and a lack of expertise). However, this critique relies on either an assumption about the status quo being neutral, fair and efficient, or an existence of a "natural" direction of technological progress, neither of which are true. Through prior and ongoing actions, governments have already been affecting business choices and the direction of technological progress, which in turn have been shaping our current reality. Government inaction now and in the future is not neutral, but is rather another decision in favor of the status quo and those who benefited from prior decisions.

That said, the above criticisms offer helpful cautions on how government can most effectively steer technological progress in support of shared prosperity (e.g., seek additional expertise for identifying new areas of government investment, partner with the private sector where mutually beneficial), but do not fundamentally challenge the need for governments to take action in this realm.

5.1.1. Government funding of AI R&D

What public investments in basic research could catalyze private sector investments into research and business models focused on complementing human workers instead of automating their tasks?

What requirements should governments set for AI R&D funding to encourage shared prosperity?

How can industrial policy be directed to support shared prosperity?

Governments around the world steer the development of technologies through conducting and funding scientific research and development (R&D), and industrial policy. Government funded research is conducted across government agencies, academia, and the private sector, and covers not only basic research (a public good), but also applied research, product development, and startup financing (Mazzucato 2011). As with private sector funding, government AI research funding has seen rapid expansion in recent years. Canada was the first to release a national AI plan in 2017, with a CAD125 million investment in AI research initiatives. Investments by the US and China are now setting pace for the globe; the US spends the most of any government on R&D broadly, but is likely slightly outspent by China on AI-specific R&D. In 2018 China spent between $1.7 and $5.7 billion on non-defense AI R&D and between $2.0 and $8.4 billion on defense AI R&D (Acharya, Arnold 2019). In 2019, the US spent around $1 billion on non-defense AI R&D and around $4 billion on defense AI R&D (Cornillie 2019). Industry observers anticipate AI defense R&D will create new non-defense capabilities, as has previously occurred with government defense research in nuclear, telecommunications, and GPS technologies.

Governments' participation in fundamental research and sizable funding commitments in AI give them substantial power to shape AI's trajectory. Governments could use this power to encourage shared prosperity-producing AI either by creating a basic research agenda aimed at creating the building blocks for shared prosperity-supporting AI (like technologies for human augmentation or human/machine teaming) and by screening AI funding proposals for the technology's potential impact on shared prosperity. The government should also explore options for how rewards from successful research can be shared with the public, like through equity stakes or citizen dividends (Mazzucato 2011).

Governments can encourage the adoption of shared prosperity-supporting technology through industrial policy, setting provisional goals in collaboration with the private sector actors, iteratively monitoring the progress, exchanging learnings, and revising strategies (Rodrik, Sabel 2019). In February 2021, President Biden signaled that the US will begin implementing a more active industrial policy (for the first time in decades) with a direct impact on AI, since his policy intends to scale up US manufacturing of semiconductors—the hardware AI runs on (Biden 2021).

5.1.2. Tax policy

How could tax codes be revised to rebalance corporate preferences for utilization of labor versus capital?

Which alternative taxes could compensate for potential revenue shortfalls in rebalanced tax structures?

How could tax incentives be used to encourage the adoption of shared prosperity targets by AI developing and deploying organizations?

Most countries presently tax labor more heavily than capital, making labor relatively more expensive than capital when completing the same task and thus incentivizing automation. Several scholars have suggested rebalancing taxation by reducing present tax codes' preference for capital (Acemoglu et al. 2020b), or by introducing an automation tax (Abbott, Bogenschneider 2018).

Governments should also consider the overall revenue impact of AI advancement and, in particular, the impact on labor tax revenues. Taxes on labor make up a significant portion of tax revenue for many countries—for example, taxes on individual laborers made up an average of 23.5% of tax revenue for OECD countries in 2018 (OECD 2020). Frictional job displacement from AI deployment could lead to drastic falls in tax revenue precisely when states may need that money to support unemployed citizens. Potential ways to address shortfalls (if they occur), include the aforementioned automation tax as well as taxes likely to be more resilient to any disruption of labor markets, including on land, carbon emissions, and capital gains/financial transactions. These proposals merit more exploration to determine their effects on innovation and shared prosperity, as well as overall government revenues.

5.2. Supporting workers navigating AI-induced changes

In addition to their role in steering AI progress, governments are a crucial source of support and protection for workers. This Agenda encourages the development of new technologies that create a sufficient amount of new tasks for humans, not mostly displace tasks, permanently lowering overall labor demand. But the task content of jobs and potentially the skills required for them will still be changing. This may create frictional unemployment, necessitating updates or expansions to social protection programs and severance payments requirements to ensure that workers are supported through the transition. Technological progress benefits society writ large, but places new skill demands on specific workers; it should not be their burden alone to upskill or retrain to find or keep work. Governments play a critical role in ensuring workers have access to retraining, and collaborating with employers to ensure that retraining actually provides skills needed for open jobs. As with past technological advances, AI is anticipated to shift the geographic distribution of work, and governments will have a role to play in moving workers to jobs and jobs to workers. The changing nature of work may require updates to labor law in order to protect the dignity, autonomy, and safety of workers, and enshrine positive rights to a share of the productivity gains from AI. As AI becomes ubiquitous in workplaces, many workers will effectively become knowledge workers, creating the data critical for training AI systems and making them useful. New legal rules and frameworks may need to be created to ensure workers are not adversely impacted by the data collection and share in the gains derived from their data.

The above is not an exhaustive list of domains where government action will be needed to adequately support workers as they navigate changes brought about by AI. There is a large and important role for governments to play to help workers gain new skills, find new jobs, transition to AI-augmented work, and face the new legal and societal challenges of a world with greater AI integration.[19] The rest of this section discusses each of these domains in greater detail.

5.2.1. Labor law

What AI-enabled workforce practices need to be regulated due to their job quality impacts?

Are present labor laws sufficient to enforce worker protections from AI harms, or are new laws, enforcement frameworks, or enforcement authorities needed?

What new legislated worker rights and benefits could come from productivity gains (e.g., 4-day work weeks, shorter working days, higher minimum wages)?

What new mechanisms of worker participation in regulation and law-making should be considered?

Artificial intelligence is already present in workplaces, with worker impacts subject to labor law. When a new technology like AI clashes with the goals of labor law—to protect worker safety and dignity—it does so in one of two ways: it causes harms already covered by labor law and meriting enforcement, or it causes harms not currently covered by labor law, revealing a need for new laws. An example of the former would be discrimination in AI-supported hiring decisions, which are already addressed by existing legal frameworks. When harms from AI arise in existing regulatory domains, adjustments may still be needed, like scaling up enforcement to address increased violation frequency, or additional education on identifying harms from novel sources.

By contrast, AI and managerial decisions it enables could also create harms to workers insufficiently regulated by present labor laws. Countries around the world are grappling with worker classification challenges created by gig work platform companies' treatment of their workers as independent contractors, which affect everything from worker compensation and benefits to their ability to unionize (Prassl 2018, Rogers 2016). Widespread and granular worker surveillance reveals another harm, novel if not in kind then by degree; before the rise of big data, companies had no cost-effective way to intensively monitor their workforce. Surveillance reduces workers' sense of autonomy and dignity, and sometimes encourages risky behavior and damages worker safety. It can also be used to prevent unionization and chill legal organizing. The US and other countries lack comprehensive privacy laws and robust worker privacy laws, leaving workers relatively unprotected. On the positive side of AI's impacts, labor law could also be used to secure some of its productivity gains for workers in the form of new positive rights—like the right to shorter work weeks or work days, or higher minimum wages.

Identification of these impacts from AI may require or be assisted by creating new avenues of worker participation in governments' labor departments. Possible options include dedicated technology teams in labor departments tasked with surfacing worker concerns, and technology councils formed by workers.

19 For a comprehensive discussion of how businesses can proactively support workers in transition through company-led initiatives, see "Automation: A Framework for Sustainable Transition" by BSR https://www.bsr.org/reports/BSR_Automation_Sustainable_Jobs_Business_Transition.pdf

5.2.2. Social protection programs

How could traditional unemployment insurance programs be updated to address fluctuations in labor demand from AI?

Does shared prosperity-supporting AI pose novel challenges to unemployment insurance systems?

What alternative systems might policymakers consider to address these challenges, and what are their comparative advantages and disadvantages?

The shared prosperity targets are designed to ensure a sufficient number of high quality jobs for the world's workforce, but the dynamism of technological advances will still likely cause frictional unemployment for some workers as has happened in major technological transformations of the past. Governments may need to update their work-related social protection programs to support their citizens through these short-term disruptions, and respond to the changing nature of work. Most countries currently address job losses through unemployment insurance (UI), which replaces some or all of unemployed workers' incomes. The level of benefits, their conditions, and their duration vary by country. On the generous side of the scale, France pays on average 65% of a worker's previous salary for 2-3 years, and workers need only have worked for four months prior to unemployment to claim benefits (Asenjo, Pignatti 2019). On the other side of the scale, the United States' system pays 38% of a worker's previous income on average and some states end benefits after 14 weeks (Thomson-DeVeaux 2020). Unfortunately, most workers around the globe are not covered by UI at all– for example, workers in informal sector jobs, and many workers in low-income countries. Effective UI systems must also ensure that displaced workers receive the other elements of a necessary dignified existence (like healthcare or retirement funds) when these are linked to jobs, and address lost hours or benefits for those in AI-enabled on-demand labor arrangements, like gig workers. UI will likely be most effective when paired with job search assistance and retraining to help workers find available jobs.

Alternatively, governments seeking to ensure their citizens have sufficient standards of living could move away from UI and towards more unconditional and long-term income support, like universal basic income (UBI). UBI proposes to routinely give all the members of a polity, like a country, an unconditional cash transfer, potentially funded via taxes on increased financial gains from AI. Proposals vary on how much this transfer should be and whether it should replace or complement other benefits. UBI could support workers displaced through frictional unemployment instead of dedicated UI systems. Depending on the amount of the UBI, it could also allow workers a genuine choice on whether they wish to work under the conditions offered by employers, potentially prompting employers to compete for workers through higher compensation and improved job quality.

5.2.3. Worker retraining

What elements contribute to successful retraining programs?

How can retraining programs shift the burden of reskilling away from individual workers?

Retraining is a commonly discussed support for workers in light of AI's current path, but it is an insufficient solution to the problem. The present technological trajectory is towards an ever-shrinking set of job tasks and associated wages for workers, making retraining a functional solution for a similarly shrinking share of workers. The shared prosperity agenda, by contrast, explicitly targets a shift, towards an abundance of well-paid, high quality AI-augmented jobs. Shared prosperity-supporting AI, however, will still change skill requirements for work. Many jobs will see workplace AI create new tasks or new approaches to completing tasks, some jobs will see their tasks wholly displaced by AI, and entirely new jobs will arise—in each instance, workers may need to learn new skills. The benefits of technological progress redound to society writ large, but specific individuals often bear the costs. Accordingly, governments should ensure workers do not bear the burden of reskilling or retraining for new work alone.

Governments have taken myriad approaches to supporting worker retraining. A common thread amongst the most successful is the creation of tight links between job seekers, the skills retraining programs give them, and the needs of employers to fill open jobs—and the way those linkages enable programs to get ahead of future disruptions and demand. Sweden's "job security councils," offer interview prep, therapy, job matching services, and financial support to attend courses to gain missing skills for laid off workers. The councils are funded by employer associations (groups representing private industry), workers pay minimal fees to join the council and insure themselves in the event they are laid-off, and the government offers the councils tax exemptions. The Swedish government offers an additional retraining and job matching program to support newcomers to the job market and the long-term unemployed. As of 2019, 9 out of 10 job-seeking clients found new employment within seven months of first contact with the councils, and nearly 7 in 10 of those jobs were at an equal or higher salary to their prior job (Eurofound 2020).

Another high-performing system, Germany, places a heavy emphasis on vocational and technical education, as well as apprenticeships throughout its education system. When its workers are in need of retraining, the German government unemployment agency provides vouchers for unemployed workers to attend classes and gain new skills, pays companies to directly retrain workers, and offers career advice and financial support for weeks spent retraining and not working.

5.2.4. Aligning workers and jobs within and across borders

What tools can governments use to address unequal labor demand within and between countries created by deployment of AI?

AI will have differentiated impacts on labor markets across geographies based on deployment timelines and their unique economic specializations, affecting the movement of workers and the strength of local economies. To help ameliorate any disruption to workers' livelihoods, governments can encourage workers to move to locations that have unmet labor demand, or attempt to move jobs to communities that need them. The difficulty of addressing this problem depends on its relation to borders. It is vastly easier for workers to move within their own country than it is to move between countries for multiple reasons, including the presence of legal barriers to immigration, language and cultural barriers, difficulties around getting one's credentials and experience recognized in a new country, etc. Relaxing migration restrictions, or introducing rotational labor mobility programs could help workers move to unfilled jobs in other countries, in turn lessening the automation pressure in the receiving geographies (Pritchett 2020). Governments can make it easier for citizens to move within countries through policies like tax breaks or vouchers for moving expenses, job placement services that match for jobs in other areas, standardizing licensing, and ensuring affordable housing and strong public infrastructure in areas with work. AI technology itself can be used for more responsive and efficient job matching and can help handle the excess complexity of looking for jobs across an entire region or country, as opposed to in one locality. Governments can also move work opportunities to areas that need them through incentives for job creation, government-sponsored work, and other place-based interventions. Jobs can be "moved" across national borders with the help of trade agreements.

5.2.5. Data ownership, privacy, and monetization

What data ownership structures could support the aims of shared prosperity?

How do the needs and challenges surrounding data rights for workers differ from those of consumers?

How can governments better protect these rights?

Is the value of worker data adequately captured in current compensation? If not, what compensation models would be best to remedy this?

Workers produce data while doing their jobs, such as video of their movements working an assembly line, written content, labeled images, and audio from customer support calls. In aggregate, this data is valuable, and provides the foundations for machine learning models, including models that could replace the workers themselves. The value and treatment of workers' data raises questions of privacy, ownership, and fair compensation. While there is an ongoing movement to compensate consumers for their data, worker data is a distinct issue with contractual agreements between employers and workers, and expectations of fair compensation for labor. The mass datafication of work creates new questions about intellectual property. For example, the routes a worker walks through a warehouse or the particular driving skills exhibited by an experienced delivery driver reflect the subtle innovations of those workers and could be considered their intellectual property. Yet, through workplace surveillance these data can be claimed by their employers and even turned into automated versions of those same workers. Employment contracts and corresponding laws and regulations used to settle ownership over data and IP have not caught up to the era of AI, in which all workers are knowledge workers, and all work is IP. Governments have the power to limit what provisions can be included in contracts, and are in a position to correct any imbalances that currently exist in how data ownership is negotiated between workers and employers.

Governments can also set the stage for fair data compensation through creating clearer privacy guidelines and empowering workers to control what information is collected about them and how it is used. One possible template comes from the California Consumer Privacy Act, which gives Californians the right to know what data is being collected from them, view and delete this data, know who it is being sold to, and be able to deny its sale. Key elements of any adequate privacy law will include specifying what data can be collected in the first place, and creating requirements for obtaining consent. Privacy not only respects workers' dignity, but also allows for the exercise of critical labor rights like the right to organize without suppression.

ML systems will produce productivity gains for their owners—and not necessarily reward the workers whose data trained them. Governments can create legislation that makes it easier for workers to be compensated for their data. The California Consumer Privacy Act, mentioned above, does so by allowing Californians to select an "authorized agent" who can help wield their newly enshrined privacy rights on their behalf which opens up the door to collective bargaining around data. Monetization of data can then be negotiated by "data unions" or other nonprofits fighting for privacy and autonomy around data. Several such organizations already exist, and reveal one method for ensuring that those who create data—whether workers or consumers—are justly compensated.

5.3. Additional challenges faced by low- and middle-income countries

As described in section 1.2.2 of the Introduction, many of the low- and middle-income economies stand to be impacted in profound ways by the technological trends originating from high-income countries, but have limited control over the direction of AI advancement. While many of the questions discussed earlier in this chapter would still be valuable to consider in the context of the low- and middle-income countries, there are additional layers of complexity associated with these countries' specific challenges.

"Low- and middle-income countries" is a broad category that very different countries can fall under. This section will mostly focus on countries that do not dominate AI development; whose population is young and growing; where learning outcomes are comparatively poor; and whose fiscal position is not strong enough to permit broad coverage of social protection or cash transfer programs, like in a few natural resources-rich countries.

5.3.1. Steering AI progress

What policies around managing technology transfers and cross-border AI deployments would be most effective for allowing low- and middle-income countries governments maximum control over the impacts on local labor markets?

What AI applications could complement the labor of local workers, especially informal and rural workers, and strengthen their income prospects?

Low- and middle-income countries' ability to steer AI in service of shared prosperity using mechanisms described above—public R&D investments and domestic tax incentives—is limited. Public R&D budgets are often comparatively small; AI development predominantly takes place abroad and therefore is less sensitive to domestic tax regimes of low- and middle-income countries. Sensitivity of AI deployment to domestic incentives, though present, also appears to be muted due to marginal costs of deployment being small compared to the initial investment in technology development.[20] Given this context, policies around technology transfers and being selective about which cross-border deployments to welcome and which to discourage become especially important.

5.3.2. Supporting working people

What mechanisms of worker empowerment could serve to give low- and middle-income countries' workers, including informal sector workers, a voice in shaping the direction of AI advancement?

What regulatory steps are necessary to prevent labor rights abuses in outsourcing and supply chain management of AI projects, while encouraging employment creation?

What new labor rights need to be recognized (e.g. around worker privacy, data access and ownership)?

What international "digital tax" regime would be globally equitable?

20 In 2019, McDonalds bought an AI startup Dynamic Yield to help power its self-order kiosks, which it claimed to have deployed in the US in response to the campaign to raise the minimum wage. The kiosks have been deployed in multiple countries beyond the US, including the Philippines and South Africa, where the unemployment at the time of the kiosks roll-out was 29%. At the time of acquisition, McDonalds announced plans to deploy self-order kiosks in 37,000 restaurants across 120+ countries (Rensi 2016).

Given the reality of the learning crisis,[21] digital divide[22] and the fiscal space constraints faced by some low- and middle-income countries, policies around workforce re-skilling and social protection programs expansion are both much needed and extra-difficult to scale there. The feasibility of those policies might further dwindle in the long term if the productive capacities hyper-concentrate in the countries dominating AI development, unless robust cross-border transfer mechanisms are developed either through a globally equitable digital tax regime or other means.

As of today, low-income countries' workforces are less likely to be able to influence AI development and deployment via unionization or other traditional means. AI is poised to impact them indirectly, for example through the loss of competitiveness of their employer. Therefore unionizing might have limited impact on the foreign company whose decisions around AI generate knock-on effects on these workers. Innovative thinking is needed around ways to endow developing countries' workers, especially informal and rural workers, with negotiation power and influence that would help channel the advancement of AI to their benefit.

AI development benefits tremendously from data collected in the developing world, including from workers who labor in online platforms, providing human judgement necessary for labeling training data and validating low-confidence algorithmic outputs. This data generation and data preparation work, critical for the development of AI, if often not recognized as such and is inadequately compensated, if compensated at all. Local governments can play a role in enforcing better working conditions in data labeling and human review services supply chains. There is also a growing need around recognizing new rights of workers needed in the digital work context, like rights around privacy, data access and data ownership.

Simultaneously, millions of informal workers who labor offline do not currently "exist" for the AI systems, forfeiting the opportunity to train models that would be complementary to these workers, making their jobs safer, less physically taxing and more productive.

21 See, for example: https://www.riseprogramme.org/blog/PISA-D_low_learning or https://www.brookings.edu/blog/education-plus-development/2019/05/17/what-have-we-learned-about-the-learning-crisis/

22 As of 2019, global internet penetration was 53.6%, most of the offline population lives in the least developed countries where internet penetration is 19.1%. In Africa it is 28.2%, in Asia/Pacific - 48.4% (ITU Publications 2019).

6: The role of international organizations, educational institutions and civil society in encouraging AI that enables shared prosperity

Steering AI progress in service of shared prosperity is a difficult, if not impossible, task for a single country's government. If a handful of governments get their regulatory environment just right, their efforts can still be undermined by "policy arbitrage", or a spread of "so-so" automation originating from other countries. Moreover, countries do not wield equal power over setting the trajectory of AI: those with the majority of AI talent, R&D, and investment capacities are more influential, while others are often subjected to decisions made outside their borders (see section 1.2.2 of the Introduction).

An analogy with international trade is instructive: when a country imposes or removes trade barriers, it impacts not only its own residents, but residents of countries trading with it. Trade agreements are created to mediate those impacts and discourage social dumping. Similarly, thoughtful agreements about AI are needed to avoid subjecting many economies to runaway consequences of decisions made in countries dominating AI development, as well as to ensure effectiveness of their domestic policy efforts.

International organizations and civil society groups can additionally influence national legislative and regulatory processes, as well as draw upon the framework of international law, treaties, and conventions to induce companies and states to live up to their existing obligations.

6.1. International bodies and agreements

What new international instruments are needed to effectively steer AI in service of expanding access to good jobs globally?

What international bodies are best placed to house agreements on AI? Is there a need for a new body?

Are there cross-border AI deployment or AI supply chain management practices that should be considered social dumping and banned, and are there those that should trigger safeguard clauses?

Economic impacts of AI advancement will be felt in every country, implying that international collaboration will be necessary to govern risks and equitably distribute benefits. While there are currently no binding treaties or international agreements regarding AI, there has been discussion, collaboration, and consideration of AI at an international level. The OECD released five ethical principles for AI in May 2019, with 46 country signatories to date (OECD 2021). The African Union launched a working group on artificial intelligence in 2019. Other institutionalized international collaborations include the Global Partnership on AI, which includes 19 countries and was developed out of the G7 and French and Canadian Presidencies, The Council of Europe's Committee on AI, and the AI for Good Summit run by The International Telecommunication Union (ITU).

The statements of these groups reveal international support for shared prosperity. For example, the first OECD principle reads that "AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being" (OECD 2021). One potential blueprint for more specific, binding agreements could be the General Agreement on Tariffs and Trade (GATT), a binding series of agreements with 164 members. GATT imposes expectations on worker safety, environmental protections, and human rights, among other topics, and punishes those members who violate its terms. While not a perfect analogy, GATT reflects a process for determining rules that countries would abide by and hold each other to, and shows a mechanism—in this case via trade—for enforcement. Alternatively, the Paris Climate Agreement, while non-enforceable, shows how the world can reach consensus and set goals and expectations for each of the 195 country signatories on a major issue of shared concern.

6.2. International human rights instruments

How are Economic, Social, and Cultural Rights (ESCRs) impacted by the advancement of AI?

What should be the role of AI businesses in not only respecting ESCRs but also proactively supporting their enjoyment, realization and fulfillment?

How could consideration of AI's impact on the labor demand and job quality be incorporated into the processes and procedures organizations already follow to uphold the Guiding Principles on Business and Human Rights (UNGPs), or into Human Rights Impact Assessments?

As discussed in Chapters 1 and 2, respect for human rights in the workplace is a basic requirement of job quality. In addition, AI-developing and deploying businesses can contribute substantially to shared prosperity by proactively advancing the enjoyment, realization and fulfillment of Economic, Social, and Cultural Rights. ESCRs include many rights that are highly relevant to shared prosperity, such as the right to social security and social protection; to share in and benefit from scientific advancement. They also include workers rights, such as the right to choose work, to adequate living wages and equal pay for equal work, to leisure and reasonable limitation of working hours, to safe and healthy working conditions, to engage in collective bargaining, to join and form trade unions, and to strike. ESCRs are codified in the Universal Declaration of Human Rights and in the International Covenant on Economic, Social and Cultural Rights, along with several international treaties,[23] regional conventions,[24] and non-binding instruments.

Human rights are interdependent: respect for civil and political rights is directly linked to the pursuit of economic, social and cultural rights. However, the attention of many corporate responsible AI efforts is focused on civil and political rights. There has been vastly more effort, investment and recognition of the responsibility to curb bias and discrimination in AI development and deployment than to prevent job displacement and proliferation of underemployment. Civil and political rights are deeply important—but they must be complemented by ESCRs to ensure AI is developed responsibly. Governments have the primary duty to protect human rights, but businesses also have a responsibility to respect them, which applies to ESCRs as well as civil and political rights. Many ESCRs relate directly to employment, which means employers are naturally at the frontlines of respecting these rights, as well as proactively contributing to their realization. They can create environments where respect for these rights is ubiquitous, and can identify violations more quickly than government enforcers.

23 Relevant UN conventions include Convention on the Elimination of All Forms of Discrimination against Women, Convention on the Rights of Persons with Disabilities, International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families

24 See, for example, the Additional Protocol to the American Convention of Human Rights in the Area of Economic, Social, and Cultural Rights (Protocol of San Salvador), and the European Convention on Human Rights.

Businesses can use human rights guidelines—especially from non-binding instruments which often set more aspirational goals—to inform their practices and to proactively recognize and remedy harms. The primary set of guidelines is the United Nations Guiding Principles on Business and Human Rights, the first global framework for anticipating, preventing and remedying adverse human rights impacts related to all stages of business activity. UNGPs encourage businesses to identify both actual and potential adverse human rights impacts which they may be involved in either directly or indirectly, and provide access to remedy if negative impacts have occurred. The UNGP framework is broadly accepted internationally and has been endorsed unanimously by the UN Human Rights Council. As such, any efforts to evaluate and mitigate the impact of AI development and deployment on labor demand and job quality should be grounded in and coherent with UNGPs. Human rights groundings would be further strengthened by the incorporation of diverse conceptions of personhood and dignity (Mhlambi 2020).

6.3. Academic training

How useful is encouraging the inclusion of modules about AI's labor impacts into AI or AI ethics curricula?

What forms of collaborations between computer science and social sciences/economics departments have been fruitful, and what new ones are needed to prepare young talent to work collaboratively on the economic and labor issues surrounding the advancement of AI?

How could education overall be more explicitly geared towards "uniquely human" skills?

In response to a growing call to increase awareness of ethical issues among AI engineers, AI ethics courses have proliferated. Many of them try to equip computer science students with technical solutions to ethical problems. This approach, as well as the idea that AI scientists and engineers are best positioned to think through these problems has been criticised (see, for example, Raji et al. 2021). Not only might this approach place unrealistic expectations on the CS toolkit and establish an unhelpful hierarchy of disciplines, it could also erase the expertise and experience of other relevant stakeholders, especially affected workers in the case of AI applications geared towards automating economically relevant tasks.

On the other hand, it is not at all uncommon for leading AI scientists and entrepreneurs to advocate for ideas about economic futures and ways to redistribute wealth generated by AI that are not economically sound. Such ideas pose risks because they provide unfounded intellectual cover for the AI field to charge ahead towards an aspired economic future while ignoring the mounting threat of entrenched inequality and disempowerment. Therefore, it is important to resolve whether thinking systematically through economic impacts should be made part of the AI curriculum, or more cross-disciplinary collaborations should be fostered between AI scientists, economists and humanistic social scientists, or both.

Lastly, should the effort to redirect AI to complement humans instead of displacing them in economically relevant tasks be accompanied by educational emphasis on "uniquely human" skills? Do those exist, and if so, what are they?

7: Conclusion

This Agenda proposes to design shared prosperity targets to be taken up by the AI industry either voluntarily or with regulatory encouragement. The targets would constitute commitments by AI companies to non-destruction of good jobs and would help develop AI that supports an expansion of access to good jobs—well-paying, stable, honored and empowered—across the globe. Broad adoption of shared prosperity targets would help stem the proliferation of "so-so automation" —technologies that eliminate jobs but fail to deliver meaningful productivity boost leading to creation of better jobs elsewhere in the economy (Acemoglu, Restrepo 2019).

The Agenda outlines questions that need to be answered in order to construct adequate metrics to be tracked under the shared prosperity targets, create blueprints of processes and disclosure requirements organizations adopting the targets should follow, as well as align the incentives, interests and relative powers of key stakeholders in pursuit of a shared prosperity-advancing path for AI. We invite all interested stakeholders to contribute new and refined questions to this list, as well as collaborate on defining the ways AI developing and deploying organizations should practically act on their responsibility to steer AI in service of shared prosperity.

Bibliography

Abbott, R. and Bogenschneider, B., 2018. Should robots pay taxes: Tax policy in the age of automation. Harv. L. & Pol'y Rev., 12, p.145.

Acemoglu, D. (1998). Why do new technologies complement skills? Directed technical change and wage inequality. Quarterly Journal of Economics, 113(4):10551089.

Acemoglu, D., Manera, A. and Restrepo, P., 2020. Does the US Tax Code Favor Automation? (No. w27052). National Bureau of Economic Research.

Acemoglu, D., Manera, A., Restrepo, P., 2020. Taxes, Automation, and the Future of Labor (Research Brief), MIT Work of the Future. Massachusetts Institute of Technology.

Acemoglu, D. and Restrepo, P., 2018. Automation and new tasks: The implications of the task content of production for labor demand. Journal of Economic Perspectives, 33(2), pp.3-30.

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.

Acemoglu, D. and Restrepo, P., 2020. The wrong kind of AI? Artificial intelligence and the future of labour demand. Cambridge Journal of Regions, Economy and Society, 13(1), pp.25-35.

Acharya, A., Arnold, Z., 2019. Chinese Public AI R&D Spending: Provisional Findings (Issue Brief). Center for Security and Emerging Technology.

AFL-CIO (2019). AFL-CIO Commission on the future of work and unions: Report to the AFL-CIO General Board

Anderson, E., 2017. Private government. Princeton University Press.

Asenjo, A., Pignatti, C., 2019. Unemployment insurance schemes around the world: Evidence and policy options (Working Paper). International Labour Office.

Athreya, B., 2020. Slaves to Technology: Worker control in the surveillance economy. ATR 82–101.

Autor, D., Dorn, D. and Hanson, G., 2018. When work disappears: Manufacturing decline and the falling marriage market value of young men. American Economic Review: Insights, 1(2), pp.161-78.

Autor, D., Mindell, D. and Reynolds, E., 2020. The Work of the Future: Building Better Jobs in an Age of Intelligent Machines. MIT Work of the Future.

Balakrishnan, T., Chui, M., Hall, B., Henke, N., 2020. Global survey: The state of AI in 2020. McKinsey & Company

Baldissarri, C., Andrighetto, L., 2021. Being Treated as an Instrument: Consequences of Instrumental Treatment and Self-Objectification on Task Engagement and Performance. Human Performance 34, 85–106.

Benanav, A., 2020. Automation and the Future of Work. Verso.

Biden, J., 2021. Remarks by President Biden at Signing of an Executive Order on Supply Chains. The White House.

Böckerman, P. and Ilmakunnas, P., 2012. The job satisfaction-productivity nexus: A study using matched survey and register data. ILR Review, 65(2), pp.244-262.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., Trench, M., 2017. Artificial Intelligence: The next digital frontier? (Discussion Paper). McKinsey Global Institute.

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.

Case, A., Deaton, A., 2020. Deaths of Despair and the Future of Capitalism. Princeton University Press.

Cornillie, C., 2019. Finding Artificial Intelligence Money in the Fiscal 2020 Budget | BGOV. Bloomberg Government. URL https://about.bgov.com/news/finding-artificial-intelligence-money-fiscal-2020-budget/ (accessed 4.2.21).

Corser, Maggie, 2019. Technology and Disruption: Workers predictions on the future of retail. Center for Popular Democracy, Fair Workweek Initiative, United for Respect.

Costanza-Chock, S., 2020. Design Justice : Community-Led Practices to Build the Worlds We Need. The MIT Press.

Delfanti, A., Frey, B., 2021. Humanly Extended Automation or the Future of Work Seen through Amazon Patents. Science, Technology, & Human Values 46, 655–682.

Design Justice Network [WWW Document], n.d. . Design Justice Network. URL https://designjustice.org (accessed 4.3.21).

Diao, X., Ellis, M., McMillan, M. and Rodrik, D., 2021. Africa's Manufacturing Puzzle: Evidence from Tanzanian and Ethiopian Firms.

Doshi-Velez, F., Kim, B., 2017. Towards A Rigorous Science of Interpretable Machine Learning. arXiv:1702.08608 [cs, stat].

Eurofound, 2020. Job security councils [WWW Document]. European Monitoring Centre on Change. URL https://www.eurofound.europa.eu/observatories/emcc/erm/support-instrument/job-security-councils (accessed 4.2.21).

Feldstein, S., 2019. The global expansion of AI surveillance (Vol. 17). Washington, DC: Carnegie Endowment for International Peace.

Genz, S., Bellmann, L. & Matthes, B. (2019). Do German Works Councils Counter or Foster the Implementation of Digital Technologies?. Jahrbücher für Nationalökonomie und Statistik, 239(3), 523-564

Global Deal for Decent Work and Inclusive Growth, 2020. Thematic Brief: Social Dialogue and the Future of Work, Together for Decent Work and Inclusive Growth. International Labour Organization.

Goldstein, A., 2017. Janesville: An American Story. Simon and Schuster.

Gray, M.L. and Suri, S., 2019. Ghost work: How to stop Silicon Valley from building a new global underclass. Eamon Dolan Books.

Gutiérrez, G. and Philippon, T., 2020. Some Facts about Dominant Firms (No. w27985). National Bureau of Economic Research.

Harter, J.K., Schmidt, F.L. and Hayes, T.L., 2002. Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. Journal of applied psychology, 87(2), p.268.

International Labour Organization, 2018. Social dialogue and the future of work. Presented at the ILO-AICESIS Conference, International Labour Organization, Athens, Greece, p. 67.

Ishikawa, J., 2003. Key Features of National Social Dialogue: a Social Dialogue Resource. International Labour Office, Geneva.

ITU Publications, 2019. Measuring Digital Development 2019. International Telecommunication Union (ITU).

Jordan, J. and Sorrel, M., 2019. Why you should create a 'shadow board' of younger employees. Harvard Business Review.

Kagan, R., Gelles, R., Arnold, Z., 2021. Corporate Investors in Top U.S. AI Startups. Center for Security and Emerging Technology. https://doi.org/10.51593/20200055

Kageyama, Y., 2017. Reverence for robots: Japanese workers treasure automation [WWW Document]. AP NEWS. URL https://apnews.com/article/c42f629c34c14754adb71d9c244b98a7

Kellogg, K.C., Valentine, M.A., Christin, A., 2020. Algorithms at Work: The New Contested Terrain of Control. ANNALS 14, 366–410. https://doi.org/10.5465/annals.2018.0174

Klinova, K. and Korinek, A., 2021. AI and Shared Prosperity. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES'21), May 19–21, 2021, Virtual Event, USA. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3461702.3462619

Korinek, A. and Stiglitz, J.E., 2019. Artificial intelligence and its implications for income distribution and unemployment (No. w24174). National Bureau of Economic Research.

Korinek, A. and Stiglitz, J.E. 2020, "Steering Technological Progress," working paper

Korinek, A. and Stiglitz, J.E., 2021a. Artificial Intelligence, Globalization, and Strategies for Economic Development (No. w28453). National Bureau of Economic Research.

Korinek, A. and Stiglitz, J.E., 2021b. Covid-19 driven advances in automation and artificial intelligence risk exacerbating economic inequality. bmj, 372.

Lane, M. and Saint-Martin, A., 2021. The impact of Artificial Intelligence on the labour market: What do we know so far?

Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., Robinson, O., 2021. The future of work after COVID-19, The postpandemic economy. McKinsey Global Institute.

Mazzucato, M., 2011. The entrepreneurial state. Soundings, 49(49), pp.131-142.

McNicholas, Shierholz and Poydoc (2021). Union Workers had more job security during the pandemic, but unionizations remain historically low. Economic Policy Institute.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., 2019. A survey on bias and fairness in machine learning. arXiv preprint arXiv:1908.09635.

Mhlambi, S., 2020. From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance. Carr Center for Human Rights Policy Discussion Paper Series, 9.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A. and Lum, K., 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. Annual Review of Statistics and Its Application, 8.

Mohamed, S., Png, M.T. and Isaac, W., 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology, 33(4), pp.659-684.

Moradi, P. and Levy, K., 2020. The Future of Work in the Age of AI: Displacement or Risk-Shifting?

OECD, 2019. OECD Employment Outlook 2019: The Future of Work, OECD Publishing, Paris, https://doi.org/10.1787/9ee00155-en.

OECD, 2020. Revenue Statistics 2020: Tax revenue trends in the OECD. OECD.

OECD, 2021. Recommendation of the Council on Artificial Intelligence [WWW Document]. OECD Legal Instruments. URL https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText

Partnership on AI, 2020. Framework for Promoting Workforce Well-being in the AI-Integrated Workplace. https://www.partnershiponai.org/workforce-wellbeing/

Pew Research Center, 2020. Are AI and job automation good for society? Globally, views are mixed. Pew Research Center.

Pfeffer, J., 2018. Dying for a paycheck: How modern management harms employee health and company performance-and what we can do about it. New York, NY: HarperBusiness, an imprint of HarperCollins Publishers.

Power, B., 2012. Are Your Employees Drivers or Victims of Process Innovations? Harvard Business Review.

Power, B., 2012. Get Your Workers to Disrupt Their Jobs. Harvard Business Review.

Prassl, J., 2018. Humans as a Service: The Promise and Perils of Work in the Gig Economy. Oxford University Press, Oxford, New York.

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.

Raji, I.D., Bender, E., Paullada, A., Denton, E., Hanna, A., 2020. AI and the Everything in the Whole Wide World Benchmark. ML Retrospectives NeurIPS Workshop proceedings.

Raji, I.D., Scheuerman, M.K. and Amironesei, R., 2021, March. You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 515-525).

Rensi, E., 2016. Thanks To "Fight For $15" Minimum Wage, McDonald's Unveils Job-Replacing Self-Service Kiosks Nationwide [WWW Document]. Forbes. URL https://www.forbes.com/sites/realspin/2016/11/29/thanks-to-fight-for-15-minimum-wage-mcdonalds-unveils-job-replacing-self-service-kiosks-nationwide/

Robinson, A.G., Schroeder, D.M., 2009. The Role of Front-Line Ideas in Lean Performance Improvement. Quality Management Journal 16, 27–40.

Rodrik, D., 2018. New technologies, global value chains, and developing economies (No. w25164). National Bureau of Economic Research.

Rodrik, D. and Sabel, C., 2019. Building a good jobs economy.

Rogers, B., 2016. Employment Rights in the Platform Economy: Getting Back to Basics. Harvard Law & Policy Review 10, 479–520.

Scholz, T., 2016. Platform Cooperativism: Challenging the Corporate Sharing Economy. Rosa Luxemburg Stiftung, New York NY USA.

Scott, J.C., 1998. Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed. Yale University Press.

Shih, J., 2002. "...Yeah, I could hire this one, but I know it's gonna be a problem": how race, nativity and gender affect employers' perceptions of the manageability of job seekers. Ethnic and Racial Studies 25, 99–119.

Siddarth, D., Weyl, G., Dick, S., Crawford, K., forthcoming. Don't Let's Talk About AI.

Thomson-DeVeaux, A., 2020. Many Americans Are Getting More Money From Unemployment Than They Were From Their Jobs. FiveThirtyEight.

Tilly, C., Agarwala, R., Mosoetsa, S., Ngai, P., Salas, C., & Sheikh, H. (2013). Informal Worker Organizing as a Strategy for Improving Subcontracted Work in the Textile and Apparel Industries of Brazil, South Africa, India and China. UCLA: Institute for Research on Labor and Employment.

Tirole, J., 2017. Economics for the common good. Princeton University Press.

Vogel, S., 2017. Addressing digital and technological change through social dialogue. Eurofound.

World Bank, 2017. World development report 2018: Learning to realize education's promise. The World Bank.

Yorke, C., Bodek, N., 2005. All You Gotta Do is Ask. PCS Inc.