Guidelines for AI and Shared Prosperity

Tools for improving AI's impact on jobs

Contents

Though this document reflects the inputs of many PAI Partners, it should not be read as representing the views of any particular organization or individual within the AI and Shared Prosperity Initiative's Steering Committee or any specific PAI Partner.

Quick Reference

Signals of Opportunity and Risk

Signals of Opportunity for Shared Prosperity

An opportunity signal (OS) is present if an AI system may:

OS1   Generate significant, widely distributed benefits

OS2   Boost worker productivity

     Caveat 1: Productivity boosts can deepen inequality

     Caveat 2: Productivity boosts can displace workers

     Caveat 3: Productivity boosts can significantly hamper job quality

OS3   Create new paid tasks for workers

     Caveat 1: Someone's unpaid tasks can be someone else's full-time job

     Caveat 2: New tasks often go unacknowledged and unpaid

OS4   Support an egalitarian labor market

OS5   Be appropriate for lower-income geographies

OS6   Broaden access to the labor market

OS7   Boost revenue share of workers and society

OS8   Respond to needs expressed by impacted workers

OS9   Be co-developed with impacted workers

OS10   Improve job quality or satisfaction

     Caveat 1: Systems can improve one aspect of job quality while harming another

     Caveat 2: AI systems are sometimes deployed to redress job quality harms created by other AI systems

Signals of Risk to Shared Prosperity

A risk signal (RS) is present if an AI system may:

RS1   Eliminate a given job's core tasks

RS2   Reallocate tasks to lower-paid or more precarious jobs

RS3   Reallocate tasks to higher- or lower-skilled jobs

RS4   Move jobs away from geographies with few opportunities

RS5   Increase market concentration and barriers to entry

RS6   Rely on poorly treated or compensated outsourced labor

RS7   Use training data collected without consent or compensation

RS8   Predict the lowest wages a worker will accept

RS9   Accelerate task completions without other changes

RS10   Reduce schedule predictability

RS11   Reduce workers' break time

RS12   Increase overall difficulty of tasks

RS13   Enable detailed monitoring of workers

RS14   Reduce worker autonomy

RS15   Reduce mentorship or apprenticeship opportunities

RS16   Reduce worker satisfaction

RS17   Influence employment and pay decisions

RS18   Operate in discriminatory ways

Responsible Practices for Organizations

Responsible Practices for AI-Creating Organizations (RPC)

RPC1 Make a public commitment to identify, disclose, and mitigate the risks of severe labor market impacts presented by AI systems you develop

RPC2 In collaboration with affected workers, perform Job Impact Assessments early and often throughout the AI system lifecycle

RPC3 In collaboration with affected workers, develop mitigation strategies for identified risks

RPC4 Source data enrichment labor responsibly

RPC5 Create and use robust and substantive mechanisms for worker participation in AI system origination, design, and development

RPC6 Build AI systems that align with worker needs and preferences

RPC7 Build AI systems that complement workers (especially those in lower-wage jobs), not ones that act as their substitutes

RPC8 Ensure workplace AI systems are not discriminatory

RPC9 Provide meaningful, comprehensible explanations of the AI system's function and operation to workers using or affected by it

RPC10 Ensure transparency about what worker data is collected, how and why it will be used, and enable opt-out functionality

RPC11 Embed human recourse into decisions or recommendations you offer

RPC12 Apply additional mitigation strategies to sales and use in environments with low worker protection and decision-making power

RPC13 Red team AI systems for potential misuse or abuse

RPC14 Ensure AI systems do not preclude the sharing of productivity gains with workers

RPC15 Request deployers to commit to following PAI's Shared Prosperity Guidelines or similar recommendations

Responsible Practices for AI-Using Organizations (RPU)

RPU1 Make a public commitment to identify, disclose, and mitigate the risks of severe labor market impacts presented by AI systems you use

RPU2 Commit to neutrality towards worker organizing and unionization

RPU3 In collaboration with affected communities, perform Job Impact Assessments early and often throughout AI system implementation and use

RPU4 In collaboration with affected communities, develop mitigation strategies for identified risks

RPU5 Create and use robust and substantive mechanisms for worker agency in identifying needs, selecting AI vendors and systems, and implementing them in the workplace

RPU6 Ensure AI systems are used in environments with high levels of worker protections and decision-making power

RPU7 Source data enrichment labor responsibly

RPU8 Ensure workplace AI systems are not discriminatory

RPU9 Procure AI systems that align with worker needs and preferences

RPU10 Staff and train sufficient internal or contracted expertise to properly vet AI systems and ensure responsible implementation

RPU11 Prefer vendors who commit to following PAI's Shared Prosperity Guidelines or similar recommendations

RPU12 Ensure transparency about what worker data is collected, how it will be used, and why, and enable workers to opt out

RPU13 Provide meaningful, comprehensible explanations of the AI system's function and operation to workers overseeing it, using it, or affected by it

RPU14 Establish human recourse into decisions or recommendations offered, including the creation of transparent, human-decided grievance redress mechanisms

RPU15 Red team AI systems for potential misuse or abuse

RPU16 Recognize extra work created by AI system use and ensure work is acknowledged and compensated

RPU17 Ensure mechanisms are in place to share productivity gains with workers

Get Involved

The Partnership on AI seeks to engage all interested stakeholders to refine, test, and drive the adoption and evolution of all parts of the Shared Prosperity Guidelines, including the Job Impact Assessment Tool, the Responsible Practices, and Suggested Uses. We also seek to curate a library of learnings, use cases and examples, as well as partner with stakeholders to co-create companion resources to help make the Guidelines easier to use for their communities.

We will pursue these goals by means of stakeholder outreach, dedicated workshops, and limited implementation collaborations. If you're interested in engaging with us on this work or want to publicly endorse the Guidelines, please get in touch.

Executive Summary

Our economic future is too important to leave to chance.

AI has the potential to radically disrupt people's economic lives in both positive and negative ways. It remains to be determined which of these we'll see more of. In the best scenario, AI could widely enrich humanity, equitably equipping people with the time, resources, and tools to pursue the goals that matter most to them.

Our current moment serves as a profound opportunity — one that we will miss if we don't act now. To achieve a better future with AI, we must put in the work today. Many societal factors outside the direct control of AI-developing and AI-using organizations will play a role in determining this outcome. However, much still depends on the choices those organizations make, as well as on the actions taken by labor organizations and policymakers.

You can help guide AI's impact on jobs

AI-creating companies, AI-using organizations, policymakers, labor organizations, and workers can all help steer AI so its economic benefits are shared by all. Using Partnership on AI's (PAI) Guidelines for AI & Shared Prosperity, these stakeholders can guide AI development and use towards better outcomes for workers and labor markets.

Included in the Guidelines are:

- a high-level Job Impact Assessment Tool for analyzing an AI system's positive and negative impact on shared prosperity
- a collection of Stakeholder-Specific Recommendations to help minimize the risks and maximize the opportunities to advance shared prosperity with AI

How to use the Guidelines

The Shared Prosperity Guidelines can be used by following a guided, three-step process.

Step 1
Learn about
the Guidelines

Step 2
Apply the Job Impact
Assessment Tool

Step 3
Follow our Stakeholder-
Specific Recommendations

This is the first version of the Guidelines, developed under close guidance from a multidisciplinary AI and Shared Prosperity Initiative's Steering Committee and with direct engagement of frontline workers from around the world experiencing the introduction of AI in their workplaces. The Guidelines are intended to be updated as the AI technology evolves and presents new risks and opportunities, as well as in response to stakeholder feedback and suggestions generated through workshops, testing, and implementation.

Learn About the Guidelines

The Need for the Guidelines

Action is needed to guide AI's impact on jobs

Artificial intelligence is poised to substantially affect the labor market and the nature of work around the globe.

- Some job categories will shrink or disappear entirely and new types of occupations will arise in their place
- Wages will be affected, with AI changing the demand for various skills and the access workers have to jobs
- The tasks workers perform at their jobs will change, with some of their previous work automated and other tasks assisted by new technologies
- Job satisfaction and job quality will shift. Benefits will accrue to the workers with the highest control over how AI shows up in their jobs. Harms will occur for workers with minimal agency over workplace AI deployments

The magnitude and distribution of these effects is not fixed or pre-ordained.[A] Today, we have a profound opportunity to ensure that AI's effects on the labor market and the future of work contribute to broadly shared prosperity.

In the best scenario, humanity could use AI to unlock opportunities to mitigate climate change, make medical treatments more affordable and effective, and usher in a new era of improved living standards and prosperity around the world. This outcome, however, will not be realized by default.[1] It requires a concerted effort to bring it about. AI use poses numerous large-scale economic risks that are likely to materialize given our current path, including:

- Consolidating wealth in the hands of a select few companies and countries
- Reducing wages and undermining worker agency as larger numbers of workers compete for deskilled, lower-wage jobs
- Allocating the most fulfilling tasks in some jobs to algorithms, leaving humans with the remaining drudgery
- Highly disruptive spikes in unemployment or underemployment[B] as workers start at the bottom rung in new fields, even if permanent mass unemployment does not arise in the medium term

> Artificial intelligence is poised to substantially affect the labor market and the nature of work around the globe.

A Example explanations of why technological change is the result of market-shaping policies (and not some "natural" or predetermined trajectory) can be found in:

Redesigning AI: Work, democracy, and justice in the age of automation

Steering technological progress

B We use the definition of underemployment from Merriam-Webster dictionary: "the condition in which people in a labor force are employed at less than full-time or regular jobs or at jobs inadequate with respect to their training or economic needs."

The Guidelines are tools for creating a better future

Partnership on AI's (PAI) Shared Prosperity Guidelines are intended to equip interested stakeholders with the conceptual tools they need to steer AI in service of shared prosperity.

All stakeholders looking to ground their decisions, agendas, and interactions with each other in a systematic understanding of labor market opportunities and risks presented by AI systems can use these tools. This includes:

AI-creating organizations   **AI-using organizations**   **Policymakers**   **Labor organizations and workers**

Origin of the Guidelines

This work comes from years of applied research and multidisciplinary input

A key output of PAI's AI and Shared Prosperity Initiative, PAI's Shared Prosperity Guidelines were developed under the close guidance of a multidisciplinary Steering Committee and draw on insights gained during two years of applied research work. This work included economic modeling of AI's impacts on labor demand,[23] engaging frontline workers around the world to understand AI's impact on job quality,[4] mapping the levers for governing AI's economic trajectory,[5] as well as a major workstream on creating and testing practitioner resources for the responsible sourcing of data enrichment labor. The plan for this multi-stakeholder applied research work was shared with the public in "Redesigning AI for Shared Prosperity: an Agenda" published by Partnership on AI in 2021, following eight months of Steering Committee deliberations.

Though this document reflects the inputs of many PAI Partners, it should not be read as representing the views of any particular organization or individual within the AI and Shared Prosperity Initiative's Steering Committee or any specific PAI Partner.

Design of the Guidelines

We offer two tools for guiding AI's impact on jobs

A high-level Job Impact Assessment Tool with:

- Signals of Opportunity indicating an AI system may support shared prosperity
- Signals of Risk indicating an AI system may harm shared prosperity

A collection of Stakeholder-Specific Recommendations: Responsible Practices and Suggested Uses for stakeholders able to help minimize the risks and maximize the opportunities to advance shared prosperity with AI.
In particular, they are written for:

- AI-creating organizations
- AI-using organizations
- Policymakers
- Labor organizations and workers

These tools can guide choices about any AI system

PAI's Shared Prosperity Guidelines are designed to apply to all AI systems, regardless of:

- Industry (including manufacturing, retail/services, office work, and warehousing and logistics)
- AI technology (including generative AI, autonomous robotics, etc.)
- Use case (including decision-making or assistance, task completion, training, and supervision)

As a whole, the Guidelines are general purpose and applicable across all existing AI technologies and uses, though some sections may only apply to specific technologies or uses.

To apply these guidelines, stakeholders should:

- For an AI system of interest, perform the analysis suggested in the Job Impact Assessment section, identifying which signals of opportunity and risk to shared prosperity are present.
- Use the results of the Job Impact Assessment to inform your plans, choices, and actions related to the AI system in question, following our Stakeholder-Specific Recommendations. For AI-creating and AI-using organizations, these recommendations are Responsible Practices. For policymakers, unions, workers, and their advocates, these recommendations are Suggested Uses.

We look forward to testing the Guidelines and refining the use scenarios together with interested stakeholders. If you have suggestions or would like to contribute to this work, please get in touch.

> PAI's Shared Prosperity Guidelines are designed to apply to all AI systems, regardless of industry, AI technology, or use case.

Our approach focuses on AI's impact on labor demand

In these Guidelines, we consider an AI system to be serving to advance the prosperity of a given group if it boosts the demand for labor of that group — since selling labor remains the primary source of income for the majority of people in the world.

We recognize that some communities advocate to advance shared prosperity in the age of AI through benefits redistribution mechanisms such as universal basic income. While a global benefits redistribution mechanism might be an important part of the solution (especially in the longer term) and we welcome research efforts and public debate on this topic, we left it outside of the scope of the current version of the Guidelines.

Instead, the Guidelines focus on governing the impact of AI on labor demand. We believe this approach will be extremely necessary at least in the short to medium term, enabling communities to have effective levers of influence over the pace, depth, and distribution of AI impacts on labor demand.

AI's impacts on labor demand can manifest themselves as:

- Changes in availability of jobs for certain skill, demographic, or geographic groups[c]
- Changes in the quality of jobs affecting workers' well-being[D]

In line with PAI's framework for promoting workforce well-being in the AI-integrated workplace and other leading resources on high-quality jobs,[678] we recognize multiple dimensions of job quality or workers' well-being, namely:

- Human rights
- Financial well-being
- Physical well-being
- Emotional well-being
- Intellectual well-being
- Sense of meaning, community, and purpose.

Thus, for the purposes of these Guidelines, **we define AI's impact on shared prosperity as the impact of AI use on availability and quality of formal sector jobs across skill, demographic, or geographic groups.**[E]

In turn, the overall impact of AI on the availability and quality of jobs can be anticipated as a sum total of changes in the **primary factors** AI use is known to affect.[91011] Those factors are:

We define AI's impact on shared prosperity as the impact of AI use on availability and quality of formal sector jobs across skill, demographic, or geographic groups.

C Groups' boundaries can be defined geographically, demographically, by skill type, or another parameter of interest.

D In other words, AI's impact on labor demand can affect both incumbent workers as well as people interested in looking for work in the present or future.

E The share of informal sector employment remains high in many low- and middle-income countries. The emphasis on formal sector jobs here should not be interpreted as treating the informal sector as out of scope of the concern of PAI's Shared Prosperity Guidelines. The opposite is the case: If the introduction of an AI system in the economy results in a reduction of availability of formal sector jobs, that reduction cannot be considered to be compensated by growth in availability of jobs in the informal sector.

- Relative productivity of workers (versus machines or workers in other skill groups)
- Labor's share of organization revenue[F]
- Task composition of jobs
- Skill requirements of jobs
- Geographic distribution of the demand for labor[G]
- Geographic distribution of the supply of labor
- Market concentration
- Job stability
- Stress rates
- Injury rates
- Schedule predictability
- Break time
- Job intensity
- Freedom to organize
- Privacy
- Fair and equitable treatment
- Social relationships
- Job autonomy
- Challenge level of tasks
- Satisfaction or pride in one's work
- Ability to develop skills needed for one's career
- Human involvement or recourse for managerial decisions (such as performance evaluation and promotion)
- Human involvement or recourse in employment decisions (such as hiring and termination)

Anticipated effects on the above primary factors are the main focus of the risks and opportunities analysis tool provided in the Guidelines. Another important focus is the **distribution of those effects**. An AI system may bring benefits to one set of users and harms to another. Take, for example, an AI system used by managers to set and monitor performance targets for their reports. This system could potentially increase pride in one's work for managers and raise rates of injury and stress for their direct reports.

When this dynamic prompts conflicting interests, we suggest higher consideration for the more vulnerable group with the least decision-making power in the situation as these groups often bear the brunt of technological harms.[12] By a similar logic, where we call for worker agency and participation, we suggest undertaking particular effort to include the workers most affected and/or with the least decision authority (for example, the frontline workers, not just their supervisors).

Key Principles for Using the Guidelines

These application principles apply independently of who is using the Guidelines and in what specific scenario they are doing so.

Engage affected workers

Make sure to engage worker communities that stand to be affected by the introduction of an AI system in the Job Impact Assessment, as well as in the development of risk mitigation strategies. This includes, but is not limited to, engaging and affording agency to workers who will be affected by the AI system and their representatives.[H] Bringing in multi-disciplinary experts will help understand the full spectrum and severity of the potential impact.

Workers may work with AI systems or have their work affected by them. In cases where one group of workers uses an AI system (for instance, uses an AI performance evaluation tool to assess their direct reports) and another group is affected by that AI system's use (in this example, the direct reports), we suggest giving highest consideration to affected workers and/or the workers with the least decision-making power in the situation (in this example, the direct reports rather than the supervisors).

Seeking shared prosperity doesn't mean opposing profits

Some of the signals of risk to shared prosperity described in the Guidelines are actively sought by companies as profit-making opportunities. **The Guidelines do not suggest that companies should stop seeking profits, just that they should do so responsibly.**

Profit-generating activities do not necessarily have to harm workers and communities, but some of them do. The presence of signals of risk indicate that an AI system being assessed, while possibly capable of generating profit for a narrow set of beneficiaries, is likely to do that at the expense of shared prosperity, and thus might be undesirable from the societal benefit perspective. We encourage companies to follow the Guidelines, developing and using AI in ways that generate profit while also advancing shared prosperity.

Signals are indicators, not guarantees

Presence of a signal should be interpreted as an early indicator, not a guarantee that shared prosperity will be advanced or harmed by a given AI system. Presence of opportunity or risk signals for an AI system being assessed is a necessary, but not sufficient, condition for shared prosperity to be advanced or harmed with the introduction of that AI system into the economy.

Many societal factors outside of the direct control of AI-creating organizations play a role in determining which opportunities or risks end up being realized. Holding all other societal factors constant, the purpose of these Guidelines is to minimize the chance that shared prosperity-relevant outcomes are worsened and maximize the chance that they are improved as a result of choices by AI-creating and -using organizations and the inherent qualities of their technology.

Signals should be considered comprehensively

Signals of opportunity and risk should be considered comprehensively. Presence of a signal of risk does not automatically mean an AI system in question should not be developed or deployed. That said, an absence of any signals of opportunity does mean that a given AI system is highly unlikely to advance shared prosperity and whatever risks it might be presenting to society are not justified.

The Guidelines do not suggest for companies to stop seeking to make profit, but merely to do it responsibly.

Signals of opportunity do not "offset" signals of risk

Presence of signals of opportunity should not be interpreted as "offsetting" the presence of signals of risk. In recognition that benefits and harms are usually borne unevenly by different groups, the Guidelines strongly oppose the concept of a "net benefit" to shared prosperity, which is incompatible with a human rights-based approach. In alignment with the UN Guiding Principles on Business and Human Rights, a mitigation strategy should be developed for each risk identified, prioritizing the risks of the most severe impacts[I] first. Mitigation strategies can range from eliminating the risk or reducing the severity of potential impact to ensuring access to remedy or compensation for affected groups. If effective mitigation strategies for a given risk are not available, it should be considered as a strong argument in favor of meaningful changes in the development, implementation, and use plans of an AI system, especially if it is expected to affect vulnerable groups.

Analysis of signals is not prescriptive

The analysis of signals of opportunity and risk is not prescriptive. Decisions around the development, implementation, and use of increasingly powerful AI systems should be made collectively, allowing for the participation of all affected stakeholders. We anticipate that two main uses of the signals analysis will include:

Informing stakeholders' positions in preparation for dialogue around development, deployment, and regulation of AI systems, as well as appropriate risk mitigation strategies

Identifying key areas of potential impact of a given AI system which warrant deeper analysis (such as to illuminate their magnitude and distribution)[13] and further action

I PAI's Shared Prosperity Guidelines use UNGP's definition of severity: an impact (potential or actual) can be severe "by virtue of one or more of the following characteristics: its scale, scope or irremediability. Scale means the gravity of the impact on the human right(s). Scope means the number of individuals that are or could be affected. Irremediability means the ease or otherwise with which those impacted could be restored to their prior enjoyment of the right(s)."

Apply the Job Impact Assessment Tool

Use the high-level Job Impact Assessment Tool to analyze a given AI system:

Go over the full list of signals of opportunity and risk

Analyze the distribution of potential benefits and harms

Repeat this process for upstream and downstream markets

Instructions for Performing a Job Impact Assessment

Assess the AI system against the full list of signals

Go over **the full list** of signals of opportunity and risk and document which signals are present in the case of the AI system being assessed. Not all signals apply for every AI system. Document those that do not apply as not applicable, but do not skip or cherry-pick signals. For each step, document the explanation for the answer for future reference.

For each signal, if you estimated the likelihood of the respective opportunity or risk materializing as a result of the introduction of the AI system into the economy to be anything but "zero," please note the respective signal as **"present."**

Certainty in likelihood estimation is not a prerequisite for this high-level assessment and is assumed to be absent in most cases. When in doubt, note the signal as "present."

Analyze the distribution of potential benefits and harms

Document in as much detail as possible your understanding of the distribution of potential benefits and harms of an AI system across skill, geographic, and demographic groups, and how it might change over time.[J] (Are today's "winners" expected to lose their gains in the future? The reverse?) The exact steps needed to perform the distribution of impacts analysis are highly case-specific. PAI is looking to engage with stakeholders to curate a library of distribution analysis examples for the community to learn from. If you would like to contribute to this, please get in touch.

[J] Relevant time period depends on how long the AI system being assessed is expected to remain in use.

Repeat this process for upstream and downstream markets

In order to take into account the possible effects on the competitors, suppliers, and clients of the AI-using organization, repeat the signal detection and analysis processes not only

for the primary market the AI system is intended to be deployed in, but also upstream and downstream markets.

Proceed to our Stakeholder-Specific Recommendations

After completing the high-level Job Impact Assessment analysis, AI-creating and AI-using organizations should implement recommended Responsible Practices (where not already in use) to improve anticipated outcomes — for instance, to eliminate or mitigate anticipated harms or increase likely benefits for workers and the economy. These Responsible Practices can be found under Step 3 of the Shared Prosperity Guidelines. (Responsible Practices will be added and refined through community testing and feedback.)

Policymakers, workers and their representatives can use the results of the high-level Jobs Impact Assessment to inform their decisions, actions, and agendas as outlined in the Suggested Uses section under Step 3 of the Shared Prosperity Guidelines. We look forward to collecting feedback on the Guidelines and curating use examples in partnership with interested stakeholders. To get involved, please get in touch.

Signals of Opportunity for Shared Prosperity

If one or more of the statements below apply to the AI system being assessed, this indicates a possibility of a positive impact on shared prosperity-relevant outcomes.

An opportunity signal (OS) is present if an AI system may:

OS1. Generate significant, widely distributed benefits

Will the AI system generate significant, widely distributed benefits to the planet, the public, or individual consumers? One of the primary motivations for investing in the research and development of AI is its potential to help humanity overcome some of our most pressing challenges, including ones related to climate change and the treatment of disease. Hence, the potential of an AI system to generate public goods or benefit the environment is a strong signal of opportunity to advance shared prosperity.

Individual consumer benefits can be more controversial as many advocates point out the growing environmental costs that frequently accompany the commodification of consumer goods. But if production and consumption are environmentally conscious, a potential to generate significant and widely distributed consumer benefits is a signal of opportunity to advance shared prosperity. Cheaper or more high-quality goods or services make consumers richer in real terms,[K] freeing up parts of their incomes to be spent to buy other goods and services, boosting the demand for labor in respective sectors of the economy.

How significant and widely distributed consumer benefits should be to justify job losses

K This is a result of the "real income effect." For the same nominal amount of money, consumers are able to buy more or higher quality goods.

is a political question,[L] but quantifying consumer gains per job lost would help sharpen up any debate about the value of an AI innovation.[M] As stated in "Key Principles for Using the Guidelines," independently of the magnitude and distribution of anticipated benefits, appropriate mitigation strategies should be developed in response to the risk of job losses or wage decreases.

OS2. Boost worker productivity

Will the AI system boost productivity of workers, in particular those in lower-paid jobs, without increasing strain? By a worker's productivity, we mean a worker's output per hour. A more productive worker is more valuable to their employer and (all other conditions remaining the same) is expected to be paid more.[N] Therefore, if an AI system comes with a promise of a productivity boost that is a positive signal. Besides, productivity growth is often the prerequisite for the creation of consumer benefits discussed in OS1.

However, three important caveats should be noted here.

Caveat 1: Productivity boosts can deepen inequality
It is quite rare for a technology to equally boost productivity for everyone involved in the production of a certain good, more often it helps workers in certain skill groups more than others. If it is helping workers in lower-paying jobs relatively more, the effect could be inequality-reducing. Otherwise, it may be inequality-deepening. Please document the distribution of the productivity increase across the labor force when assessing the presence of this opportunity signal.

Caveat 2: Productivity boosts can displace workers
Even if productivity of all workers involved in the production of a certain good is boosted equally by an AI system, fewer of them might find themselves employed in the production of that good once the AI system is in place. This is because fewer (newly more productive) worker-hours[O] are now needed to create the same volume of output. For production of the good in question to require more human labor after AI deployment, **two conditions must be met:**

- Productivity gains of the firm introducing AI need to be shared with its clients (such as consumers, businesses, or governments) in the form of lower-priced or higher-quality products — something which is less likely to happen in a monopolistic environment

- Clients should be willing to buy sufficiently more of that lower-priced or higher-quality product

If the first condition is met but the second is not, the introduction of the AI system in question might still be, on balance, labor-demand boosting if it induces a "productivity effect" in the broader economy. When productivity gains and corresponding consumer benefits are sufficiently large, consumers will experience a real income boost generating new labor demand in the production of complementary goods. That new labor demand might be sufficient to compensate for the original loss of employment due to an introduction of an AI system. Issues arise when the productivity gains are too small like in the case of "so-so" technologies[14] or are not shared with consumers. If that is the case, please document OS2 as **"not present"** when performing the Job Impact Assessment.

L For example, in 2011, the US government imposed tariffs to prevent job losses in the tire industry. Economic analysis later showed that the tariffs cost American consumers around $0.9 million per job saved. It seems implausible that such large consumer costs are worthwhile, relative to the job gains.

M In this paper, Brynjolfsson et al. estimate the value of many free digital goods and services. They do so by proposing a new metric called GDP-B, which quantifies their benefits rather than costs, and then estimating consumers' willingness-to-pay for free digital goods and services in terms of GDP-B.

N As emphasized in Key Principles for Using the Guidelines, signals of opportunity are not guarantees: It is possible that the introduction of a new technology into the workplace boosts workers' productivity but does not lead to wage growth because, in practice, workers' productivity is only one of the factors determining their wage. Other factors include how competitive the market is and how much bargaining power workers have. In fact, a large number of countries have been experiencing productivity-wage decoupling in recent decades. This points to a diminishing role of productivity in determining wages, but it remains non-zero and hence has to be accounted for by the Guidelines.

O The impact of a productivity-enhancing technology can manifest itself as a reduction of the size of the workforce, or a reduction in hours worked by the same-size labor force. Either option can negatively impact shared prosperity.

Caveat 3: Productivity boosts can significantly hamper job quality
Introduction of an AI system can lead to productivity enhancement through various routes: by allowing workers to produce more output per hour of work at the same level of effort or by allowing management to induce a higher level of effort from workers. If productivity boosts are expected to be achieved solely or mainly through increasing work intensity, please document OS2 as "not present" when performing the Job Impact Assessment.

Lastly, frontline workers[15] reported appreciation for AI systems that boosted their productivity by assisting them with core tasks. Conversely, technologies that boosted productivity by automating workers' core tasks were associated with a reduction in job satisfaction.[16] Hence, pursuit of productivity increases through technologies that eliminate non-core tasks is preferred over paths that involve eliminating core tasks. Examples of technologies that assist workers on their core tasks include:

- Training and coaching tools
- Algorithmic decision support systems that give users additional information, analytics, or recommendations without prescribing or requiring decisions

OS3. Create new paid tasks for workers

Will the AI system create new tasks for humans or move unpaid tasks into paid work? Technological innovations have a great potential for benefit when they create new formal sector jobs, tasks, or markets that did not exist before. Consider, for example, the rise of social media influencers and content creators. These types of jobs were not possible before the rise of contemporary media and recommendation technologies. It has been estimated that, in 2018, more than 60 percent of employees were employed in occupations that did not exist in 1940.[17]

> **Technological innovations have a great potential for benefit when they create new formal sector jobs, tasks, or markets that did not exist before.**

Caveat 1: Someone's unpaid tasks can be someone else's full-time job
It is important to keep in mind that technologies seemingly moving unpaid tasks into paid ones might, upon closer inspection, be producing an unintended (or deliberately unadvertised) effect of shifting tasks between paid jobs — often accompanied by a job quality downgrade. For example, a technology that allows people to hire someone to do their grocery shopping might convert their unpaid task into someone else's paid one, but also reduce the demand for full-time domestic help workers, increasing precarity in the labor market.

Caveat 2: New tasks often go unacknowledged and unpaid
Sometimes the introduction of an AI system adds unacknowledged and uncompensated tasks to the scope of workers. For example, the labor of smoothing the effects of machine malfunction remains under the radar in many contexts,[18] creating significant unacknowledged burdens on workers who end up responsible for correcting machine's errors (without being adequately positioned to do that).[19]

When performing the Job Impact Assessment, please explicitly document the applicability of these two caveats associated with OS3 for the AI system being assessed and its deployment context.

OS4. Support an egalitarian labor market

Will the AI system support a more egalitarian labor market structure? A superstar labor market structure is a situation where a relatively small number of workers dominate the market or satisfy most of the labor demand that exists in it. The opposite is an "egalitarian" labor structure where each worker's output is small relative to the output of all other

workers in the industry. The key factor that makes a labor market's structure egalitarian is the presence of a need to invest an additional unit of worker time to serve an additional consumer. For example, the rise of the music recording industry has made its labor market structure less egalitarian for musicians. Today, to satisfy the demand for music from an additional customer, musicians do not need to physically get in front of them or do any additional work.

OS5. Be appropriate for lower-income geographies

Will the AI system be appropriate for lower-income geographies? Capital and labor of various skill types can be relatively more or less abundant in different countries. Technologies that take advantage of the factor of production (capital or labor of a certain skill type) that is relatively more abundant in a given country and do not require much of a factor that is relatively scarce there are deemed appropriate for that country.

Generally, capital is relatively more abundant in the higher-income countries while labor is relatively more abundant in the lower-income countries, many of which also struggle with poor learning outcomes limiting the training the workforce receives.[20] Therefore, capital-intensive labor-saving AI systems are generally inappropriate for lower-income countries whose main comparative advantage is relatively abundant labor.[21] Such technologies being adopted by high-income countries can hurt economic outcomes in lower-income countries because competitive forces in the export industries force the latter to adopt those technologies to remain competitive.[22,23]

Consequently, lower-income countries would greatly benefit from access to technologies that would allow them to stay competitive by leveraging their abundant labor resources and creating gainful jobs that do not require high levels of educational attainment.

When assessing the presence of this signal, please also document if and how the relative abundance of capital and labor of various skill types is expected to change over time.

> Capital-intensive labor-saving AI systems are generally inappropriate for lower-income countries.

OS6. Broaden access to the labor market

Will the AI system broaden access to the labor market? AI systems that allow communities with limited or no access to formal employment to get access to gainful formal sector jobs are highly desirable from the perspective of broadly shared prosperity. Examples include AI systems that:

- Assist the disabled
- Make it easier to combine work and caregiving responsibilities
- Enable work in languages the worker does not have a fluent command of

OS7. Boost revenue share of workers and society

Will the AI system boost workers' and society's share of an organization's revenue?

Workers' share of revenue is the percentage of an organization's revenue spent on workers' wages and benefits. For the purposes of these Guidelines, we suggest excluding C-suite compensation when calculating workers' share.

If, following the introduction of an AI system, workers' share of organization's revenue is expected to grow or at least stay constant, it is a very strong signal that the AI system in question will serve to advance shared prosperity. The opposite is also true. If, following the introduction of an AI system, workers' share of organization's revenue is expected to shrink, it is a very strong signal that the AI system in question will harm shared prosperity.

Please note that worker benefits are included in workers' share of an organization's revenue. For example, consider an organization that adopts a productivity-enhancing AI system which allows it to produce the same or greater amount of output with fewer hours of work needed from human workers. That organization can decide to retain the same size of the workforce and share productivity gains with it (for example, in the form of higher wages, longer paid time off, or shorter work week at constant weekly pay), keeping the workers' share of revenue constant or growing. That would be a prime example of using AI to advance shared prosperity.

Lastly, if an organization was able to generate windfall gains from AI development or usage and is committed to sharing the gains not only with workers it directly employs but the rest of the world's population as well, that can be a great example of using AI to advance shared prosperity. While some have proposed this,[24] more research is needed to design mechanisms for making sure windfall gains are distributed equitably and organizations can be expected to reliably honor their commitment to distribute their gains.

OS8. Respond to needs expressed by impacted workers

Did workers who will use the AI system or be affected by it (or their representatives) identify the need for the system? AI systems created from a worker's idea or identified need build in workers' job expertise and preferences from the outset, making it more likely the AI systems will be beneficial or useful to workers affected by them and welcomed as such. Much of the current AI development pipeline starts with advances in research and development, only later identifying potential applications and product-market fit. The market for workplace AI technology is largely composed of senior executives and managers, creating a potential misalignment between needs perceived by budget holders and managers and the needs perceived by the workers who use or are most affected by the technology. AI systems emerging from the ideas and needs of workers who use or are most affected by them (or their representatives, who represent the collective voice of a given set of workers, not just the perspective of an individual worker) reduce this potential for misalignment.[25]

OS9. Be co-developed with impacted workers

Were workers who will ultimately use or be affected by the AI system (or their representatives) included and given agency in every stage of the system's development? Workers are subject matter experts in their own tasks and roles, and can illuminate opportunities and challenges for new technologies that are unlikely to be seen by those with less familiarity with the specifics of the work. The wisdom of workers who use or are most affected by AI systems introduced throughout development can smooth many rough edges that other contributors might only discover after systems are in the market and implemented. Where relevant worker representatives exist, they should be brought into the development process to represent collective worker interests from start to finish.

Fully offering affected workers agency in the development process requires taking the time to understand their vantage points, and equip them or their representatives with enough knowledge about the proposed technology to meaningfully participate. They also must be afforded sufficient decision-making power to steer projects and, if necessary, end them in instances where unacceptable harms cannot be removed or mitigated. This also necessitates protecting their ability to offer suggestions freely without fear of repercussions. Without taking these steps, participatory processes can still lead to suboptimal outcomes — and possibly create additional harms through covering problems with a veneer of worker credibility.

> Workers are subject matter experts in their own tasks and roles, and can illuminate opportunities and challenges for new technologies.

OS10. Improve job quality or satisfaction

Was the AI system intended to improve job quality or increase job satisfaction? AI technology has the potential to improve many aspects of job quality and job satisfaction, from increasing occupational safety to providing personalized coaching that leads to career advancement. This requires taking job quality, worker needs, and worker satisfaction seriously.

Two important caveats are required for this signal.

Caveat 1: Systems can improve one aspect of job quality while harming another
For example, many AI technologies positioned as safety enhancements are in reality invasive surveillance technologies. Though safety improvements may occur, harms to human rights, stress rates, privacy, job autonomy, job intensity, and other aspects of job quality may occur as well. Other AI systems purport to improve job quality by automating tasks workers dislike (see RS1 for more detail on the risks of task elimination).

When a system enhances one aspect of job quality while endangering another, this signal can still be counted as "present," but the need to consider the rest of the opportunity and risk signals is particularly important.

Caveat 2: AI systems are sometimes deployed to redress job quality harms created by other AI systems
For example, some companies have introduced AI safety technologies to correct harms resulting from the prior introduction of an AI performance target-setting system that encouraged dangerous overwork.[26]

When this is the case, the introduction of the new AI system to redress the harms of the old does not count for this signal and should be marked as "not present."

Instead of introducing new AI systems with their own attendant risks, the harms from the existing systems should be addressed in line with the Responsible Practices provided by the Guidelines for AI-using organizations and additional case-specific mitigations.

⚠️ Signals of Risk to Shared Prosperity

If one or more of the statements below apply to the AI system being assessed, this indicates a possibility of a negative impact on shared prosperity-relevant outcomes.

Some of the signals of risk to shared prosperity described in the Guidelines are actively sought by companies as profit-making opportunities. **The Guidelines DO NOT suggest that companies should stop seeking profits, just that they should do so responsibly.**

Profit-generating activities do not necessarily have to harm workers and communities, but some of them do. The presence of signals of risk indicate that an AI system being assessed, while possibly capable of generating profit for a narrow set of beneficiaries, is likely to do that at the expense of shared prosperity, and thus might be undesirable from the societal benefit perspective. We encourage companies to follow the Guidelines, developing and using AI in ways that generate profit while also advancing shared prosperity.

For-profit companies might feel pressure from investors to cut their labor costs no matter the societal price. We encourage investors and governments to join civil society in an effort to incentivize responsible business behavior with regards to shared prosperity and labor market impact.

Some practices or outcomes included in this section are illegal in some jurisdictions, and as such are already addressed in those locations. We include them here due to their legality in other jurisdictions.

A risk signal (RS) is present if an AI system may:

RS1. Eliminate a given job's core tasks

Will the AI system eliminate a significant share of tasks for a given job? A lot of technological innovations eliminate some job tasks that were previously done by human workers. That is not necessarily an unwelcome development, especially when those technologies also create new paid tasks for humans (see OS3), boost job quality (see OS10), or bring significant broadly distributed benefits (see OS1). For example, it can be highly desirable to automate tasks posing unmitigable risks to workers' physical or mental health. Primary research conducted by the AI and Shared Prosperity Initiative indicated that frontline workers often experience automation of their non-core tasks as helpful and productivity-boosting.[27]

> Profit-generating activities do not necessarily have to harm workers and communities, but some of them do.

However, if an AI system is **primarily** geared towards eliminating core paid tasks without much being expected in terms of increased job quality or broadly shared benefits, nor in terms of new tasks for humans being created in parallel, then it warrants further attention as posing a risk to shared prosperity. The introduction of such a system will likely lower the demand for human labor, and thus wage or employment levels for affected workers.[28] Automation of core tasks can also be experienced by workers as directly undermining their job satisfaction since workers' core responsibilities are closely tied to their sense of pride and accomplishment in their jobs. For workers who see their jobs as an important part of their identity, core tasks are a major aspect of how they see themselves in the world.[29] Automation of core tasks can also lower the skill requirements of a job and reduce the formation of skills needed to advance to the next level.[30]

Please note that to evaluate the share of a given job's tasks being eliminated, those tasks should be weighted by their importance for the production of the final output. We consider task elimination above 10% significant enough to warrant attention.

RS2. Reallocate tasks to lower-paid or more precarious jobs

Will the AI system enable reallocation of tasks to lower-paid or more precarious jobs or informal or unpaid labor? Often, while not eliminating human tasks on balance, AI technology enables shifting tasks from full-time jobs to unpaid or more precarious labor. The latter can happen, for example, through the "gig-ification" of work: technologically enabled separation of "time on task" and "idle time" which leads to unstable and unpredictable wages as well as the circumvention of minimum wage laws.

Paid tasks can also be converted into unpaid when new technology enables them to be performed by customers. Examples of that are self-checkout kiosks or automated customer support.[31]

RS3. Reallocate tasks to higher- or lower-skilled jobs

Will the AI system enable the reallocation of tasks to jobs with higher or lower specialized skills requirements? Jobs with higher specialized skills requirements generally are better compensated, hence an AI system shifting tasks into such jobs will likely lead to a positive effect of more of them being opened up. However, those jobs might not be accessible to people affected by task reallocation because those people might not possess the newly required specialized skills. Retraining and job matching support programs can help here, though those often fall short. Word processor is an example of a technology that reallocated typing-related tasks away from typists to managers. Generative AI applications are an example of a recent technology anticipated to induce broad-reaching shifts in skill requirements of large swaths of jobs.[32] [33] [34]

Importantly, AI-induced reallocation of tasks to jobs with lower specialized skills requirements may be positive but is still a risk signal warranting further attention, because

Automation of core tasks can be experienced by workers as directly undermining their job satisfaction.

lowering specialized skill requirements can lower not only the barriers to entry to the occupation, but also prevailing wages.

RS4. Move jobs away from geographies with few opportunities

Will the AI system move job opportunities away from geographies where there would be few remaining? Due to associated costs and excessive immigration barriers, labor mobility remains low, both within and between countries. As a result, changes that move job opportunities from one area to another can harm workers in the losing area. Research suggests that disappearance of stable, well-paying jobs can profoundly re-shape regions, leading to a rise in "deaths of despair," addictions, and mental health problems.[35][36] Impacted communities might be able to bounce back from job loss if comparable alternative job opportunities are sufficiently available in their area. But even when those exist, the presence of labor market frictions make it important to invest in creating support programs to help workers move into new jobs of comparable quality.

In addition to jobs disappearing as the direct effect of labor-saving technology being introduced in a region, please note that this effect can also be an indirect result of labor-saving technology initially introduced in a completely different region or country. Due to excessive immigration barriers, AI developers based in high-income countries face massively inflated incentives to create labor-saving technologies far in excess of what would be socially optimal given the world's overall level of labor supply/demand for jobs.[37] Once that technology is developed in the high-income countries it gets deployed all over the world, including countries facing a dire need of formal sector jobs.[38]

Disappearance of stable, well-paying jobs can profoundly re-shape regions.

RS5. Increase market concentration and barriers to entry

Will an AI system increase market concentration and barriers to market entry? An increase in market concentration is a signal of a possible labor market impact to come for at least two reasons:

- It increases the risk of job cuts by competing firms
- It makes it less likely that the winning firm shares efficiency gains with workers in the form of better wages/benefits or with consumers in the form of lower prices/higher-quality products

Therefore, in a monopolistic market, any benefits brought on by AI are likely to be shared by few, while the harms might still be widely distributed. Similarly, job impacts that might occur in upstream or downstream industries due to an AI-induced increase in market concentration need to be accounted for as well.

RS6. Rely on poorly treated or compensated outsourced labor

Will the AI system rely on, for either model training or operation, outsourced labor deprived of a living wage and decent working conditions? The process of building datasets for

model training can be highly labor-intensive. It often requires human workers (whom we will refer to as data enrichment professionals) to review, classify, annotate, and otherwise manage massive amounts of data. Despite the foundational role played by data enrichment professionals, a growing body of research reveals the precarious working conditions that they face, which include:[39]

- Inconsistent and unpredictable compensation for their work
- Unfairly rejected and therefore unpaid labeling tasks
- Long, ad-hoc working hours
- Lack of means to contest or get an explanation for the decisions affecting their take-home pay and ratings
- Lack of transparency around data enrichment labor sourcing practices in the AI industry exacerbate this issue.

RS7. Use training data collected without consent or compensation

Will the AI system be trained using a dataset containing data collected without consent and/or compensation? AI systems can be trained on data that embeds the economically-relevant know-how of people who generated that data, which can be especially problematic if the subsequent deployment of that AI system reduces the demand for labor of those people. Examples include but are not limited to:

- Images created by artists and photographers that are used to train generative AI systems
- Keystrokes and audio recordings of human customer service agents used to create automated customer service routines
- Records of actions taken by human drivers used to train autonomous driving systems

RS8. Predict the lowest wages a worker will accept

Will the AI system be used to predict the lowest wage a given worker would accept? It has been documented that workers can experience the impact of AI systems used for workforce management as effectively depriving them of being able to predict their take-home wages with any amount of certainty.[40] An AI system allowing predictions about the lowest wages an individual worker would accept is analogous to a system allowing for perfect price discrimination of consumers. Price discrimination, while always driven by monopoly power and thus inefficient, is considered acceptable in certain situations, such as reduced price of museum admission for seniors and students. However, that acceptability is predicated on the transparency of the underlying logic. A possibility of using an algorithmic system to create take-home pay "personalization," especially based on logic that is opaque to the workers or ever-changing, should serve as a strong signal of a potential negative impact on shared prosperity. A related risk for informal workers is the use of AI to reduce their bargaining power relative to those they contract with. Information asymmetries created through AI use by purchasers of their work are an emerging risk to workers in the informal sector.[41]

RS9. Accelerate task completions without other changes

Will the AI system accelerate task completion without meaningfully changing resources, tools, or skills needed to accomplish the tasks? Some AI systems push workers to higher performance on goals, targets, or KPIs without modifying how the work is done. Examples of this include speeding up the pace with which workers are expected to complete tasks or using AI to set performance goals that are just out of reach for many workers. When this occurs without additional support for workers in the form of streamlining, simplifying, or otherwise improving the process of completing the task, it risks higher stress and injury rates for workers.

RS10. Reduce schedule predictability

Will the AI system reduce the amount of advance notice a worker receives regarding changes to their working hours? Schedule predictability is strongly tied to workers' physical and mental health.[42][43] Automated, last-minute scheduling software can harm workers':

- Emotional well-being through increased stress
- Occupational safety and health through sleep deprivation/unpredictability and the physical effects of stress
- Financial well-being through missed shifts and increased need for more expensive transit (for example, ride-hailing services at times when public transit isn't frequent or safe).

Recent AI technology designed to lower labor costs by reducing the number of people working during predicted "slow" times has disrupted schedule predictability, with workers receiving minimal notice about hours that have been eliminated from or added to their schedules.

RS11. Reduce workers' break time

Will the AI system infringe on workers' breaks or encourage them to do so? Workers' breaks are necessary for their recovery from physically, emotionally, or intellectually strenuous or intense periods of work, and are often protected by law. Some AI systems billed as productivity software infringe on workers' breaks by sending them warnings based on the time they've spent away from their workstations or "off-task," even during designated breaks or while they are using allotted break time.[44] Others implicitly encourage workers to skip breaks by setting overly ambitious performance targets that pressure workers to work through downtime to meet goals. These systems can foster higher rates of injury or stress, undermine focus, and reduce opportunities to form social relationships at work.

RS12. Increase overall difficulty of tasks

Will the AI system increase the overall difficulty of tasks? When AI systems are used to automate less demanding tasks (for example, the most straightforward, emotionally

neutral customer requests in a call center), workers may be left with a higher concentration of more demanding tasks, effectively increasing the difficulty of their job.[45] Difficulty increases may take the form of more physically, emotionally, or intellectually demanding tasks. The higher intensity may also place them at higher risk of burning out. While some workers may welcome the added challenge, the above concerns merit caution, especially if workers are not compensated equitably for the increased difficulty.

RS13. Enable detailed monitoring of workers

Will the AI system monitor something other than the pace and quality of task completion? The use of AI to monitor workers is just the latest entry in the long history of the technological surveillance of labor.[46] However, AI capabilities have increased the frequency, comprehensiveness, and intensiveness of on-the-job monitoring. This use of AI often extends beyond monitoring of workers' direct responsibilities and outputs, including information as varied as their time in front of their computer or time spent actively using their computer, their movements through an in-person worksite, and the frequency and content of communications with other workers. This detailed monitoring risks:

- Increasing stress and anxiety
- Harming their privacy
- Causing them to feel a lack of trust from their employer
- Undermining their sense of autonomy on the job
- Lowering engagement and job satisfaction
- Chilling worker organizing, undermining worker voice.[47] [48]

Monitoring systems can have legitimate uses, but even good systems can be abused.

While monitoring systems can have legitimate uses (such as enhancing worker safety), even good systems can be abused, particularly in environments with low worker agency or an absence of regulations, monitoring, and enforcement of worker protections.[49]

RS14. Reduce worker autonomy

Will the AI system reduce workers' autonomy, decision-making authority, or control over how they complete their work? Autonomy, decision-making authority, job control, and the exercise of discernment in performing one's job are correlated with high job quality and job satisfaction.[50] Reducing scope for these activities could also be a sign of a shift from a "high-road" staffing approach (where experience and expertise is valued) to a "low-road" approach (where less training or experience is needed and thus workers hold less bargaining power and can be more easily replaced). In the informal sector, this may appear as a reduction in the scope for design and creativity by artisans and garment workers.[51]

RS15. Reduce mentorship or apprenticeship opportunities

Will the AI system reduce workers' opportunities for mentorship or apprenticeship? Automated training, automated coaching, and automation of entry-level tasks may

lower workers' opportunities for apprenticeship and mentorship. Apprenticeship is an important way for workers to learn on the job, and develop the skills they need to advance.[52] Mentorship and apprenticeship can help workers develop social relationships and community with peers and supervisors. Additionally, mentors can help workers learn to navigate unspoken rules and norms in the workplace, and assist them with career development within and beyond their current workplace.

RS16. Reduce worker satisfaction

Will the AI system reduce the motivation, engagement, or satisfaction of the workers who use it or are affected by it? While this test directly speaks to meaning, community, and purpose, it is also a proxy for other aspects of worker well-being. Demotivation and disengagement are signs of lowered job satisfaction and serve as indications of other job quality issues.

RS17. Influence employment and pay decisions

Will the AI system make or suggest decisions on recruitment, hiring, promotion, performance evaluation, pay, wage penalties, and bonuses? The decisions outlined in this signal are deeply meaningful to workers, meriting heightened attention from employers. Automation of these decisions should raise concern, as automated systems might lack the complete context necessary for these decisions and risk subjecting workers to "algorithmic cruelty."[53] They also risk introducing additional discriminatory bases for decisions, beyond those already existent in human decisions.[54] In instances where AI systems are used to suggest (rather than decide) on these questions, careful implementation focused on increasing decision accuracy and transparency can benefit workers. However, human managers using these systems often find it undesirable or difficult to challenge or override recommendations from AI, making the system's suggestions more binding than they may initially appear and meriting additional caution in these uses.

RS18. Operate in discriminatory ways

Will the AI system operate in ways that are discriminatory? AI systems have been repeatedly shown to reproduce or intensify human discrimination patterns on demographic categories such as gender, race, age, and more.[55][56][57][58] Workplace AI systems should be rigorously tested to ensure that they operate fairly and equitably.

Workplace AI systems should be rigorously tested to ensure that they operate fairly and equitably.

Follow Our Stakeholder-Specific Recommendations

Foster shared prosperity by enacting best practices and suggested uses:

For AI-creating organizations

For AI-using organizations

For policymakers

For labor organizations and workers

Responsible Practices for AI-Creating Organizations (RPC)

Use of workplace AI is still in early stages, and as a result information about what should be considered best practices for fostering shared prosperity is still preliminary. Below is a list for AI-creating organizations of starter sets of practices aligned with increasing the likelihood of benefits to shared prosperity and decreasing the likelihood of harms to it. The list is drawn from early empirical research in the field, historical analogues for transformative workplace technologies, and theoretical frameworks yet to be applied in practice. For ease of use, the lists of Responsible Practices are organized by the earliest AI system lifecycle stage where the practice can be applied.

AT AN ORGANIZATIONAL LEVEL

RPC1. Make a public commitment to identify, disclose, and mitigate the risks of severe labor market impacts presented by AI systems you develop

Multiple AI-creating organizations aspire (according to their mission statements and responsible AI principles) to develop AI that benefits everyone. Very few of them, however, currently publicly acknowledge the scale of labor market disruptions their AI systems might bring about or make efforts to help communities that stand to be affected have a say in the decisions determining the path, depth, and distribution of labor market disruptions. At the same time, AI-creating organizations are often best positioned to anticipate labor market risks well in advance of those becoming apparent to other stakeholders, thus making risk disclosures by AI-creating organizations a valuable asset for governments and societies.

The public commitment to disclose severe risks* should specify the severity threshold considered by the organizations to warrant disclosure, as well as explain how the threshold level of severity was chosen and what external stakeholders were consulted in that decision.

Alternatively, an organization can choose to set a threshold in terms of an AI system's anticipated capabilities and disclose all risk signals which are present for those systems. For example, if the expected return on investment from the deployment of an AI system is a multiple greater than 10, or more than one million US dollars were spent on training compute and data enrichment, its corresponding risks would be subject to disclosure.[P]

DURING THE FULL AI LIFECYCLE

RPC2. In collaboration with affected workers, perform Job Impact Assessments early and often throughout the AI system lifecycle

Run opportunity and risk analyses early and often in the AI research and product development process, using the data available at each stage. Update as more data becomes available (for example, as product-market fit becomes clearer or features are built out enough for broader worker testing and feedback). **Whenever applicable, we suggest using AI system design and deployment choices to maximize the presence of signals of opportunity and minimize the presence of signals of risk.**

Always solicit the input of workers that stand to be affected — both incumbents as well as potential new entrants — and a multi-disciplinary set of third-party experts when assessing the presence of opportunity and risk signals. Make sure to compensate external contributors for their participation in the assessment of the AI system.

Please note that the analysis of opportunity and risk signals suggested here is different from red team analysis suggested in RPC13. The former identifies risks and opportunities created by an AI system working perfectly as intended. The latter identifies possible harms if the AI system in question malfunctions or is misused.

RPC3. In collaboration with affected workers, develop mitigation strategies for identified risks

In alignment with UN Guiding Principles for Business and Human Rights, a mitigation strategy should be developed for each risk identified, prioritizing the risks primarily by severity of potential impact and secondarily by its likelihood. Severity and likelihood of potential impact are determined on a case-by-case basis.[Q]

Mitigation strategies can range from eliminating the risk or reducing the severity of potential impact to ensuring access to remedy or compensation for affected groups. If effective mitigation strategies for a given risk are not available, this should be considered a strong argument in favor of meaningful changes in the development plans of an AI system, especially if it is expected to affect vulnerable groups.

P These thresholds are used for illustrative purposes only: AI creating organizations should set appropriate thresholds and explain how they were arrived at. Thresholds need to be reviewed and possibly revised regularly as the technology advances.

Always solicit the input of workers that stand to be affected.

Q An algorithm described here is very useful for determining the severity of potential quantitative impacts (such as impacts on wages and employment), especially in cases with limited uncertainty around the future uses of the AI system being assessed.

Engaging adequately compensated external stakeholders in the development of mitigation strategies is critical to ensure important considerations are not being missed. It is especially critical to engage with representatives of communities that stand to be affected.

RPC4. Source data enrichment labor responsibly

Key requirements for the responsible sourcing of data enrichment services (such as, data annotation and real-time human verification of algorithmic predictions) include:

- Always paying data enrichment workers above the local living wage
- Providing clear, tested instructions for data enrichment tasks
- Equipping workers with simple and effective mechanisms for reporting issues, asking questions, and providing feedback on the instructions or task design

In collaboration with our Partners, PAI has developed a library of practitioner resources for responsible data enrichment sourcing.

DURING SYSTEM ORIGINATION AND DEVELOPMENT

RPC5. Create and use robust and substantive mechanisms for worker participation in AI system origination, design, and development

Workers who will use or be affected by AI hold unique perspectives on important needs and opportunities in their roles. They also possess particular insight into how AI systems could create harm in their workplaces. To ensure AI systems foster shared prosperity, these workers should be given agency in the AI development process from start to finish.

This work does not stop at giving workers a seat at the table throughout the development process. Workers must be properly equipped with knowledge of product functions, capabilities, and limitations so they can draw meaningful connections to their role-based knowledge. Additionally, care must be taken to create a shared vocabulary on the team, so that technical terms or jargon do not unintentionally obscure or mislead. Workers must also be given genuine decision-making power in the process, allowing them to shape product functions and features, and be taken seriously on the need to end a project if they identify unacceptable harms that cannot be resolved.

> Workers who will use or be affected by AI hold unique perspectives on important needs and opportunities in their roles.

RPC6. Build AI systems that align with worker needs and preferences

AI systems welcomed by workers largely fall into three overarching categories:

- Systems that directly improve some element of job quality
- Systems that assist workers to achieve higher performance on their core tasks
- Systems that eliminate undesirable non-core tasks (See OS3, RS1, and RS2 for additional detail)

Starting with one of these objectives in mind and creating robust participation mechanisms for workers throughout the design and implementation process is likely to

result in win-win-wins for AI creators, employers who implement AI, and the workers who use or are affected by them.

RPC7. **Build AI systems that complement workers (especially those in lower-wage jobs), not ones that act as their substitutes**

A given AI system complements a certain group of workers if the demand for labor of that group of workers can be reasonably expected to go up when the price of the use of that AI system goes down. A given AI system is a substitute for a certain group of workers if the demand for labor of that group of workers is likely to fall when the price of the use of that AI system goes down.

Note that the terms "labor-augmenting" technology and "labor-complimentary" technology are often erroneously used interchangeably. "Labor-augmenting technology" is increasingly being used as a loose marketing term which frames workplace surveillance technology as worker-assistive.[59]

Getting direct input from workers is very helpful for differentiating genuinely complementary technology from the substituting kind. Please also see the discussion of the distinction between core and non-core tasks and the acceptable automation thresholds in RS1.

RPC8. **Ensure workplace AI systems are not discriminatory**

In general, AI systems frequently reproduce or deepen discriminatory patterns in society, including ones related to race, class, age, and disability. Specific workplace systems have shown a propensity for the same. Careful work is needed to ensure any AI systems affecting workers or the economy do not create discriminatory results.

> AI systems frequently reproduce or deepen discriminatory patterns in society.

BEFORE SELLING OR DEPLOYING THE SYSTEM

RPC9. **Provide meaningful, comprehensible explanations of the AI system's function and operation to workers using or affected by it**

The field of explainable AI has advanced considerably in recent years, but workers remain an underrepresented audience for AI explanations.[60] Providing workers explanations of workplace AI systems tailored to the particulars of their roles and job goals enables them to understand the tools' strengths and weaknesses. When paired with workers' existing subject matter expertise in their own roles, this knowledge equips workers to most effectively attain the upsides and minimize the downsides of AI systems, meaning AI systems can enhance their overall job quality across the different dimensions of well-being.

RPC10. Ensure transparency about what worker data is collected, how and why it will be used, and enable opt-out functionality

Privacy and ownership over data generated by one's activities are increasingly rights recognized inside and outside the workplace. Respect for these rights requires fully informing workers about the data collected on them and inferences made, how they are used and why, as well as offering them the ability to opt out of collection and use.[61] Workers should also be given the opportunity to individually or collectively forbid the sales of datasets that include their personal information or personally identifiable information. In particular, system design should follow the data minimization principle: collect only the necessary data, for the necessary purpose, and hold it only for the necessary amount of time. Design should also enable workers to know about, correct, or delete inferences about them. Particular care must be taken in workplaces, as the power imbalance between employer and employee undermines workers' ability to freely consent to data collection and use compared to other, less coercive contexts.[62]

RPC11. Embed human recourse into decisions or recommendations you offer

AI systems have been built to hire workers, manage them, assess their performance, and promote or fire them. AI is also being used to assist workers with their tasks, coach them, and complete tasks previously assigned to them. In each of these decisions allocated to AI, the technologies have accuracy as well as comprehensiveness issues. AI systems lack the human capacity to bring in additional context relevant to the issue at hand. As a result, humans are needed to validate, refine, or override AI outputs. In the case of task completion, an absence of human involvement can create harms to physical, intellectual, or emotional well-being. In AI's use in employment decisions, it can result in unjustified hiring or firing decisions. Simply placing a human "in the loop" is insufficient to overcome algorithmic bias: demonstrated patterns of deference to the judgment of algorithmic systems. Care must be taken to appropriately position the strengths and weaknesses of AI systems and empower humans with final decision-making power.[63]

RPC12. Apply additional mitigation strategies to sales and use in environments with low worker protection and decision-making power

AI systems are less likely to cause harm in environments with:

- High levels of legal protection, monitoring, and enforcement for workers' rights (such as those related to health and safety or freedom to organize)
- High levels of worker voice and negotiating ability (due to strong protections for worker voice or high demand for workers' comparatively scarce skills), especially those where workers have meaningful input into decisions regarding the introduction of new technologies

These factors encourage worker-centric AI design. Workers in such environments also possess a higher ability to limit harms from AI systems (such as changing elements of an implementation or rejecting the use of the technology as needed), including harms outside

direct legal protections. This should not, however, be treated as a failsafe for harmful technologies, particularly when AI systems can easily be adopted in environments where they were not originally intended.[64] In environments where workers lack legal protection and/or decision-making power, it is especially important to scrutinize uses and potential impacts, building in additional mitigations to compensate for the absence of these worker safeguards. Contractual or licensing provisions regarding terms of use, rigorous customer vetting, and geofencing are some of the many steps AI-creating organizations can take to follow this practice. Care should be taken to adopt fine-grained mitigation strategies where possible such that workers and economies can reap the gains of neutral or beneficial uses.

RPC13. Red team AI systems for potential misuse or abuse

The preceding points have focused on AI systems working as designed and intended. Responsible development also requires comprehensive "red teaming" of AI systems to identify vulnerabilities and the potential for misuse or abuse. Adversarial ML is increasingly a part of standard security practice. Additionally, the development team, workers in relevant roles, and external experts should test the system for misuse and abusive implementation.

RPC14. Ensure AI systems do not preclude the sharing of productivity gains with workers

The power and responsibility to share productivity gains from AI system implementation lies mostly with AI-using organizations. The role of AI-creating organizations is to make sure the functionality of an AI system does not fundamentally undermine opportunities for workers to share in productivity gains, which would be the case if an AI system de-skills jobs and makes workers more likely to be viewed as fungible or automates a significant share of workers' core tasks.

RPC15. Request deployers to commit to following PAI's Shared Prosperity Guidelines or similar recommendations

The benefit to workers and society from following these practices can be meaningfully undermined if organizations deploying or using the AI system do not do their part to advance shared prosperity. We encourage developers to make adherence to the Guidelines' Responsible Practices a contractual obligation during the selling or licensing of the AI system for deployment or use by other organizations.

> The role of AI-creating organizations is to make sure the functionality of an AI system does not fundamentally undermine opportunities for workers to share in productivity gains.

Responsible Practices for AI-Using Organizations (RPU)

Use of workplace AI is still in early stages, and as a result information about what should be considered best practices for fostering shared prosperity is still preliminary. Below is a list for AI-using organizations of starter sets of practices aligned with increasing the likelihood of benefits to shared prosperity and decreasing the likelihood of harms to it. The list is drawn from early empirical research in the field, historical analogues for transformative workplace technologies, and theoretical frameworks yet to be applied in practice. For ease of use, the lists of Responsible Practices are organized by the earliest AI system lifecycle stage where the practice can be applied.

AT AN ORGANIZATIONAL LEVEL

> **RPU1.** Make a public commitment to identify, disclose, and mitigate the risks of severe labor market impacts presented by AI systems you use

Labor practices and impacts are increasingly a part of suggested, proposed, or required non-financial disclosures. These disclosures include practices affecting human rights, management of human capital, and other social and employee issues. Regulatory authorities have suggested, proposed, or required these disclosures as material to investor decision-making,[R] as well as for the benefit of the broader society. We recommend that AI-using organizations identify, disclose, and mitigate the risks of severe labor market impacts for the same rationales, as well as to provide both prospective and existing workers with the information they need to make informed decisions about their own employment.

The public commitment to disclose severe risks[S] should specify the severity threshold considered by the organization to warrant disclosure, as well as explain how the threshold level of severity was chosen and what external stakeholders were consulted in that decision.

Alternatively, an organization can choose to set a threshold in terms of an AI system's marketed capabilities and disclose all risk signals which are present for systems meeting that threshold. For example, if an organization's expected return on investment from the use of an AI system under assessment is a multiple greater than 10, its corresponding risks would be subject to disclosure. In instances where organizational impact is driven by a series of smaller system implementations, the organization could choose to disclose all risk signals present once the cumulative cost decrease or revenue increase exceeds 5%.[T]

R See, for instance, the Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework or the United States Securities and Exchange Commission's 2023 agenda, as reported in Reuters.

S PAI's Shared Prosperity Guidelines use UNGP's definition of severity: an impact (potential or actual) can be severe "by virtue of one or more of the following character-istics: its scale, scope or irremediability. *Scale* means the gravity of the impact on the human right(s). *Scope* means the number of individuals that are or could be affected. *Irremediability* means the ease or otherwise with which those impacted could be restored to their prior enjoyment of the right(s)."

T A recent study of corporate respondents showed roughly one quarter of respondents were able to achieve a 5% improvement to EBIT in 2021. As AI adoption becomes more widespread, we anticipate more organizations will meet this threshold.

RPU2. Commit to neutrality towards worker organizing and unionization

As outlined in the signals of risk above, AI systems pose numerous risks to workers' human rights and well-being. These systems are implemented and used in employment contexts that often have such comprehensive decision-making power over workers that they can be described as "private governments."[65] As a counterbalance to this power, workers may choose to organize to collectively represent their interests. The degree to which this is protected, and the frequency with which it occurs, differs substantially by location. Voluntarily committing to neutrality towards worker organizing is an important way to ensure workers' agency is respected and their collective interests have representation throughout the AI use lifecycle if workers so choose (as is repeatedly emphasized as a critical provision in these Guidelines).

RPU3. In collaboration with affected communities, perform Job Impact Assessments early and often throughout AI system implementation and use

Run opportunity and risk analyses early and often across AI implementation and use, using the data available at each stage. Update as more data becomes available (for example, as objectives are identified, systems are procured, implementation is completed, and new applications arise). **Whenever applicable, we suggest using AI system implementation and use choices to maximize the presence of signals of opportunity and minimize the presence of signals of risk.**

Solicit the input of workers that stand to be affected[U] and a multi-disciplinary set of independent experts when assessing the presence of opportunity and risk signals. Make sure to compensate external contributors for their participation in the assessment of the AI system.

Please note that the analysis of opportunity and risk signals suggested here is different from red team analysis suggested in RPU15. The former identifies risks and opportunities created by an AI system working perfectly as intended. The latter identifies possible harms if the AI system in question malfunctions or is misused.

RPU4. In collaboration with affected communities, develop mitigation strategies for identified risks

In alignment with UN Guiding Principles for Business and Human Rights, a mitigation strategy should be developed for each risk identified, prioritizing the risks primarily by severity of potential impact and secondarily by its likelihood. Severity and likelihood of potential impact are determined on a case-by-case basis.[V]

Mitigation strategies can range from eliminating the risk or reducing the severity of potential impact to ensuring access to remedy or compensation for affected groups.

U It is frequently the case that workers who stand to be affected by the introduction of an AI system include not only workers directly employed by the organization introducing AI in its own operations, but a wider set of current or potential labor market participants. Therefore it is important that not only incumbent workers are given the agency to participate in job impact assessment and risk mitigation strategy development.

V An algorithm described here is very useful for determining the severity of potential quantitative impacts (such as impacts on wages and employment), especially in cases with limited uncertainty around the future uses of the AI system being assessed.

If effective mitigation strategies for a given risk are not available, this should be considered a strong argument in favor of meaningful changes in the development plans of an AI system, especially if it is expected to affect vulnerable groups.

Engaging workers and external experts as needed in the creation of mitigation strategies is critical to ensure important considerations are not being missed. It is especially critical to engage with representatives of communities that stand to be affected. Please ensure that everyone engaged in consultations around assessing risks and developing mitigation strategies is adequately compensated.

RPU5. Create and use robust and substantive mechanisms for worker agency in identifying needs, selecting AI vendors and systems, and implementing them in the workplace

Workers who will use or be affected by AI hold unique perspectives on important needs and opportunities in their roles. They also possess particular insight into how AI systems could create harm in their workplaces. To ensure AI systems foster shared prosperity, these workers should be included and afforded agency in the AI procurement, implementation, and use process from start to finish.[66]

Workers must be properly equipped with knowledge of potential product functions, capabilities, and limitations, so that they can draw meaningful connections to their role-based knowledge (see RPU13 for more information). Additionally, care must be taken to create a shared vocabulary on the team, so that technical terms or jargon do not unintentionally obscure or mislead. Workers must also be given genuine decision-making power in the process, allowing them to shape use (such as new workflows or job design) and be taken seriously on the need to end a project if they identify unacceptable harms that cannot be resolved.

> Workers who will use or be affected by AI hold unique perspectives on important needs and opportunities in their roles.

RPU6. Ensure AI systems are used in environments with high levels of worker protections and decision-making power

AI systems are less likely to cause harm in environments with:

- High levels of legal protection, monitoring, and enforcement for workers' rights (such as those related to health and safety or freedom to organize)
- High levels of worker voice and negotiating ability (due to strong protections for worker voice or high demand for workers' comparatively scarce skills), especially those where workers have meaningful input into decisions regarding the introduction of new technologies

These factors encourage worker-centric AI design. Workers in such environments also possess a higher ability to limit harms from AI systems (such as changing elements of an implementation or rejecting the use of the technology as needed), including harms outside direct legal protections. This should not, however, be treated as a failsafe for harmful technologies: other practices in this list should also be followed to reduce risk to workers.

RPU7. Source data enrichment labor responsibly

Key requirements for the responsible sourcing of data enrichment services (such as, data annotation and real-time human verification of algorithmic predictions) include:

- Always paying data enrichment workers above the local living wage
- Providing clear, tested instructions for data enrichment tasks
- Equipping workers with simple and effective mechanisms for reporting issues, asking questions, and providing feedback on the instructions or task design

In collaboration with our Partners, PAI has developed a library of practitioner resources for responsible data enrichment sourcing.

RPU8. Ensure workplace AI systems are not discriminatory

In general, AI systems frequently reproduce or deepen discriminatory patterns in society, including ones related to race, class, age, and disability. Specific workplace systems have shown a propensity for the same. Careful vetting and use is needed to ensure any AI systems affecting workers or the economy do not create discriminatory results.

WHEN IDENTIFYING NEEDS, PROCURING, AND IMPLEMENTING AI SYSTEMS

RPU9. Procure AI systems that align with worker needs and preferences

AI systems welcomed by workers largely fall into three overarching categories:

- Systems that directly improve some element of job quality
- Systems that assist workers to achieve higher performance on their core tasks
- Systems that eliminate undesirable non-core tasks (See OS2, OS9, RS1, and RS2 for additional detail)

Starting with one of these objectives in mind and creating robust participation mechanisms for workers throughout the design and implementation process is likely to result in win-win-wins for AI creators, employers who implement AI, and the workers who use or are affected by them.

RPU10. Staff and train sufficient internal or contracted expertise to properly vet AI systems and ensure responsible implementation

As discussed throughout, AI systems raise substantial concerns about the risks of their adoption in workplace settings. To understand and address these risks, experts are needed to vet and implement AI systems. In addition to technical experts, this includes sociotechnical experts capable of performing the Job Impact Assessment described above to the level of granularity necessary to fully identify and mitigate risks of a specific system in a given workplace.

The importance of this practice increases with AI system customization or integration. In situations where systems are developed by organizations who follow the Shared Prosperity

> AI systems frequently reproduce or deepen discriminatory patterns in society.

Guidelines or similar recommendations, disclose potential labor impacts, and design these systems to be used off-the-shelf, less internal expertise may be required from users. However, when systems are more customized or integrated into workplaces, specifics related to the organization and worksite more heavily influence labor impacts arising from the particulars of system use, requiring additional expertise.

RPU11. Prefer vendors who commit to following PAI's Shared Prosperity Guidelines or similar recommendations

The benefit to workers and society from following these practices can be meaningfully undermined if organizations designing and selling the AI system do not do their part to advance shared prosperity. We encourage users to make developer adherence to PAI's Guidelines or similar recommendations a priority when selecting vendors and systems for use.

RPU12. Ensure transparency about what worker data is collected, how it will be used, and why, and enable workers to opt out

Privacy and ownership over data generated by one's activities are increasingly rights recognized inside and outside the workplace. Respect for these rights requires fully informing workers about the data collected on them and inferences made, how they are used and why, as well as offering them the ability to opt out of collection and use.[67] Workers should also be given the opportunity to individually or collectively forbid the sales of datasets that include their personal information or personally identifiable information. Depending on use, generative AI may present novel privacy risks, through extracting information about worker practices and sharing with managers and colleagues. System design and use should follow the data minimization principle: collect only the necessary data, for the necessary purpose, and hold it only for the necessary amount of time. Design should also enable workers to know about, correct, or delete inferences about them.[68]

Particular care must be taken in workplaces, as the power imbalance between employer and employee undermines workers' ability to freely consent to data collection and use compared to other, less coercive contexts. In practice, data use decisions by employers often shift over time, making it especially important for AI-using organizations to explicitly and transparently inform workers regarding each new use of their data and its implications, and request consent for each new use or repurposing.[69]

RPU13. Provide meaningful, comprehensible explanations of the AI system's function and operation to workers overseeing it, using it, or affected by it

The field of explainable AI has advanced considerably in recent years, but workers remain an underrepresented audience for AI model explainability efforts.[70] Providing managers and workers explanations of workplace AI systems tailored to the particulars of their roles and job goals enables them to understand the tools' strengths and weaknesses. When paired

with workers' existing subject matter expertise in their own roles, this knowledge equips managers and workers to most effectively attain the upsides and minimize the downsides of AI systems, meaning AI systems can enhance overall job quality across the different dimensions of well-being.

RPU14. Establish human recourse into decisions or recommendations offered, including the creation of transparent, human-decided grievance redress mechanisms

AI systems have been built to hire workers, manage them, assess their performance, and promote or fire them. AI is also being used to assist workers with their tasks, coach them, and complete tasks previously assigned to them. In each of these decisions allocated to AI, the technologies have accuracy as well as comprehensiveness issues. AI systems lack the human capacity to bring in additional context relevant to the issue at hand. As a result, humans are needed to validate, refine, or override AI outputs. In the case of task completion, an absence of human involvement can create harms to physical, intellectual, or emotional well-being. In AI's use in employment decisions, it can result in unjustified hiring or firing decisions. Simply placing a human "in the loop" is insufficient to overcome algorithmic bias: demonstrated patterns of deference to the judgment of algorithmic systems. Care must be taken to appropriately position the strengths and weaknesses of AI systems and empower humans with final decision-making power.

> AI systems lack the human capacity to bring in additional context relevant to the issue at hand.

RPU15. Red team AI systems for potential misuse or abuse

The preceding points have focused on AI systems working as designed and intended. Responsible development also requires comprehensive "red teaming" of AI systems to identify vulnerabilities and the potential for misuse or abuse. Managers, workers in relevant roles, and external experts should test the system for misuse and abusive implementation.

RPU16. Recognize extra work created by AI system use and ensure work is acknowledged and compensated

The above practice of red-teaming addresses intentional misuse or abuse. More routinely, AI systems fail to work as marketed or intended in ways big and small, creating additional tasks for workers to absorb. New tasks generated by the gap between AI system expectations and realities often go unrecognized, leaving workers to shoulder extra responsibilities or work without providing them additional time to complete these tasks or compensation for doing so.[71][72] Address this issue by holding routine reviews with the workers who use or oversee systems to identify areas of new work and adjust accordingly.

RPU17. Ensure mechanisms are in place to share productivity gains with workers

The power and responsibility to share productivity gains from AI system implementation lies largely with AI-using organizations. AI-using organizations hold final decisions about wages, benefits, working hours, job design, worker retraining and reskilling, and more. To the extent that AI systems deliver cost savings and/or higher revenues via increased worker productivity, AI-using organizations hold authority over how to allocate increased margins. As highlighted in OS7, AI systems present a major opportunity to improve workers' well-being, financial and otherwise, through maintaining or increasing their share of revenue without decreasing absolute returns to owners or shareholders.

Suggested Uses for Policymakers

We currently anticipate two primary ways in which the Guidelines can be used by policymakers, described below. If you have feedback, suggestions, or would like to explore using the Guidelines in your work, please get in touch.

1. Policymakers can integrate the Job Impact Assessment steps suggested by the Guidelines into existing or emerging standards, risk management frameworks, and conformity assessments to encourage AI-creating and AI-using organizations to assess and disclose their anticipated impacts on shared prosperity and abide by Responsible Practices suggested by the Guidelines. This can be done either as a part of "horizontal" or sectoral AI regulation or by making existing worker protection laws better fit the age of rapid adoption of AI throughout the economy.

2. Policymakers can perform the Job Impact Assessment Tool's risk and opportunities analysis themselves to better identify the possible impacts of AI uses of interest on shared prosperity. Such analysis can be relevant in multiple contexts, including:

   - Considering the need for new regulation or modification of existing regulation in light of emergence of new uses of AI
   - Informing good jobs creation strategy at the local, regional, or state level
   - Making decisions about whether to provide tax breaks or other incentives to attract specific industries into the region with the goal of strengthening the local labor market
   - Ensuring sustainability of social protection mechanisms in the context of changing technological landscape, anticipating the pace and timing of increases in unemployment benefits claims, and declines in labor income tax revenue.

Suggested Uses for Labor Organizations and Workers

We currently anticipate four ways in which the Guidelines can be used by unions, worker organizations, worker representatives, and workers, described below. If you have feedback, suggestions, or would like to explore using the Guidelines in your work, please get in touch.

1. The Job Impact Assessment Tool and Responsible Practices can be used to audit or assess existing or prospective AI systems and offer a foundation for dialogue or negotiation over system need identification, purchases, implementation, and use. Such dialogues or negotiations could consider existing or potential impacts on workers, as well as transparency and consent in workplace data collection and use. Where opportunities exist for workers and their representatives to have agency in AI system design, the tools provided in the Guidelines can be used to identify areas for further analysis and improvement.

2. The Guidelines offer ideas for potential provisions to be included in collective bargaining agreements or other mechanisms for advancing employer workplace policies. Some jurisdictions explicitly delineate technology as an area for collective worker input and decision-making, while in others it is voluntary. Not all signals or responsible practices will be applicable to all AI systems or workplaces, but they can serve as an inventory for negotiators to include or draw inspiration from as they consider risks in their own workplaces.

3. The Guidelines outline issues that unions and worker organizations may wish to cover in trainings or educational sessions with members. The Job Impact Assessment Tool offers guidance on potential harms to watch out for, as well as possible benefits that workers can advocate for. Additionally, familiarizing workers with the Responsible Practices for AI-using organizations can equip them for advocacy for better workplace AI use within their teams, worksites, or organizations.

4. The Guidelines can be used to inform positions in policy discussions. As unions and worker organizations consider their policy objectives and goals, this tool can support informed engagement to shape the future of work.

Acknowledgements

The content of the Guidelines was iteratively developed by PAI's Labor and Economy team (Katya Klinova, Stephanie Bell, and Sonam Jindal) under comprehensive guidance of the AI and Shared Prosperity Initiative's Steering Committee. We thank the Steering Committee members for their steadfast commitment to the work of the Initiative and their generosity in investing their time, expertise, and effort to advance this effort.

We thank the frontline workers in the US, India, and Sub-Saharan Africa whose participation centrally informed the content of the Guidelines. We are grateful to Andre Barbe and Albert Tanjaya for excellent research support, and to Hudson Hongo and Neil Uhl for their thought partnership, crucial editorial support, and user experience design. We thank all past and present PAI colleagues and Partners who enabled and supported the work of the Initiative from its inception. We thank Ed Bayes, Elaine Chang, and Christina Colclough, for reviewing early drafts of the Guidelines and providing valuable feedback.

We are grateful to the Ford Foundation's Future of Work(ers) Program for financial support of the AI and Shared Prosperity Initiative, and personally to Ritse Erumi for valuable support and guidance.

The Guidelines build on the work of many exceptional scholars and advocates cited throughout this document. We are deeply grateful for their leadership.

AI and Shared Prosperity Initiative's Steering Committee

Abbie Langston, Director, Equitable Economy at PolicyLink

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

Andrea Dehlendorf, Senior Advisor and Co-Founder at United for Respect

Andrew Kortina, Co-founder of Venmo and Fin.com

Anton Korinek, Professor of Economics at University of Virginia and David M. Rubenstein Fellow at the Brookings Institution

Arturo Franco, Senior Vice President, Center for Inclusive Growth at Mastercard

Damon Silvers, Senior Advisor at AFL-CIO

Daron Acemoglu, Institute Professor of Economics at MIT

Dean Carignan, Senior Director of Product Management, Office of the Chief Scientist at Microsoft

Deborah Greenfield, Former Deputy Director-General for Policy at the International Labour Organization

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

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

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

Jody Medich, CEO, Founder at Superhuman-X

Jordan Usdan, Senior Director, Strategy & Innovation, Office of the CTO at Microsoft

Juana Catalina Becerra Sandoval, Visiting Scholar, Responsible & Inclusive Tech at IBM

Lama Nachman, Intel Fellow & Director, Anticipatory Computing Lab at Intel

Pamela Mishkin, Policy Staff Member at OpenAI

Rahul Panicker, Head of Product, Robotics Applications at Intrinsic

Reema Nanavaty, Executive Director at Self Employed Women's Association

Ryan Gerety, Acting Director at Athena Coalition

Sarah Treuhaft, Senior Director of Policy and Partnerships at the Institute on Race, Power, and Political Economy

Shakir Mohamed, Senior Staff Scientist at DeepMind

Endorsements

"Without an explicit intention to develop and use artificial intelligence toward enhancing the livelihoods of all of us, it will accelerate and reinforce the most unequal power dynamics in our society. These Guidelines reflect the field's current best thinking on evaluating whether a particular use enhances shared prosperity or fuels wealth concentration off the backs of working people. We hope to use these Guidelines together with workers to evaluate corporate employer practices and fight for better working conditions."

United for Respect

"AI has enormous potential to change how we work and–like any powerful technology—must be deployed responsibly and incorporate feedback from a wide variety of stakeholders. We welcome the Guidelines as an important step in ensuring AI benefits all of humanity, and we are pleased to work with PAI to help refine and operationalize these guidelines to help ensure that everyone can share in the economic prosperity unleashed by new AI technologies."

Pamela Mishkin
Policy Staff Member, OpenAI

"I highly recommend the Guidelines for AI and Shared Prosperity for AI developers and deployers. It's our responsibility to assess the economic and job quality impacts of our innovations. With these tools, we can make well-informed choices and avoid causing more harm than good."

Anton Korinek
Professor of Economics, University of Virginia

"Developing AI that genuinely complements workers and improves business processes is a difficult challenge we've been working hard on at Intel. The Guidelines for AI and Shared Prosperity are a helpful resource on that journey. I'm glad to have guided their development and look forward to helping test the Guidelines—I encourage leaders and researchers at other AI companies to join this effort."

Lama Nachman
Intel Fellow & Director, Anticipatory Computing Lab, Intel

"Our decisions about how to develop, use, and govern AI will reshape our society and determine who benefits and who is left behind. PAI's strong research, stakeholder engagement, and practical guidelines are all essential tools for policymakers, developers, and companies adopting these technologies, to ensure that they truly complement human effort. We can build a world that balances productivity with opportunity."

Arturo Franco
Senior Vice President, Mastercard Center for Inclusive Growth

"A just society would not allow AI systems to degrade job quality and wages for the most marginalized workers in the name of greater efficiency and growth that benefits the already-prosperous. I applaud PAI's Guidelines for AI and Shared Prosperity for prioritizing the needs of workers with the least power to protect themselves from AI harms. The Guidelines are a crucial resource for policymakers, civil society, labor organizers, and anyone else interested in ensuring AI creates equitable outcomes for all workers."

Sarah Treuhaft
Senior Director of Policy and Partnerships, Institute on Race, Power, and Political Economy

"I'm delighted that the Shared Prosperity Guidelines include recommendations for "AI using organizations" in addition to "AI creating organizations". Companies deploying AI systems make essential choices that determine the impact of AI on jobs and workers, and these Guidelines provide a new and much-needed resource for responsible AI governance and decision making."

Dunstan Allison-Hope
Vice President, Business for Social Responsibility

"A future of work that embraces technology as a tool for equitable, inclusive and sustainable growth depends upon efforts like PAI's Guidelines for AI and Shared Prosperity. The initiative's worker-centered foundation provides a powerful tool for trade unions and other advocates to understand the promises and risks of AI, engage in meaningful dialogue with those who develop and disseminate it, and harness its power on behalf of labor."

Deborah Greenfield
Former Deputy Director-General for Policy, International Labour Organization

"Without swift and careful action, artificial intelligence may cause substantial harms to workers around the globe. I welcome the release of PAI's Guidelines for AI & Shared Prosperity, and appreciate their focus on ensuring AI will have positive impacts for all workers, including workers in low- and middle-income countries. The Guidelines are an essential tool for any AI-developing or AI-using company, and offer helpful guidance for policymakers, workers, unions, and civil society around the world."

Grace Mutung'u
Centre for Intellectual Property & Information Technology at Strathmore University

"If corporations productively deploy AI, they'll see a boost in the efficiency of information work processes. Given corporate incentive structures and (lack of) retraining agility, the easiest way to realize benefits of improved efficiency will be through a reduction in workforce size. I hope some policy makers are anticipating this possibility and evaluating ways to redistribute corporate profits to workers who will consequently need to search for new jobs or new industries."

Andrew Kortina

Co-founder, Venmo & fin.com

"Today, over 60% of the workforce globally and over 90% of the workforce in the Global South countries is in the informal economy. These workers generally form the bottom and the lower-middle sector of the global value chains where the risks are concentrated. Emerging technologies like AI are going to have maximum effect on the jobs of these workers—displacing them, pushing them out of the workforce, making their existing skills redundant. These Guidelines will serve as an important tool for these poor informal sector workers and their organizations to prepare them for the upcoming impacts of AI and help build their resilience against the changing world of work."

Reema Nanavaty
Director, Self Employed Women's Association (SEWA)

"New technological and AI tools in the workplace are having an undeniable impact on workers and industries, from low-wage gig work to creative industries like film and journalism. However, the public discourse has been dominated by stories of the inevitability of technology and not enough attention has been placed on the decisions that went into getting us to this point, namely who gets to reap the benefits and who assumes the risks? The Guidelines for AI and Shared Prosperity offer an opportunity for all stakeholders to make transparent how those risks and benefits are allocated, and reflect or even change on those decisions, because this is how we have a real dialogue about the benefits of AI."

Aiha Nguyen
Program Director, Labor Futures Initiative Data & Society Research Institute

"The Partnership on AI has done an outstanding job developing these important recommendations for how we as a society should deploy AI so that it can benefit all. These practical and commonsense guidelines for developers of AI, impacted companies and workers, and policy makers, are an important step towards ensuring true shared prosperity."

Rahul Panicker
Head of Product, Robotics Applications, Intrinsic

Sources

1 Acemoglu, D., & Johnson, S. (2023). Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity. *Public Affairs*, New York.

2 Klinova, K., & Korinek, A. (2021). AI and shared prosperity. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 645-651).

3 Korinek, A. (2022). How innovation affects labor markets: An impact assessment.

4 Bell, S. A. (2022). AI and Job Quality: Insights from Frontline Workers. Available at SSRN 4337611.

5 Klinova, K. (2022). Governing AI to Advance Shared Prosperity. In Justin B. Bullock et al. (Eds.), The Oxford Handbook of AI Governance. Oxford Handbooks.

6 International Labour Organization. (n.d.). *Decent work.* https://www.ilo.org/global/topics/decent-work/lang--en/index.htm.

7 US Department of Commerce and US Department of Labor. (n.d.). *Department of Commerce and Department of Labor Good Jobs Principles, DOL.* http://www.dol.gov/general/good-jobs/principles.

8 Institute for the Future of Work. (n.d.). *The Good Work Charter.* https://www.ifow.org/publications/the-ifow-good-work-charter.

9 Klinova, K., & Korinek, A. (2021). AI and shared prosperity. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 645-651).

10 Bell, S. A. (2022). AI and Job Quality: Insights from Frontline Workers. Available at SSRN 4337611.

11 Partnership on AI, 2021. Redesigning AI for Shared Prosperity: an Agenda. https://partnershiponai.org/paper/redesigning-ai-agenda/

12 Negrón, W. (2021). Little Tech is Coming for Workers. Coworker.org. https://home.coworker.org/wp-content/uploads/2021/11/Little-Tech-Is-Coming-for-Workers.pdf.

13 Korinek, A., 2022. How innovation affects labor markets: An impact assessment.

14 Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3-30.

15 Bell, S. A. (2022). AI and Job Quality: Insights from Frontline Workers. Available at SSRN 4337611.

16 Valentine, M., & Hinds, R. (2022). How Algorithms Change Occupational Expertise by Prompting Explicit Articulation and Testing of Experts' Theories. https://doi.org/10.2139/ssrn.4246167.

17 Autor, D. (2022). *The labor market impacts of technological change: From unbridled enthusiasm to qualified optimism to vast uncertainty* (No. w30074). National Bureau of Economic Research.

18 Mateescu, A., & Elish, M. (2019). AI in context: the labor of integrating new technologies.

19 Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction (pre-print). *Engaging Science, Technology, and Society* (pre-print).

20 World Bank. (2017). *World development report 2018: Learning to realize education's promise.* The World Bank.

21 Korinek, A., & Stiglitz, J.E. (2021). Artificial intelligence, globalization, and strategies for economic development (No. w28453). National Bureau of Economic Research.

22 Diao, X., Ellis, M., McMillan, M. S., & Rodrik, D. (2021). *Africa's manufacturing puzzle: Evidence from Tanzanian and Ethiopian firms* (No. w28344). National Bureau of Economic Research.

23 Rodrik, D. (2022). 4 Prospects for global economic convergence under new technologies. *An inclusive future? Technology, new dynamics, and policy challenges*, 65.

24 O'Keefe, C., Cihon, P., Garfinkel, B., Flynn, C., Leung, J., & Dafoe, A. (2020, February). The windfall clause: Distributing the benefits of AI for the common good. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 327-331).

25 Bell, S. A. (2022). AI and Job Quality: Insights from Frontline Workers. Available at SSRN 4337611.

26 Scherer, M., & Brown, L. X. (2021). Warning: Bossware May Be Hazardous to Your Health. *Center for Democracy & Technology.* https://cdt.org/wp-content/uploads/2021/07/2021-07-29-Warning-Bossware-May-Be-Hazardous-To-Your-Health-Final.pdf

27 Bell, S. A. (2022). AI and Job Quality: Insights from Frontline Workers. *Available at SSRN 4337611*.

28 Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica*, 90(5), 1973-2016.

29 Valentine, M., & Hinds, R. (2022). How Algorithms Change Occupational Expertise by Prompting Explicit Articulation and Testing of Experts' Theories. https://doi.org/10.2139/ssrn.4246167.

30 Nurski, L., & Hoffmann, M. (2022). *The Impact of Artificial Intelligence on the Nature and Quality of Jobs.* Working Paper. Bruegel. https://www.bruegel.org/sites/default/files/2022-07/WP%2014%202022.pdf.

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

32 Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

33 Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4375283.

34 Korinek, A. (2023). *Language models and cognitive automation for economic research* (No. w30957). National Bureau of Economic Research.

35 Case, A., & Deaton, A. (2020). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

36 Gihleb, R., Giuntella, O., Stella, L., & Wang, T. (2022). Industrial robots, workers' safety, and health. *Labour Economics*, 78, 102205.

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

38 Pritchett, L. (2023). Choose People. *LaMP Forum*. https://lampforum.org/2023/03/02/choose-people/

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

40 Dubal, V. (2023). On Algorithmic Wage Discrimination. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331080.

41 Bell, S. A. (2022). AI and Job Quality: Insights from Frontline Workers. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4337611.

42 Schneider, D., & Harknett, K. (2017, April). Schedule Instability and Unpredictability and Worker and Family Health and Well-being. In *PAA 2017 Annual Meeting*. PAA.

43 Williams, J. et al. (2022). Stable scheduling study: Health outcomes report. https://ssrn.com/abstract=4019693

44 Bell, S. A. (2022). AI and Job Quality: Insights from Frontline Workers. Available at SSRN 4337611.

45 Dzieza, J. (2020). Robots aren't taking our jobs — They're becoming our bosses. The Verge. https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon.

46 Levy, K. (2022). *Data Driven: truckers, technology, and the new workplace surveillance*. Princeton University Press.

47 Moore, P.V. (2017). *The quantified self in precarity: Work, technology and what counts*. Routledge.

48 Scherer, M., & Brown, L. X. (2021). Warning: Bossware May Be Hazardous to Your Health. *Center for Democracy & Technology*. https://cdt.org/wp-content/uploads/2021/07/2021-07-29-Warning-Bossware-May-Be-Hazardous-To-Your-Health-Final.pdf.

49 Brand, J., Dencik, L. & Murphy, S. (2023). *The Datafied Workplace and Trade Unions in the UK*. Data Justice Lab. https://datajusticeproject.net/wp-content/uploads/sites/30/2023/04/Unions-Report_final.pdf.

50 Nurski, L., & Hoffmann, M. (2022). *The Impact of Artificial Intelligence on the Nature and Quality of Jobs*. Working Paper. Bruegel. https://www.bruegel.org/sites/default/files/2022-07/WP%2014%202022.pdf.

51 Nanavaty, R. (2023). Interview with Reema Nanavaty, Self-Employed Women's Association.

52 Beane, M. (2022). Today's Robotic Surgery Turns Surgical Trainees into Spectators: Medical Training in the Robotics Age Leaves Tomorrow's Surgeons Short on Skills. IEEE Spectrum, 59(8), 32-37. https://doi.org/10.1109/MSPEC.2022.9852406.

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

54 (Center for Democracy & Technology et al. 2022)

55 Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

56 Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.

57 Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on human-computer interaction*, 2(CSCW), 1-22.

58 Rosales, A., & Fernández-Ardèvol, M. (2019). Structural ageism in big data approaches. *Nordicom Review*, 40(s1), 51-64.

59 Klinova, K. (2022) Governing AI to Advance Shared Prosperity. In Justin B. Bullock et al. (Eds.), *The Oxford Handbook of AI Governance*. Oxford Handbooks.

60 Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021, May). Human-AI interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

61 Bernhardt, A., Suleiman, R., & Kresge, L. (2021). Data and algorithms at work: the case for worker technology rights. https://laborcenter.berkeley.edu/wp-content/uploads/2021/11/Data-and-Algorithms-at-Work.pdf.

62 Colclough, C.J. (2022). Righting the Wrong: Putting Workers' Data Rights Firmly on the Table. https://doi.org/10.7551/mitpress/13835.003.0023.

63 Pasquale, F. (2020). *New Laws of Robotics*. Harvard University Press.

64 Rodrik, D. (2022). 4 Prospects for global economic convergence under new technologies. An inclusive future? *Technology, new dynamics, and policy challenges*, 65.

65 Anderson, E. (2019). Private Government: How Employers Rule Our Lives (and Why We Don't Talk about it). Princeton University Press.

66 Institute for the Future of Work. (2023). *Good Work Algorithmic Impact Assessment Version 1: An approach for worker involvement*. https://www.ifow.org/publications/good-work-algorithmic-impact-assessment-an-approach-for-worker-involvement.

67 Bernhardt, A., Suleiman, R., & Kresge, L. (2021). Data and algorithms at work: the case for worker technology rights. https://laborcenter.berkeley.edu/wp-content/uploads/2021/11/Data-and-Algorithms-at-Work.pdf.

68 Colclough, C.J. (2022). Righting the Wrong: Putting Workers' Data Rights Firmly on the Table. https://doi.org/10.7551/mitpress/13835.003.0023.

69   Brand, J., Dencik, L. & Murphy, S. (2023). The Datafied
     Workplace and Trade Unions in the UK. Data Justice Lab.
     https://datajusticeproject.net/wp-content/uploads/
     sites/30/2023/04/Unions-Report_final.pdf.

70   Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021, May).
     Human-AI interaction in human resource management:
     Understanding why employees resist algorithmic
     evaluation at workplaces and how to mitigate burdens. In
     Proceedings of the 2021 CHI Conference on Human Factors in
     Computing Systems (pp. 1-15).

71   Mateescu, A., & Elish, M. (2019). AI in context: the labor of
     integrating new technologies.

72   Elish, M. C. (2019). Moral crumple zones: Cautionary tales
     in human-robot interaction (pre-print). *Engaging Science,
     Technology, and Society (pre-print).*