

**PARTNERSHIP ON AI**

Partnership on AI response to the USAID Request for Information (RFI) on the Global AI Research Agenda

Background

Partnership on AI (PAI) is a non-profit partnership of academic, civil society, industry, and media organizations creating solutions to ensure that AI advances positive outcomes for people and society. PAI studies and formulates sociotechnical approaches aimed at achieving the responsible development of artificial intelligence (AI) and machine learning (ML) technologies. Today, we connect over 100 partner organizations in 14 countries to be a uniting force for the responsible development and fielding of AI technologies.

PAI sets research and action agendas, and develops tools, recommendations, and other resources by inviting multistakeholder voices from across the AI community and beyond to identify areas of interest regarding AI's social impacts, define problems and opportunities in actionable granularity, build alignment on solutions, and share insights that can be synthesized into actionable guidance. We then work to promote adoption in practice, inform public policy, and advance public understanding. We are not an industry or trade group nor an advocacy organization. We aim to advance understanding, change practice, and inform policy.

The information in this document is provided by PAI and is not intended to reflect the view of any particular Partner organization of PAI. The comments provided herein are intended to provide evidence-based information, based on PAI's research, in response to USAID's RFI.

Executive Summary

Partnership on AI welcomes the Global AI Research Agenda to be published by USAID and the State Department. We especially welcome the focus on the global human impacts of AI, and labor impacts in particular, as areas that are critically important but often under addressed in safe and responsible AI efforts. Partnership on AI works across many topics relevant to this [RFI](#). This work includes our [Guidance for Safe Foundation Model Deployment](#) (which contains specific guidance for ensuring safety and responsibility in research releases, relevant to the RFI question regarding research best practices for foundation model developers and downstream users¹), our [Global Task Force for Inclusive AI](#), a [white paper](#) and [article](#) on safe publication norms for AI, and [recommendations for how AI conferences can create cultures of responsibility](#), written in collaboration with international partners at CIFAR (Canada) and The Ada Lovelace Institute (United Kingdom). Each of these resources offers “currently used criteria and frameworks” on the RFI’s question regarding best practices to ensure AI research is “safe, ethical, and sensitive to global contexts.”

For this RFI, we have chosen to focus on our extensive work on AI’s labor and economic impacts, and make a number of recommendations about the research priorities for the Agenda.

Summary of Recommendations

Below we have set out recommendations to inform the Global AI Research Agenda, with a particular focus on human impacts:

1. **Based on findings from PAI’s research, the Global AI Research Agenda should prioritize in-depth research of the data supply chain to:**
 - a. extensively map various types of employment models that data enrichment workers fall under
 - b. assess the prevalence of each employment model
 - c. analyze the diverse actors involved across the global supply chain
 - d. investigate the impact of these different configurations on workers engaged in this global and digitally arbitrated labor market.

Developing a more comprehensive understanding of the dynamics of this nascent labor market and addressing its opacity will enable policymakers, NGOs, and other key stakeholders to strategically target interventions aimed at improving conditions for workers and their communities.

2. **The Global AI Research Agenda should prioritize empirical studies examining whether and to what extent the circumstances surrounding dataset construction affect their**

¹ We define research releases for foundation models as “Models released in a restricted manner to demonstrate research concepts, techniques, demos, fine-tuned versions of existing models. The release is meant to share knowledge and allow others to build upon it and excludes small-scale individual projects.”

effectiveness, reliability, and safety measures. By linking the development process to safety considerations, this research agenda has the potential to influence decision-making within the broader AI ecosystem—including the decisions of developers, vendors, service providers, data brokers, and procurers of AI models/datasets. Importantly, this research has the potential to demonstrate to actors across the AI supply chain that there is a need to emphasize and formalize the workforce responsible for enriching data used to train and fine-tune AI models, and help to define and measure the effects on other impacted stakeholders, geographies and environments.

3. **The Global AI Research Agenda should include a focus on social science and economic research that can be used to design markets that promote fairer outcomes for those contributing to the data/knowledge economy that powers AI development and maintenance.** Investment in research to shape our understanding of how innovative economic models, fiscal policies, and trade policies could yield more socially desirable outcomes could promote safe, responsible, and rights-affirming development and deployment of AI abroad, consistently with section 11(c) of the [Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence](#).
4. **The Global AI Research Agenda should include job quality impacts in addition to job availability and wage impacts in its focus on labor market impacts, and should also include the following, under-addressed areas:**
 - a. AI impacts on informal labor markets.
 - b. Anticipated AI exposure in low and middle income economies, by country.
 - c. Labor market impacts beyond direct displacement, including cross-border effects.
 - d. Potential next order effects produced by rapid rises in economic inequality and disruption of economic growth pathways, and mitigation approaches.
 - e. Applications and use cases that support the UN Sustainable Development Goals.
 - f. Mechanisms to strengthen worker voice and worker protection.
5. **It is critical for the Global AI Research Agenda to be both interdisciplinary and inclusive of voices outside academia.** The US AI Executive Order specifically asks for the Global AI Research Agenda to address AI's 'labor-market implications across international contexts, including by recommending risk mitigations'. **In order to achieve this goal in a meaningful and equitable way, it is important to provide opportunities for non-academic participation and engagement, with a specific focus on the inclusion of perspectives of workers and impacted communities.** Lived experience and knowledge should play a central role in the agenda to achieve the EO mandate and impactful research. See PAI's [brief for the United Nations Department of Economic and Social Affairs' 2023 Multi-Stakeholder Forum](#) on Science, Technology, and Innovation for the SDGs on the importance of this approach.
6. **Finally, we encourage the United States Agency for International Development, the U.S. Department of State and The National Science Foundation (NSF) to work with organizations such as Partnership on AI as they design and implement the Global Research Agenda, as this is likely crucial for high quality research.** We recognize the

challenge of trying to deliver a global research agenda, while ensuring it is not ‘too top-down’. Working closely with organizations, in the design and delivery of the agenda, will ensure that innovative and novel research approaches are reflected in the agenda, the needs of the field are embedded, and that the agenda targets the right questions while being guided by the views of different stakeholders across industry, civil society and academia. **PAI would be happy to collaborate and support this initiative.**

Transparency for Data Supply Chains – and the Labor that Powers these Supply Chains

Building safe and ethical models requires greater transparency and scrutiny over the data supply chains that enable AI models.

As the impressive capabilities of AI have captured the world’s attention, calls for research on how to ensure AI models are safe and beneficial for society have also grown.² These initiatives tend to focus on addressing the potentially harmful *impacts* or *consequences* of AI models once deployed. A focus on impact is logical given the range of risks and concerns that arise due to the application of AI, from lost jobs³ and the spread of disinformation⁴ to expanded surveillance capabilities⁵ and privacy violations.⁶

However, advancing safe, ethical, and beneficial AI models requires that we also examine the data supply chains that underlie these models – and the labor that powers these supply chains. This is a significantly absent part of the AI Safety discussion that warrants greater attention, and we advise NSF and USAID to hone in on using its AI research agenda.

AI models are built on large datasets. Understanding the conditions under which these datasets are created is critical to being able to assess whether the resulting AI models are safe, reliable, and applicable to a particular use case. Currently, there is very little transparency over the data supply chain and the AI production process. In addition to having important implications for assessing the quality and applicability of a particular AI model, developing greater transparency over the production process is also important to ensure that AI models are built in a way that respects human rights.

Poor working conditions may result in low quality data, and consequently unreliable AI or harmful models.

Advancing AI safety requires improving labor conditions for data enrichment workers by increasing transparency and accountability across the data supply chain.

² <https://sciencebusiness.net/news/ai/eu-and-us-work-together-ai-guidelines> ; <https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute>

³<https://jobs.washingtonpost.com/article/ai-and-job-displacement-the-realities-and-harms-of-technological-unemployment/>

⁴<https://apnews.com/article/artificial-intelligence-elections-disinformation-chatgpt-bc283e7426402f0b4baa7df280a4c3fd>

⁵ <https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847>

⁶ <https://www.reuters.com/legal/legalindustry/privacy-paradox-with-ai-2023-10-31/>

Improving working / human impact conditioning through industry practice: A PAI case study with DeepMind

In our case study⁷ with DeepMind, we found that adoption of worker-centric guidelines⁸ and related internal governance mechanisms, such as an ethics review process that standardized various teams data enrichment practices, can improve labor conditions and enhance transparency across the data sourcing pipeline.

Adopting these types of practices can promote an “ecosystem of responsibility”⁹ within a company and make relevant decision makers accountable to their workers.

Furthermore, we found anecdotal evidence that implementation of these accountability mechanisms improved data quality, which will enable the long-term development of safer and more equitable AI models. While future studies should be done to quantify the impact of adopting these guidelines on data quality, these early findings highlight the need to prioritize data enrichment workers in advancing AI safety efforts.

The RFI specifically asks ‘what considerations are most important for safe and ethical research into the human impacts of AI systems’? **Building on the findings from PAI’s research, the Global AI Research Agenda should consider and prioritize:**

- The role that internal governance mechanisms can play in improving labor conditions and enhancing transparency
- Promoting empirical studies examining how the circumstances surrounding dataset construction can affect their effectiveness, reliability, and safety measures

Given the central role of datasets in building AI models, the lack of attention and transparency over the data supply chains makes it difficult to understand and assess characteristics of the resulting AI models that may be of interest to the broader society; this includes both harms associated with the model itself, as well as those associated with its development. Developing greater transparency and accountability of the underlying datasets can better position us to understand, critique, and improve AI models for society.

The Impact of Model Development on Global Workers

When trying to better assess the impact of AI systems on humans, it is critical to evaluate the way in which humans are impacted both in the development and deployment of those models. More research is needed to:

- 1) **Map how this labor market is developing** – including the prevalence of different employment models and their respective impact on data enrichment workers.

⁷Jindal, Sonam. "Implementing Responsible Data Enrichment Practices at an AI Developer." (2022).

⁸ <https://partnershiponai.org/wp-content/uploads/2022/11/data-enrichment-guidelines.pdf>

⁹Jindal, Sonam. Ibid.

2) Explore how different economic models and policies may lead to more equitable outcomes for those contributing to the training of these AI models.

The sensationalized narrative around building technology that is potentially capable of human-like reasoning largely glosses over the fact that human intelligence and reason are central to training, building and maintaining useful AI models. Millions of people from around the world, known as “data enrichment workers,” prepare, clean, label, annotate, and otherwise enrich the data that is used to train AI models or provide human review of algorithmic outputs. **Some estimate that nearly 163 million people worldwide are involved in the type of data enrichment work necessary to develop AI.**¹⁰

Yet the central role of these human contributors in the development process is at odds with how they are treated and compensated. These workers are often not recognized for this critical role and instead, face precarious working conditions. Companies, researchers, and labs training AI models ought to be more intentional about ensuring that workers are treated fairly. More research is needed to understand how decisions are made across the data supply chain and how these various decisions impact the experience of these workers and the resulting AI models. Rather than celebrating and recognizing the critical importance of human intelligence in fuelling the AI advances that have captured our imagination, data enrichment work remains undervalued, underpaid and underappreciated.

Consistent with broader outsourcing trends, much of this work is done in low-income countries in the Global South, where lower wages can be paid. In addition to low pay and wage uncertainty, data enrichment workers face a lack of benefits, psychological harm from reviewing toxic content, lack of power to contest their conditions, unpredictable streams of work, high transaction costs for equipment and other forms of support to enable their work, and overall precarious conditions.^{11 12} The contributions of data enrichment workers at the heart of model development are structurally devalued.¹³

Currently, the data production process is often treated as an inconvenient means to the higher priority objective of building AI models. In turn, the data enrichment workers at the heart of model development are overlooked, often face precarious working conditions, lack basic labor protections, are compensated poorly, and are hidden from consumers of AI products.^{14 15} A crucial first step towards improving conditions for these workers is centering their contributions within the data supply chains that enable artificial intelligence.

¹⁰ Kässi O, Lehdonvirta V and Stephany F. How many online workers are there in the world? A data-driven assessment [version 1; peer review: 2 approved, 2 approved with reservations]. Open Res Europe 2021, 1:53 (<https://doi.org/10.12688/openreseurope.13639.1>)

¹¹ Mark Graham, Isis Hjorth, and Vili Lehdonvirta. 2017. Digital labor and development: impacts of global digital labor platforms and the gig economy on worker livelihoods. Transfer: European Review of labor and Research 23, 2 (May 2017), 135–162. <https://doi.org/10.1177/1024258916687250>

¹² Mary L. Gray and Siddharth Suri. 2019. Ghost work: how to stop Silicon Valley from building a new global underclass. Houghton Mifflin Harcourt, Boston.

¹³<https://partnershiponai.org/wp-content/uploads/2021/08/PAI-Responsible-Sourcing-of-Data-Enrichment-Services.pdf>

¹⁴ Mary L. Gray and Siddharth Suri. 2019. Ghost work: how to stop Silicon Valley from building a new global underclass. Houghton Mifflin Harcourt, Boston.

¹⁵ Graham, Hjorth, and Lehdonvirta. 2017. Digital labor and development.

Beyond this devaluation, there are several additional factors that contribute to the mistreatment of data enrichment workers, such as the incentives created by venture capitalists and broader economic trends towards deindustrialisation and the resultant creation of surplus labor.¹⁶ Availability of this undervalued labor at the margins, alongside the lack of attention placed on dataset construction¹⁷, has led to the creation of many on-demand service models that allow data enrichment projects to be spun up as needs arise. The global, digitally arbitrated nature of this labor market has meant that data enrichment workers lack formal protections and there is little oversight over their treatment. As a result, there is **limited transparency over the AI community's approach to data enrichment** and a lack of field-wide standards on how to do so.

In the short-term, AI developers can implement basic worker-centric [guidelines](#) when setting up a project involving data enrichment. These guidelines are based on the input from a multi-stakeholder community of data enrichment companies, civil society, AI practitioners, and leading researchers in the space. See [Partnership on AI Data Enrichment Sourcing Guidelines](#) for further information, along with some additional [resources](#) for AI developing companies.

In the long-term, we need concrete mechanisms to ensure transparency and accountability across the data supply chain. To enable greater accountability in the future, we are encouraging AI developing companies and data enrichment vendors/platforms to begin adopting the guidelines, develop internal governance mechanisms to create consistency in practices across their organizations, promote consistent practices across their supply chains, publish transparency reports with details about their practices and supply chain, and thoughtfully develop inclusive methods to have workers/worker representatives shape future iterations of best practices.

We can only achieve meaningful progress on the human impacts of AI systems by first engaging with the needs of the humans who power these systems, specifically the data enrichment workers who are routinely overlooked in the AI lifecycle. Much of PAI's work in this area has focused on actions AI companies can take today to improve conditions for workers and improve alignment with workers to yield higher quality datasets.

However there is still a great deal of opacity in the data supply chain and the data labor market. In order to enable policymakers, NGOs, worker organizations, and others to promote better conditions for data enrichment workers, it is absolutely critical to develop a more comprehensive understanding of the extended data supply chain. The Global AI Research Agenda can build on existing research in the field to more extensively map out the

¹⁶ Lilly Irani. 2015. Difference and Dependence among Digital Workers: The Case of Amazon Mechanical Turk. *South Atlantic Quarterly* 114, 1 (2015), 225–234. <https://doi.org/10.1215/00382876-2831665> ; Phil Jones. 2021. *Work without the worker: labor in the age of platform capitalism*. Verso, London. OCLC: on1263817278.

¹⁷ Ben Hutchinson, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes, and Margaret Mitchell. 2021. *Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure*. <https://doi.org/10.48550/arXiv.2010.13561> arXiv:2010.13561.

nature of these complex supply chains, the different actors across this supply chain, the prevalence of various types of employment models, trends in this labor market, and how these various factors shape workers' experiences.

Furthermore, as human contributions drive the growth of the AI industry, we must adapt our understanding of what constitutes work in the AI economy and how different types of labor generate value and should be valued. If we build an AI ecosystem that appropriately values these human contributions, we have the opportunity to build a more equitable economy in which more people benefit from AI advances. While AI certainly has the potential to transform the global economy, we have the power to design an economy that will enable AI development to better serve society's interests. The Global AI Research Agenda can explore how to better design markets such that these contributions to the AI economy are treated equitably.

Priorities to integrate into the Global AI Research Agenda for further work

To improve understanding of the effects of AI development and deployment on data workers, the Global AI Research Agenda should prioritize research on the data supply chain, including:

- Mapping the various types of employment models that data enrichment workers fall under
- Assessing the prevalence of these employment models
- Analyzing the diverse actors involved across the global AI supply chain, including their roles, geographical distribution and responsibilities
- Investigating the impact of the factors above on workers engaged in this global and digitally arbitrated labor market.

The Global AI Research Agenda could also benefit from investing in economic research that can be used to design markets that promote fairer outcomes for those contributing to the data/knowledge economy that powers AI development and maintenance.

Other Key Labor Market Issues For Inclusion in the Global AI Research Agenda

Partnership on AI welcomes the Global AI Research Agenda's inclusion of labor market impacts. We suggest that the research agenda, USAID, and other agencies acting to execute it follow a broad definition of "labor market impacts," and examine AI's effects on workers' ability to access high quality jobs. This definition of labor market impacts includes impacts on job availability, wages, and job quality, as set forth in Partnership on AI's [Redesigning AI for Shared Prosperity: an Agenda](#), report on [AI & Job Quality: Lessons from Frontline Workers](#), and [Guidelines for AI & Shared Prosperity](#) (a high level job impact assessment

including all of the specified dimensions, and sets of responsible practices for AI developing organizations and AI implementing organizations to respectively follow).

These reports and recommendations were created with the input and advice of a global, multistakeholder Steering Committee, including labor advocates and experts, academic economists, and AI industry practitioners, and with the intentional inclusion of representatives from the Global South.

Through the creation of this work, we identified several major gaps in existing research worthy of focus in the Global AI Research Agenda as it considers labor market impacts in Global South contexts. We recommend the Global AI Research Agenda include the following topics in its scope and suggested areas of research:

1. **Impact on informal labor markets.** 60% of the world's workers work in informal rather than formal employment.¹⁸ Initial reports indicate that AI is reshaping the conditions of those markets in ways that further advantage well-capitalized buyers over informal workers selling their products or services, by rebalancing the information dynamics within these marketplaces.¹⁹ Initial evidence on the ground suggests that informal workers are offered lower prices and made to take on higher levels of risk due to buyer confidence in predictions made by AI models, making it even more difficult for an already marginalized group of workers to make a living.
2. **Anticipated AI exposure in low and middle income economies, by country.** Task level assessments of AI exposure have become a common method to anticipate potential areas of labor market disruption.²⁰ While directional and partial (see next point), this work is beneficial for policymakers attempting to anticipate economic effects and which worker groups may benefit from reskilling and retraining. This method can be easily adopted in areas where there is sufficient data on occupations; where this data does not exist, alternative or proxy methods need to be developed.
3. **Labor market impacts beyond direct displacement, including cross-border effects.** Labor market displacement effects are often considered at the level of direct automation effects. However, as outlined in Partnership on AI's research paper "AI and Shared Prosperity," AI's impacts also include demand effects, vertical effects, horizontal effects, and factor reallocation, all of which need to be explored in more depth to better understand impacts on labor markets.²¹ These effects often take place across firms (e.g., using AI systems instead of BPO firms) and across borders, and are particularly important to understand given initial applications of AI (entry-level coding and data analysis, customer support) are heavily overlapping with intentional value-added or service economy development strategies undertaken by numerous middle-income countries.

¹⁸ International Labour Office. "Women and Men in the Informal Economy: A Statistical Picture (Third Edition)." International Labour Office, 2018. http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_626831.pdf.

¹⁹ Bell, Stephanie A., AI and Job Quality: Insights from Frontline Workers, 2022, 12. https://partnershiponai.org/wp-content/uploads/dlm_uploads/2022/09/PAI_paper_ai-job-quality-1.pdf

²⁰ See, for instance, Eloundou, Tyna, et al. "Gpts are gpts: An early look at the labor market impact potential of large language models." arXiv preprint arXiv:2303.10130 (2023).

²¹ Katya Klinova and Anton Korinek. 2021. AI and Shared Prosperity. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21). <https://doi.org/10.1145/3461702.3462619>

4. **Potential next order effects produced by rapid rises in economic inequality and disruption of economic growth pathways, and mitigation approaches.** AI has the potential to substantially increase economic inequality within and between countries. Major economic shocks have the potential to affect democratic stability and international stability, among other dynamics. We recommend a focus on understanding these potential disruptions as well as ways to avert or remedy them in more granularity; we offer an initial exploration of these impacts on international democracies and potential mitigations in our article “AI’s Economic Peril.”²²
5. **Applications and use cases that support the UN Sustainable Development Goals.** Current development trajectories for AI systems are heavily focused on finding market-sustainable applications for AI use within high income countries, a dynamic which not only leads to the disruption of economic development pathways abroad as described above, it monopolizes resources (AI talent, compute) away from potentially transformative applications of AI in pursuit of positive social impact. Identifying promising projects and applications in support of achieving the SDGs and meaningfully resourcing them offers an alternative, beneficial path for these increasingly capable technologies.
6. **Mechanisms to strengthen worker voice and worker protection.** Worker voice is critical to ensure AI use does not degrade worker well-being; moreover, worker input can help business leaders identify new and effective ways to improve business outcomes. However, AI is already repeating the harms of technologies past, including physical and mental harm through work intensification. A more comprehensive list of these harms and the drivers that produce them is included in our AI and Job Quality report.²³ Further research into locally feasible methods of protecting and enhancing worker voice and well-being, and ways to integrate these protections into local laws is critical to ensure workers are protected from the harms of AI and able to steer its use towards beneficial applications. In addition to legal provisions, research on effective mechanisms and methods to include worker voice in AI design, development, and use is nascent and largely focused on workers in high income countries. We recommend that the research agenda include further research into mechanisms for global worker voice in these areas of AI creation and implementation.

Contact

For any further information and questions related to this submission, contact Stephanie Bell stephanie@partnershiponai.org and Sonam Jindal sonam@partnershiponai.org (and please copy John Howell john@partnershiponai.org and policy@partnershiponai.org).

²² Bell, Stephanie A., and Anton Korinek. "AI's Economic Peril." *Journal of Democracy* 34.4 (2023).

²³ Bell, AI and Job Quality: Insights from Frontline Workers.