The Global **AI** Research Agenda

TABLE OF CONTENTS

GLOSSARY OF TERMS

AI: Artificial Intelligence

R&D: Research and Development

GAIRA: Global Artificial Intelligence Research Agenda

G7: Group of Seven

OECD: Organisation for Economic Co-operation and Development

UNGA: United Nations General Assembly

RFI: Request For Information

E.O.: Executive Order

U.S.: United States

G20: Group of Twenty

NIST: National Institute of Standards and Technology

NSF: National Science Foundation

UNESCO: United Nations Educational, Scientific and Cultural Organization

GBV: Gender-based Violence

TFGBV: Technology-facilitated Gender-based Violence

LMICs: Low- and Middle-Income Countries

SDGs: United Nations Sustainable Development Goals

AgTech: Agricultural Technology

LLMs: Large Language Models

LGBTQ+: Lesbian, Gay, Bisexual, Transgender, Queer, and Intersex

ISO: International Organization for Standardization

IEC: International Electrotechnical Commission

SDOs: Standards Developing Organizations

HICs: High-Income Countries

SMEs: Small- and Medium-sized Enterprises

PTSD: Post-traumatic Stress Disorder

B Corp: Benefit Corporation

ILO: International Labour Organization

STEM: Science, Technology, Engineering, and Math

EU: European Union

PPPs: Public-Private Partnerships

USAID: United States Agency for International Development

SECTION 1: EXECUTIVE SUMMARY

The rapid development of AI technologies is taking place in an interconnected landscape, where funding, data, talent, information, and computing resources flow across borders. Despite this global relevance, AI research and development (R&D) policies are most commonly driven by domestic priorities; funding silos and other barriers to collaboration can contribute to isolated networks that undermine the potential benefits of international scientific cooperation. The global research ecosystem should be founded on the scientific values of reciprocity, openness, and transparency, including for the study of AI systems.

The Global AI Research Agenda (GAIRA) recommends principles, priorities, and practices for AI research and development (R&D) to advance safe, secure, and trustworthy development of AI systems in international contexts. It aims to strengthen collaboration in researching the interactions between individuals, communities, and society with AI systems, foster innovation, and support equitable access to the benefits of AI. The conclusions presented serve as a starting point to align a global research vision, in which research communities continuously assess the state of AI research, review current publications addressing the presented priorities, and identify gaps to guide future research, focusing on global needs.

There is no international consensus on guiding principles for AI research, reflecting the still evolving state of AI research communities. In consideration of this lack of consensus, this agenda will be guided by three recommended principles for AI research: 1) Inclusion and Equity, 2) Responsible Research Conduct, and 3) Partnership and Collaboration. These principles emphasize the importance of building AI systems that benefit individuals, communities, and society, upholding the highest standards for preserving autonomy, dignity, and privacy, involving impacted people as collaborators, and striving for sustainability of systems and research. They resonate broadly with prior work outlined by researchers, funders, and multilateral forums and reflect a collective commitment to create an AI ecosystem that prioritizes human well-being and societal values, including respect for human rights. When designed, developed, deployed and used in a rights-respecting manner, AI can propel technological advances that benefit societies and individuals. These principles were developed through thematic analysis of both U.S. domestic and international guiding documents – namely Executive Order 14110, the G7 Common Values and Principles on Research Integrity and Research Security, the Organization for Economic Cooperation and Development (OECD) AI Principles, and the United Nations General Assembly Resolution on "Seizing the Opportunities of Safe, Secure, and Trustworthy AI Systems for Sustainable Development."
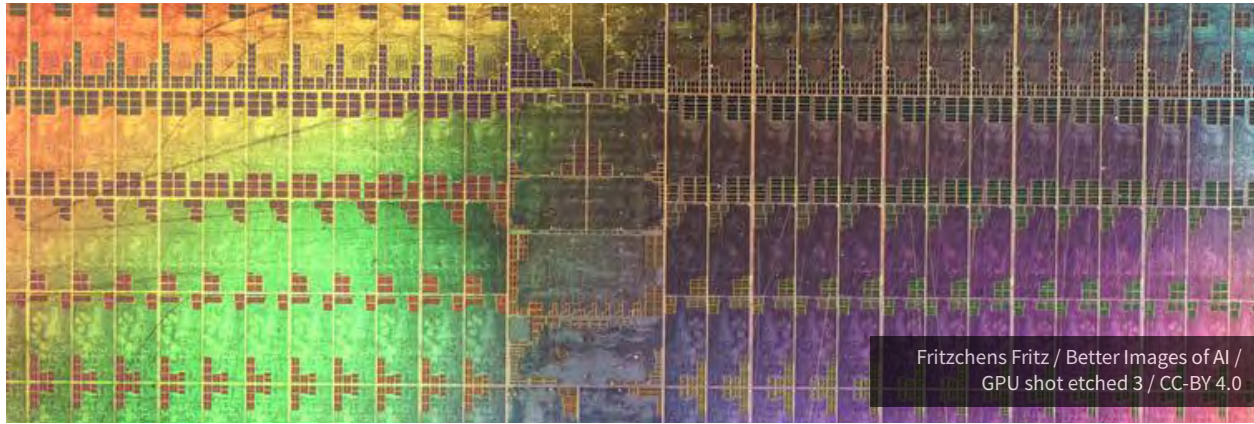
The **research priorities to advance safe, secure, inclusive, and trustworthy AI** presented in this document broadly align with the following themes:

- *Sociotechnical Research* - There is an urgent need for research that can deepen understanding of interactions between technology and society and guide the design and deployment of AI systems that enhance human well-being. The need to understand AI systems as sociotechnical systems—one in which technology interacts with and influences human behavior, culture, and institutions—is even more acute in a global context in which technologists and researchers may develop or deploy AI tools in unfamiliar environments. Research to monitor and evaluate the impacts of AI development and deployment in real-world settings is important to identify tangible benefits and risks of AI applications and promote and protect human rights and fundamental freedoms for all.

- *Inclusive Research Infrastructure* - A critical element for AI research progress is enabling researcher access to data, compute power, and research platforms, with appropriate safeguards to protect privacy and international human rights, that can support innovation in AI technologies and systems. Research infrastructure should be available for researcher access to foster innovation, high-quality research, and the dissemination of findings. Diverse perspectives will drive important discoveries, applications, and sensitivities critical to advancing AI technology around the world.

- *Research to Support AI for Global Challenges* - Prioritizing AI applications to address global challenges can promote environmental stewardship, economic resilience, and social well-being. Additionally, collaborative efforts can make AI systems, equipment, and infrastructure more sustainable, and performing applied research can advance a safer, healthier, and more prosperous world.

- *Fundamental Research on AI, including AI Safety, Security, and Trustworthiness* - AI is still in an early stage of development and more technical advancements are needed to advance the state-of-the-art and drive towards safe, secure, and trustworthy systems. To support these developments and contribute to international discussions on AI governance, researchers must also address issues related to information integrity, cybersecurity, and measurement and evaluation, among others.

- *Research on AI's Global Labor Market Implications* - AI's potential to automate processes and increase labor efficiency calls for research into the labor market impacts of AI. This section outlines opportunities to address research gaps, understand disruptions, and propose risk mitigation measures that fit within different global and economic contexts. Researchers should consider labor market shifts, the emergence of new types of work, and impacts on wealth distribution and income inequality.

The last section in this document lists **recommended practices for AI research** based in part on input received from the [Request For Information (RFI),](#)[1] informed by respondents from various regions, industries, and backgrounds. It provides a global perspective on what implementation practices could look like for research funders, research ecosystem hubs, and research teams. This section can serve as a guide for stakeholders to align actions with the agenda's goals.

- **For Research Funders:**
  - Elevate Transparency
  - Open Funding Calls to Support International AI Research Collaboration
  - Promote Research Infrastructure Accessibility for Diverse Demographics
  - Offer Connection Opportunities
  - Pursue Public-Private Partnerships

- **For Research Ecosystem Hubs:**
  - Incentivize Reproducibility of Research
  - Foster Broadly Accepted Practices Around Research Publications
  - Coordinate and Collaborate Around AI Research Guidelines
  - Support Research Beyond Topics Prioritized by the Private Sector
  - Provide Guidance to Navigate the Evolving Regulatory Landscape

- **For Research Teams:**
  - Prioritize Multidisciplinary Teams
  - Collaborate with Local Researchers
  - Make Use of Sociotechnical Methodologies and Research Design
  - Incorporate Risk Assessments

This research agenda, as part of a broad body of national and [international](#) literature,[2] underscores the importance of AI research collaboration, enabling stakeholders to leverage resources and competitive advantages to facilitate cutting edge AI research that upholds fundamental values, such as international human rights and the protection of privacy. By focusing efforts on both the technical advancements of AI systems and the interactions between AI and communities, the global research community can foster safe, secure, and trustworthy AI development. Progress in creating this balanced ecosystem should be based upon promotion of these priority areas within international research networks and enduring conversations with global research stakeholders as the field transforms.

---

1   "Global AI Research Agenda." 2024. Federal Register. March 14, 2024. https://www.federalregister.gov/documents/2024/03/14/2024-05357/global-ai-research-agenda.
2   Kerry, Cameron F., Joshua P. Meltzer, and Andrea Renda. "AI Cooperation on the Ground: AI Research and Development on a Global Scale." Brookings, November 4, 2022. https://www.brookings.edu/articles/ai-cooperation-on-the-ground-ai-research-and-development-on-a-global-scale/.

SECTION 2: INTRODUCTION

2.1 Executive Order on Safe, Secure, and Trustworthy Development and Use of AI

The Global AI Research Agenda (GAIRA) is published pursuant to Section 11(c)(ii) of Executive Order 14110,[3] Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. The Executive Order outlines a whole-of-government strategy, promoting global practices through multilateral fora, strengthening partnerships to address shared AI challenges, and supporting global capacity building. The Executive Order states: "[t]he Secretary of State and the Administrator of the United States Agency for International Development, in collaboration with the Secretary of Energy and the Director of NSF, shall develop a Global AI Research Agenda to guide the objectives and implementation of AI-related research in contexts beyond United States borders. The Agenda shall:

(A) include principles, guidelines, priorities, and best practices aimed at ensuring the safe, responsible, beneficial, and sustainable global development and adoption of AI; and

(B) address AI's labor-market implications across international contexts, including by recommending risk mitigations."

Complementary to the concepts outlined in this Global AI Research Agenda is another E.O deliverable, the AI in Global Development Playbook,[4] which aims at promoting the responsible use of AI for global good while mitigating potential risks and challenges.

---

3   "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." 2023. Federal Register. November 1, 2023. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.
4   "U.S. Agency for International Development. AI in Global Development Playbook. Washington, D.C.: USAID, 2024. http://www.usaid.gov/digital-development/ai-global-development-playbook

2.2 National AI R&D Strategic Plan

The GAIRA is meant to supplement the U.S. [National AI R&D Strategic Plan,][5] updated in May 2023, by exploring tangible opportunities to advance international research collaboration for safe, secure, and trustworthy AI systems in global contexts. Namely, this agenda explores more in-depth the opportunities to galvanize the research community towards the following priorities:

- **Strategy #1:** Make Long-Term Investments in Fundamental and Responsible AI Research

- **Strategy #2:** Develop Effective Methods for Human-AI Collaboration

- **Strategy #5:** Develop Shared Public Datasets and Environments for AI Training and Testing

- **Strategy #9:** Establish a Principled and Coordinated Approach to International Collaboration in AI Research

U.S. Government science funding agencies, such as the National Science Foundation and Department of Energy, actively contribute to global AI research, in line with the National AI R&D Strategic Plan. The publication of the GAIRA is not accompanied by any additional U.S. Government funding, but it does serve as an outline and expansion of key areas for AI research already funded by science agencies. The GAIRA emphasizes the need for international collaboration and coordination to address these areas and to derive the positive benefits of developing safe, secure, and trustworthy AI systems through existing U.S. research funding commitments.

2.3 International AI Research Efforts

Principle 2.1[6] of the [OECD Principles for Trustworthy AI][7] calls for governments to facilitate public and private investment in AI research and development, emphasizing the importance of basic research and acknowledging the extended timeframes involved in research policy. The principle highlights long-term public investment, investment in open science, and interdisciplinary efforts as essential to spur innovation in trustworthy AI, as scientific breakthroughs enabled by AI could help solve societal challenges and create entirely new industries. This type of investment, paired with evidence-based understanding of the risks, capabilities, and limitations of AI systems, and when tailored to the appropriate [level of AI-preparedness,][8] is essential to drive and shape trustworthy AI innovation and ensure beneficial outcomes for all.

Many organizations, fora, and countries are actively pursuing research priorities, and the GAIRA is complementary to a mosaic of existing research globally. This includes research outputs as directed and published by the United Nations, OECD, Global Partnership on AI, G20, G7, and the International Scientific Report on the Safety of Advanced AI, among others.

5   Select Committee On Artificicial Intelligence, National Science and Technology Council, Arati Prabhakar, Kei Koizumi, Laurie Locascio, Sethuraman Panchanathan, Geraldine Richmond, et al. 2023. "National Artifical Inelligence Research and Development Strategic Plan 2023 Update." https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf.
6   "Investing in AI Research and Development (OECD AI Principle) - OECD.AI." n.d. https://oecd.ai/en/dashboards/ai-principles/P10.
7   "AI Principles Overview - OECD.AI." n.d. https://oecd.ai/en/ai-principles.
8   International Monetary Fund. "Artificial Intelligence Potential Index (AIPI)." IMF DataMapper. Accessed September 12, 2024..

2.4 Defining the Global Artificial Intelligence Research Agenda

The Global Artificial Intelligence Research Agenda (GAIRA) aims to inform the objectives and implementation of AI-related research, globally. To achieve this ambition, it has four interrelated subgoals:

1. **Advance multifaceted research collaborations and leverage them to promote the safe, secure, and trustworthy development of AI technologies around the world in a manner that is both inclusive and sustainable.** Building stronger international partnerships to address global priorities and promoting responsible research practices can positively influence AI development and help achieve this subgoal.

2. **Outline important areas of inquiry, including the impacts of the development and use of AI systems on economies and societies around the world.** Researchers are still in the early stage of understanding many technical aspects of recent AI advances and how they may interact with and influence economies, societies, and cultures. Since the development and use of these technologies can impact across national borders, these inquiries need to take into consideration language, culture, geography, environment, socioeconomic development, economic competitive dynamics, international human rights and fundamental freedoms, and conflict dynamics. AI's global spread will also be shaped by digital divides, in which the costs and benefits of AI will accrue differently to people with different levels of connectivity, access, and digital literacy.

Shutterstock Standard License / Photo contributor Boijonell Prod

3. **Elevate opportunities for academics and researchers globally to contribute to a body of evidence to guide policy and decisionmakers towards promoting equitable access to the benefits of AI and support implementation of the AI in Global Development Playbook.** The AI in Global Development Playbook offers a roadmap to develop the capacity, ecosystems, frameworks, partnerships, and institutions to leverage safe, secure, and trustworthy AI for sustainable development and foster good governance regimes for AI. While the Playbook recommendations are geared towards governmental, private sector, and development donor and philanthropic stakeholders, the diffuse, fast-changing nature of AI technology means no small group of stakeholders alone will be successful in delivering the benefits of AI or mitigating its risks. Collaboration between researchers and academia globally can bolster a robust body of research necessary to inform evidence-based decision-making amongst these stakeholders long-term.

4. **Outline needed areas of research on the labor market implications of AI systems across international contexts.** AI's impacts on employment and labor may lead to unpredictable changes in the quantity, profitability, safety, quality, and nature of work, driving new concerns about the well-being of workers, particularly in diverse global settings. These changes are impacting different economies around the world in disparate ways, as AI can also create job opportunities and drive economic growth, particularly in developing regions, by enabling access to new markets and fostering innovation. Research into the varied aspects of how AI is impacting the labor market and workers can offer a more nuanced analysis of the benefits and risks of applications.
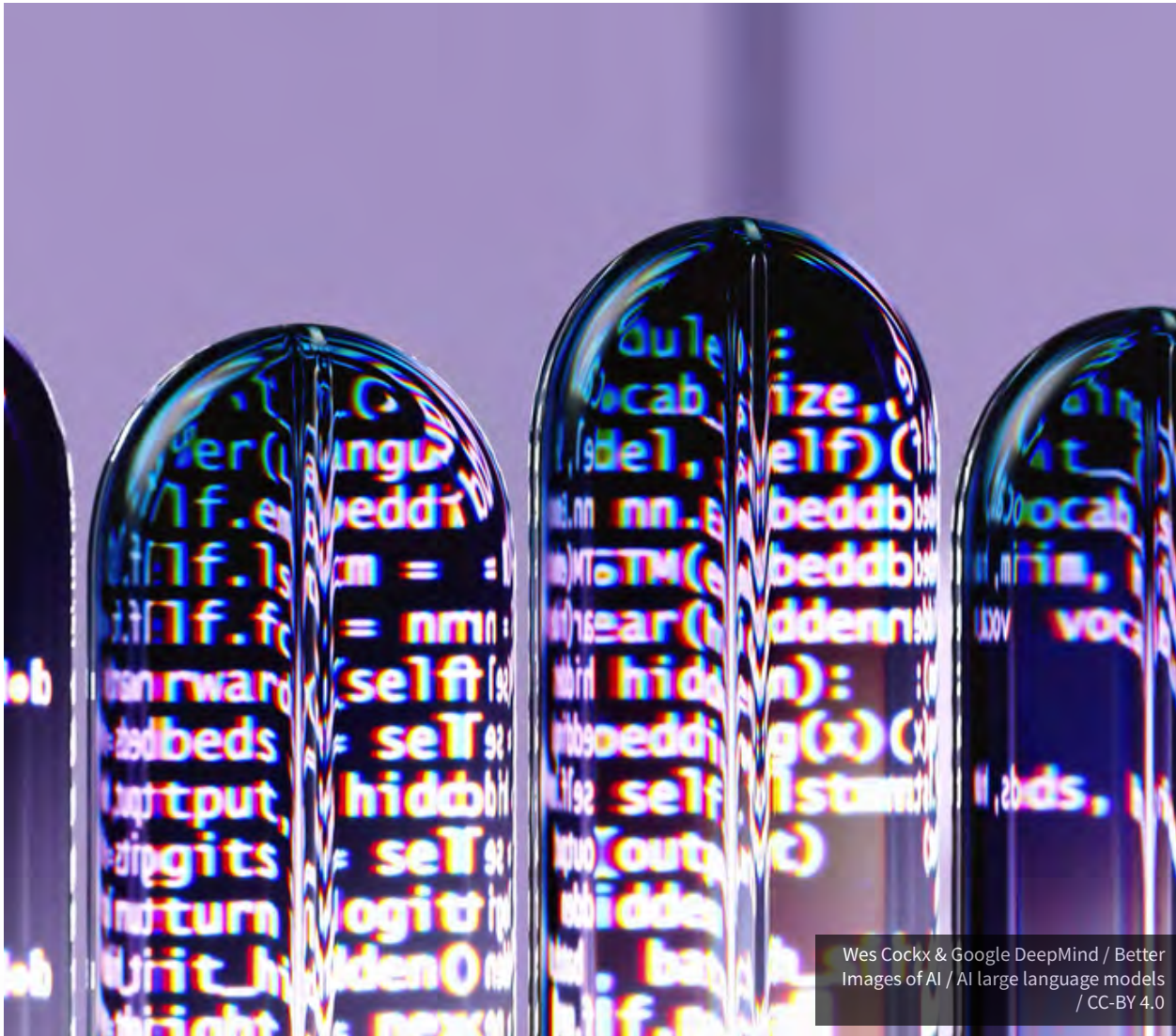
2.5 Focus and Audience

The rapid advancement of AI technologies holds significant implications for economies and societies, globally. Uneven international uptake of AI can be detrimental if societal impacts and implications are overlooked. This research agenda proposes a holistic plan to advance AI's interaction with human societies and technical systems addressing gaps in understanding AI's impacts, promoting evidence-based assessments, and encouraging international collaboration around the technical underpinnings of AI systems. The audience of the GAIRA includes research funders, policymakers, and practitioners globally, aiming to promote responsible AI design and adoption through research partnerships and to deploy AI for positive economic benefit.

The GAIRA drafting team invited the scientific community and global stakeholders to help shape the principles, guidelines, priorities, and recommended practices outlined in this research agenda and sought stakeholder feedback via an RFI. Doing so ensured the GAIRA content was informed by both local and global perspectives, fostering international collaboration and promoting cooperative practices across regions.

SECTION 3: RECOMMENDED PRINCIPLES FOR AI RESEARCH

Principles can help to inform research priorities as well as behaviors and actions of the research community. Since 2016, numerous documents[9] articulating principles for responsible and ethical AI have been released by companies, governments, professional organizations, multilateral forums, and nonprofits.

Researchers have identified some thematic convergence[10] around high-level principles, and many governments have begun to adopt internationally-negotiated principles such as the OECD's Principles for Trustworthy AI. Research integrity and security principles are also an active focus of scientific stakeholders, and of countries domestically and internationally. At the same time, there is still a lack of consensus on exactly what unique principles should apply to AI research--, as opposed to AI systems themselves or the research process writ large.

The GAIRA will be guided by a set of three recommended principles that resonate broadly with existing frameworks in this space.

- **Inclusion and Equity** -- Research should focus on developing AI systems that are representative, flexible, and fit for their context. Technology systems impact the social contexts into which they are deployed; to support holistic research evaluation and analysis, and equitable access to the benefits of AI systems, researchers should strive for inclusive research processes.

- **Responsible Research Conduct** -- AI research should uphold the highest standards for preserving the autonomy, dignity, well-being, and privacy of people who are directly and indirectly impacted by the research and the development of any resulting systems, including data subjects, platform testers, end users, and workers engaged through data enrichment or crowdsourcing platforms. Researchers should implement frameworks to systematically evaluate and optimize practices, to help ensure the process and outputs are both sustainable and responsible.

- **Partnership and Collaboration** -- AI researchers should seek ways to involve impacted individuals, communities, and organizations as collaborators early in the research process, especially in international contexts where local stakeholders have necessary and valuable perspectives about potential use cases, benefits, and risks of the research. Tools, guidelines, and training for participatory AI design[11] and research are urgently needed to make these practices more accessible for AI practitioners.

9 "AI Ethics Guidelines Global Inventory - AlgorithmWatch." n.d. AlgorithmWatch. https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/.

10 Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." Nature Machine Intelligence 1 (9): 389–99. https://doi.org/10.1038/s42256-019-0088-2.

11 Madaio, Michael A., Shivani Kapania, Rida Qadri, Ding Wang, et al. "Learning about Responsible AI On-The-Job: Learning Pathways, Orientations, and Aspirations." In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2024. https://doi.org/10.1145/3630106.3658988.

These recommended principles draw from converging areas discussed in relevant national and international documents, such as the following:

- **Executive Order 14110:** At the national level, the President of the United States signed E.O. 14110 "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" in October 2023, which in addition to driving federal actions, also outlines a set of guiding principles to guide the federal government in addressing AI. These include promoting responsible innovation, competition, and collaboration; advancing equity and civil rights; protecting privacy and civil liberties; among others.

- **G7 Common Values and Principles on Research Security and Research Integrity:** In the 2021 G7 Research Compact[12] governments committed to the promotion of international research cooperation and the conditions of freedom, independence, openness, reciprocity, and transparency under which it flourishes. The G7 Principles of Research Security[13] subsequently outlined by the G7 provide a common framework to guide responses to research security risks in domestic and international research and recognizes the foundational importance of research integrity to academic discovery. Importantly, the framework notes that openness and security are not contradictory, but complementary and mutually reinforcing.

- **OECD AI Principles:** At the multilateral level, in 2024, the OECD revised its AI Principles[14] which were initially released in 2019. These reflect important considerations for risk management, as outlined in the NIST AI Risk Management Framework,[15] and define a values-based approach to the creation and management of AI systems which should also drive the global policy and research agenda. These include inclusive growth, sustainable development and well-being; human rights and democratic values, including fairness, civil liberties, and privacy; transparency and explainability; robustness, security, and safety; and accountability.

- **UNGA AI Resolution:** At the global level, in 2024, The United Nations General Assembly 2024 resolution A/78/L.49[16] on "Seizing the Opportunities of Safe, Secure, and Trustworthy AI Systems for Sustainable Development" fostered a global vision to inform the design, development, and use of AI. The resolution outlines a shared vision for AI systems that are: (1) human-centric, (2) reliable, (3) explainable, (4) ethical, (5) inclusive, (6) privacy preserving, and (7) responsible. By combining a "sustainable development oriented" view of AI "in full respect, promotion, and protection of human rights and international law," this document affirmed the compatibility of human rights-respecting AI and sustainable development, and that nations support pursuit of both together.

12   "G7 Common Values and Principles on Research Security and Research IntegrityFootnote 1." 2023. September 13, 2023. https://science.gc.ca/site/science/en/safeguarding-your-research/general-information-research-security/international-research-security-resources/g7-common-values-and-principles-research-security-and-research-integrity.

13   G7 Working Group on the Security and Integrity of the Global Research Ecosystem (SIGRE). 2022. "G7 Common Values and Principles on Research Security and Research Integrity." https://www.bmbf.de/SharedDocs/Downloads/de/2022/220812-g7-sigre-paper.pdf?__blob=publicationFile&v=2

14   "OECD Legal Instruments." n.d. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

15   U.S. Department of Commerce, National Institute of Standards and Technology. 2023. "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST AI 100-1. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

16   "General Assembly Adopts Landmark Resolution on Artificial Intelligence." 2024. UN News. March 21, 2024. https://news.un.org/en/story/2024/03/1147831.

SECTION 4: RESEARCH PRIORITIES TO ADVANCE SAFE, SECURE, INCLUSIVE, AND TRUSTWORTHY AI

The following section covers four broad priorities to advance AI research fields towards safe, secure, and trustworthy AI. These priorities encompass methodologies, research infrastructure needs, applied research, and fundamental research. International cooperation is crucial to facilitate the sharing of resources and best practices across borders, fostering a collaborative environment that accelerates innovation and supports equitable access to AI benefits.

4.1 Sociotechnical Research

As AI tools become more widespread, they will become more integrated within economies and societies. Sociotechnical research methods can enhance understanding of the interactions between people and technology, improving researchers' ability both to design and use AI to advance human well-being and to support effective AI governance and policymaking. The following sociotechnical research priorities help to advance these goals.

4.1a Interactions between AI, People, and Societal Systems

AI technologies do not operate in a vacuum; rather they emerge from and constantly interact with people and societal structures. These systems reflect the values, perspectives, biases, and decisions of their developers. In turn, people are directly and indirectly impacted and influenced by AI systems: they are present (or not) in the underlying datasets; they act on and react to system outputs; and their lives can be positively or negatively shaped even by systems they did not interact with. Understanding AI's benefits and risks thus requires understanding the broader interactions between individuals, technology, and the contexts in which they operate. It also requires understanding what inequalities or biases AI tools may reflect—not simply quantifying bias but assessing how technologies impact marginalized people in different contexts, and the broader and longer-term implications of AI and automation more generally on equity, democracy, and human rights and fundamental freedoms. Effective research into these interactions will benefit from collaboration between technologists and experts in social sciences and the humanities fields, such as psychology, design science, law, sociology, anthropology, political science, and economics, among others.

Global Research Opportunities:

- Understand the needs of global stakeholders for AI systems and document the experiences of AI users and other individuals and communities across a variety of contexts, with particular emphasis on improving well-being for marginalized groups.

- Develop human-AI interaction paradigms that are useful for diverse users and help people build any necessary mental models about the systems. This includes paradigms to help people understand how AI systems arrive at their outputs, when they can be relied upon, and situations in which they should not be used, and to build resiliency to automation bias.

- Develop methodologies and testbeds for evaluating AI systems that go beyond computer modeling performance evaluation to investigate real world interactions with individuals and groups.

- Study the impacts of AI tools on social systems across cultures, such as measuring family formation, religious belief, and community cohesion, as well as individual and community-based health outcomes such as for mental health.

- Characterize how people around the world make sense of and react to AI system outputs, such as predictions, recommendations, or content, including how their reactions vary depending on system design and context and how those reactions relate to trust.

- Identify how AI may create systemic risks to economies, societies, cultures, and governments, including systemic harms that may arise from effects that are beneficial on the individual level.

4.1b Global Perspectives on AI Misuse

People around the world are grappling with the risks of AI misuse— ranging from individual-level fraud and impersonation scams to unlawful and arbitrary surveillance, information manipulation, and censorship. State and nonstate actors are increasingly using AI tools to undermine democracies through false and misleading information, influence in international institutions, and surveillance of opponents. Globally, AI misuse can exacerbate repression and contribute to active conflicts. Research into AI misuse and its impacts can provide policymakers, lawmakers, and technology leaders with insight into how to build effective legal, normative, and technical safeguards to mitigate misuse.

Global Research Opportunities:

- Document the ways and the extent to which AI tools are being used for repressive or illegal aims, and how patterns of misuse are affected by technological and political change.

- Understand which legal, normative, and technical safeguards and interventions are most effective at preventing the misuse of AI by governments and state-affiliated actors, during periods of both active conflict and relative political stability.

- Explore which aspects of AI system design can make those systems more difficult to misuse. This could include protections related to data (such as privacy-enhancing technologies) or emerging techniques designed to constrain the behavior of generative AI systems.

Adrien Limousin / Better Images of AI / Non-image / CC-BY 4.0

4.1c Gender Implications and & Gender-Based Violence

Women and girls are common targets of AI-facilitated abuse-- the vast majority[17] of synthetic content videos created by AI involve sexual depictions of women and girls. In addition, a 2024 UNESCO report[18] highlights a variety of ways in which generative AI tools could be used to facilitate gender-based violence[19] (GBV).

Technology-facilitated gender-based violence (TFGBV)[20] is a subset of gender-based violence that describes any act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or other digital tools, that results in or is likely to result in harm, or other infringements of rights and freedoms. These problems are likely to become even more acute as the use of AI increases globally, intersecting with the gender digital divide.[21]

Global Research Opportunities:

- Determine which legal, normative, and technical safeguards and interventions are most effective at preventing AI systems from being used to facilitate gender-based violence and how robust these safeguards are across languages and models.

- Explore the extent to which common risk management strategies, such as algorithmic auditing and gender-inclusive design teams, result in AI systems that more effectively promote gender equity and safe access to and use of AI systems.

- Document the extent of technology-facilitated gender-based violence globally and the prevalence and impacts of that violence, with particular attention to intersectional aspects such as ethnicity, religion, disability, sexual orientation, gender identity, and socioeconomic status.

- Evaluate opportunities of AI systems to address gender inequality and the perpetuation of gender biases, with particular attention to intersectional aspects such as ethnicity, nationality, religion, disability, sexual orientation, gender identity, and socioeconomic status.

---

17 "2023 State of Deepfakes: Realities, Threats, and Impact." n.d. https://www.securityhero.io/state-of-deepfakes/.
18 UNESCO, Rumman Chowdhury, and Dhanya Lakshmi. 2023. "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI. Vol. World trends in freedom of expression and media development. https://unesdoc.unesco.org/ark:/48223/pf0000387483.
19 "United States Strategy to Prevent and Respond to Gender-Based Violence Globally 2022 - United States Department of State." 2023. United States Department of State. February 27, 2023. https://www.state.gov/reports/united-states-strategy-to-prevent-and-respond-to-gender-based-violence-globally-2022/.
20 "Frequently Asked Questions: Tech-facilitated Gender-based Violence." n.d. UN Women – Headquarters. https://www.unwomen.org/en/what-we-do/ending-violence-against-women/faqs/tech-facilitated-gender-based-violence.
21 Ramundo Media for DAI and Miles Sedgwick, Rana Labs. 2020. The Gender Digital Divide Primer. https://www.usaid.gov/sites/default/files/2022-05/DAI-1089_GDD_Primer-web_rev1_9.6.21.pdf

4.2 Inclusive Research Infrastructure

Trusted international AI research infrastructure refers to a trusted global network of facilities, resources, and standards that support AI research and development technologies. Trusted international research infrastructure is crucial for enabling researchers everywhere to participate in cutting-edge AI research, and can access resources, best practices, and knowledge across borders on voluntary and mutually agreed terms. Building robust and interconnected research infrastructures on a global scale will also help promote interoperable and globally accepted practices from the bottom-up, promoting consistency and trust in AI systems worldwide. Accessible and widely available research infrastructure can also facilitate the participation of researchers from marginalized groups into the AI research and development ecosystem, ensuring that diverse perspectives are represented and potentially reduce bias in the development and training of AI systems.

4.2a Open-Source AI Software, Widely Available Model Weights, Open Datasets, and Open Source Enabling Tools

Multiple forms of openness in AI models, training data, and enabling software have long been an essential part of the AI research and development ecosystem. Open-source software used to train, evaluate, and deploy models in industry and academia makes robust and well-documented design patterns in machine learning accessible to range of researchers. The wide availability of training data is also useful in advancing the state of the art, including by allowing researchers to compare approaches, algorithms, and models across common benchmarks, though in some cases this utility needs to be carefully balanced against data privacy and confidentiality concerns. Fully open and reproducible AI models -- where associated code, training data, and model weights are all made public -- allow a wide set of researchers to approach otherwise impossible or prohibitively expensive research questions, such as questions related to the training procedures

Jazmin Morris & AI4Media / Better Images of AI / Braided Networks 1 / CC-BY 4.0

of large models. These forms of openness clearly foster innovation and collaboration in the AI community, allowing researchers to share their work globally and accelerate scientific progress. Researchers can leverage existing tools and software libraries reducing redundancy, development time, cost, and lowering barriers to entry. Openness also furthers transparency and reproducibility, enables more fulsome peer review and validation, and enhances the credibility and reproducibility of research results. Many open-source projects encourage continuous improvement and updates from the research community, keeping open-source products and tools up-to-date and at the state-of-the-art. Supporting the continued development of open and reproducible AI research means developing a better understanding of how openness and confidentiality across training data, model weights, and tooling interact.
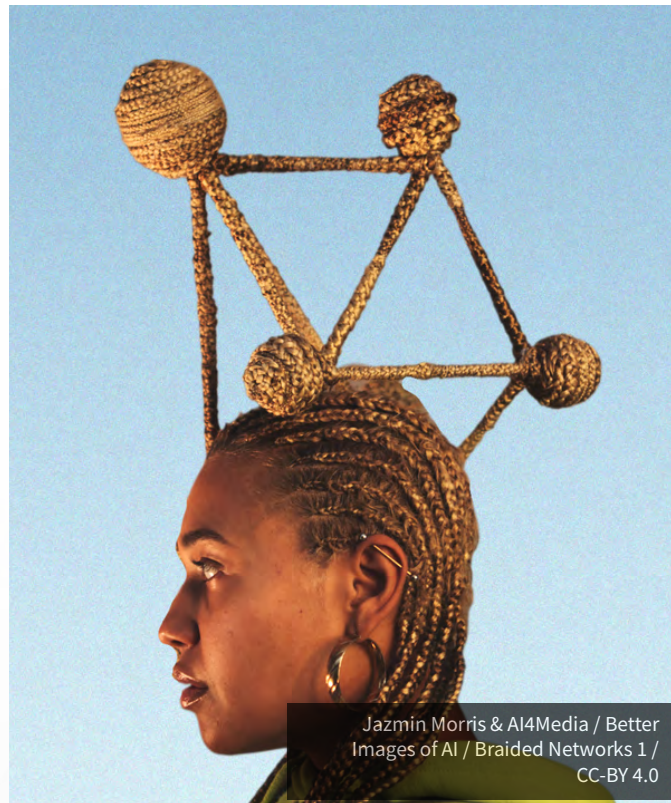
Global Research Opportunities:

- Develop new tools, training, and communities of practice to make open models, datasets, and libraries more collaborative, accessible, and transparent to researchers across and within a variety of application areas.

- Characterize the relationship between training data and model behavior, including risks related to training on low quality data, data specific to the local context where the AI tool is being used, or on the wide release of model weights trained on confidential data.

- In AI application areas that facilitate scientific discovery, enhance open science practices and access to relevant datasets amongst researchers.

- Advance efforts to meaningfully evaluate and review open models across dimensions, including safety, security, trustworthiness, capabilities, limitations, and risk.

4.2b Linguistic, Cultural, and Geographic Contexts

While leading generative AI tools respond to prompts in various languages, today's models are often trained on data from majority-spoken languages and images from dominant cultures. As a result, they tend to produce more accurate, less biased, and higher-quality responses to prompts in relation to these representations. For AI to be globally inclusive, it must perform well across the world's linguistic, cultural, and geographic diversity, including those languages and cultures that are not well-represented on the Internet, or where there have been fewer resources devoted to training and refining AI systems in the past. Governments, academia, and the private sector should collaborate to ensure AI tools perform well across a wide variety of languages and cultural contexts, including those from smaller communities. Voluntary partnerships across sectors and across borders may improve the availability of data from more cultural contexts. This may include enhancing multilingual and multicultural capacity of mainstream AI models, or embarking on collaborative smaller, targeted applications. Incorporating more assorted languages and culturally accurate images into AI systems—and evaluating system performance within those contexts—is crucial for technological enhancement, social equity, and economic inclusivity.

Global Research Opportunities:

- Identify good practices for language and image data collection and attribution in collaboration with affected communities, recognizing that cultural and social norms around data sharing may vary widely. Consider enhancing capacity-building of these communities so they can be stewards of their linguistic and cultural resources and help to ensure that benefits and results from AI research reach their communities.

- Establish and support consortia dedicated to advancing AI technologies that enhance linguistic and image-based cultural diversity, involving state actors, stakeholder communities, language speakers, prominent AI research bodies, academic institutions, and private-sector firms, including through building more representative datasets in partnership with affected communities.

- Investigate gaps in AI models' performance for culturally diverse inputs, and the extent to which these gaps can be mitigated through data curation, data development, translation efforts, fine-tuning, algorithmic advances, or other means.

- Improve AI modeling techniques for low-resource languages, focusing on making the most of limited data and capturing the nuances of dialects and local variations.
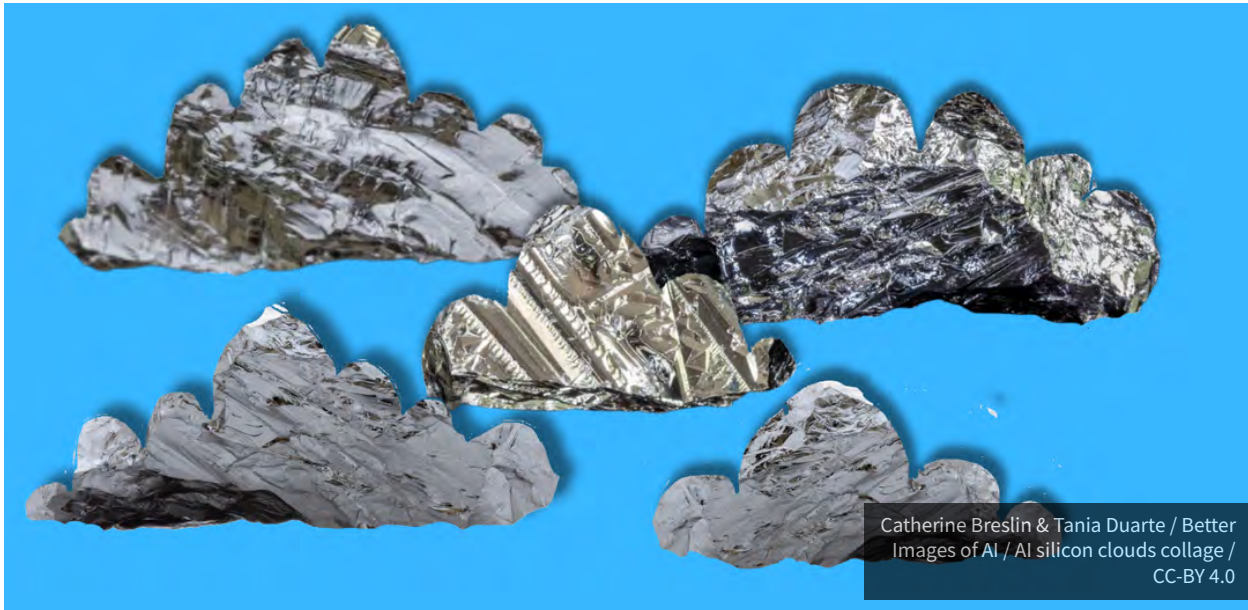
4.2c Increasing Access and Inclusion in AI Research Infrastructure

With appropriate safeguards, increasing access to data, compute, and other research infrastructure is essential for AI system and tool development, and can strengthen the capabilities of the global AI workforce while enabling inclusivity in the AI research ecosystem. AI research faces challenges due to varying data sharing requirements and data localization laws. Data is important for developing AI systems and applications, and researchers ultimately benefit from trusted cross-border data sharing. Under-resourced institutions, especially in low- and middle-income countries (LMICs), may face challenges related to access for data, compute, testing resources, and software. To support these areas, the AI research community should advance the availability of compute, datasets, testbeds, testing frameworks, and open-source libraries and toolkits.

Global Research Opportunities

- Develop international guidance on how to conduct ethical and privacy-preserving research, particularly for the development of appropriate safeguards around the use of shared research infrastructure.

- Digitize scientific and historical data still stored in analog formats to expand and improve the quality and diversity of training data to ensure it is sufficiently representative across different dimensions.

- Explore linking trusted research infrastructure in national, regional, and international contexts to advance researcher access to diverse datasets, knowledge bases, testbeds, testing frameworks, and open-source libraries and toolkits.

- Evaluate the efficiency and scalability of computational resources in handling large-scale data processing tasks and develop optimization techniques that enhance performance and reduce resource consumption.

- Define globally accepted, machine-readable formats for scientific data to increase the interoperability of datasets internationally and their use for training and fine-tuning AI systems and tools to facilitate AI data readiness.

4.3 Research to Support AI for Global Challenges

AI can revolutionize the production and utilization of knowledge to tackle global challenges, particularly those related to the United Nations Sustainable Development Goals (SDGs), by designing new tools and conducting applied research to create a safer, healthier, more prosperous world. Research to support AI tools, that address global challenges, should be a focus of the international research community, as it fosters collaboration and entails pooling resources and expertise across borders to tackle pressing issues, such as climate change, global health, and food insecurity. This section explores ways in which international cooperation in applied AI research can leverage innovative solutions in pursuit of global sustainability.

4.3a Climate, Energy, and the Environment

Joint research initiatives can contribute to reduction of the negative impacts of AI on climate, energy, and the environment, particularly by leading to breakthroughs which could reduce electricity consumption associated with AI. International research can also accelerate the development and deployment of AI-driven solutions and AI-informed climate models and can facilitate data sharing that will help empower countries, communities, and individuals to address climate, environmental, and energy challenges. New AI methods can contribute[22] tools towards responding to the climate crisis by enhancing climate modeling, prediction,[23] adaptation, resilience, and mitigation efforts. Additionally, the clean energy transition requires modernizing and decarbonizing global energy systems. AI can advance this by managing energy use, incentivizing customer behavior, integrating renewable sources, discovering new energy materials, developing clean energy technologies, assisting in obtaining permits for these technologies, and accelerating the resilience of power systems for severe weather. AI research and collaboration can also drive novel insights from disparate data sources to improve the decision-making process and promote a wider range of mitigation, adaptation, and resilience strategies to support more coordinated efforts in tackling the impacts of climate change to build a more resilient future. Increased capabilities in AI can help integrate local and Indigenous knowledges and can incorporate context-specific data to ground-truth assumptions and validate model predictions. International cooperation is essential to share best practices, technologies, and innovations, minimizing AI risks and maximizing its benefits for the international community, particularly in developing countries.

Global Research Opportunities:

- Explore pathways for international collaboration to enable the development of novel AI systems to predict and combat climate change. This includes establishing comprehensive and detailed digital twins of the planet, modeling climate scenarios and response options for warming, sea-level rise, flooding, wildfires, and other slow-onset and extreme weather events, as well as assessing climate impacts on human migration and biodiversity.

- Share best practices and develop technologies to reduce AI's resource demands, including energy, water, and critical minerals, through hardware, software, and efficiency measures. This holistic approach to sustainable AI considers performance, reliability, usability, and sustainability. Additionally, accelerate efforts to quantify and reduce the global environmental effects of AI, particularly energy and water consumption.

---

22   "Climate Change AI - Interactive Summaries." n.d. Climate Change AI. https://www.climatechange.ai/summaries.
23   "AI4ESP | Artificial Intelligence for Earth System Predictability." n.d. https://ai4esp.org/.

Shutterstock Standard License / Photo contributor FOTO Eak

- Focus a subset of research on the next generation of energy efficient AI-enabled supercomputing platforms and unlocking new clean energy sources, optimizing energy production, improving grid resilience, and building tomorrow's advanced energy economy.

- Develop AI tools to evaluate scenarios to reduce greenhouse gas emissions while balancing their deployment with the protection of the world's natural and cultural resources, such as promoting climate-smart agriculture, sustainable land use, and biodiversity protection (see also section on Agriculture and Food Security below).

- Develop and define tools that leverage global geospatial data, satellite imagery, climate data, and text to enhance situational awareness and design responses for natural disasters enhancing multi-hazard early warning systems, while also improving machine learning methodologies to handle the complex structured and unstructured data types central to climate work to aid in mitigation and recovery efforts.

- Develop tools that leverage climate and other relevant data to provide early warning for climate-sensitive disease outbreaks and other impacts to more effectively mitigate risks to the well-being of individuals and populations.

4.3b Health

AI has the potential to accelerate the impact of digital health systems and initiatives. It can positively impact global health by supporting health systems, healthcare delivery, and health science. AI could drive down healthcare costs by optimizing market efficiencies, such as streamlining administrative processes and reducing waste. Additionally, AI could support prevention and public health strategies by enabling early detection of diseases and personalized health interventions, ultimately reducing the need for expensive treatments. A recent United States Agency for International Development analysis[24] of AI applications in the global health context highlighted opportunities across four domains—population health, individual health, health systems, and pharmaceutical and medical technologies—and identifies AI use cases for addressing aspects in each of those areas. The use of AI techniques, such as computer vision and natural language processing, can streamline data entry and improve data quality and analytics for better and faster decision-making. AI-enabled tools can assist with clinical decision support, ensure treatment regimen compliance, improve diagnosis accuracy, and provide personalized health coaching and other innovative wellness services. AI-based technologies and services can help patients better understand their medical information to enable deeper engagement in their health care.

As researchers look to use AI systems and develop AI tools to advance health outcomes, it is important that local perspectives are prioritized to identify relevant and impactful use cases, and that locally relevant data is available and incorporated to train these models for optimal performance, relevance, and sustainability. Safety, security, and privacy approaches to health-relevant data and outputs is of utmost importance and approaches may vary significantly by country. Developing health-related benchmark datasets in multiple languages and applicable to various geographies will be key to supporting AI research in health. AI tools should be designed to integrate into existing health systems, digital health architectures, and healthcare workflows, improving efficiencies and enhancing capacities rather than replacing them. Finally, it is essential to identify ways to measure the impact of AI systems on health systems and outcomes to identify and prioritize tools for scaling and integration.

Fritzchens Fritz / Better Images of AI / GPU shot etched 1 / CC-BY 4.0

---
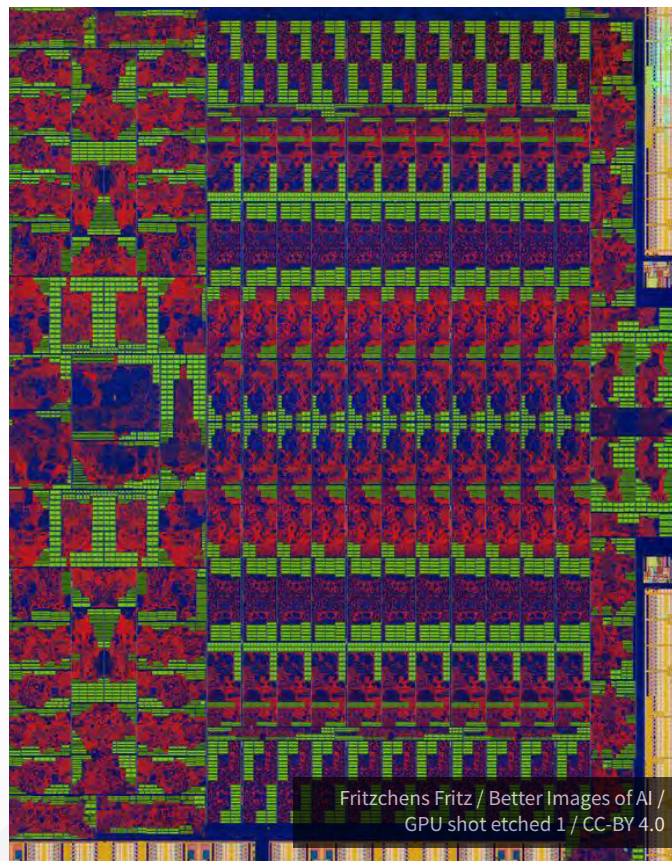
24   USAID's Center for Innovation and Impact (CII), Bill & Melinda Gates Foundation, The Rockefeller Foundation, and Boston Consulting Group. 2021. "Artificial Intelligence in Global Health: Defining a Collective Path Forward." CII's Innovating for Impact Series. USAID's Center for Innovation and Impact (CII). https://www.usaid.gov/sites/default/files/2022-05/AI-in-Global-Health_webFinal_508.pdf.

Global Research Opportunities:

- Design AI-based health applications specifically for and with LMICs, to help ensure such applications address unique challenges and needs, rather than simply applying existing applications across all countries.

- Develop common AI standards, reporting, and terminology to align with broader digital health and health data standardization and interoperability efforts, improving data management and curation for AI systems and tools development.

- Advance efforts to prepare for, detect and monitor, and respond to, emerging global health threats, including pandemics, combining disparate and localized data to protect population and individual health and global health security.

- Explore methods of protecting patient data privacy and meaningful consent and address ethical and integration challenges of AI deployment in global health settings.

- Examine approaches for vetting the quality and safety of AI applications used by providers and patients including misinformation, disinformation, and malicious uses.

- Build an implementation science evidence base to support understanding for the factors that promote effective health worker and patient uptake of AI-based tools to improve health outcomes. Build off the extensive lessons learned, research outcomes, implementation science and principles developed for the broader digital health ecosystems.

Photo Credit: Riaz Jahanpour for USAID

4.3c Agriculture & Food Security

In high-income countries, AI applications for industrial-scale <u>agriculture</u>[25] are used to maximize yields, minimize inputs, and manage environmental impact. Increasingly in the developed world, agricultural technology (AgTech) innovators and some government agencies are adapting precision agriculture for smallholder farmers, who produce around a third of the world's food consumption and are concentrated in LMICs. These include yield forecasting systems which can help offer tailored advice to individual farmers, low-cost distributed sensor networks, supply chain resiliency systems, robotic systems to augment farm labor support, animal health monitoring systems, and smartphone-based systems for image-based diagnosis of plant diseases and pests. Beyond precision agriculture, AI is being used to supplement inadequate extension systems with agriculture advisory chatbots and to improve credit and insurance modeling, customizing financial services to make them more attractive for smallholder farmers.
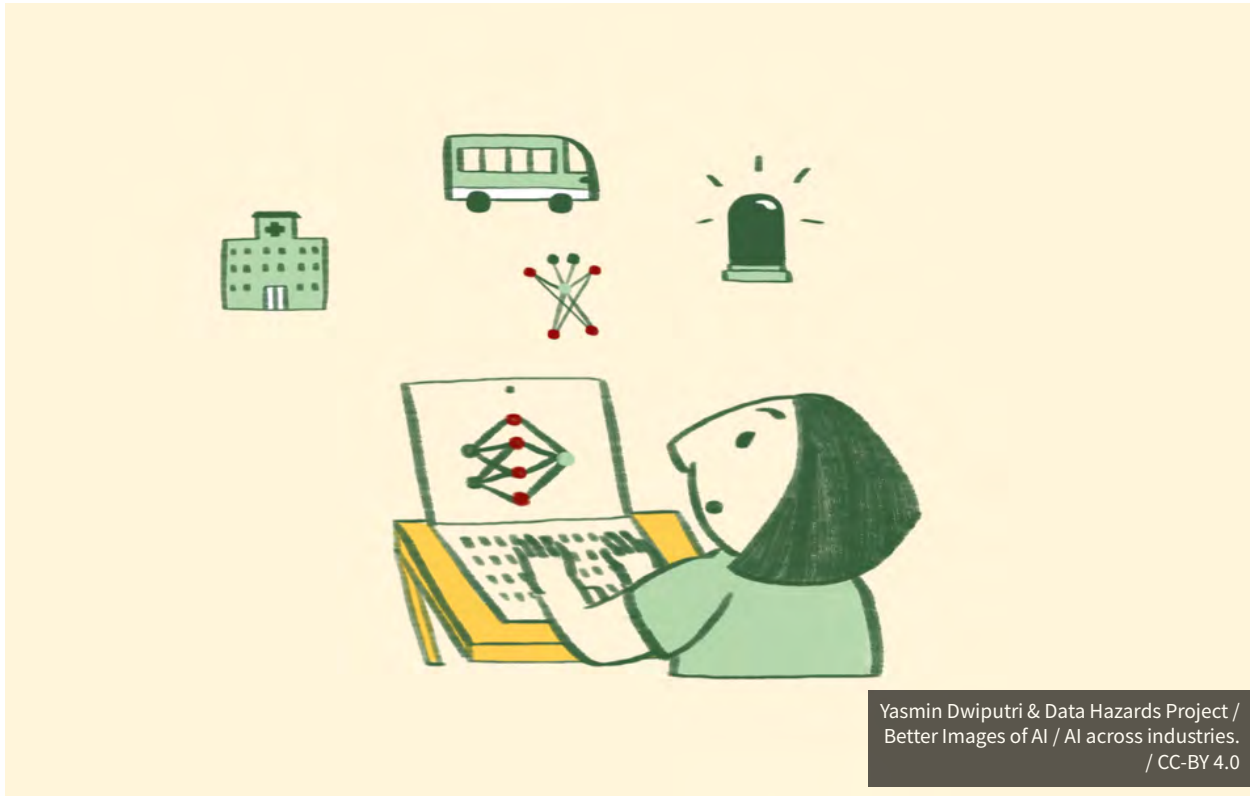
Global Research Opportunities:

- Increase efforts to protect and support smallholder farmers, who produce around a third of the world's food, by understanding AI tools applicable in low-resource environments to boost yields, engagement in formal economies, and optimized distribution.

- Integrate data sources such as climate modeling, satellite imagery, and affordable in situ sensors to build more effective models to predict and improve farming practices and yields.

- Work with agricultural scientists, behavioral scientists, and others to develop AI-enabled agricultural decision support tools that can help farmers effectively use model outputs.

- Develop participatory design practices that can be used to include rural actors, including smallholder farmers, in the design and governance of AI tools and datasets.

- Promote AI applications that can be used to improve food and agricultural resilience, especially in the context of mitigating the effects of and adapting to increased prevalence of natural disasters and other consequences of climate change.

4.3d AI in Science and Engineering

AI is already enhancing and improving scientific and engineering workflows,[26] from autonomous laboratories creating new materials to deep learning algorithms analyzing how pathogen evolution affects disease spread. New AI techniques could support the continued growth and expansion of science infrastructure, especially when handling large and complex data. Increasingly, scientific applications are shifting from human-in-the-loop to AI-driven design, discovery, and evaluation. Scientists are more frequently leveraging AI to structure processes, optimize large-scale simulations, and operate next-generation instruments. International collaboration is essential to share AI advances and integrate diverse expertise, accelerating global scientific progress. Joint efforts, when paired with appropriate risk mitigation strategies, can help standardize AI methodologies and ensure that innovations benefit a wide range of scientific domains and researchers across different countries.

---

25   "Can AI Inclusively Advance Agri-food Systems?" n.d. https://www.genesis-analytics.com/reports-and-other-documents/can-ai-inclusively-advance-agri-food-systems.
26   President's Council of Advisors on Science and Technology. 2024. "Supercharging Research: Harnessing Artificial Intelligence to Meet Global Challenges." Report to the President. https://www.whitehouse.gov/wp-content/uploads/2024/04/AI-Report_Upload_29APRIL2024_SEND-2.pdf.

Global Research Opportunities:

- Expand secure and responsible access of anonymized government datasets for critical research needs with appropriate protections and safeguards.

- Support basic, translational, and applied research in AI that involves collaborations across academia, industry, government laboratories, and agencies.

- Adopt accountability and other methods designed to promote responsible, transparent, and trustworthy AI use throughout all stages of the scientific research process to manage the risks of inaccurate or non-replicable findings from scientific uses of AI.

- Encourage innovative approaches to integrating AI assistance into scientific workflows to support researchers conducting high quality science that utilizes AI assistance responsibly.

4.4 Fundamental Research on AI, including AI Safety, Security, and Trust

AI is still in its early stages of development, and fundamental advances in AI technology and governance are necessary before some AI systems can be considered safe or trustworthy. Recent generative models, such as large language models (LLMs), introduce or exacerbate various risks[27] including issues of robustness, information integrity, and privacy preservation that impact many other types of AI systems (likely including future generations of AI technology). This section focuses on four priority areas where international research cooperation is particularly crucial.

---

27   National Institute of Standards and Technology. 2024. "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile." NIST AI 600-1. NIST Trustworthy and Responsible AI. U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.600-1.

4.4a Synthetic content and Information Integrity

While generative AI may bring many benefits, the potential misuse of generative AI to create misleading or harmful content remains a concern - images, audio, video, and text that have been generated or manipulated by AI – can be challenging to identify and understand its lineage. When AI is misused to spread dis- or misinformation, it can undermine public health, individual well-being, and trust in institutions, governance structures, news industries, and media, ultimately threatening information integrity in the global digital ecosystem. AI-manipulated content can be particularly dangerous because it can be quickly customized, created, and deployed to reach and influence audiences on a large scale. This could have widespread implications. As the use of AI to create or manipulate digital content becomes more prevalent, it is important for people all over the world to be aware of digital content's origins and ability to detect whether it was manipulated or altered. Research is vital to manage these risks.

Research can help policymakers understand public perceptions and the societal impacts of an increased presence of and exposure to synthetic and authentic content, and identify techniques to enhance digital content transparency, media literacy, and information integrity globally. Addressing this challenge will require a suite of localized approaches, deployed at scale and informed by research, to safeguard the efficacy and authenticity of information online, and empower global audiences to better understand where digital content came from and whether and how it was manipulated. This suite includes algorithms for labeling, watermarking, metadata recording, and synthetic content detection.

Global Research Opportunities:

- Identify mechanisms for enhancing societal awareness, digital media literacy, and transparency about the origins of digital content, including synthetic content. Ensure that these mechanisms work in low-literacy contexts and for people not explicitly or regularly exposed to AI.

- Identify reliable techniques for disclosing and tracking the provenance of digital content (e.g., AI-manipulated images, audio, videos and text), to include watermarking and metadata, that can be applied at scale across platforms, across languages, and across institutions. Assess the robustness, security, and other tradeoffs of these techniques across content modalities.

- Identify data-driven best practices for labeling digital content regarding its provenance in a way that builds trust, promotes transparency, and considers different languages and cultures. Also explore designing effective user experiences beyond labels that enhance digital content transparency.

- Explore and assess the effectiveness of techniques for the reliable detection of synthetic content, including voice clones, along with safeguards to mitigate the consequences of false positives or negatives. Explore how user interface design can help users respond appropriately to uncertainty in detection results.

- Conduct research to monitor and understand the sociotechnical effects of increased exposure to synthetic content, labeling synthetic and authentic content, and the resulting impacts on the information environment, as well as implications for democratic processes, including by the tactical use of synthetic content against women and LGBTQ+ political and public figures by foreign state and non-state actors and further investigate future harmful uses of synthetic media as technology evolves.

4.4b AI and Cybersecurity

International cooperation on AI is essential for improving cybersecurity, including cybersecurity of AI systems and the use of AI for cyber defense. Malicious cyber activity against one country can quickly extend to networks across borders, posing threats to security and stability. Recognizing cyber incidents frequently have global impacts, preventing and responding to these incidents also requires international cooperation. AI is particularly powerful in international social engineering attempts because it provides the ability to tailor cyberattacks in the recipient's native language. Through efforts like the U.S.-led International Countering Ransomware Initiative,[28] global cooperation to address cyber threats is growing. However, more coordination is needed on cybersecurity research and development to identify emerging threats that are unique to or exacerbated by AI systems and to leverage AI to develop effective ways to mitigate security threats.

Global Research Opportunities:

- Develop AI-driven systems for real-time threat detection, automated response to cyber incidents, and proactive defense measures, leveraging global data sharing, collaboration, and advanced analytics to improve accuracy, effectiveness, compliance with global cybersecurity regulations, and resilience against evolving cyber threats.

- Research methods to enhance the security and robustness of AI systems, including protection from cyberattacks,[29] to ensure confidentiality, integrity, and availability.

- Explore AI techniques that enhance cybersecurity while preserving user privacy to ensure data protection across international borders.

- Advance applications of AI to secure critical infrastructure sectors, such as energy, transportation, and healthcare, against sophisticated cyber threats. This includes developing AI solutions tailored to the specific needs and vulnerabilities of these sectors.

4.4c Evaluate AI Systems Using Benchmarks, and Develop Technical Standards based on Pre-standardization Research and Collaboration

Benchmarks, frameworks, and standards for AI as part of efforts to enable greater safety, security, and trustworthiness are still nascent. Benchmarks are standardized tests that can measure AI system performance. To date, most AI benchmarks have used data drawn from a select set of sources, creating a risk that future AI systems may underperform for users and applications in different contexts. In recent years there has been progress in using more representative benchmark datasets (i.e., curated data collections used to evaluate AI capabilities); however, significant work remains. Such work is essential, as researchers are incentivized to demonstrate the performance of new systems and approaches using robust methods. If benchmarks can be made more robust and representative, then diversity (and accuracy) of output may be a valued feature of new AI models, not an afterthought. Similarly, AI standards for testing, evaluation, verification, and validation methods will be most useful if they respond to the needs of diverse demographics around the world.[30] AI has been an important focus of activity within standards

28   The White House 2023. "International Counter Ransomware Initiative 2023 Joint Statement." The White House. November 2, 2023. https://

www.whitehouse.gov/briefing-room/statements-releases/2023/11/01/international-counter-ransomware-initiative-2023-joint-statement/.
29   Vassilev, Apostol, Alina Oprea, Alie Fordyce, Hyrum Anderson, National Institute of Standards and Technology, Northeastern University, and Robust Intelligence, Inc. 2024. "Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations." NIST Trustworthy and Responsible AI. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2023.pdf.
30   U.S. Department of Commerce, Gina M. Raimondo, and Laurie E. Locascio. 2024. "A Plan for Global Engagement on AI Standards." NIST

development organizations, such as the work being done by Subcommittee 42 of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) Joint Technical Committee 1 (ISO/IEC JTC 1/SC 42). In support of efforts conducted within Standards Developing Organizations (SDOs), researchers have an important role to play in pre-standardization activities. Pre-standardization research and stakeholder collaboration lay the foundation for future standards development efforts. Robust standards development processes often reflect the widely-accepted principles of transparency, openness, impartiality and consensus, effectiveness and relevance, and coherence. Researchers may also wish to contribute their technical expertise directly to standards projects.

Global Research Opportunities:

■ Foster AI standards development through evidence-based practices that are scientifically sound, open, transparent, and reflect the needs and inputs of diverse global stakeholders.

■ Build researcher contributions to standardization work into research funding programs.

■ Prioritize pre-standardization research into topics where standardization is urgently needed, including terminology and taxonomy, risk mitigation, bias and discrimination, provenance, generated content, and transparency standards about system and data characteristics.

■ Build more globally representative benchmark datasets, including datasets that support the implementation of standards.

Shutterstock Standard License / Photo contributor Gorodenkoff

4.4d Fundamental Developments in AI System Design and Implementation

There is significant opportunity for improvements in AI systems' functionality, reliability, and performance to enable use-cases beyond today's state-of-the-art designs. Maximizing the potential benefits of AI technologies will require continuous research into fundamental scientific and engineering dimensions of AI system design and implementation. The improvements this research can drive are crucial for addressing the diverse needs of various applications and industries and have the potential to address constraints faced by researchers in low- and middle-income countries (LMICs). Interoperability, accessibility, and reproducibility of data and research products should also be considered in fundamental developments of AI system design.

Global Research Opportunities:

- Identify avenues for improving the multi-modal performance of AI systems, enabling systems to perform reliably in mixed-media contexts, such as those combining audio, video, and textual inference tasks.

- Explore topics relating to the performance and computational efficiency of AI systems, such as designs for improved data-caching, speculative decoding, or parallelism during training and inference.

- Advance the development of AI systems that better preserve the privacy of users and individuals represented in training datasets, such as through differential privacy and federated learning techniques.

- Improve methods to detect and mitigate potential bias in AI systems and datasets.

- Advance transfer learning methods that enable greater performance in low-data languages, geographic or cultural settings, or other underrepresented application contexts.

4.5 Research on AI's Global Labor Market Implications

Adopting human-centered, socially responsible, safe, secure, and trustworthy AI can deliver substantial economic benefits and accelerate progress on many of the SDGs, including improved labor productivity,[31] management processes,[32] safety, health,[33] fair and equitable hiring, and job quality.[34] AI can positively augment work by replacing and automating repetitive tasks or assisting with routine activities and decisions, which may reduce the burden on some workers and allow them to better perform other tasks. At the same time, labor-related AI implementation carries risks[35] including increased work intensity, mental distress, and reduced agency in managerial decisions related to hiring, evaluation and dismissal, and fair remuneration. Gender disparities, including the gender digital divide, and other asymmetries could be exacerbated, with a range of disproportionate consequences across diverse sectors and regions. Unequal access to and use of AI in labor markets could also lead to increased imbalances among firms and workers, within and across countries. Additionally, certain uses of AI in the workplace – including for increased worker monitoring – have the potential to undermine collective bargaining and other fundamental principles and rights at work, including freedom of association, non-discrimination, and a safe and healthy working environment.

31   Kane, Aidan T., and Martin Neil Baily. 2024. "How Will AI Affect Productivity?" Brookings, May 2, 2024. https://www.brookings.edu/articles/how-will-ai-affect-productivity/.

32   Nieto-Rodriguez, Antonio. 2023. "How AI Will Transform Project Management." Harvard Business Review. October 3, 2023. https://hbr.org/2023/02/how-ai-will-transform-project-management.

33   Reinhold, Karin, Marina Järvis, Tallin University of Technology, Aleksandr Christenko, Vaida Jankauskaitė, Agnė Paliokaitė, Visionary Analytics, et al. 2022. "Artificial Intelligence for Worker Management: Implications for Occupational Safety and Health." European Agency for Safety and Health at Work. https://doi.org/10.2802/76354.

34   Khattar, Rose. 2024. "Will AI Benefit or Harm Workers?" CAP 20. 2024. https://www.americanprogress.org/article/will-ai-benefit-or-harm-workers/.

35   Lane, Marguerita, Morgan Williams, and Stijn Broecke. 2023. "The Impact of AI on the Workplace: Main Findings From the OECD AI Surveys of Employers and Workers." OECD Social Employment and Migration Working Papers. https://doi.org/10.1787/ea0a0fe1-en.

This section explores research directions touching on several aspects of AI's impacts on labor. In addition to highlighting important research needs, this section closes with strategies to mitigate potential negative labor market implications of AI that are grounded in current evidence and can be refined by further research.

4.5a Job Augmentation and Automation

AI, like many past technological advances, has potential to increase productivity, add complexity to jobs, change the nature of some tasks, and fully automate others. Generative AI, in particular, could broaden the scope of the types of tasks that can be augmented or automated, though there is still much uncertainty about how these changes will develop. Effects of AI will vary between countries, but also within countries by occupation, gender, income level, and other factors. Recent research suggests certain types of AI may lead to greater augmentation and automation of tasks in HICs, compared to LMICs, due in part to the prominence of digitized workplaces in these countries.[36] However, more research is needed to examine potential impacts of different types and uses of AI on job tasks, globally, and how this may change with future advancements[37].

Global Research Opportunities:

- Forecast employment trends by studying AI's impacts on tasks in specific sectors and occupations, and how these trends may change based on projections of how AI capabilities and limitations might evolve in the future.

- Conduct analyses of AI adoption and forecast its labor market impacts – including, for example, whether AI may decrease the availability of middle-skilled occupations – on various countries, regions, sectors, and occupations, considering various aspects of identity and social factors including gender, age, geography, language, sexual orientation, race, disability, immigration status, documentation status, ethnicity, and nationality.

- Map AI-related risks and benefits for job design and distribution, task complexity and receptiveness, productivity and wage distribution across contexts. Assess how these risks and benefits may impact job recruitment, satisfaction, and retention, as well as how changes in these factors will impact the larger global economy.

- With particular focus on women and marginalized groups, examine patterns of worker transitions between occupations and industries in response to AI, including whether they are transitioning to higher- or lower-wage jobs, if they are transitioning to jobs in sectors with higher or lower degrees of formality, and if internal or cross-border labor migration is involved in their transitions.

- Perform gap analyses of national labor market data collection systems, especially in LMICs, to determine whether countries can collect robust and accurate data related to AI and work, and to inform the design of indicators of labor market impacts from AI development, deployment, and use.

- Use gender-responsive research methods from different fields of study, including ethnography, anthropology, sociology, and social work, to ensure meaningful inclusion of diverse perspectives in AI labor market research.

---

36   Cazzaniga, Mauro, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J. Panton, Carlo Pizzinelli, Emma Rockall, and Marina M. Tavares. 2024. "Gen-AI: Artificial Intelligence and the Future of Work." Staff Discussion Notes No. 2024/001. International Monetary Fund. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379?cid=bl-com-SDNEA2024001.
37   Gmyrek, Paweł, Janine Berg, David Bescond, and International Labour Organization. 2023. "Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality." ILO Working Paper 96. International Labour Office. https://doi.org/10.54394/FHEM8239.

Photo Credits: USAID

4.5b New Jobs Resulting from AI Development, Deployment, and Use

As AI continues to develop, so does its value chain and the type and amount of human labor required to support it. Globally, millions of data enrichment workers—who label and validate data to train AI models—work in pay-per-task environments. These workers are primarily from LMICs,[38] recruited for this work due to lower costs to technology companies in HICs. While some data enrichment workers are employed full-time in settings similar to call centers, most work through task-based online platforms.[39] Basic facts about these workers, such as their numbers, locations, working time on the platform, and tasks are unknown. In some cases, these workers may be tasked with reviewing toxic and inappropriate content. As demand for data enrichment grows, more research is needed to understand this sector and develop strategies and frameworks to ensure workers' protections, benefits, and grievance mechanisms.

Beyond data enrichment, it is likely other new jobs and industries will emerge in response to AI, as seen with previous technological advancements.[40] More research is needed to anticipate where these jobs and industries will emerge.

---

38   United Nations. n.d. "Valuing Data Enrichment Workers: The Case for a Human-Centric Approach to AI Development | United Nations." https://www.un.org/en/un-chronicle/valuing-data-enrichment-workers-case-human-centric-approach-ai-development.
39   Tan, Rebecca, and Regine Cabato. 2023. "Behind the AI boom, an army of overseas workers in 'digital sweatshops.'" The Washington Post, August 28, 2023. https://www.washingtonpost.com/world/2023/08/28/scale-ai-remotasks-philippines-artificial-intelligence/.
40   Agrawal, Ajay, J., Joshua Gans, and Avi Goldfarb. n.d. "An Economic Framework for Understanding Artificial Intelligence." In Chapter 7, 243–45. https://www.whitehouse.gov/wp-content/uploads/2024/03/ERP-2024-CHAPTER-7.pdf.

Global Research Opportunities:

- Map the AI value chain to identify jobs that support the development and deployment of AI systems, those created as a result of AI systems, and those negatively impacted by AI systems. Document within these jobs their locations, industries, quality, tasks, impact on worker well-being, relationship to AI, and to whom they are accessible, prioritizing direct engagement with workers.

- Analyze how AI systems are shaped and influenced by the contributions of data enrichment workers of various educational, cultural, linguistic backgrounds, employment classifications, and under different labor conditions and workplace exposure to AI. Assess the effectiveness of unionization efforts and labor regulations to protect and empower these workers.

- Forecast potential global, national, and regional economic impacts of jobs created by, or related to the AI value chain.

- Project how new jobs resulting from AI design, development, and use, including data enrichment work, will change over time, including by studying whether these jobs may require increasingly skilled or specialized labor.

4.5c Effects on Inequality

AI development may impact income distribution differently across different groups, countries, regions, and sectors. One study suggests AI deployment may reduce intra-occupational inequalities[41] by benefitting less experienced workers. Overall, more research is needed to understand AI's contribution to global inequality and its distribution along the value chain.

High Income Countries – As different tasks in HICs are automated and augmented, there may be a polarizing effect on labor markets, increasing income inequalities. However, mature industries, strong digital infrastructures, and non-routine occupations may help HICs benefit from AI early on, potentially softening harmful labor market impacts. The development of AI-complementary industries in HICs could mitigate the negative effects of automation in these countries.

Low- to Middle-Income Countries – AI will affect each LMIC in different ways, depending on its current ability to develop and deploy AI. Recent research[42] estimates LMICs that rely on manual labor and more traditional industries may benefit less from AI deployment than HICs or other LMICs. The offshoring of work by companies in HICs to lower-cost operations in LMICs – ranging from call centers to manufacturing – has contributed to economic development in many LMICs. While it's possible that offshored jobs may be augmented or new types of jobs created because of AI, the technology could potentially disrupt development offshoring by automating certain tasks, rendering some of these jobs obsolete. Additionally, since many LMICs may lack the digital infrastructure to utilize AI technologies to complement certain tasks, it is possible that companies may reshore labor to technologically advanced countries, potentially widening inequalities between LMICs and HICs. More research is needed to determine how AI may impact jobs that have been offshored by companies in HICs.

41 Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2023. "Generative AI at Work." https://danielle-li.github.io/assets/docs/GenerativeAIatWork.pdf.
42 Cazzaniga, Mauro, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J. Panton, Carlo Pizzinelli, Emma Rockall, and Marina M. Tavares. 2024. "Gen-AI: Artificial Intelligence and the Future of Work." Staff Discussion Notes No. 2024/001. International Monetary Fund. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379?cid=bl-com-SDNEA2024001.

Global Research Opportunities:

- Evaluate how AI may impact income inequality and wealth distribution within and between countries, among various worker demographics, sectors, occupations, and regions.

- Assess the impact of AI on small- and medium-sized enterprises (SMEs) and examine factors that may enhance their effective and ethical AI use.

- Research the prevalence and patterns of AI-related outsourcing and nearshoring across and within countries, regions, and sectors. Examine how changes in these practices caused by AI will impact economies on national, regional, and global levels.

4.5d Risks for Workers

AI's impact on workers' rights may be significant, with risks spanning the global labor market and potentially affecting certain groups of workers, countries, sectors, and occupations more prominently. More research is needed to identify which of the following risks may impact workers around the world — especially for vulnerable populations.

Algorithmic Management and Decision-Making – Algorithmic management uses AI to support, inform, or automate management decisions or tasks, often resulting in increased collection of data and expanded monitoring of workers. Algorithmic management is regularly used[43] in digital platforms and is being increasingly used in warehouses, retail and hospitality, and in manufacturing. This practice can lead to increased work intensity, loss of worker autonomy, reduced job satisfaction, discrimination, and privacy infringement.

Shutterstock Standard License / Photo contributor Eva Corbella

Deskilling – As AI augments tasks, some workers' jobs can become simpler and more repetitive, leading to a downward pressure on wages[44] and stalled career growth, especially if workers lack access to reskilling programs.

Wellbeing – AI system deployment may lead to negative health effects[45] for workers in particular sectors and in certain countries and regions due to increased pressure for productivity, diminished human interaction, anxiety related to job security, the assignment of fewer meaningful tasks, and downward pressure on wages. Jobs in the AI value chain, such as content moderation, also present mental health challenges to workers. For example, PTSD caused by constant exposure to violent and graphic subject matter has been documented[46].

---

43   Aloisi, Antonio, and De Stefano Valerio. 2022. "Introduction — Your Boss Is an Algorithm. Artificial Intelligence, Platform Work and Labour." July 14, 2022. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4139319.
44   "OECD Employment Outlook 2023." 2023. OECD. July 11, 2023. https://www.oecd.org/en/publications/oecd-employment-outlook-2023_08785bba-en.html.
45   Stella, Luca, Osea Giuntella, Rania Gihleb, and Tianyi Wang. 2023. "Keeping Workers Safe in the Automation Revolution." Brookings, September 12, 2023. https://www.brookings.edu/articles/keeping-workers-safe-in-the-automation-revolution/.
46   Newton, Casey. 2020. "Facebook Will Pay $52 Million in Settlement With Moderators Who Developed PTSD on the Job." The Verge, May 12, 2020. https://www.theverge.com/2020/5/12/21255870/facebook-content-moderator-settlement-scola-ptsd-mental-health.

Discrimination in Hiring Practices – The use of AI in employment decisions [raises concerns](#)[47] about how potential biases in the technology may lead to discrimination in hiring, promotions, termination, pay increases, and other areas. Individuals may also [experience discrimination](#) from AI models when searching for jobs online.[48]

Exploitation in New Value Chains – Opaque, international data enrichment value chains have limited regulations, giving rise to global labor abuses[49] like nonpayment of wages, absence of grievance mechanisms, and excessive monitoring.[50]

Misclassification – As algorithmic and remote management rise, companies increasingly [outsource jobs](#) to subcontractors.[51] Companies may seek to outsource labor to avoid paying the full cost of employment potentially by misclassifying workers as independent contractors, or through other non-standard forms of work, preventing the workers from receiving the same workplace protections as employees.

Effects on Freedom of Association – AI could potentially weaken workers' ability to exercise their right to freedom of association and right to collective bargaining, leaving them without key advocacy mechanisms to ensure fair deployment and use of AI in the workplace. The increasing use of non-standard forms of employment could either exclude large numbers of workers from organizing and bargaining or significantly increase the legal and practical barriers to these activities. Some workplaces may use AI to monitor, and potentially act against, organizing or collective bargaining activity in the workplace. Many workers are also unaware of AI and automation technology in their workplace, limiting their power to negotiate[52] the introduction and acceptable use of AI with their employers.[53]

Global Research Opportunities:

- Evaluate existing and emerging uses of AI in the workplace internationally and employ longitudinal research to analyze how these uses impact workplaces and the labor market over time.

- Conduct sector-specific comparative studies in LMICs and HICs to determine which sectors and populations will be at the highest risk of algorithmic management, deskilling, misclassification, discrimination, and other risks of AI on workers, with a focus on vulnerable groups.

- Study AI's impact on workers and their communities, with particular focus on mental and physical health.

- Identify country-specific regulatory gaps for all workers, with special attention to non-standard workers, to exercise rights to freedom of association and collective bargaining in the AI value chain or in the AI-impacted workplace.

- Map legal frameworks governing the deployment of algorithmic management in the workplace across countries, focused on the most protective and least protective policies.

47   Bryant, Kalina. 2023. "How AI Will Impact the Next Generation Workforce." Forbes, May 31, 2023. https://www.forbes.com/sites/kalinabryant/2023/05/31/how-ai-will-impact-the-next-generation-workforce/.
48   Wall, Sheridan. 2021. "LinkedIn's Job-matching AI Was Biased. The Company's Solution? More AI." MIT Technology Review, June 25, 2021. https://www.technologyreview.com/2021/06/23/1026825/linkedin-ai-bias-ziprecruiter-monster-artificial-intelligence/.
49   Gray, Mary L., and Siddharth Suri. Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Boston: Houghton Mifflin Harcourt, 2019.
50   Perrigo, Billy. 2022. "Inside Facebook's African Sweatshop." TIME, February 17, 2022. https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/.
51   Bernhardt, Annette, Lisa Kresge, Reem Suleiman, and UC Berkeley Labor Center. 2021. "Data and Algorithms at Work: The Case for Worker Technology Rights." UC Berkeley Labor Center. https://laborcenter.berkeley.edu/wp-content/uploads/2021/11/Data-and-Algorithms-at-Work.pdf.
52   Johnston, Hannah, Chris Land-Kazlauskas, INWORK, and International Labour Organization. 2019. Organizing On-demand: Representation, Voice, and Collective Bargaining in the Gig Economy. Conditions of Work and Employment Series. https://webapps.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---travail/documents/publication/wcms_624286.pdf.
53   Hadwiger, Felix and International Labour Organization. 2022. "Realizing the Opportunities of the Platform Economy Through Freedom of Association and Collective Bargaining." ILO Working Paper 80. International Labour Organization. https://doi.org/10.54394/VARD7939.

4.5e Mitigating AI-Related Risks in the Global Labor Market

The following section outlines potential avenues for mitigating AI's potential negative effects on the global labor market, including value chain transparency and oversight mechanisms, policy interventions, and worker empowerment efforts.

Value Chain Transparency and Oversight

Supply Chain Tracing of AI Value Chains – Consultations with a variety of stakeholders revealed a primary research need at the nexus of labor and AI—tracing AI value chains from design to data collection, training, testing, deployment, and use. Little is known about these value chains, and research is still needed to map the business relationships and the complexities within them. Mapping these value chains will include determining how many workers are engaged and where, the types of employment involved, conditions of their work, benefits afforded (or not afforded) to them, and their access to grievance and remediation mechanisms. Additionally, there is a need for tracing the value chain downstream, beyond deployment, to the productive use of AI systems. Mapping digital systems in workplaces and companies can explore where, how, and why AI is being used, and the extent of its costs or benefits.

Enforcement Mechanisms – Ethical frameworks must be accompanied by enforcement mechanisms and accountability systems to be effective[54]. Organizations have recommended that countries identify paths to ensure all workers are protected by labor laws, including laws related to wages and hours, child labor and forced labor, human trafficking, occupational safety and health, workplace safety, and the right to organize[55]. A recent OECD report[56] says that countries will likely want to have specialized training for labor inspectors to ensure labor law compliance in the AI value chain. Research could determine potentially effective mechanisms and strategies for enforcing labor laws in these complex supply chains, spotlighting gaps in enforcement frameworks around the world to guide policymakers. Researchers may evaluate effective enforcement in similar supply chains.

Voluntary Commitments – Voluntary impact assessments, social impact labels, and audits can be found in various sectors to validate companies' commitments and ongoing actions to various social, environmental, and ethical causes. Some labels include Fair Trade, B Corp, GoodWeave, and Rainforest Alliance. AI auditing is a new and growing field and does not always focus on implications for workers. Incorporating labor considerations into existing or new voluntary assessments, or into requirements for ethical AI labels or certifications, could show consumers and other companies that certified AI systems are utilizing ethical labor practices in their value chain and have considered the system's impacts on the labor market and their workforce.

Corporate Reporting of AI Value Chains – Corporate reporting of the AI value chain can increase transparency, worker voice[57], and allow for more insightful research regarding corporate operations and ethical AI production and use. Corporate research could include information regarding the use of recruiters and contractors in AI value chains, and how their company mitigates risks of AI-related labor abuses, job displacement, and poor working conditions among suppliers.

---

54   ARTICLE 19. 2019. "Governance With Teeth: How Human Rights Can Strengthen FAT and Ethics Initiatives on Artificial Intelligence." ARTICLE 19. http://creativecommons.org/licenses/by-ncsa/2.5/legalcode.
55   ILO, ISSA and OECD. 2023. "Providing Adequate and Sustainable Social Protection for Workers in the Gig and Platform Economy." https://www.ilo.org/wcmsp5/groups/public/---dgreports/---ddg_p/documents/publication/wcms_867535.pdf.
56   Lane, Marguerita. 2020. "Regulating Platform Work in the Digital Age." No. 1. OECD Going Digital Toolkit Notes. OECD Publishing. https://doi.org/10.1787/181f8a7f-en.
57   "Worker Voice." n.d. DOL. https://www.dol.gov/agencies/ilab/worker-voice.

Global Research Opportunities:

- Trace the value chains of AI systems and include information about the types of work involved (e.g., piece-rate or task-rate work, hourly wage work), how prevalent each type is, the number of workers engaged in this work, worker location, benefits, pay, hours worked, as well as the collective bargaining environment in the countries in which these workers are employed.

- Study the potential roles of monitoring bodies, supervisory authorities, and self-reporting in impact assessments, audits, and evaluation of the use of automated systems in the world of work.

- Consider how and where AI may pose unique or significant challenges to labor-focused regulatory bodies – including, for example, the enforcement impacts of business process outsourcing in AI value chains – when exploring effective labor standards and transparency requirements.

- Develop and analyze the effectiveness of impact assessments focused on ethical labor practices in AI value chains, as well as reporting mechanisms for labor transparency and accountability in AI value chains, considering varying data collection and assessment capacities among different countries. Map and evaluate existing corporate due diligence efforts in AI value chains.

- Conduct multidisciplinary research to identify procedures, systems, laws, and entities to hold AI value chain actors accountable for labor abuses. Perform comparative case studies using successful supply chain accountability systems and models from other sectors, including systems that apply to various employment models as well as work in the informal economy.

Policy Interventions

Social Safety Nets – To protect and support workers whose job tasks have been automated by the deployment of AI, some organizations have suggested the expansion of access to benefits such as unemployment insurance or adjustment assistance.

Decent Work in Digital Labor Platforms – The structure of digital platforms has introduced new challenges for workers, and at the same time has served to intensify existing challenges. The ILO Governing Body has included an upcoming agenda item focused on standard-setting on decent work in the platform economy. Countries including Spain, Italy, the United States, and the United Kingdom have started to develop guidance, policies, and/or legislation related to digital labor platforms[58] and worker misclassification, but there is still much work to do[59].

Education and Curriculum Development – Investing in STEM education and in education globally will help prepare the future workforce for AI-related skills and roles, as well as the potential augmentation and displacement of jobs. Curriculum development can focus on information technology skills, computer literacy, coding, online safety, and ethical technology, as well as communication, critical thinking, teamwork, and interdisciplinary collaboration. Education systems should update curricula based on what skills are in demand to effectively prepare students for the labor market[60] (UNESCO), and should ensure accessibility of the educational infrastructure (e.g., digital textbooks, learning management systems) and curriculum content to ensure a diverse workforce.

---

58   "Cloudwork (Online Work) Principles." n.d. https://fair.work/en/fw/principles/cloudwork-principles/.
59   International Labour Organization. n.d. "Extending social security to workers in the informal economy: Policy Resource Package." https://www.social-protection.org/gimi/Emodule.action;jsessionid=QQmVpeOYZajx06_Pu3FFCG74kzFAyx-ps-rJoD-6gX2UQ1emlzjj!-1857443667?id=65.
60   United Nations Educational, Scientific and Cultural Organization. 2022. "Recommendation on the Ethics of Artificial Intelligence." Report. United Nations Educational, Scientific and Cultural Organization. United Nations Educational, Scientific and Cultural Organization. http://creativecommons.org/licenses/by-nc-sa/3.0/igo/.

Strengthening Collective Bargaining Rights – Collective bargaining plays a critical role in ensuring [workers' voices](#) are heard.[61] Without sufficient laws advancing collective bargaining, workers may experience unhealthy working conditions, wage theft, discrimination, harassment, and an absence of grievance and remediation mechanisms. Some stakeholders have recommended strengthening collective bargaining safeguards, to ensure workers are able to negotiate terms of AI use in the workplace and the impacts of AI-related industry changes on their jobs. To effectively bargain, [experts](#)[62] stress that workers must know which technologies are being used in the workplace and how.

LMIC-Specific Policies – Funding local research and development in the local AI industry may allow LMICs to minimize reliance on large technology companies based in high-income countries and may contribute to domestic economic growth. Additionally, for LMICs to take advantage of increased investments in STEM education and AI research and development, and to increase greater attraction of investment capital in AI research and development, they may invest in infrastructure to support these endeavors, including adequate computing power, internet access, and electricity.

Global Research Opportunities:

- Assess the ability of existing policy approaches to mitigate AI-specific labor market disruptions (e.g., EU's AI Act, the Corporate Due Diligence Directive).

- Evaluate the readiness of countries' social protection systems to support workers that may be affected by AI-related disruptions and consider the feasibility of potential tax system restructuring to extend social safety nets to all affected workers.

- Develop AI tools to enhance productivity in critical sectors for LMICs (e.g., agriculture, small-scale manufacturing).

- Identify and evaluate effectiveness of proactive industry policies and practices intended to retain staff as tasks are automated and augmented.

- Review countries' privacy and intellectual property laws and explore how they may impact the use and transparency of AI in the workplace.

Worker Empowerment

Skills Training Programs – Upskilling programs, which help workers enhance existing skills or cultivate new skills, and reskilling programs, which aim to help workers transition to new jobs, are both valuable tools for mitigating job disruption caused by AI. Companies can support worker upskilling and reskilling by designing and implementing new training programs, with the involvement of workers, workers' organizations, and social dialogue to ensure that they are robust and effective. Research shows that workers particularly vulnerable to AI-related labor market effects may need targeted incentives to participate in these programs. Research should focus on incentivizing participation and analyzing the effectiveness of these programs in different countries [and sectors](#)[63] (OECD, "[Skills Matter](#)" 2019). Upgrading the training infrastructure and training program content is necessary to ensure accessible and inclusive upskilling opportunities. Upgrading the training infrastructure and training program content is necessary to ensure accessible and inclusive upskilling opportunities.

61   Anner, Mark, Matthew Fischer-Daly, Center for Global Workers' Rights, The Pennsylvania State University, ICF Macro, Inc., and United States Department of Labor, Bureau of International Labor Affairs. 2023. "Worker Voice: What It Is, What It Is Not, and Why It Matters." https://www.dol.gov/sites/dolgov/files/ILAB/Worker-Voice-Report-Final-3-6-24.pdf

62   Center for Labor and a Just Economy at Harvard Law School. 2024. "Worker Power and Voice in the AI Response." https://clje.law.harvard.edu/app/uploads/2024/01/Worker-Power-and-the-Voice-in-the-AI-Response-Report.pdf

63   "Skills Matter." n.d. https://www.oecd-ilibrary.org/education/skills-matter_1f029d8f-en.

Data Transparency – Algorithmic management has given rise to increased collection of worker data. Workers' organizations[64] have recommended that workers should be able to access data collected about them in the workplace, and that there are processes to have data reviewed, edited, and/or deleted. Advocates also recommend that source code and AI models remain accessible, and that research prioritize the development of standards and techniques to facilitate the understanding of AI systems by workers and workers' organizations.

Collective Bargaining and Worker Voice – Collective bargaining is crucial for protecting and empowering workers during technological shifts. Workers and their representatives, especially from underserved communities, should have genuine input in the design, development, testing, training, use, and oversight of AI systems in the workplace. Doing so could help ensure workers benefit from AI and are protected from its negative effects. Worker voice improves working conditions and is **associated** with more effective adoption of advanced technologies.[65] Understanding how worker representation impacts labor market outcomes, working conditions, and job responsibilities is increasingly important as firms continue to adopt AI.

Global Research Opportunities:

- Determine whether worker representation in decision-making moderates or exacerbates AI's impact on working conditions, job design complexity, labor market outcomes, and mediating factors (e.g. AI literacy of worker representatives). Develop case studies in different occupations and sectors globally to determine how and where worker organizations' efforts have been effective in empowering workers in various supply chains.

- Evaluate current levels of AI literacy across countries, sectors, and demographics and identify future beneficial skills for workers in various sectors, occupations, and countries.

- Analyze the effectiveness and best practices of participatory upskilling and reskilling models across different demographics and regions, and at various levels of learning with attention to funding requirements, including tax implications, as well as scalability and applicability in LMICs.

- Through stakeholder consultations, develop, or evaluate existing, frameworks for achieving worker protections related to AI, including for organized workers, workers not engaged in collective bargaining, and workers engaged in platform work in the AI value chain, with attention to achieving worker protections related to data collection and use, and workers' mental and physical health concerns.

- Analyze the effectiveness of worker grievance and remediation mechanisms for algorithmic management and examine common worker and union grievances in the context of use of AI in the workplace.

---

64   UNI Global Union. 2021. "Top 10 Principles For Workers' Data Privacy and Protection." https://www.dol.gov/sites/dolgov/files/ILAB/Worker-Voice-Report-Final-3-6-24.pdf
65   "Bringing Worker Voice Into Generative AI | MIT Sloan." 2023. MIT Sloan. December 21, 2023. https://mitsloan.mit.edu/centers-initiatives/institute-work-and-employment-research/bringing-worker-voice-generative-ai.

SECTION 5: RECOMMENDED RESEARCH PRACTICES

The practices included in this section can inform and support engagement with diverse sets of international partners. Such efforts will be most effective when tailored to the level of AI-preparedness of global AI partners. These proposed practices also provide an avenue for the implementation of the previously proposed principles that guide this research agenda. The following practices are separated by categories of global stakeholders engaged with the research landscape with a focus on research funders, research hubs, and research teams.

5.1 Research Funders

Research funders encompass organizations, institutions, or entities that provide financial support for scientific, academic, or technological research projects. They allocate resources to researchers and research institutions to facilitate the advancement of knowledge, innovation, and development in various fields. Research funders can encourage international cooperation on responsible AI research through the following:

Elevate Transparency -- Transparency in AI research can include disclosing input materials, including training data sources, compute usage, hardware requirements, resource consumption, and third-party data enrichment. While research funders and publishers can introduce transparency requirements, a comprehensive approach will require broad cooperation, such as coordinating practices across international research communities and institutions. To the extent that computing resources are supported by research funders, data on energy use, water use, and environmental impacts of the resources should be shared with researchers.

Open Funding Calls to Support International AI Research Collaboration -- Funding calls are often restrictive regarding eligible countries, and individual research projects are typically carried out bilaterally. Differences between legal requirements can underlie challenges to collaboration between researchers from different countries. New mechanisms could help to support deeper collaborations in AI research with countries in LMICs, where traditional arrangements can be inadequate. Exploring new mechanisms will help to ensure AI research and development in partnership with these countries is engaged in and benefits from global advancements.

Promote Research Infrastructure Accessibility for Diverse Demographics -- With the increasing size of datasets and complex computational resources required to participate in AI research, it is important to explore whether and how access to research infrastructure could be expanded. Providing state-of-the-art infrastructure across the research community (including open-source software and open datasets as appropriate) could improve access to responsible and transparent resources to support knowledge, innovation, and development in AI.

Offer Connection Opportunities -- Productive collaborative research depends on researchers knowing and trusting each other. While international conferences offer some opportunities for connection, funding for focused meetings on specific topics specifically help facilitate building connections between diverse global researchers, further enhancing these collaborations. A multinational research consortium[66] allowing researchers to pair with foreign partners could simplify procedures and reduce barriers to collaboration.

Pursue Public-Private Partnerships -- Public-Private Partnerships (PPPs) enable research funders to leverage the resources and expertise of the private sector through risk-sharing arrangements. These arrangements are essential in the AI space, as private sector parties have more resources to dedicate to experimentation and fundamental research and development. When properly designed and executed, PPPs can offer innovation in project design, incorporation of global expertise, and access to new sources of capital.

Catherine Breslin / Better Images of AI / Silicon Closeup / CC-BY 4.0

---

66   Trillion Parameter Consortium (TPC). 2024. "Trillion Parameter Consortium (TPC) - Generative AI for Science and Engineering." Trillion Parameter Consortium (TPC) - Generative AI for Science and Engineering. August 9, 2024. https://tpc.dev/?r34icsym=202409.

5.2 Research Ecosystem Hubs

Conferences, journals, and professional societies provide key hubs for international engagement and the codification of AI research practices. Leading AI conferences are experimenting with mechanisms for encouraging international engagement and strengthening a focus on responsible research. For example, some conferences include a required ethics statement for submitted papers, and in a few cases, these have led to papers being removed from consideration. Research ecosystem hubs have the potential to set the basis for ethical norms for research globally and should consider the following good practices.

Incentivize Reproducibility of Research -- The growing interest in AI and the rise of preprint servers, like arXiv, have improved access to recent research and enhanced the accessibility of scientific papers still awaiting peer review. However, it also poses challenges to scientific integrity. Academic journals and conferences typically select submissions that are of high quality and significance. Research ecosystem hubs should create mechanisms that incentivize the reproducibility of published research and elevate reproducibility as a global best practice to enhance scientific integrity in this space.

Foster Broadly Accepted Best Practices Around Research Publications -- It is important to address the surge in paper submissions driven by differing researcher assessment methods. Differences between how countries assess the success of research via publication seems to be driving a surge in the volume of papers submitted to top academic conferences. Tying salaries and promotions to publication numbers encourages a "flood the zone" strategy and unethical behaviors like collusion rings. Promoting best practices for peer-review quality can mitigate these issues.

Coordinate and Collaborate Around AI Research Guidelines – Research is informed by researcher background, data and resources, methodology selection, planned dissemination and a multitude of other factors. International coordination and collaboration on research guidelines can assist research teams in fostering inclusive, relevant, rigorous, conflict-of-interest-free, and trustworthy AI development, global collaboration, and equitably distributed AI benefits.

Support Research Beyond Topics Prioritized by the Private Sector -- Collaboration with the private sector, including through mechanisms like Cooperative Research and Development Agreements (CRADAs) with national and international partners at facilities such as DOE national laboratories, is valuable. Academic researchers are highly aware of industry interests (such as issues surrounding the maintenance of data centers). Competition for resources can lead researchers to be less engaged on research issues that expose them to diverse communities and contribute to maximizing AI's global benefit.

Provide Guidance to Navigate the Evolving Regulatory Landscape -- Differences in regulatory regimes pose challenges for researchers. Professional societies and international convening bodies can guide researchers in navigating the evolving regulatory landscape and work to promote ethical standards throughout the entire research community.
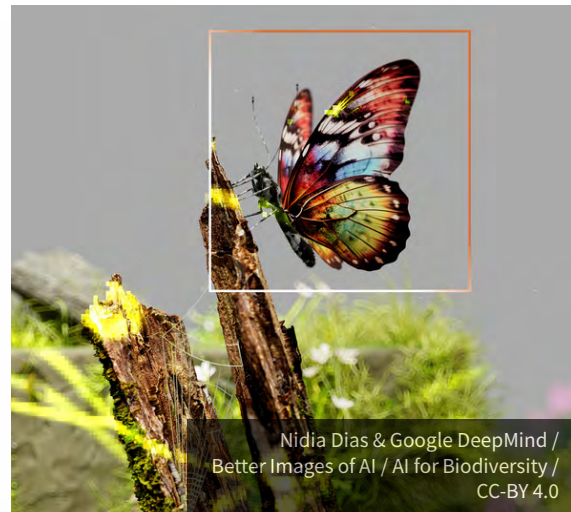
5.3 Research Teams

International collaboration on AI research is, ultimately, implemented through and by research teams. The make-up of research teams will determine which perspectives and expertise are represented, which methodologies are employed, and what research resources are accessible to the team, itself. Researchers should consider the following good practices when designing a team for a research project:

Prioritize Multidisciplinary Teams -- Multidisciplinary teams (including researchers in data science, computer science, and engineering, as well as those from social sciences and humanities) are key to holistically unpacking social impacts and opportunities of AI technologies. Encouraging multidisciplinary teams, while also supporting productive multidisciplinary engagements, can enhance outputs of applied research. AI used in global challenges, such as agriculture or for climate modeling, will require multidisciplinary teams for model validation and successful integration into applied workflows.

Collaborate with Local Researchers -- Funds can go to researchers from various universities, organizations, and locations; however, it is essential to encourage partnership with local researchers—those situated as close as possible to the populations being studied or impacted by the results of the research. Doing so leads to organic partnerships, long-term network building, and a pooling of resources and expertise.

Make Use of Sociotechnical Methodologies and Research Design -- It can be difficult for researchers to assess or audit AI models and tools and their impacts, particularly in settings that are novel, unfamiliar, or outside of the researchers' domestic context. Sociotechnical methods for system design and evaluation consider multiple factors outside of the technical, including human, social, and organizational, and typically deliver better value to stakeholders. Using these methodologies helps to ensure the research considers, from inception, the context within which it will be conducted and context in which results may translate into action.

Philipp Schmitt / Better Images of AI / Data flock (digits) / CC-BY 4.0

Risk Assessments -- AI has significant potential to enhance the ways in which scientific research is done and to make progress on difficult research problems but also introduces potential for new risks and harms. Testing and evaluation of AI systems, techniques for quantifying, tracking, and reducing risks of harm, and studies on community impact are all areas where additional funding for research, development, experimentation, and innovation are needed. Furthermore, international research collaborations offer the opportunity to disseminate best practices, and promising advancements.

SECTION 6: CONCLUSION

The Global AI Research Agenda (GAIRA) proposes a guiding framework for research funders, policymakers, and practitioners to contribute to a balanced international AI research ecosystem where technological progress and human well-being advance hand in hand.

The recommended principles and priorities outlined in this agenda aim to align AI development with fundamental values, promote human rights and human well-being, and to help ensure responsible AI development, globally. By implementing this agenda, the international research community can foster safe, secure, and trustworthy AI development, global collaboration, and equitable access to the benefits of AI, preventing misuse and aligning AI deployment with societal values. Progress in creating this balanced ecosystem would be characterized by the effective promotion of research into the priority areas within international research networks. Doing so involves continuing conversations with global research stakeholders to maintain engagement and ensure that each research opportunity receives the necessary attention. The combined effort in each area would lead to more successful collaboration, demonstrating that the whole is greater than the sum of its parts. Given that the priorities and objectives can be too numerous for any single entity to tackle alone, this endeavor necessitates a global call to action, fostering collaboration and burden-sharing across the international research community. Ultimately, this collective effort will pave the way for a future where AI technologies will contribute positively to our global community.

The Global **AI** Research Agenda