

ACM SIGCAS Computers and Society

SPECIAL ISSUE

*Navigating the Influences
of Human-Centered Computing
on Policy Innovation*

TABLE OF CONTENTS

General Information

- 1 Introduction to SIGCAS and Membership Benefits
- 2 SIGCAS Executive Committee
- 3 From the Editor

SIGCAS Updates

- 4 Call for Short Pieces

General Article Features

- 5 Is It Worth It? Considering The Cost and Value of Generative AI

Special Edition Article Features

- 9 Balancing Global Majority AI Ambitions with Environmental Considerations
- 13 Towards an Ecosystem Approach to Quantum Computing Ethics
- 18 Toward Genuine Policy Engagement in Human-Centered Computing
- 21 AI Governance and the Fundamentals of Data: The Relational, Political, and Material
- 31 Community Voice in Policy Action

Introduction to SIGCAS

SIGCAS Computers and Society is the ACM Special Interest Group that addresses the social aspects and ethical consequences of widespread computer usage. The main goals of SIGCAS are to raise awareness about the impact that technology has on society and to support and advance the efforts of those who are involved in this important work. Our members are computer professionals from both industry and academia, as well as ethicists, psychologists, sociologists, and others. We welcome students from a variety of disciplines. Our areas of involvement include computer ethics, universal access to computer technology, security, privacy, and reliability. We collaborate with other ACM bodies engaged in related work, such as COPE, USACM, SIGITE, and SIGCSE. ACM *Computers & Society* is an online publication accessible via the ACM Digital Library and the SIGCAS website. The newsletter aims to be an effective communication vehicle between the members of the group and the outside world.

SIGCAS Computers and Society Readers and writers are invited to join and participate actively in this Special Interest Group. Membership is open to all for \$25 USD per year and to students for \$10 USD per year. The link to join can be found on our website at www.sigcas.org.

Contribute. The Editor-in-Chief invites contributions of all types of written material (i.e., articles, working papers, news, interviews, reports, book reviews, bibliographies of relevant literature, and letters) on all aspects of computing that have a bearing on society and culture.

Please note that it is *NOT* a peer-reviewed publication. Submissions are checked for relevance, accessibility, and basic suitability by the editors but not fully peer reviewed.

For the latest Call(s) for Submissions, or instructions regarding formatting guidelines and copyright policy, please see the website: <http://www.sigcas.org/>. Submissions may be sent to editors_sigcas@acm.org.

Copyright & Notice to Contributing Authors to SIG Newsletters

By submitting your article or other material for distribution in this Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights:

- to publish in print on condition of acceptance by the editor
- to digitize and post your article or other material in the electronic version of this publication
- to include the article or other material in the ACM Digital Library and any Digital Library-related services
- to allow users to make a personal copy of the article or other material for non-commercial, educational, or research purposes.

However, as a contributing author, you retain copyright to your article or other material, and ACM will refer requests for republication directly to you.

Membership Benefits

- Subscription to the online publication ACM SIGCAS Computers and Society, which is published three to four times a year.
- Access to the full archive of the online publication and its printed predecessor in the ACM DL.
- Discounted registration fees for SIGCAS-sponsored conferences and workshops. “In cooperation” sponsor of several ACM and non-ACM conferences related to SIGCAS’ interests, including LIMITS.
- SIGCAS presents two annual awards: The Making a Difference Award and the SIGCAS Outstanding Service Award.
- SIGCAS-ANNOUNCE mailing list: includes regular announcements of upcoming conferences and calls for participation. SIGCAS-Talk mailing list: enables member-to-member interactions, and the committee will seek to stimulate discussion on this list amongst members. Subscription to the list is restricted to SIGCAS members and is optional.

Computers and Society Editorial Board

Chinasa T. Okolo, Technecultura

Computers and Society Executive Committee

Michelle Trim — SIGCAS Chair

Samuel Mann — SIGCAS Vice-Chair

Chinasa T. Okolo — Member-at-Large and Editor-In-Chief, *Computers and Society*

Brian Krupp — Member-at-Large

Doug Schuler — Member-at-Large

From the Editor

by Chinasa T. Okolo

Welcome to the latest edition of ACM SIGCAS *Computers and Society*! I'm pleased to publish this second edition in my capacity as Editor-in-Chief. Earlier this year, I commissioned my first special issue of *Computers and Society*, which focuses on the influences of human-centered computing on policy innovation. As a computer scientist working in policy, I aimed to develop a special issue that would enable this newsletter to build upon the rich history of topics covered by prior contributors and dive into the rapidly emerging landscape of digital policy.

We received numerous submissions and are pleased to publish pieces from a diverse set of contributors, covering topics such as environmental considerations, policy engagement, participatory research methods, the role of data in AI governance, and the ethics of quantum computing. These submissions represent the vast intricacies of human-centered computing and the role that emerging technologies play in influencing the design, development, and implementation of regulatory mechanisms.

Over the next year, I plan to deepen collaborations with the SIGCAS Executive Committee to further expand our impact. As SIGCAS continues its discussion on the best strategies for the organization, we look forward to finding new avenues to increase our engagement with the community. Please get in touch if you'd like to volunteer or have any questions or concerns.

Chinasa T. Okolo, Ph.D.

Editor-in-Chief, *ACM SIGCAS Computers and Society*

Call for Short Pieces

The SIGCAS Publications Committee aims to encourage more voices and varied perspectives on topics relevant to computers and society. In addition to longer pieces, we seek short pieces that are relevant, provocative, diverse, and unexpected for our issues of *Computers and Society*. Your short piece could raise arguments, issues, critical questions, resource needs, current work, research, reviews, discussions, etc. We currently plan to run several short pieces per issue.

To be considered for the next newsletter, please submit your short piece to the SIGCAS *Computers and Society* Editor-in-Chief, Chinasa T. Okolo, through email at editors_sigcas@acm.org. Submissions are due on March 15th, 2026 (the subsequent issue will be published in April 2026). Please include "Short Piece" in the subject line. Submissions should be a maximum of 1,200 words, drafted in Microsoft Word or Google Docs format, and references formatted in [IEEE style](#).

SIGCAS Editors may be contacted at editors_sigcas@acm.org with questions about the *Computers and Society* newsletter submission process.

SIGCAS *Computers and Society* Areas of Interest

Advanced Computing	For or From The Archives	Practitioner Perspectives
Art, Media, Computing, and Society	Gender & Women's Studies	Public Opinion
Climate, Biodiversity, & the Environment	Global Development Studies	Recent News & Reviews (Articles, Books, Movies, etc.)
Computing Ethics Education	Health & Medicine	Religion & Spirituality
Conflict & Militarization	History of Computing	Social Justice
Civic Technology	Human Computer Interaction	Student Voices
Digital Privacy	Human Rights & Immigration	Systemic Racism & Marginalization
Economic Development	Information Security & Cybersecurity	Urban Computing & Smart Cities
Ethics & Responsible Computing	Labor & the Workforce	User Experience & Design
	Peace & Conflict Studies	
	Policy & Computing	

ARTICLE

Is It Worth It? Considering The Cost and Value of Generative AI

by Brian Krupp

Case Western Reserve University

The Current State

The use and advancement of generative AI has exploded in the past few years. This has resulted in large tech companies seeking alternative approaches we would not have previously considered to supply the power-hungry processes that require training these advanced models. These approaches include not only placing data centers strategically near power plants, but also creating their own nuclear reactors. Quantifying the power consumption of these services is difficult, and skeptics of the value of generative AI could argue that quantification is intentionally difficult. One analysis compared the generation of an image to powering a smartphone [3]. Another analysis compared generating 100 words consumes up to 3 bottles of water [6]. There have already been warnings that energy consumption from data centers, AI, and cryptocurrency could double in just four years [1]. While this new demand is affecting some of big tech's climate goals, what is clear is that they are also looking for a competitive edge in obtaining power at a lower cost and are willing to take approaches we would not have previously considered.

The Value

Meanwhile, within the CS discipline and throughout society, some have questioned the

value that some of these services provide. From hallucinations to fake citations and even harmful therapy provided by generative AI [8], it can be challenging to understand the overall benefit to society that generative AI provides. Certainly, there are valuable ways of using these services. Generative AI can be used in areas like medicine to find new treatments or to aid in discovering new solutions to existing problems. One well-known example is the use of DeepMind to discover proteins in the protein-folding problem [9]. However, what most consumers are being sold is the ability to generate images or text for frivolous use or to become more "efficient". During the 2024 Olympics, Google's advertisement showed that a child could write a letter to an Olympic athlete they admired [4]. Microsoft's ad showed how a person could quickly create a presentation in 15 minutes. I can't help but think as a human, do I really want to read a letter that another human did not actually write? Do I really want someone to sit in a meeting and watch a presentation that someone did not create? I do not, nor do I imagine that many people would say yes to either of those things.

I can't help but question, as we see these commercials, are we so willing to give up what makes us human? Should we slow down and let ourselves write a letter that has grammar mistakes or that is sincerely from an admiring child? Or should we let the real constraints of

time make us question if rushing a presentation made by generative AI is worth having our colleagues sit in yet another meeting where they could spend that time doing something more creative?

It can be easy to sound as a modern-day Luddite with these concerns. But, as a computer scientist, I still believe that our discipline can have a profound positive impact on society. Within our own research group, CASLab¹, we have worked on projects that have a direct impact on the community, such as reducing food waste, reducing food insecurity, and improving the ability for communities to have a better understanding of their air quality. There are dedicated conferences, such as GoodIT, where researchers are focused on solutions that have an overall benefit to society. Within our professional organization, ACM, we have a dedicated special interest group (SIG) whose primary interests are the intersection of computing and society (SIGCAS).

Climate Change

As a college professor, I see people walk around campus scrolling through their phones, ignoring the scenery and sounds around them. This is not just on college campuses, but you can observe the same in many public spaces. As I observe this, it becomes easier to understand why we ignore the warning signs from our climate. We are breaking records of days with extreme heat, and this past year, 2024, was the warmest year on record [2]. Ice sheets in Antarctica are breaking off at a rapid pace, and the ice in Greenland is melting at an alarming rate [10]. Hurricanes are becoming stronger,

¹ <https://caslab.case.edu>

more frequent, and areas that were previously thought to be immune to climate change are being impacted by severe drought and flooding.

There is certainly irony in how big tech companies broadcast sustainability and environmental goals; however, they are willing to build their own nuclear power plants or bring back online a nuclear power plant that previously had a partial meltdown. Yet, as the world literally burns and temperatures increase globally, the same efforts could have occurred prior to the explosion of generative AI to help provide energy that produces fewer greenhouse gases. These companies only seemed interested in such efforts when the market provided a competitive advantage to do so. Big oil is also looking to take advantage of this increase in energy demands from data centers [5]. While some have offered "cleaner" alternatives and carbon capture, given the current state of the climate crisis, is it worth burning additional fossil fuels? Are we really benefiting from the AI race between big tech companies? Are these tools making us more efficient, or in their current state, being used to solve real-world problems? Or are we using them to write stories for other people to read, images for other people to see that have no real benefit to society?

What Can We Do?

What can we do then? At times, we may feel powerless as the world warms and companies seem to have more control over our individual lives. But, as leaders in our discipline, there are things that we can do.

Within our own research community, we travel thousands of miles to share ideas, where the infrastructure that our discipline helped create

allows us to share those ideas instantaneously in a more equitable and sustainable way. Moshe Vardi, in a previous column, brought attention to this issue, in how some of the travel requires trans-oceanic travel, where the cost is inconsistent with our goals of a professional society in serving the public good [11]. As researchers, we should also consider the impact of not only the outputs of our research, but also how we conduct our research and the use of various models that have used considerable amounts of resources from the environment to train these models. As teachers in the classroom, we can raise awareness of the energy and environmental impact of using generative AI. In my own experience, most students were not aware of the issues or the more relatable findings of power and water use in generative AI.

Within our personal communities, we can advocate for policies that align with more responsible use of resources in the environment. An example of where this was effective was in Virginia, home to the nation's largest data center market. Previously, there was a debate about running data centers on carbon-emitting diesel generators during power shortages. However, after facing significant public pushback from environmental groups, the area utility is exploring other options [7].

We should be leading the way, not when it is economically convenient, but just as we have goals in our discipline for the future, the impacts of our discipline, and the sustainability of our discipline, must be at the forefront. We cannot afford to ignore the impact that our research, teaching, and service can have on the environment.

Final Thoughts

The direction we choose does not mean we halt advancement in our field. Being mindful of how we invest our energy and other resources can put us on a track where we can still make scientific discoveries, but in a more sustainable way. However, we need to pause, take a step back, and think about what it is that we are truly trying to achieve and what benefit does this have for society? What is the cost? Not just financially, but on society and the environment?

We are at an inflection point in computing and also within the ACM. Is the work in generative AI really benefiting society? Is it worth the power that it requires? I understand that it may be impractical to analyze each decision or step within our own research at this level, but it seems clearer than ever that the power cost is too high given the current state of the climate, and the benefit to society does not justify the means.

We are heading down a path that large tech companies have paved that does not provide much benefit for society. Collectively, we have also ignored the debts being accrued when we do not consider the costs of the advancement of our field. That debt is having an existential impact on the future generation. We can no longer afford to ignore the impacts, while our markets have put profit over purpose. We are using vast amounts of energy and talent to create things that distract us and do not have a larger benefit to society. There are undoubtedly benefits to the advancement of artificial intelligence. But, in its current state and investment, we are on an unsustainable path for our planet and society.

References

- [1] C. Crownhart. 2024. AI is an energy hog. This is what it means for climate change. MIT Technology Review. [Online]. Available: <https://www.technologyreview.com/2024/05/23/1092777/ai-is-an-energy-hog-this-is-what-it-means-for-climate-change>.
- [2] R. Bardan. 2025. Temperatures Rising: NASA Confirms 2024 Warmest Year on Record. NASA. Accessed: January 10th, 2025. [Online]. Available: <https://www.nasa.gov/news-release/temperatures-rising-nasa-confirms-2024-warmest-year-on-record>
- [3] M. Heikkilä. 2023. Making an image with generative AI uses as much energy as charging your phone. MIT. Accessed: December 1st, 2023 from <https://www.technologyreview.com/2023/12/01/1084189/making-an-image-with-generative-ai-uses-as-much-energy-as-charging-your-phone>
- [4] L. Holms. 2024. The antithesis of the Olympics: Using AI to write a fan letter. NPR. Accessed: August 3rd, 2024. [Online]. Available: <https://www.npr.org/2024/07/30/nx-s1-5056201/google-olympics-ai-ad>
- [5] S. Khan and S. Awasthi. 2024. Big Oil eyes powering Big Tech's data center demand. Reuters. Accessed: December 12, 2024. [Online]. Available: <https://www.reuters.com/business/energy/chevron-working-supply-power-data-centers-executive-says-2024-12-11>
- [6] S. T. P. Verma. 2024. A bottle of water per email: The hidden environmental costs of using AI chatbots. Washington Post. Accessed: September 18, 2024. [Online]. Available: <https://www.washingtonpost.com/technology/2024/09/18/energy-ai-use-electricity-water-data-centers/>
- [7] L. Ramadan and S. Brownstone. 2024. Data centers demand a massive amount of energy. Here's how some states are tackling the industry's impact. Ars Technica. Accessed: August 4th, 2024. [Online]. Available: <https://arstechnica.com/science/2024/08/data-centers-demand-a-massive-amount-of-energy-heres-how-some-states-are-tackling-the-industrys-impact/>
- [8] E. Roth. 2024. Character.AI and Google sued after chatbot-obsessed teen's death. The Verge. Accessed: October 23rd, 2024. [Online]. Available: <https://www.theverge.com/2024/10/23/24277962/character-ai-google-wrongful-death-lawsuit>
- [9] M. Sparkes. 2022. DeepMind's protein-folding AI cracks biology's biggest problem. NewScientist. Accessed: July 28th 2022. [Online]. Available: <https://www.newscientist.com/article/2330866-deep-minds-protein-folding-ai-cracks-biologys-biggest-problem>
- [10] C. Tabachnick. 2024. World Greenland's ice sheet melting faster than scientists previously estimated, study finds. CBS News. Accessed: January 17th, 2024. [Online]. Available: <https://www.cbsnews.com/news/greenland-ice-sheet-melting-faster-than-scientists-previously-estimate>
- [11] M. Y. Vardi. 2023. ACM for the Public Good. Commun. ACM 66, 5 (April 2023), 5. <https://doi.org/10.1145/3589609>

Balancing Global Majority AI Ambitions with Environmental Considerations

by Chinasa T. Okolo

Technecultura

Keywords: AI Policy, Sustainable Development Goals, Artificial Intelligence, Environmental Impact

Introduction

The promise of artificial intelligence as a developmental accelerator for Global Majority countries has captured considerable attention in policy circles, with proponents suggesting these technologies could help nations "fast-track" progress toward the Sustainable Development Goals. Yet this optimistic narrative obscures a more complex reality where AI's transformative potential intersects with significant risks already manifesting globally, with notable impacts in Global Majority countries. From exploitative labor practices in data annotation to algorithmic bias that deepens existing inequalities, communities worldwide are confronting the uneven distribution of AI's benefits and harms.

The Environmental Burden of Digital Progress

The ecological footprint of AI development presents one of the most pressing yet underexamined challenges in this landscape. The computational infrastructure underlying modern AI systems demands enormous quantities of natural resources, particularly energy and water for cooling data centers, training large language models, and supporting inference operations. Recent research from UC Riverside and UT Arlington projects that global AI demand could require between 4.2 and 6.6

billion cubic meters of water withdrawal by 2027, which is equivalent to four to six times Denmark's total annual water consumption [1]. While initiatives like the AI Energy Score, a collaborative framework led by Hugging Face and Salesforce, attempt to standardize energy efficiency metrics across AI models, a comprehensive assessment of the environmental impacts of AI systems remains elusive [2]. While insightful, many efforts to measure the natural resource and energy consumption of AI models represent conservative estimates. These efforts are constrained by the opacity that characterizes much of the AI industry's environmental reporting, given the lack of transparency provided by tech companies and other large model developers on the data, computing architecture, and other factors used to train AI systems. Overall, this lack of transparency creates particular challenges for measuring and mitigating AI's environmental impact. However, research on the environmental impact of AI represents a crucial step toward accountability, offering researchers and developers tools to compare energy consumption across different AI architectures and consumers with awareness of how their AI usage consumes precious resources. For Global Majority countries facing electricity limitations and constrained cloud computing infrastructure, such benchmarking

tools could prove especially valuable in optimizing AI workflows within existing resource constraints.

Climate Change and Algorithmic Harm

The environmental costs of AI development compound existing climate vulnerabilities that disproportionately affect Global Majority communities [3]. These populations face heightened exposure to climate-induced extreme weather events, which generate cascading effects including resource conflicts, food insecurity, and forced displacement. The rapid expansion of AI development and increasing investments in AI infrastructure to support AI scaling threaten to intensify these challenges while simultaneously introducing new forms of harm through biased algorithms, discriminatory decision-making systems, and the proliferation of misinformation [4]. The underrepresentation of Global Majority communities in AI training datasets exacerbates these risks [5]. When AI systems are deployed in contexts without being trained on relevant cultural, linguistic, or contextual data, they often perpetuate or amplify existing inequalities. This dynamic is particularly problematic in automated content moderation systems, where the absence of multilingual and culturally nuanced training data can lead to the systematic marginalization of non-English speaking communities [6]. As governments and organizations increasingly turn to automated solutions, and in some cases, abandon human fact-checking for content governance, the risk of scaling these biases across entire information ecosystems becomes acute [7]. These intersecting vulnerabilities underscore the need to adequately address climate change,

necessitating that governments acknowledge the unique challenges faced by Global Majority communities and implement policies that prioritize their specific needs and perspectives. This includes providing financial and technical assistance to help them adapt to climate change, supporting their efforts to develop and adopt sustainable technologies, and addressing the systemic inequalities that exacerbate communities' vulnerability to climate-related impacts.

Bridging Climate Policy and AI Regulation

Despite the interconnected nature of climate and AI challenges, policy frameworks have struggled to comprehensively address their intersection. One hundred ninety-eight countries participate in the United Nations Framework Convention on Climate Change [8], and 196 countries, territories, and multilateral bodies (including the European Union) have drafted or enacted climate legislation, according to the Climate Change Laws of the World database [9]. However, the Climate Action Tracker, an independent scientific project tracking government climate action against goals outlined in the Paris Agreement, indicates that no individual country has enacted policies or made commitments consistent with the Paris Agreement's goal of limiting global warming to 1.5 degrees [10]. This governance gap becomes more pronounced when considering AI's climate impact, which remains largely absent from existing regulatory frameworks. Paradoxically, Global Majority countries demonstrate some of the strongest climate leadership within current international frameworks. Nine of the ten countries rated "Almost Sufficient" by the

Climate Action Tracker—Bhutan, Chile, Costa Rica, Ethiopia, the Gambia, Kenya, Morocco, Nepal, and Nigeria—are in the Global Majority. While there is still much progress to be made towards global climate action, the Climate Action Tracker identifies notable data gaps across Africa, the Caribbean, Central Asia, and Oceania. Nevertheless, Global Majority communities have already demonstrated strong leadership in developing local climate solutions [11]. This leadership suggests significant potential for developing climate-conscious AI solutions that prioritize smaller models optimized for limited internet connectivity, small datasets, and the high mobile penetration rate across these regions.

Balancing Development Imperatives with Responsible AI Innovation

Sufficiently addressing climate concerns related to AI should be a priority for all governments worldwide. However, the tension between development aspirations, climate considerations, and responsible AI deployment creates complex policy challenges for Global Majority governments. Many of these nations prioritize AI adoption as a solution to persistent socioeconomic development problems, which could lead to insufficient attention on longer-term risks and unintended consequences. While a vast number of AI solutions have shown promise in helping to address problems such as flood forecasting [12], humanitarian aid distribution [13], conservation [14], and poverty mapping, many AI tools are being rapidly adopted for more sensitive domains such as asylum application processing [15], refugee management [16], and public surveillance [17]. This nearsighted focus

on AI adoption and rush to implement AI solutions without adequate safeguards risks not only wasting limited resources on ineffective technologies but also causing substantial harm to vulnerable populations. This pattern reflects what scholars have termed "technosolutionism," the tendency to view complex social problems as amenable to technological fixes without addressing underlying structural inequalities, an issue that has been increasingly present in efforts focused on sustainable development [18].

Pathways Towards Robust Climate and AI Governance

Addressing these interconnected challenges requires approaches that center local expertise and context-specific research. The predominant focus of AI fairness, labor implications, and climate impact research on Western, high-income countries creates knowledge gaps that leave Global Majority nations without adequate frameworks for understanding and mitigating risks within their specific contexts. Developing regional research capacity across the Global Majority and supporting locally-led investigations into AI's climate and sociotechnical impacts represents a crucial step toward more robust, equitable, and effective governance. The path forward requires recognizing that climate action and responsible AI development are not competing priorities but interconnected challenges that demand integrated solutions. By leveraging local expertise, supporting context-relevant research, and prioritizing community needs in technology design, countries in the Global Majority can contribute to both climate mitigation and equitable AI governance. This approach offers

the potential to transform constraints into innovations, creating AI systems that serve developmental goals and environmental sustainability while respecting the rights and dignity of all communities.

References

- [1] P. Li, et al. Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models. *Commun. ACM* 68, 7, July 2025, 54–61. <https://doi.org/10.1145/3724499>.
- [2] AI Energy Score. Accessed: Jun. 19, 2025. [Online]. Available: <https://huggingface.co/AIEnergyScore>
- [3] B.S. Levy and J. A. Patz. "Climate change, human rights, and social justice." *Annals of global health* 81.3 (2015): 310-322.
- [4] C. T. Okolo. "African democracy in the era of generative disinformation: Challenges and countermeasures against AI-generated propaganda." arXiv preprint arXiv:2407.07695, 2024.
- [5] "International AI Safety Report." Accessed: Jun. 18, 2025. [Online]. Available: <https://www.gov.uk/government/publications/international-ai-safety-report-2025/international-ai-safety-report-2025>
- [6] M. Elswah, "Moderating Maghrebi Arabic Content on Social Media," Center for Democracy and Technology. Accessed: Jun. 18, 2025. [Online]. Available: <https://cdt.org/insights/moderating-maghrebi-arabic-content-on-social-media/>
- [7] L. McMahon, Z. Kleinman & C. Subramanian, "Facebook and Instagram get rid of fact checkers," Accessed: Jun. 18, 2025. [Online]. Available: <https://www.bbc.com/news/articles/cly74mpy8klo>
- [8] United Nations Climate Change, "Parties to the United Nations Framework Convention on Climate Change." Accessed: Jun. 19, 2025. [Online]. Available: <https://unfccc.int/process/parties-non-party-stakeholders/parties-convention-and-observer-states>
- [9] Climate Change Laws of the World. Accessed: Jun. 18, 2025. [Online]. Available: <https://climate-laws.org/?q=Climate+framework+laws&y=2024&y=2025>
- [10] Climate Action Tracker. Accessed: Jun. 19, 2025. [Online]. Available: <https://climateactiontracker.org/>
- [11] UNDP, "The Global South have mastered locally led climate adaptation solutions. It's time to scale them up." Accessed: Jun. 18, 2025. [Online]. Available: <https://www.adaptation-undp.org/global-south-have-mastered-locally-led-climate-adaptation-solutions-its-time-scale-them>
- [12] G. Nearing, et al. "Global prediction of extreme floods in ungauged watersheds." *Nature* 627.8004 (2024): 559-563.
- [13] E. Aiken, et al. "Machine learning and phone data can improve targeting of humanitarian aid." *Nature* 603.7903, 2022: 864-870.
- [14] L. Xu, et al. "Stay ahead of Poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations." 2020 IEEE 36th international conference on data engineering (ICDE). IEEE, 2020.
- [15] M. Forster. "Refugee protection in the artificial intelligence era," Chatham House. Accessed: Jun. 19, 2025. [Online]. Available: <https://www.chathamhouse.org/2022/09/refugee-protection-artificial-intelligence-era>
- [16] S. Abboud. "Artificial Humanitarianism—The Data-Driven Future of Refugee Responses," MERIP. Accessed: Jun. 19, 2025. [Online]. Available: <https://merip.org/2025/01/artificial-humanitarianism>
- [17] K. Hao and H. Swart. South Africa's private surveillance machine is fueling a digital apartheid. Accessed: Jun. 19, 2025. [Online]. Available: <https://www.technologyreview.com/2022/04/19/1049996/south-africa-ai-surveillance-digital-apartheid/>
- [18] H.S. Sætra. "Technology and sustainable development: The promise and pitfalls of techno-solutionism." Taylor & Francis, 2023.

Towards an Ecosystem Approach to Quantum Computing Ethics: Implications for the Global Majority

by Shamira Ahmed

Delft University of Technology

Keywords: Responsible Research and Innovation (RRI), Quantum Computing (QC), Innovation Ecosystems, Ethics, Global Majority

Introduction

Quantum Computing (QC) represents a class of technologies that harness fundamental principles of quantum mechanics, particularly superposition and entanglement, to process, transmit, and secure information in ways unattainable by classical systems [1]. Positioned at the core of Quantum Information and Communication Technologies (QICT) [2], QC is central to the ongoing Second Quantum Revolution, driving innovation infrastructures and redefining capabilities in computation, networking, and secure information exchange at fundamental levels [3]. Unlike the first quantum revolution, which focused on physics and laboratory experiments, today's Quantum technology (QT) applications are inherently digital, dependent on and deeply integrated within the digital stack, relying on cloud platforms, high-performance computing, advanced telecommunications, and potentially converging with other emerging digital technologies (EDT) such as artificial intelligence (AI), blockchain, big data, the Internet of Things (IoT) [4], [5], offering new paradigms for the digital economy and society [6]. While there remains considerable uncertainty regarding their practical, scalable applications, the disruptive potential of QC is increasingly recognized [7]

with profound ethical, legal, social, and policy implications [8] and a need for practical approaches to their governance [9]. QC is not a standalone technology, but operates as a complex innovation ecosystem where research, policy, markets, and ethics co-evolve [10].

Rationale for an Ethics-Based Approach

As an EDT, QC cannot be regarded as intrinsically good or bad; their impacts are determined by how they are conceived, developed, and applied within specific social, cultural, and political settings [5], [11]. Innovation ecosystem outcomes are not neutral but reflect the values, governance frameworks, and power structures embedded in their design and deployment [12], [13]. In this respect, QC should be understood as sociotechnical systems, where technical capabilities and societal contexts co-evolve and mutually shape one another [4], [11]. Therefore, an ethics-based approach is essential to guide governance across the QC innovation ecosystem. Ethics for innovation ecosystems must not be treated as an external constraint, but rather as an embedded element of research, design, and policymaking, ensuring that QC development and deployment align with societal goals such

as accountability, justice, and intergenerational responsibility [5], [9], [10], [11].

Navigating Ethics from Human-Centered Computing (HCC) towards a Responsible Research and Innovation (RRI) Ecosystems approach

Human-Centered Computing (HCC) emphasizes inclusivity, usability, and fairness in human-machine interaction. HCC has provided effective frameworks for addressing design-level shortcomings in emerging technologies and ensuring that systems are accessible and responsive to diverse user needs to address complex entangled systems [13]. However, given the systematic risks and complexity of dimensions associated with QC [10], [14], HCC provides a limited scope when applied to QC; threats of QC extend far beyond interface-level concerns [15] and require transdisciplinary knowledge integration across disciplinary, institutional, and societal boundaries [16]. For example, the potential breaking of public-key cryptography would disrupt financial infrastructures, communication systems, and governance mechanisms worldwide, challenges that cannot be solved only through user-centered design principles alone [4], [10].

To address broader challenges, QC ethics requires a shift from HCC to Responsible Research and Innovation (RRI). RRI provides a framework that embeds anticipation, inclusion, reflexivity, and responsiveness into ethics-related discourse on information systems and broader innovation ecosystems for EDT [5], [15], [17]. RRI offers a systemic approach that expands ethics beyond artefacts or interfaces to

encompass research agendas, commercialization pathways, infrastructures, power, regulation, and policy frameworks [9], [10], [15], [18]. For QC, this means ethics must be integrated across its entire lifecycle that shapes cryptographic standards, telecommunications integration, and international internet governance processes, to name a few [6], [9], [10].

Recent scholarship highlights the need for responsible innovation in QC, since RI approaches remain underdeveloped, with ethical concerns often sidelined in favor of technical progress [13], [18]. Similarly, there is consensus that failing to embed ethics at the innovation ecosystem level risks exacerbating inequalities, geopolitical asymmetries, and legitimacy crises associated with EDT [15]. By embedding RRI principles, ethical considerations for QC evolve from being a constraint to becoming a constitutive element of innovation quality control, ensuring that QC serves societal well-being, sustainability, and justice rather than narrow technical or commercial interests [9], [13], [18].

An innovation ecosystem approach is central to the ethical analysis in this paper for several reasons [23]; First, QC raises unique and amplified ethical challenges, at a systems level, particularly in relation to privacy, security, inequality, and responsible QC [7], [10], [11], [13]. Second, QC's unprecedented computational capacity directly threatens classical encryption systems, raising urgent concerns around large-scale privacy breaches, data protection, and the security of critical infrastructures [4]. Third, as QC requires significant financial, technical, and human

capital investments, access will likely be restricted to advanced nations and global corporations, a dynamic that risks deepening digital divides unless more equitable innovation ecosystems are fostered [5]. Fourth, QC could enable new forms of surveillance and profiling, intensifying the need for ethical and legal safeguards to prevent misuse and protect human rights in [19]. Finally, the energy intensity and specialized hardware demands of QC highlight pressing sustainability and environmental ethics considerations that must inform its development [20].

Beyond technical milestones, transitioning towards the quantum advantage (QA) carries profound ethical and geopolitical implications. The reality where QC can solve specific problems faster or more efficiently than the best available classical supercomputers could disrupt global cybersecurity, reshape industrial competitiveness, and widen digital divides if access is limited to a small number of powerful actors [4], [9]. This paper proposes positioning QC as an urgent and illustrative focal point for evolving from HCC approaches to RRI ecosystems discourses for informing ethics-related research to ensure that digital transformation towards QA remains human-centric and rights-oriented [13], [15].

In addition, the existing frameworks for assessing technological opportunities and risks are often insufficient and require substantial re-evaluation to effectively address the novel challenges and possibilities associated with QC [16]. Furthermore, QC will not operate independently but will be closely integrated with classical computing technologies and information systems which may amplify existing

challenges found in classical computing such as cybersecurity risks, system compatibility, and resource demands [9], [13], [21]. Another critical consideration is that QC demands considerable investments in resources and infrastructure, raising issues of equitable access and distribution of benefits and risks, which are particularly acute for under-resourced nations and stakeholder groups [4], [6].

Emerging Risks: Focus on the Global Majority

QC decryption risks are uneven across Global Majority regions due to differences in digital infrastructure, governance, and reliance on foreign providers [22].

Emerging QC ELSPI and governance frameworks reflect Global North priorities [14], with limited consideration of the Global Majority. While these initiatives illustrate strong political will and significant investment in QC governance, they reflect the priorities of high-income economies and risk marginalizing the Global Majority, where infrastructures are fragmented, and policy frameworks remain underdeveloped [22].

Imposing Global North sociotechnical imaginaries risks misalignment where local priorities differ such as prioritizing broadband access or regional cloud sovereignty [22], [23], the imbalance of priorities raises ethical concerns about reinforcing global asymmetries, particularly if powerful states dominate the early design of QC standards and infrastructures, thereby shaping global governance trajectories in ways that may not reflect the priorities of lower- and

middle-income countries. From an ethical standpoint, the concentration of power risks entrenching digital dependency and undermining data sovereignty in the Global Majority. Ethical considerations, therefore, demand that QC governance frameworks move beyond narrowly defined national security imperatives to address inclusivity, justice, and equity at global scales [5], [9], [24].

RRI offers a framework to counter by embedding anticipation, inclusion, reflexivity, and responsiveness into transnational governance [5]. In practice, this means engaging citizens and civil society in discussions of quantum security, ensuring that post-quantum cryptography (PQC) is accessible across different economies, and aligning QC strategies with regional priorities [10], [15]. By reframing QC ethics through RRI, ethics becomes a constitutive part of ecosystem resilience in the Global Majority. This approach ensures that QC innovation pathways not only mitigate ethical risks but also advance self-determined digital futures, sustainable development, and intergenerational justice [5].

Future Implications and Conclusion

Lessons from AI offer key insights that highlight risks of reactive regulation, the need for multi-stakeholder governance, and the importance of international cooperation. An ethical approach to QC innovation ecosystems must balance innovation with security while maintaining public trust. HCC methodologies ensure policies remain transparent and accessible, while RRI offers a framework to counter by embedding anticipation, inclusion,

reflexivity, and responsiveness into transnational governance.

By reframing QC ethics through RRI at an ecosystem level, ethics becomes a constitutive part of ecosystem resilience for the Global Majority and offers a powerful lens to rethink innovation ecosystems, particularly how we design, who we design with, and whose expertise is valued in developing QC innovation ecosystem pathways, not only to mitigate risks but also to advance self-determined digital futures, sustainable development, and intergenerational justice.

References

- [1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [2] A. Manzalini and L. Artusio, "The rise of quantum information and communication technologies," *Quantum Reports*, vol. 6, pp. 29–40, 2024, doi: 10.3390/quantum6010003.
- [3] J. P. Dowling and G. J. Milburn, "Quantum Technology: The Second Quantum Revolution," arXiv preprint arXiv:quant-ph/0206091, 2002. [Online]. Available: <https://arxiv.org/abs/quant-ph/0206091>
- [4] World Economic Forum, *Quantum Computing Governance Principles*. Geneva, Switzerland: WEF, 2022. [Online]. Available: https://www3.weforum.org/docs/WEF_Quantum_Computing_2022.pdf
- [5] B. C. Stahl, *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*. Cham, Switzerland: Springer, 2021.
- [6] Organisation for Economic Co-operation and Development (OECD), *Quantum Technologies as a New Paradigm for Digital Economies and Societies*, Paris, France: OECD, 2025. [Online]. Available: https://www.oecd.org/en/publications/quantum-technologies-as-a-new-paradigm-for-digital-economies-and-societies_e6664d58-en.htm
- [7] F. S. Khan and D. La Torre, "Quantum information technology and innovation: A brief history, current state and future perspectives for business and

- management,” *Technology Analysis & Strategic Management*, vol. 33, no. 11, pp. 1281–1289, 2021, doi: 10.1080/09537325.2021.1991576.
- [8] M. Kop, “Quantum-ELSPI: A novel field of research,” *DISO*, vol. 2, p. 20, 2023, doi: 10.1007/s44206-023-00050-6.
- [9] E. Perrier, “The Quantum Governance Stack: Models of governance for quantum information technologies,” *DISO*, vol. 1, p. 22, 2022, doi: 10.1007/s44206-022-00019-x.
- [10] M. Kop et al., “Towards responsible quantum technology: Safeguarding, engaging and advancing quantum R&D,” *UC Law Science and Technology Journal*, vol. 15, no. 1, pp. 63–94, 2024. [Online]. Available:
- [11] B. C. Stahl, G. Eden, M. Jirotko, and M. Coeckelbergh, “From computer ethics to responsible research and innovation in ICT: The transition of reference discourses informing ethics-related research in information systems,” *Information & Management*, vol. 51, no. 6, pp. 810–818, 2014.
- [12] L. M. Possati, “Ethics of Quantum Computing: An Outline,” TU Delft Repository, 2023. [Online]. Available: https://repository.tudelft.nl/file/File_424d9ca8-4ae0-4a9e-8333-6ba0e58890c6
- [13] B. Shneiderman, *Leonardo’s Laptop: Human Needs and the New Computing Technologies*. New York, NY, USA: ACM, 2005. doi: 10.1145/1099554.1099555.
- [14] M. Golec, “Quantum cloud computing: Trends and challenges,” *ScienceDirect*, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2949948824000271>
- [15] A. K. Kar, W. He, F. C. Payton, V. Grover, A. S. Al-Busaidi, and Y. K. Dwivedi, “How could quantum computing shape information systems research – An editorial perspective and future research directions,” *International Journal of Information Management*, vol. 80, p. 102776, 2025, doi: 10.1016/j.ijinfomgt.2024.102776
- [16] A. Unger et al., “From inter- to transdisciplinarity via research labs: A format for joining diverse perspectives and expertise in HCI research,” in *Proc. Mensch und Computer 2025 (MuC ’25)*, New York, NY, USA: ACM, 2025, pp. 599–604, doi: 10.1145/3743049.3748577.
- [17] F. Arute et al., “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, Oct. 2019, doi: 10.1038/s41586-019-1666-5.
- [18] P. Inglesant, C. Ten Holter, M. Jirotko, and R. Williams, “Asleep at the wheel? Responsible Innovation in quantum computing,” *Technology Analysis & Strategic Management*, vol. 33, no. 11, pp. 1364–1376, 2021, doi: 10.1080/09537325.2021.1988557.
- [19] O. van Daalen, “Developing a human rights compatible governance framework for quantum computing,” *Research Directions: Quantum Technologies*, vol. 2, p. e1, 2024, doi: 10.1017/qut.2024.2.
- [20] N. Arora and P. Kumar, “Sustainable Quantum Computing: Opportunities and challenges of benchmarking carbon in the quantum computing lifecycle,” *arXiv preprint arXiv:2408.05679v2*, Aug. 2024.
- [21] Kaspersky, “Kaspersky highlights top risks of Quantum Computing,” Jul. 14, 2025. [Online]. Available: <https://me-en.kaspersky.com/about/press-releases/kaspersky-highlights-top-risks-of-quantum-computing>
- [22] European Centre for International Political Economy (ECIPE), “Adapting under pressure: The Global South’s quantum journey,” Jul. 23, 2025. [Online]. Available: <https://ecipe.org/blog/the-global-souths-quantum-journey/>
- [23] M. Smolka and S. Böschen, “Responsible innovation ecosystem governance: Socio-technical integration research for systems-level capacity building,” *Journal of Responsible Innovation*, vol. 10, no. 1, pp. 1–18, 2023, doi: 10.1080/23299460.2023.2207937.
- [24] S. Umbrello, Z. C. Seskir, and P. E. Vermaas, “Communities of quantum technologies: Stakeholder identification, legitimation, and interaction,” *Int. J. Quantum Inf.*, pp. 1–25, 2024, doi: 10.1142/S0219749924500126.

SPECIAL ISSUE ARTICLE

Beyond “Policy-Washing”: Toward Genuine Policy Engagement in Human-Centered Computing

by Michal Luria

Center for Democracy & Technology

Keywords: Research Practices, Policy Engagement, Academia

Introduction

In recent years, researchers in human-centered computing and adjacent fields have increasingly found themselves encouraged to highlight “real-world impacts” in their work; by funders, by their departments, and by publication venues. But given low incentives to do so in practice, and especially compared to publication and other academic-focused requirements, most academics tend to spend time and effort within their own scholarly ecosystems, publishing at key venues and presenting to familiar faces. On the one hand, this produces deep, rigorous scholarship and a sense of professional community. On the other hand, the research itself also stays within the boundaries of the immediate academic community.

Within the various channels that push for “real world” or “broader” impacts, funding agencies are significant actors, encouraging researchers to identify and emphasize contributions to society when submitting research grant proposals [1]. While well-intentioned, researchers often treat the requirement for assessing societal impact as a secondary priority, sometimes drafting it in a rush as the last checkbox before a submission deadline, which also explains why scholars have found

that “broader impacts” tend to decline throughout a research project cycle [2].

One of the common areas invoked here is ‘policy’ and ‘policymaking’ — which is perceived as a significant domain of broader impact, as it very well might be. But fast forward to the publication of proposed broader impact in the policy landscape, and abstracts are riddled with mentions of policy and significance to policymakers. Unfortunately, many leave it at that. In conducting a meta-review on the intersection of conversational AI and policy, my colleagues and I found tens of papers that initially included in the paper corpus made claims about potential impact on policy in the abstract, but never revisited or extended those claims in the body of the work [3]. In fact, almost a quarter (24%) of all papers mentioned “policy” or “policymakers” only once, in the abstract, with no additional reference. We refer to this practice as “policy-washing”: a superficial invocation of policy that gestures towards policy but provides no substantial engagement.

Here too, the problem is unlikely to be mal-intentioned, but rather motivated by researchers wanting their work to matter, and perhaps pressure to make their research appear

policy-relevant. However, these statements result in diluting the credibility of the field and the use of ‘policy’ related terms, and make it harder for substantive policy-engaged research to stand out and get into the hands of policymakers and other policy experts.

If researchers wish to make genuine contributions to policymaking, a single, “policy-washing” declarative sentence within a paper abstract is not enough. Below, I offer three ways in which researchers, academic leaders, and funders can take steps that could add clarity to the use of policy-related terms in research papers, and in appropriate cases, make research more accessible to policy audiences.

Researchers can identify and highlight actionable takeaways from research. Within one’s own work, spending more time on narrowing the research-policy gap [4], [5] may be worthwhile in an attempt to make policy contributions more evident. This can be done by thoroughly considering what concrete insights a policy expert or policymaker could reasonably use or find valuable; technical results should be translated into plain language; statistical analysis should be interpreted; and complicated and nuanced findings should lay out key arguments. Takeaways should be highlighted within a publication in a way that would be easy for stakeholders to locate and engage with. Alternatively, these takeaways could be a separate and shareable output that researchers make use of when policy-related opportunities arise (see [6] as an example).

One strategy to make takeaways more accessible to policy audiences is to incorporate scenarios; instead of vaguely suggesting that

policymakers should “consider” your findings, illustrate how the issue might arise in a real-world context. For example, a school deploying chatbots for student support, or a healthcare provider making use of LLMs for diagnosis support. By framing research in terms of concrete applications, findings can be more legible to various audiences.

Leaders in academia should incentivize deeper engagement with policy implications.

To shift away from a policy-washing mindset, academics who serve as chairs and editors of conference and journal publications could require authors to include a paragraph or two on policy implications of their work at submission. Doing so would, in turn, incentivize researchers to critically reflect on their work and draw out contributions to policy discussions. This would be especially appropriate for conferences and journals that target the intersection of human-centered computing and policy disciplines.

Funders can ask about the consideration of policy in the process of selecting research.

Instead of solely relying on the gap in academic literature, as is traditionally the case for proposing new academic research, funders could emphasize and legitimize the proposal of a research study that is based on its urgency and prominence in the policy landscape. By doing so, researchers can opt into proposing research that, from the onset, is designed to inform policy debates. This approach would mitigate the tendency to identify a genuine gap in academic literature, only to artificially force “policy relevance” later in the process — it would not only encourage researchers to ground their proposals in current policy debates,

but also diversify the criteria by which research is qualified for funding.

Ultimately, avoiding policy-washing is not just about improving scholarly output and aligning expectations; it is an opportunity to encourage researchers to think through how their work and findings can make an impact on policy. When human-centered computing researchers and academics make the effort to directly engage

with policy — whether through clear takeaways in publications, collaborations with civil society groups, or direct contributions to policy processes — they not only make policy more evidence-based, but also more human-centered. This is crucial to shape a future in which the broader public is prioritized in the development of technology, alongside corresponding policies.

References

- [1] NSF. (N.d). Broader Impacts. National Science Foundation.
<https://www.nsf.gov/funding/learn/broader-impacts>
- [2] Woodson, T., & Boutilier, S. From intent to impact—The decline of broader impacts throughout an NSF project life cycle. *Research Evaluation*, 32(2), 348-355, 2023.
- [3] Luria, M., & Grybos, E. Policy Considerations for Socially Interactive AI Agents: A Systematic Literature Review. Available at SSRN 5375608, 2025.
- [4] McKee, M. Bridging the Gap Between Research and Policy and Practice: Comment on" CIHR Health

- System Impact Fellows: Reflections on 'Driving Change' Within the Health System". *International Journal of Health Policy and Management*, 8(9), 557, 2019.
- [5] Scott, T., Crowley, M., Long, E., Balma, B., Pugel, J., Gay, B., ... & Noll, J. Shifting the paradigm of research-to-policy impact: Infrastructure for improving researcher engagement and collective action. *Development and psychopathology*, 36(5), 2324-2337, 2024.
- [6] Luria, M., & Bhatia, A. Teen and Parent Perspectives on Approaches to Age Verification. Center for Democracy & Technology, 2025.

SPECIAL ISSUE ARTICLE

AI Governance and the Fundamentals of Data: The Relational, Political, and Material

by Julian Posada
Yale University

Keywords: Data, AI Governance, Regulation, Artificial Intelligence, Digital Labor

Introduction

Over the past decade, the technology industry has promoted the integration of artificial intelligence (AI) across a wide range of economic sectors. From early efforts focused on machine learning-based applications in predictive governance and healthcare, we now witness the adoption of generative AI in media, education, business, the arts, and various other industries. As a result, the discourse around the social implications of this technology has evolved considerably, sparking critical debates among policymakers, technologists, scholars, grassroots organizers, artists, and other stakeholders about ethical and legal approaches to its advancement.

Today, competing forces aim to influence the direction of AI development. Some stakeholders, particularly large corporations that develop AI systems, advocate for fewer regulatory frameworks and the relaxation of legal restrictions such as copyright laws, which they view as obstacles to business, research, and development. Conversely, others strive to establish robust regulatory systems ensuring that AI development remains accountable and does not cause harm to marginalized communities. Regrettably, some of the stakeholders and areas of action most profoundly affected by AI are overlooked during

the formation of these regulatory frameworks. This trend has persisted since machine learning-based AI began gaining societal and economic prominence over the past decade.

The past few years have seen the emergence of various international, national, and local regulatory structures to address the expansion of AI technology. However, these initiatives often overlook the process by which AI is produced and how the technology comes to be. The focus tends to be on the end users of AI products and the consequences of their deployment, ignoring the issues that arise during the technology's development. One way to explore these issues is by examining the material, political, and relational aspects of data, which are the essential components driving advancements in machine learning. By considering data as a sociotechnical construct through these three lenses, new governance mechanisms can address issues arising from both the development and deployment of AI.

The first theme, the relational nature of data, emphasizes its value through aggregation. This is especially true in the current AI paradigm, where the “bigger is better” approach relies on vast amounts of data. A key concern in this context is the protection of the rights of data subjects, from the individuals whose data are

fed into algorithms, to the workers who curate and annotate this data. The second theme, the political nature of data, looks at how data are constructed as an abstraction of reality, which is a fundamental issue in information science. It also focuses on how decision-making processes, social dynamics, and economic incentives influence data creation. The third theme, the material nature of data, pertains to the infrastructure that supports AI technology. This includes internet systems that facilitate data flow, data centers that store and process information, and the natural resources required to power and maintain these systems. It also raises important questions about the environmental impact of this infrastructure.

This article explores these three themes by examining specific governance frameworks informed by critical academic research on AI infrastructures and the environmental impacts of this technology. It also draws from my own research on the outsourced labor underpinning the AI industry, which is crucial for generating and annotating data for machine learning, as well as for verifying the algorithmic outputs of AI models [1].

For this analysis, I focus on recent cases of AI governance frameworks emerging from Canada, the European Union, the United Kingdom, and the United States. These include legislative measures ranging from proposals to freshly approved bills, executive actions, and examples from research and the media that support these cases and the overarching argument. In each of these cases, I will also provide a discussion on these three topics, tied to my fieldwork in Latin America and research that has emerged in the region. This chapter aims to advocate for the

effectiveness of emerging AI governance frameworks by focusing on these three often overlooked themes, especially at a time when AI is profoundly impacting countless lives worldwide.

Data as Relational: The Human Cost of AI Systems

Data can be collected from user interactions with internet-connected devices, sensors, and haptics; sourced from works such as literature, film, and other media; and captured from the environment. These data points can then be aggregated to provide meaningful insights. Many information scientists argue that “raw data is an oxymoron” due to the inherently social process of datafication, which involves the collection, curation, and exploitation of data [2, p. 184], [3]. This datafication process, which makes data insightful through aggregation, involves multiple actors.

Data work is the best example of the relational process of data production and the power dynamics that shape it. As part of data procurement in the AI industry, companies and research institutions often outsource the labor-intensive tasks of data generation, annotation, and model evaluation to external firms [4]. These firms employ workers either on-site, in facilities like call centers, or through digital labor platforms that resemble gig economy models such as Uber and DoorDash [5]. These workers contribute to data generation through tasks like data entry, image capture, and audio recording. Annotation tasks, which provide meaning to data, include transcribing or translating text and labeling objects in images to train self-driving cars. Verification tasks,

usually associated with reinforcement learning with human feedback, require workers to evaluate model outputs and sometimes perform adversarial actions, known as red teaming [6].

Data work has raised considerable concerns in recent decades due to its outsourced nature, which aims to reduce AI production costs but has had a significant negative impact on working conditions. Research has revealed significant power differentials between workers and managers, along with issues such as unfair contracts, low wages, poor working conditions, inadequate compensation, and lack of worker representation [7], [8]. There are also documented instances of psychological trauma, sexual assault, and child labor exploitation among these workers [9], [10]. Their contributions are often erased and dismissed by some players in the AI industry.

When examining these issues, it becomes clear that they should be addressed through AI regulation. For AI to truly benefit humanity, it cannot be developed using methods that do not respect human dignity in the workplace. To address this issue, regulatory initiatives like the failed Consumer Privacy Protection Act section of Canada's Bill C-27 focused on data gathered through internet-based services, although it did not consider data sources beyond individual user input. While such initiatives are important for data protection and privacy, they primarily focus on individuals rather than on groups, such as workers. However, data are networked, distributed, and deeply embedded in social relations, making it crucial to recognize these dynamics.

Data can also transcend national boundaries. For example, data collected in France by a US-based company might be processed by workers in Madagascar, with several intermediary steps [11]. This complexity makes it more difficult to alert broader audiences, including regulators and policymakers, to the issues faced by data workers. Furthermore, the challenges faced by these workers do not necessarily align with the interests of the citizens whom many of these laws are intended to protect. As a result, they are often excluded from national and regional regulatory processes designed to safeguard citizens, particularly those of wealthier nations.

The Artificial Intelligence and Data Act, also part of the failed Bill C-27 in Canada, had the potential to address some of these issues, particularly through its definition of harm as a key factor in evaluating the beneficence of AI systems. However, the proposed definitions of harm, which included the economic, physical, and mental impacts of AI, would only have been effective if applied across the entire AI production process rather than at the point of deployment only. If future legislation continues to focus solely on the harmful effects of AI during deployment, it will fail to address the relational nature of data production and the exploitation that accompanies it, including the use of copyrighted or non-consensual material.

Many of these issues are evident in my fieldwork, which focuses on data work in Latin America. At the beginning of the pandemic, numerous workers, notably in Venezuela, joined digital platforms to generate, annotate, and verify data. Many of these workers were initially attracted by bonuses provided by the platforms.

However, these bonuses were later removed, a tactic widely recognized in platform theory as a means to create network effects, or sufficient demand and supply in the market [12].

This situation led to many workers leaving the platforms, although some, dependent on the platform for various reasons, remained. These individuals included single mothers nursing their children and immunocompromised individuals who had no alternative ways to earn income in their local labor markets. Additionally, similar to previously documented cases of mental harm in content moderation, several workers reported mental health issues related to the content they encountered while working for the platforms.

These examples show how governance that genuinely addresses the harmful effects of AI should recognize that its impact begins during development and continues throughout production, including the conditions under which data are procured, the associated ethical considerations, and the labor conditions of data workers. Furthermore, regulations should acknowledge the transnational nature of data, which poses challenges but does not necessarily limit national and local initiatives from influencing it. Such regulations can help prevent the deployment of unethical technologies within their jurisdictions until both development- and deployment-related harms are addressed.

Data as Political: The Problem of Bias and Neutrality

Bowker [2] critiqued the perception of data as a raw material, emphasizing the importance of the circumstances surrounding its capture and

handling, which he refers to as the “cooking” process. This includes decisions related to the collection of data points, their classification, cleaning, methods of analysis, storage, and many other factors involved in datafication as a decision-making process. In this context, the political nature of data, as the outcome of such processes, should not be overlooked in favor of treating data as neutral, objective, or equivalent to observed reality.

This understanding is important in the context of governance. One example is the European Union (EU) AI Act, which is arguably the most significant regulation on AI due to the EU’s market size and the comprehensive scope of the legislation. It covers AI across many sectors, excluding only national security and military applications. The EU AI Act categorizes AI systems based on risk, ranging from unacceptable to high and minimal risk. Unlike other AI regulations, the EU AI Act does not conform to the usual consequentialist approach, which focuses only on deployment. Instead, it includes auditing and quality assessments of datasets as tools to help ensure less harmful AI systems. Research in AI ethics and governance has yielded cautionary but promising perspectives, viewing audits as opportunities for transparency [13], [14] and documentation as a way of making this transparency visible from the outset [15], [16].

An important consideration is not to misjudge the politics of data and AI, which is a data-derived technology. Measuring, capturing, or collecting data is an abstract process. Datafication, involving decision-making processes, is never objective or neutral. Why collect one data point and not another? Why

classify it in one way and not another? Where are the lines and boundaries drawn? Considering the risks of AI means acknowledging and mitigating the biases generated by datafication, but it does not mean viewing bias as something that can be eliminated. This would mean regarding data as an objective measure which can exist devoid of context and action. Instead of striving to generate so-called neutral datasets, it is crucial to emphasize the importance of making the context and perspectives that shape datafication more visible. Moreover, mitigating potential harms should not rely exclusively on technical solutions; it should also involve exploring ways to address those harms. This includes gaining a deeper understanding of the perspectives that shape how data are encoded, the worldviews implicated, and considering who is included, marginalized, and erased.

When considering data as always “cooked” and never “raw,” and incorporating a reflection on the inherently political nature of data-driven AI, governance mechanisms that ensure the mitigation of harm can focus on the broader picture. My colleagues and I encouraged the computer science community to examine power dynamics as a key focus in mitigating social issues related to AI. We advocated a deeper understanding of historical inequities, labor conditions, and epistemological standpoints inscribed in data, rather than focusing solely on correcting biases that deviate from perceived ground truths [5]. We concluded by suggesting that one should not be content with technical fixes when unbiasing data, but should “interrogate the set of power relations that inscribe specific forms of knowledge in machine

learning datasets” to understand the inherently political nature of data.

Our research in Latin America, where we studied hundreds of annotation instructions written by AI developers, primarily in the United States, and distributed to data workers in the region, informed our findings [1]. The meaning-making process was evident from this early stage in the development of machine learning models. We observed that the industry commonly outsourced three types of tasks: data generation, data annotation, and model evaluation. Each of these tasks carried specific meanings that workers were required to reproduce. This included capturing specific types of data, such as voice and images, using taxonomies to label the data in particular ways, and providing feedback on specific streams of data while ignoring others.

Noteworthy examples included labeling individuals according to predefined social characteristics or conducting content moderation on speech based on legal definitions, in both cases using taxonomies prominent in the United States but not elsewhere. This suggested how difficult it is to scale up systems that are coded under specific cultural contexts but intended for a wider global audience [17]. These examples illustrate that the development of technology based on data is inherently a political process, deeply intertwined with societal considerations and contexts.

A commendable initiative of the EU AI Act is its prohibition of AI use in public biometric identification, social scoring, and other cases where AI models enable the manipulation of

human behavior and agency. However, AI is fundamentally a tool that sorts and classifies the world through data. Birhane (2021) argues that the fluid and complex characteristics of human nature and society cannot be easily categorized without generating political friction. She cautions against framing AI-based automation as neutral and objective when it is, in reality, shaped by political decision-making processes.

The biggest pitfall of ignoring the political nature of data when addressing social issues related to AI is the assumption that technical solutions can resolve these underlying issues and that such solutions can be scaled up and implemented across different contexts. Instead, AI governance should recognize that the entire datafication process is inherently political and find ways to reduce harm, ensure social justice, and benefit humanity. This requires focusing not only on the technicalities of the process but also on its social implications. Specifically, it means examining who benefits from these processes and at what cost.

Data as Material: Infrastructures and Environmental Impact

The prevailing paradigm of AI can be summarized as “bigger is better.” For instance, OpenAI’s flagship model, the Generative Pre-trained Transformer (GPT), featured 1.76 trillion parameters, which are numerical values that determine how a large language model processes information and generates outputs. While the company has not disclosed details for its most recent iteration at the time of writing, the forthcoming GPT-4.5, expected in 2025, is

advertised as the “largest and best model for chat yet” [18].

Like other digital technologies, large language models and other types of AI rely on physical infrastructure for data capture, processing, and transmission. Data centers are central to these data flows and have been criticized in academic circles for their significant consumption of electricity and water [19], [20]. Recently, the International Energy Agency (IEA, 2025) reported that data centers consumed 415 terawatt-hours (TWh) of electricity in 2024, constituting approximately 1.5% of global electricity consumption. The IEA estimates that this figure will more than double to 945 TWh by 2030, roughly equivalent to Japan’s current annual electricity consumption.

The material aspects of AI are closely related to hardware, especially with the recent push to onshore microchip manufacturing to the United States. Recently, the Taiwan Semiconductor Manufacturing Company (TSMC) announced plans to invest an estimated USD 100 billion into developing its manufacturing sites in Phoenix, Arizona, and Camas, Washington, in addition to its corporate facilities in Austin, Texas [21]. Like data centers, these facilities require significant amounts of water and electricity for their operation. This raises concerns, particularly in water-scarce regions like Arizona, where other data centers are also located [22].

These pressing concerns regarding the environmental costs of AI are currently being overlooked by the federal administration; however, this was not always the case. The now-rescinded Executive Order 14110 on the

Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, promulgated by the previous Democrat-led administration, was a rare attempt to address AI's environmental impact. This directive was issued on October 20, 2023, and aimed to tackle issues related to the implementation of generative AI, including discrimination, misinformation, and "existential risks."

Although much of the executive order focused on national security and the potential misuse of AI by foreign actors, it notably included provisions addressing the environmental impact of AI. Specifically, it promoted the use of AI to "streamline permitting and environmental reviews while improving environmental and social outcomes" and to "mitigate climate change risks." However, the order fell short of addressing the environmental impacts of AI development itself. While AI's potential applications in mitigating climate change are commendable, governance mechanisms like this executive order should not overlook the ways in which the development of AI exacerbates environmental issues.

Comprehensive AI governance should therefore address not only how the outputs of the technology can aid current regulatory efforts, but also how the resources it consumes and infrastructures it relies upon can be transformed to mitigate their environmental impact. Resources consumed by data centers and by the manufacturing of microchips and other components, as well as the location of these physical infrastructures and their impact on local communities, should be scrutinized further by comprehensive legislation and executive actions

on AI at both the national and international levels.

A prominent example from Latin America is the emergence of data centers in places like Querétaro, Mexico. While these centers promise employment creation and improved revenues for the communities, they also face protests due to the diversion of water from the local population, leading to potential depletion and restricted access to this vital resource [23]. Another example involves the Likan Antay indigenous community in the Atacama Desert, where ancestral land has been used for the extraction of lithium, a mineral crucial for the development of batteries [24].

These instances demonstrate that regions such as Latin America are not only users of technology but also points of origin for its development. They contribute not only labor, as previously mentioned, but also minerals for device development, and land and natural resources like water for the maintenance and operation of data processing centers. In all these cases, local populations play a crucial role and are significantly impacted by the extraction and use of natural resources and land.

Therefore, governance mechanisms should consider not only the use of resources and the harms caused by waste and pollutants, but also the social aspects involved. Questions remain regarding how these infrastructures are physically implemented in communities. Although their establishment offers prospects for increased job access and local income through taxation and other revenue sources, it raises concerns about the displacement of locals, their access to decent living standards,

and the potential environmental changes that could impact their livelihoods. These issues add to the complexities of the material nature of data and illustrate how AI development can affect many facets of life—even before the technology is ready to be deployed.

AI Governance that Prioritizes the Human Well-Being and Environmental Justice

This article has provided examples of contemporary governance, ranging from national and supranational legislation to executive orders, addressing some of the societal and environmental issues posed by the development and deployment of AI.

The first characteristic of data, its relational aspect through social interactions involved in datafication, was exemplified by the role of data workers. These workers are partly responsible for data generation, dataset curation, and the evaluation and feedback provided to AI agents. Their labor is facilitated by the internet infrastructure, which enables this type of work and underpins AI development.

Besides concerns surrounding dignified working conditions in the sector, data work exemplifies the political nature of data. Evidence suggests that although workers are often seen as potential sources of bias due to the possibility of transmitting their errors and opinions when working on data, the reduction of their agency and increased surveillance often result in data that reflect the political positions of their employers [1]. Therefore, the question is not whether data can be biased, but what decision-making processes led to their

production and what meanings are embedded within them.

The infrastructures that enable data work and underpin AI are also material. They occupy space and affect the environment from which materials and resources are extracted and on which their operation depends. Data centers, cables, mines, factories, and other physical sites impact the surrounding communities. While they have the potential to bring jobs and income to those areas, there are concerns about displacement and the degradation of living conditions for those who call those territories home. Thus, the material aspects of data are grounded in geographies and societies.

Addressing these three dimensions of data requires a paradigm shift in AI governance. Data originates from interconnected individuals and territories, making it inherently relational. It is never raw but always cooked through long processes involving many actors, making it inherently political and shaped by various perspectives and interests. AI is not ethereal or dependent on some abstract cloud but grounded in infrastructure and reliant on resources and land, making it material.

Only by recognizing and addressing these fundamentals can AI governance overcome the problems disproportionately impacting marginalized communities. These include people targeted by faulty recognition technologies, victims of social scoring, data workers whose dignified working conditions are being compromised, communities in areas of resource extraction and exploitation, and countless others whose work is incorporated into models but whose communities are erased

under a facade of neutrality. AI is not merely about numbers, calculations, and technicalities; it is about people. Its impacts transcend the boundaries of advanced economies and increasingly affect the global majority.

Embracing this multidimensional nature of AI and data is essential to ensure responsible and equitable AI governance for all.

The emerging governance landscape for AI represents a crucial advancement beyond the solely ethical debates that have proliferated in academic circles over the past decade. These debates are often criticized for their lack of practical implementation and incorporation of marginalized voices. This is particularly true for communities in the global majority affected by the technology's development—its subjects rather than its users. As we move forward, it is imperative that legislation and governance mechanisms prioritize the human and environmental dimensions of AI, rather than focusing solely on its technical aspects. They should address issues related to AI not only at the development stage but throughout the production process.

The future of AI will not be determined by regulatory frameworks alone, but by its tangible impacts on labor, environmental sustainability, and political equity. Ultimately, AI will be judged by its real-world consequences on those who work within its sphere, endure its environmental footprint, and navigate its political implications.

Acknowledgements

A shorter essay version of this chapter was published by the Just Tech Platform under the title “On a More Comprehensive Governance of Artificial Intelligence.” The author gratefully

acknowledges the editorial contributions of Rodrigo Ugarte for the original short essay and of S. Menon and A. Neilton for the longer book chapter.

Funding

This research was supported by grants from the International Development Research Centre of Canada and the Social Science Research Council, with support from the McArthur, Ford, and Surdna foundations.

References

- [1] M. Miceli and J. Posada, “The data-production dispositif,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, 2022.
- [2] G. C. Bowker, *Memory practices in the sciences*. In *Inside technology*. Cambridge, Mass: MIT Press, 2005.
- [3] U. A. Mejias and N. Couldry, “Datafication,” *Internet Policy Review*, vol. 8, no. 4, Nov. 2019, doi: 10.14763/2019.4.1428.
- [4] P. Tubaro, A. A. Casilli, and M. Coville, “The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence,” *Big Data & Society*, vol. 7, no. 1, p. 205395172091977, Jan. 2020, doi: 10.1177/2053951720919776.
- [5] M. Miceli, J. Posada, and T. Yang, “Studying up machine learning data,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. GROUP, pp. 1–14, 2022, doi: 10.1145/3492853.
- [6] A. Q. Zhang et al., “The Human Factor in AI Red Teaming: Perspectives from Social and Collaborative Computing,” in *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*, San Jose Costa Rica: ACM, Nov. 2024, pp. 712–715. doi: 10.1145/3678884.3687147.
- [7] J. Muldoon, C. Cant, M. Graham, and F. Ustek Spilda, “The poverty of ethical AI: impact sourcing

- and AI supply chains,” *AI & Soc.*, Dec. 2023, doi: 10.1007/s00146-023-01824-9.
- [8] J. Posada, “Deeply embedded wages: Navigating digital payments in data work,” *Big Data & Society*, vol. 11, no. 2, p. 20539517241242446, June 2024, doi: 10.1177/20539517241242446.
- [9] B. Perrigo, “Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer,” *TIME*. Accessed: Jan. 13, 2024. [Online]. Available: <https://time.com/6247678/openai-chatgpt-kenya-workers/>
- [10] N. Rowe, “Underage Workers Are Training AI,” *Wired*, 2023. Accessed: Apr. 07, 2025. [Online]. Available: <https://www.wired.com/story/artificial-intelligence-data-labeling-children/>
- [11] P. Tubaro and A. A. Casilli, “Human Listeners and Virtual Assistants: Privacy and Labor Arbitrage in the Production of Smart Technologies,” in *Digital Work in the Planetary Market*, M. Graham and F. Ferrari, Eds., The MIT Press, 2022, pp. 175–190. doi: 10.7551/mitpress/13835.003.0014.
- [12] J. Posada, “Embedded reproduction in platform data work,” *Information, Communication & Society*, vol. 25, no. 6, pp. 816–834, 2022, doi: 10.1080/1369118X.2022.2049849.
- [13] A. Birhane, R. Steed, V. Ojewale, B. Vecchione, and I. D. Raji, “AI auditing: The Broken Bus on the Road to AI Accountability,” Jan. 25, 2024, arXiv: arXiv:2401.14462. Accessed: May 07, 2024. [Online]. Available: <http://arxiv.org/abs/2401.14462>
- [14] V. Ojewale, R. Steed, B. Vecchione, A. Birhane, and I. D. Raji, “Towards AI Accountability Infrastructure: Gaps and Opportunities in AI Audit Tooling,” Mar. 14, 2024, arXiv: arXiv:2402.17861. Accessed: June 28, 2024. [Online]. Available: <http://arxiv.org/abs/2402.17861>
- [15] T. Gebru et al., “Datasheets for datasets,” *Commun. ACM*, vol. 64, no. 12, pp. 86–92, Dec. 2021, doi: 10.1145/3458723.
- [16] M. Miceli et al., “Documenting data production processes: A participatory approach for data work,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, 2022.
- [17] A. Hanna and T. M. Park, “Against Scale: Provocations and Resistances to Scale Thinking,” Nov. 20, 2020, arXiv: arXiv:2010.08850. Accessed: Jan. 29, 2024. [Online]. Available: <http://arxiv.org/abs/2010.08850>
- [18] OpenAI, “Introducing GPT-4.5.” Accessed: Apr. 15, 2025. [Online]. Available: <https://openai.com/index/introducing-gpt-4-5/>
- [19] P. Brodie, “Climate extraction and supply chains of data,” *Media, Culture & Society*, vol. 42, no. 7–8, pp. 1095–1114, Oct. 2020, doi: 10.1177/0163443720904601.
- [20] M. Hogan, “Data flows and water woes: The Utah Data Center,” *Big Data & Society*, vol. 2, no. 2, p. 205395171559242, Dec. 2015, doi: 10.1177/2053951715592429.
- [21] TSMC, “TSMC Intends to Expand Its Investment in the United States to US\$165 Billion to Power the Future of AI,” Taiwan Semiconductor Manufacturing Company Limited. Accessed: Apr. 15, 2025. [Online]. Available: <https://pr.tsmc.com/english/news/3210>
- [22] K. Hao, “AI Is Taking Water From the Desert,” *The Atlantic*. Accessed: Apr. 15, 2025. [Online]. Available: <https://www.theatlantic.com/technology/archive/2024/03/ai-water-climate-microsoft/677602/>
- [23] A. Valdivia, “The supply chain capitalism of AI: a call to (re)think algorithmic harms and resistance through environmental lens,” *Information, Communication & Society*, pp. 1–17, Oct. 2024, doi: 10.1080/1369118X.2024.2420021.
- [24] S. Lehedé and A. Valdivia, “Peripheries on the Rise: Eco-Imperialism in the Race for Technology Resourcing,” 2025, SSRN. doi: 10.2139/ssrn.5085112

SPECIAL ISSUE ARTICLE

Community Voice in Policy Action: Leveraging Participatory Research Methods to Address Representational Bias in AI Systems

by Agrima Seth
University of Michigan

Keywords: Large Language Models, Participatory Research, Natural Language Processing, Sensemaking, Algorithmic Bias

Introduction

Large language models (LLMs) are becoming embedded in systems that shape daily life—from content moderation and search results to educational tools and customer service chatbots. Yet as these technologies cross geographical and cultural boundaries, they carry with them a troubling pattern of systematic underrepresentation and erasure of marginalized communities.

Our systematic audit of GPT-4 Turbo across Indian subcultures revealed the depth of this problem. When prompted to generate stories about birth, death, and wedding rituals across four Indian states, the model consistently overrepresented socially dominant groups far beyond their statistical presence. In Uttar Pradesh, for instance, General Castes appeared in 76% of generated stories despite comprising only 20% of the state's population. Muslims, who constitute 14% of India's population and have substantial representation in the country's books, films, and cultural production, appeared in less than 2% of stories about religious practices. Scheduled Castes, Scheduled Tribes, and Other Backward Classes, groups that together comprise 64% of India's population,

approximately 900 million people, remained systematically underrepresented even when we explicitly prompted the model to diversify the outputs to other identities [2].

This shows that bias in generative AI is not merely a data problem; it's an algorithmic one. The Web, while inequitable, does contain substantial content about some of the minoritized communities, and our findings reveal a "winner-take-all" dynamic wherein LLMs amplify even marginal statistical advantages in training data into near-total dominance in outputs.

This amplification of bias suggests that current policy approaches of mandating "diverse training data", one of the most common policy prescriptions, will prove to be fundamentally insufficient without making changes at the algorithm and model level. The policy challenge before us - How do we create regulatory frameworks that ensure AI systems represent the full diversity of human cultures, not just those most prevalent on the English-language web?

This opinion piece argues that the solution lies in community-centered participatory research—moving beyond extractive data collection to genuine partnership with

communities in defining cultural knowledge, creating evaluation benchmarks, and shaping AI policy. Drawing from our work developing DOSA (Dataset of Social Artifacts), I propose concrete policy mechanisms that embed participatory methods into AI development, evaluation, and governance.

Real-World Harms of Representational Bias

The underrepresentation of marginalized cultures in AI systems is not merely a matter of technical accuracy; it produces tangible harms. First, it perpetuates **cultural erasure** by reinforcing the idea that certain communities, practices, and knowledge systems are less important or less real. When an AI writing assistant can fluently describe a Western wedding but struggles with an Indian wedding, it sends a clear message about whose culture counts.

Second, it places a **disproportionate communication burden** on members of non-dominant cultures. Users from underrepresented communities must either accept AI outputs that misalign with their cultural context or invest extra effort to prompt, correct, and guide the model—an additional tax that users from dominant cultures do not pay.

Third, in high-stakes applications like content moderation or educational tools, these biases can lead to **discriminatory outcomes**. Educational tools may provide students from dominant cultures with rich, contextualized information while offering stereotyped or generic content about and to those from minoritized identities.

Finally, as LLMs increasingly mediate information access, biased outputs can create **feedback loops that further marginalize underrepresented perspectives**. If AI-generated content becomes a significant portion of the web (as is already occurring), and if that content disproportionately represents dominant cultures, future models trained on this data will inherit even deeper biases.

A Participatory Approach for AI alignment: DOSA case study

Traditional methods of scraping existing digital sources or having researchers define categories and templates perpetuate the same power dynamics that created representational biases in the first place. They assume that cultural knowledge can be adequately captured by researchers or by existing documentation, when in fact much cultural knowledge is tacit, contextual, and held collectively by communities.

Participatory research methods, rooted in the principle that people affected by technology should have a say in its development and evaluation, offer an alternative. These approaches treat community members not as data sources to be mined, but as collaborators who bring essential expertise. In the context of AI and cultural representation, this means involving communities in:

1. Defining what matters: Letting communities identify which artifacts, practices, and knowledge are significant rather than imposing researcher-defined categories
2. Capturing implicit knowledge: Using methods that surface tacit understanding

and shared concepts rather than only formal documentation

3. Ensuring semantic alignment: Validating that descriptions match community sensibilities rather than outside interpretations
4. Maintaining ongoing input: Creating mechanisms for continuous community participation, not one-time extraction

In our work titled *Dosa: A dataset of social artifacts from different Indian geographical subcultures*, [1] we demonstrate how these principles can be operationalized at scale. We engaged 260 participants from 19 different Indian geographic subcultures through a two-phase participatory process that combined surveys with a gamified knowledge elicitation framework.

Phase 1: Community-Driven Artifact Identification

Rather than defining categories of cultural artifacts in advance (food, clothing, rituals, etc.), we used open-ended surveys to ask participants: "What social artifacts are important to your cultural identity and known to a reasonable number of people who share it?" We deliberately kept instructions broad, providing diverse examples but not constraining responses. This bottom-up approach ensured that the resulting dataset reflected community priorities rather than researcher assumptions.

We recruited participants through multiple channels, such as educational institutions with affirmative action policies, NGOs working with tribal communities, and digital work platforms reaching economically disadvantaged Indians, to

maximize diversity within regions. This yielded 267 artifacts spanning food, handicrafts, dance forms, textiles, rituals, landmarks, and important figures. The free-form nature of responses allowed artifacts to be listed in ways that reflected how community members actually refer to them, rather than how they might appear in formal documentation.

Phase 2: Collective Sensemaking Through Gameplay

For the second phase, we recruited a new set of participants from each of the states to play a modified version of the game *Taboo*.

Participants were paired with someone from their same cultural community. One player (the "clue giver") provided clues to help their partner guess social artifacts, with rules designed to elicit the most distinctive and important information.

This game structure created a natural mechanism for collective validation—if one community member could successfully convey an artifact to another using particular clues, those clues reflected shared understanding. By having multiple participants describe each artifact, we could identify consistent patterns in how communities conceptualize and communicate about their cultural knowledge.

The gameplay revealed rich information absent from typical web documentation, allowing us to capture contextual details that elicit the community-centered knowledge rather than formal definitions found in encyclopedias.

After completing the game, we asked participants about their experience, where one

participant noted: "*The game made us remember all these objects and things that are a part of us, and we do not necessarily think about them as being different from us.*"

Policy Implications and Recommendations

The DOSA methodology demonstrates that community-centered participatory research is feasible at scale and that such methodologies serve not just as a data collection tool, but as an intervention that activates community knowledge. The playful, collaborative format of the game in our study created space for participants to surface tacit knowledge that might not emerge in a standard survey or interview. But to translate these insights into policy impact, we need concrete regulatory and institutional mechanisms. I propose the following key recommendations for different stakeholder groups:

1. *AI companies* must conduct participatory evaluations with diverse cultural communities before deploying models in new regions. This requires genuine collaboration, not checkbox surveys. Companies should be required to: (a) test models against evaluation benchmarks co-created with community members, not researcher-defined templates; (b) publish model performance results disaggregated by cultural community to make visible disparities between dominant and marginalized populations; and (c) repeat evaluations periodically as models are updated.
2. *Companies* should maintain and publish "community participation statements"

analogous to data statements [3], documenting: (a) which communities participated in dataset creation, evaluation, and testing; (b) how participants were recruited and compensated; (c) methods used to ensure representative participation (not just convenient samples); (d) how community feedback shaped—or failed to shape—model development or deployment decisions; and (e) plans for ongoing engagement post-deployment.

3. Our findings on "winner-take-all" dynamics reveal that strategies focused solely on training data diversity are insufficient. *Companies* must (a) test whether models amplify dominant group representation beyond training data proportions; and (b) implement algorithmic interventions (not just data augmentation) when amplification is detected.
4. *Government agencies* developing AI auditing schemes (NIST AI RMF, EU AI Act implementation) should: (a) establish principles and quality criteria for meaningful community engagement; and (b) define participatory methods—particularly with marginalized groups—as mandatory for assessing cultural bias and representational harms.
5. Meaningful participatory research requires resources that Global Majority communities often lack. *Governments* should: (a) establish mechanisms for ethical data stewardship, ensuring communities retain ownership of datasets they create; and (b) create grant programs enabling research organizations to prioritize community-led AI evaluation projects, sustain longitudinal engagement with communities, and

translate findings into policy-actionable recommendations.

6. *Research communities* and policymakers should collaborate to develop and disseminate practical, open-source toolkits enabling genuine collaboration and ethical data stewardship, moving beyond focus groups to genuine co-design.
7. DOSA examined geographic subcultures, but culture is shaped by intersecting dimensions—caste, gender, religion, sexuality, disability, and more. Future participatory research must develop methodologies that capture intersectionality and recognize that no single dataset can capture culture's full complexity.

Challenges and Future Directions

Scaling participatory methods faces real challenges. Time and resources are required—our relatively modest study of 18 regions involved 260 participants and took months to complete. Conducting truly comprehensive participatory evaluations across all cultures where an AI system might deploy would be enormously expensive and time-consuming. Yet this cost must be weighed against the harms of biased systems and the current practice of deploying undertested models globally.

Power dynamics present another challenge. How do we ensure participatory research genuinely empowers communities rather than becoming extractive consultation? This requires:

- Community ownership of data and findings
- Fair compensation for participation

- Ongoing relationships beyond one-time data collection
- Capacity building that benefits communities
- Transparency about how input influences (or doesn't influence) decisions

There's also the question of scale and representativeness. No participatory project can include everyone from a culture, so whose voices are centered? We must be explicit about sampling strategies and acknowledge that even participatory research involves choices about representation.

Finally, participatory methods must evolve as technology evolves. The approaches useful for evaluating text-based LLMs may need adaptation for multimodal models, embodied AI, or future paradigms we haven't yet imagined. Participatory research is not a fixed protocol but an ongoing commitment to centering affected communities.

Conclusion

The systematic underrepresentation of marginalized cultures in AI systems is not inevitable—it is a policy choice. When we accept systems trained solely on web-scraped data, evaluated using researcher-defined benchmarks, and deployed without community input, we choose to perpetuate existing inequalities.

Participatory research offers an alternative path. DOSA demonstrates that community-centered methods can operate at scale, producing datasets that capture cultural knowledge in ways web scraping never could. More

importantly, it shows that communities have essential expertise that technical experts and policymakers cannot access without genuine collaboration.

As AI systems increasingly shape access to information, opportunities, and resources, the stakes of getting this right are immense. The path forward requires resources and genuine commitment to power-sharing. It requires

acknowledging that technical expertise alone is insufficient, that the web is not a neutral source of knowledge, and that communities know things about themselves that cannot be scraped, surveyed, or synthesized by outside observers. Most fundamentally, it requires recognizing that inclusive AI is not just a technical challenge but a matter of democratic participation and justice.

References

- [1] A. Seth, S. Ahuja, K. Bali, and S. Sitaram, "DOSAs: A Dataset of Social Artifacts from Different Indian Geographical Subcultures," ACL Anthology, pp. 5323–5337, May 2024, Accessed: Oct. 15, 2025. [Online]. Available: <https://aclanthology.org/2024.lrec-main.474>
- [2] A. Seth, M. Choudhary, S. Sitaram, K. Toyama, A. Vashistha, and K. Bali, "How Deep Is Representational Bias in LLMs? The Cases of Caste

and Religion," arXiv.org, 2025.

<https://arxiv.org/abs/2508.03712>.

- [3] E. M. Bender and B. Friedman, "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science," Transactions of the Association for Computational Linguistics, vol. 6, pp. 587–604, Dec. 2018, doi: https://doi.org/10.1162/tacl_a_00041.

ACM Computers and Society

Chinasa T. Okolo, Editor-in-Chief

Volume 53 • Number 2 • Winter 2025 • www.sigcas.org

ACM Computers & Society (ISSN 0095-2737) is published by the Association for Computing Machinery (ACM), 1601 Broadway, 10th Floor, New York, NY 10019-7434.

Want to contribute to Computers and Society? The door is open, please walk through! We're all volunteers. Don't have time for a column? Send your ideas, comments, or suggestions to: editors_sigcas@acm.org.

