Global Coalition for Tech Justice

Proposal to the Artificial Intelligence Action Summit

10 and 11 February 2025, France

AI Working Group on AI and Information Integrity
Global Coalition for Tech Justice (GCTJ)
November 2024

Thematic tracks: Trust in AI, Global AI Governance and Public Interest AI

These proposals have been prepared by the AI Working Group of the Global Coalition for Tech Justice.

The Global Coalition for Tech Justice has 250 member organizations and experts across 55 countries, and aims to ensure Big Tech plays its role in protecting democracy and human rights across the world, particularly in the global majority where companies have been negligent in dealing with the impacts of the tech platforms and technologies.

Individual co-signatories are listed, but the proposal will remain open for signature until February 2025.

Digital Action is the convenor and organiser of the Global Coalition for Tech Justice. Since 2019, Digital Action has been mobilising a global network of partners to demand better standards from the governments and corporations responsible for our digital environments.

Table of contents

I. Introduction

From design to deployment, policymaking to accountability, there is a deep global inequity at the heart of Artificial Intelligence (AI), which is increasingly shaping the future of information integrity, democracy, and human rights.[1] This inequity represents a fundamental challenge to global justice and democratic governance in the digital age. As AI systems become more deeply embedded in critical public institutions and infrastructure[2]—from healthcare and education to judicial systems and public services—this power imbalance threatens to exacerbate existing global disparities and create new forms of technological dependence that undermine national autonomy and self-determination. The rapid acceleration of AI development, concentrated in a few global centres of power,[3] risks cementing these inequities into the foundation of our shared digital future.

AI is principally designed in the Global North or China and deployed across regions with minimal consideration for local contexts or consequences. Harms in the Global Majority go systematically unaddressed, whilst there remains low capacity and expertise for rights-respecting AI policymaking in these regions. These harms range from algorithmic bias[4] to the displacement of local decision-making systems. The impacts manifest in multiple ways: AI systems failing to recognise local languages and cultural nuances, automated decision-making systems trained on Western data making inappropriate determinations in Global South contexts, and AI-driven content moderation systems inadvertently suppressing legitimate political discourse.[5] The lack of local expertise and resources for identifying and addressing these harms, coupled with companies' systematic exclusion of Global South knowledge and expertise from AI development and deployment, creates a self-reinforcing cycle of technological dependency and marginalisation.

There is a profound asymmetry in enforcement capacity and jurisdictional reach, so most Global Majority countries cannot effectively shape AI systems deployed in their information spaces even if they did regulate. This power imbalance undermines national sovereignty and democratic governance in the digital realm. Even when countries develop comprehensive AI regulations, they face significant challenges in enforcing these rules against powerful multinational technology companies. The transnational nature of AI systems, combined with the concentration of technical and legal expertise in the Global North, creates a situation where Global Majority nations often

---

[1] United Nations, 'Urgent Action Needed over Artificial Intelligence Risks to Human Rights' (UN News, 17 September 2021) https://news.un.org/en/story/2021/09/1099972 accessed 5 May 2024.

[2] General Purpose Technologies (GPTs) 'are technologies that, throughout history, have changed the entire economy and, therefore, have the potential to implement drastic changes in society with an impact on pre-existing economic and social structures'. André Guidetti, Artificial Intelligence as General Purpose Technology: An Empirical and Applied Analysis of its Perception (Master's Thesis, Università della Valle d'Aosta - Université de la Vallée d'Aoste 2020), p.1 https://univda.unitesi.cineca.it/bitstream/20.500.14084/428/1/ETI_104_Guidetti_André.pdf accessed 7 October 2024.

[3] Anu Bradford, 'The Race to Regulate Artificial Intelligence' (Foreign Affairs, 27 June 2023) https://www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence-sam-altman-anu-bradford accessed 25 October 2024

[4] United Nations, 'Impact of New Technologies on the Promotion and Protection of Human Rights in the Context of Assemblies, Including Peaceful Protests' (2020) https://undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2F44%2F24&Language=E&DeviceType=Desktop&LangRequested=False accessed 8 October 2024.

[5] Frederik Zuiderveen Borgesius, 'Discrimination, Artificial Intelligence, and Algorithmic Decision-Making' (Council of Europe, 2018) https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 accessed 8 October 2024.

must accept whatever AI systems and policies are imposed upon them, regardless of local laws or social norms.

Citizens and civil society are marginalized or excluded in parts of the process -- design, deployment, accountability, and policy-making. They need access as well as sustained support to build capacity and expertise for meaningful inclusion. This exclusion perpetuates a cycle of technological dependency and democratic deficit. Civil society organizations, which traditionally play crucial roles in protecting the public interest and promoting democratic participation, often lack the technical expertise and resources to engage with AI governance effectively. The complexity of AI systems and deliberate opacity in their development and deployment create substantial barriers to meaningful public participation. This systematic exclusion of citizen voices means that AI systems are developed without crucial input from the communities they will most affect.

This is precisely why transparency and explainability must be cornerstone principles in ethical AI development. People have a fundamental right to understand how AI systems affect their lives and to receive clear explanations of automated decisions.[6] Transparency serves two critical functions: it enables the public to understand how AI systems work, and more importantly, it provides the necessary foundation for holding developers and platforms accountable for their technologies' impacts. Without such transparency, meaningful citizen participation and effective oversight remain impossible, further entrenching the power imbalance between AI developers and the communities their systems impact.

One of the consequences of global inequity, as described, is the embedding of racial inequity in the design,[7] deployment,[8] access to accountability, and policymaking. This systemic bias[9] compounds existing social injustices and threatens to entrench historical patterns of discrimination. The lack of diversity[10] in AI development teams, training data, and testing processes leads to systems that not only fail to address existing racial inequities but actively reinforce them. From facial recognition systems that perform poorly on darker skin tones[11] to language models that perpetuate harmful stereotypes, the racial implications of current AI development practices are profound and far-reaching. The absence of effective accountability and risk analysis mechanisms means these biases often go undetected and unaddressed until significant harm has occurred.

---

[6] Gabriela Arriagada Bruneau, Los sesgos del algoritmo: La importancia de diseñar una inteligencia artificial ética e inclusiva [The Biases of the Algorithm: The Importance of Designing an Ethical and Inclusive Artificial Intelligence] (La Pollera, 2024) https://lapollera.cl/libros/sesgos-algoritmo-ia-etica/ accessed 28 October 2024.

[7] 'Every dataset used to train machine learning systems, whether in the context of supervised or unsupervised machine learning, whether seen to be technically biased or not, contains a worldview'. Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021) 139

[8] Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in *Proceedings of Machine Learning Research*, Conference on Fairness, Accountability, and Transparency (2018) 81:1–15 http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf accessed 21 October 2024.

[9] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016) 10

[10] Forum on Information and Democracy, 'AI as a Public Good: Ensuring Democratic Control of AI in the Information Space' (February 2024) https://informationdemocracy.org/wp-content/uploads/2024/03/ID-AI-as-a-Public-Good-Feb-2024.pdf accessed 10 October 2024 p.24

[11] Morgan Meaker, 'This Student Is Taking On "Biased" Exam Software' (WIRED, 5 April 2023) https://www.wired.com/story/student-exam-software-bias-proctorio/ accessed 24 October 2024

The High-level Advisory Body on Artificial Intelligence has laid out crucial principles and recommendations for global AI governance[12], emphasising that these principles cannot be effectively realised without addressing fundamental inequities, particularly between the Global North and South. While the Advisory Body's vision of inclusive, rights-based governance provides an essential foundation, transforming these principles into reality requires specific, actionable mechanisms for redistributing power in the AI ecosystem. Our proposal for an *AI Commons* offers a practical pathway to implement these principles, ensuring that historically marginalized voices can meaningfully participate in shaping AI's development and deployment. This approach directly addresses the Advisory Body's concerns about representation gaps and implementation challenges in current international AI governance initiatives, providing tangible solutions for building genuine global equity in AI governance.

In today's digital world, information integrity is fundamentally about empowering people to take control of their information — what they access, what they consume, how it is presented and how they evaluate it. However, we're facing a critical challenge: a handful of Big Tech companies currently hold unprecedented power over our information ecosystem, determining what billions of people see and how they see it. This concentration of power isn't just a business issue — it directly influences and often undermines democratic discourse.

Information integrity requires three essential elements: transparency in how information is curated and distributed, accountability from those who control these systems, and a rich plurality of trustworthy information sources. While the UN Global Principles[13] for Information Integrity provides an important framework — emphasizing societal trust, healthy incentives, public empowerment, independent media, and research transparency — their Global North-centric perspective requires critical examination. From a Global South[14] perspective, we must address deeper structural challenges: digital sovereignty in the face of platform resistance to local governance, the need to promote journalism as an ethical practice rather than just "independent media," and the reconstruction of common social spaces rather than simply building resilience against external threats. We need to move beyond a world where Big Tech platforms and their algorithms dominate information curation while acknowledging that meaningful public empowerment requires addressing fundamental social, racial, and gender inequalities. Democratic states must take active roles in preventing market concentration and ensuring equitable access to data and research opportunities, particularly for Global South researchers facing systemic barriers. This transformation requires not only technical innovation and democratic oversight[15], but also a commitment to addressing the structural inequalities that shape our information ecosystem.

---

[12] United Nations, Governing AI for Humanity: Final Report (2024) https://un.org/ai-advisory-body accessed 13 November 2024

[13] United Nations, 'Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action' (2023) https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf accessed 23 October 2024.

[14] Nina Santos, 'Five Brazilian Principles for the Integrity of the Information Ecosystem' (Tech Policy Press, 2 November 2023) https://www.techpolicy.press/five-brazilian-principles-for-the-integrity-of-the-information-ecosystem/ accessed 11 November 2024

[15] Democratic oversight of AI systems means ensuring meaningful public participation and accountability through multi-stakeholder governance structures that include diverse voices, particularly from historically marginalised communities and the Global South. This involves creating transparent mechanisms where civil society, affected communities, human rights defenders, and local experts can monitor, assess, and influence AI systems' development and deployment. The oversight should balance technical expertise with community input, incorporate regular public consultations and independent audits, and maintain clear documentation of decision-making processes. Rather than relying solely on regulatory authority, effective democratic oversight builds legitimacy through rigorous methodology,

Our proposal addresses global inequity and its myriad components, notably geographic, racial and social inequity, through coordinated and systematic intervention at multiple levels. This comprehensive approach recognises that addressing AI inequity requires simultaneous action across technical, social, and political domains. It demands the creation of new institutions and frameworks that can effectively redistribute power in the AI ecosystem while building local capacity for meaningful participation in AI governance. By addressing the interconnected nature of these inequities, we aim to create sustainable change that empowers communities to shape the AI systems that affect their lives.

It's not just about creating new regulations — it's about fundamentally redistributing power in the digital information space and ensuring that technology serves democracy and human rights rather than undermining them. For these purposes, we propose the following actionable deliverables by the AI Global Summit in February 2025, designed to create immediate impact while building long-term capacity for democratic governance of AI systems.

II. AI Commons: Democratising AI Through Global Citizen Power

We propose the launch of an “AI Commons” initiative with four implementation pillars. Imagine a global network where citizens, especially those who haven't had a seat at the table before, get to help shape how AI is developed and utilised. Through four interconnected programmes — training centres across the Global South, citizen councils that review AI designs and policies, laboratories where people can experiment with new AI policies, and a comprehensive oversight system ensuring accountability — putting the power of AI in everyone's hands. Each pillar plays a vital role: the Equity Lab Network builds capacity, the Citizens’ Design Council enables direct community input, the Policy Innovation Lab allows for safe experimentation, and the Multi-stakeholder Oversight System ensures the whole process stays accountable and transparent. It's not just about making AI fairer; it's about making sure it works for all of us, not just a select few. This isn't some distant vision for the future — it can happen now to give people real tools to ensure AI helps rather than hinders our democratic societies.

A. AI Equity Lab Network

Think of the **AI Equity Lab Network** as a global school for the future shapers of AI policy. Still, with a twist — it's specifically designed to empower voices from communities that haven’t traditionally had a say in how technology develops. Through physical hubs spread across the Global Majority and a robust online platform, it's creating spaces where people can get hands-on experience with AI systems while learning how to guide their development in the right direction.

What makes this network special is its comprehensive approach. It's not just about technical training — participants go through a year-long fellowship program where they learn everything from auditing AI systems for fairness to crafting policies that protect their communities' interests. By 2026, the network aims to have trained 1,000 new AI policy leaders from across Africa, Asia,

transparent processes, and genuine redistribution of power in AI governance while avoiding mission creep into censorship or judicial functions. Such oversight focuses on AI systems that significantly impact information flows, human rights, and public services, ensuring they remain accountable to affected communities while protecting information integrity and human rights.

the Middle East and Latin America who understand AI's technical and social aspects, creating a powerful force for positive change in the global AI landscape.

Funding And Partnership Model

The financial sustainability of the AI Equity Lab Network rests on a carefully crafted multi-stakeholder funding model. Rather than relying on a single funding source, we've designed a balanced approach that distributes resources and responsibility across different sectors. Government participation from host countries provides crucial infrastructure support and legitimacy. Their investment demonstrates a commitment to developing local AI policy expertise and ensures the program aligns with national development goals.

Corporate partnerships bring more than just financial support. Leading tech companies provide essential technical resources, mentorship opportunities, and real-world case studies. However, we've structured these partnerships to maintain the network's independence and ability to evaluate AI technologies and their impacts critically.

International institutions are vital in ensuring the program's global relevance and sustainability. Their involvement helps maintain high standards and facilitates knowledge sharing across regions. The funding model includes mechanisms for long-term sustainability, including an endowment fund and revenue-generating activities that support the network's growth while maintaining its core mission.

Curriculum & Training Structure

The AI Equity Lab Network's curriculum represents an approach to bridging the gap between technical expertise and policy understanding in AI governance. At its core, the program acknowledges that effective AI policy leaders need a comprehensive understanding that spans both technical and social dimensions. To achieve this, we've developed a sophisticated dual-track system that adapts to participants' backgrounds while ensuring everyone gets a multidisciplinary skill set.

The "AI Systems & Policy Design" track for policy and legal professionals begins by demystifying AI technology. Participants start with hands-on experience in basic programming and machine learning concepts, moving beyond theoretical understanding to practical application. Through interactive labs and real-world projects, they learn to evaluate AI systems critically, understand their limitations, and assess their societal impacts. By the program's end, these participants can effectively communicate with technical teams and make informed policy decisions based on genuine technical understanding.

Technical professionals entering the program follow the "Policy, Human Rights & Ethics Integration" track, transforming their technical expertise into policy-relevant knowledge. This track emphasises the regulatory landscape, international digital rights frameworks, and the nuanced ways AI impacts different communities. Participants learn to translate their technical expertise into policy recommendations while considering diverse cultural contexts and societal needs.

Both tracks converge in a common core curriculum that develops crucial leadership and advocacy skills. This shared experience creates a powerful network of professionals who can bridge the traditional divide between technical and policy domains.

Selection And Regional Distribution

The selection process for the AI Equity Lab Network is designed to build a diverse and impactful community of future AI policy leaders. Understanding that different regions face unique challenges and opportunities in AI development, we've established a balanced quota system that ensures representation across the Global Majority. This isn't just about numbers – it's about creating a rich dialogue between different perspectives and experiences.

Our selection criteria go beyond traditional metrics. While professional experience is important, we particularly value candidates who demonstrate a deep understanding of their regional contexts and show potential for catalysing change in their communities. We look for individuals who can bridge global AI developments and local needs, considering factors like their engagement with community initiatives and their ability to navigate complex stakeholder relationships.

The physical hub structure is crucial to our vision. Each hub – whether in Nairobi, Jakarta, or São Paulo – serves as a center of excellence for its region, adapted to local contexts while maintaining global standards. These hubs aren't just training centers; they're incubators for regional AI policy innovation designed to foster collaboration between participants and local stakeholders.

Implementation and Governance Framework

The implementation strategy reflects our commitment to building a lasting institution that can adapt and grow. Our governance structure combines global oversight with regional autonomy, ensuring programs remain relevant to local contexts while maintaining international standards. The International Advisory Board plays a crucial role in strategic direction, bringing diverse perspectives from technical experts, policy specialists, and civil society leaders.

Quality assurance is built into every aspect of the program. Regular curriculum reviews, participant feedback mechanisms, and external audits ensure the network continues to meet the evolving needs of AI policy development. Impact assessment goes beyond traditional metrics to evaluate how policy fellows influence AI policy in their regions and create positive change in their communities.

The timeline for implementation is ambitious but realistic. Starting with pilot regions allows us to refine our approach before scaling. By 2026, our goal of training 1,000 AI policy leaders isn't just about numbers – it's about creating a critical mass of expertise in regions that have historically been underrepresented in global AI governance discussions.

B. Citizens' Design Council (CDC)

The Citizens' Design Council is revolutionising how AI gets designed by bringing everyday people into the development process. With major regional hubs located in Africa, Latin America, Asia and the Middle East, it's creating a structure where communities — especially those often overlooked in tech development — have a real say in how AI systems are built and deployed in their regions.

This isn't just about feedback after the fact — the CDC involves communities at every stage of AI development. Before any system is built, they assess its cultural impact and community needs. During development, they test prototypes and monitor impacts. And after deployment, they're continuously checking how these systems affect real people's lives. It's about ensuring AI works for everyone, not just the tech-savvy few.

C. AI Policy Innovation Lab (APIL)

Imagine a space where diverse voices — from policymakers and civil society activists to affected communities, human rights defenders, academia, and private entities — can collaboratively see and experience how AI policies affect fundamental human rights and people's daily lives before implementation. That's what the AI Policy Innovation Lab offers. Its cutting-edge visualisation tools and simulation spaces bring together technologists, human rights experts, community leaders, business representatives and policymakers to transform abstract policy ideas into tangible scenarios that demonstrate real-world impacts on privacy, freedom of expression, non-discrimination, and other essential human rights.

The lab combines high-tech tools with human rights-centered policymaking through an innovative multi-stakeholder approach. In its collaborative workstations and virtual reality policy rooms, Indigenous leaders work alongside digital rights advocates, grassroots organisations partner with government officials, and academic experts join forces with youth representatives to test how different AI policies might impact vulnerable communities and fundamental freedoms. Through immersive simulations, this diverse group can experience first-hand how decisions might impact access to information, privacy protection, or amplify discrimination. Special attention is paid to intersectional impacts, with affected communities leading the conversation about how policies might differently impact people based on their gender, ethnicity, economic status, or geographical location. It’s like having a policy playground with a conscience and collective wisdom — where ideas can be safely tested and refined through multiple perspectives to ensure they protect and enhance human rights before affecting millions of lives. The lab’s participatory approach ensures that policies aren’t just created for communities but with communities, helping prevent unintended consequences that might compromise human dignity or exacerbate existing inequalities.

D. Multi-stakeholder Oversight System (MOS)

The current AI landscape presents a critical paradox. While companies like OpenAI, Meta, and regional tech giants rapidly deploy AI systems globally, oversight mechanisms remain fragmented and often ineffective. We've witnessed this in stark contrasts: China's strict control over ChatGPT access versus the US's largely self-regulatory approach, or Europe's rights-centric regulations compared to developing regions' emerging frameworks (constrained by the equity challenges described above). These disparities highlight the urgent need for a balanced, adaptable oversight system to bridge these approaches while prioritising human rights and information integrity.

The composition of these bodies is carefully balanced. Technical experts work alongside human rights defenders, legal specialists collaborate with journalists, and civil society representatives ensure community voices remain central to all decisions. This diversity isn't just about representation – it's about bringing together the skills needed to understand both the technical implications of AI systems and their real-world impact on communities.

The system acknowledges that AI governance faces different challenges across regions. In Uganda, where the government is reviewing its AI strategy, MOS could provide a framework for meaningful oversight while supporting local innovation. In regions like Southeast Asia, where data localisation and state interests play a significant role, the system could offer flexible mechanisms that respect sovereignty while ensuring human rights protection.

This system’s ability to adapt while maintaining core principles makes it particularly powerful. In regions where self-regulation dominates, it provides structured oversight mechanisms. In areas with strong state control, it offers channels for community input and rights protection. This

adaptability ensures the system remains relevant and effective across different regulatory environments.

Consider a practical example: When a major AI company wants to deploy a new language model in West Africa, the regional oversight body would evaluate not just technical specifications but also cultural implications, data privacy concerns, and potential impacts on local information ecosystems. This assessment isn't a mere checkbox exercise – it's a comprehensive evaluation that can lead to meaningful modifications or even deployment restrictions if necessary.

The MOS ensures that AI development and deployment remain transparent and accountable to all communities it affects, with human rights protection and information integrity at its core. It creates a comprehensive framework where regional oversight bodies, comprised of local representatives, human rights defenders, journalists, fact-checkers, independent media organisations, civil society members, and affected communities, have real power to monitor, assess, and influence how AI systems impact both human rights and information integrity in their regions.

The MOS will oversee both deployed and pre-deployment AI systems that significantly impact information flows and human rights in society, with a specific focus on four critical areas: (1) Large Language Models and generative AI systems that influence public discourse and information creation, (2) Content recommendation and moderation systems that shape information distribution and access, (3) Automated decision-making systems that affect fundamental rights and public services, and (4) AI-powered surveillance and monitoring technologies. This oversight encompasses comprehensive assessment of these systems' impacts on information integrity - including their role in amplifying or mitigating disinformation, their effects on media pluralism, and their influence on algorithmic bias in information distribution. Simultaneously, it maintains rigorous human rights accountability through mandatory impact assessments, binding grievance mechanisms, and community-led monitoring programs. The system employs a dual-track approach: proactive oversight through pre-deployment assessments and continuous monitoring through post-deployment evaluation, with particular attention to systems deployed by both major technology companies and government entities. The oversight process is anchored in transparent documentation requirements, regular public hearings, and clear enforcement mechanisms, ensuring that AI development and deployment remain accountable to affected communities while protecting both information integrity and human rights.

The MOS transforms oversight through five key approaches aligned with UN principles[16] on information integrity:

1. **Trust Building:** Establishing verification mechanisms for AI-generated content and promoting transparency in algorithmic systems
2. **Incentive Restructuring:** Moving platforms away from engagement-based metrics toward information quality metrics
3. **Public Empowerment:** Supporting digital literacy and creating tools for public oversight of AI systems
4. **Media Protection:** Safeguarding journalistic independence and diversity in the age of AI

[16] United Nations, 'Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action' (2023) https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf accessed 23 October 2024.

5. **Research Access:** Ensuring researchers can meaningfully study AI's impact on information ecosystems

Through its integrated "Rights First" approach, MOS ensures that AI systems respect human rights and contribute to a healthy, reliable, and diverse information environment. The accountability framework includes enforcement mechanisms ranging from public warnings and fines to operating restrictions for serious violations of human rights standards or information integrity principles.

- **A new model for democratic AI oversight**

The MOS represents an approach to AI oversight distinctly positioned between civil society observatories and formal regulatory bodies. Unlike courts that make binding decisions or regulators that enforce rules, MOS serves as a transparent observatory that sheds light on how AI systems affect our information ecosystem and human rights. Its power lies not in enforcement but in its ability to gather evidence, surface patterns, and enable informed public discourse about AI's societal impact. The MOS's observatory role aligns with emerging models of non-regulatory AI oversight, similar to the *voluntary certification mechanism for public interest AI* that have proven effective in other domains. As highlighted in the Forum on Information and Democracy's research[17], such mechanisms can help address information asymmetries between AI providers and the public while creating positive incentives for responsible development. Like successful certification bodies that maintain independence from both industry and government while fostering transparency and accountability, the MOS can serve as a trusted intermediary that enables meaningful stakeholder engagement and public scrutiny. Its position as an independent observatory allows it to document and analyze AI systems' societal impacts while avoiding the risks of regulatory capture or political influence that can affect more formal oversight bodies. This approach enables the MOS to foster trust and legitimacy through rigorous assessment and public documentation rather than through enforcement powers.

- **Democratic Legitimacy through Structure and Process**

At its core, MOS legitimacy stems from its deeply democratic structure. Leadership rotates among representatives from different regions and stakeholder groups, with strict term limits preventing any single perspective from dominating. A transparent appointment process ensures diverse representation, while clear conflict of interest policies and multiple funding sources protect against capture by powerful interests. Regular independent audits of the MOS's own operations ensure it practices the transparency it advocates for in AI systems.

- **Building Reliability Through Rigorous Methodology**

The MOS assessments gain credibility through methodological rigour rather than regulatory authority. Its evaluation frameworks emerge from public consultation and undergo peer review, ensuring they reflect diverse perspectives and current research. Before publishing any findings, multiple independent teams must verify the results, with full documentation of their methods made public. This approach creates reliable insights about AI systems' societal impacts while maintaining clear boundaries between observation and prescription.

[17] Forum on Information and Democracy, A Voluntary Certification Mechanism for Public Interest AI: Exploring the Design and Specifications of Trustworthy Global Institutions to Govern AI (Research Paper, September 2024).

- **Safeguards Against Mission Creep and Power Concentration**

To prevent MOS from becoming a *de facto* censor or parallel judiciary, its charter explicitly prohibits content moderation powers and limits its scope to systemic issues rather than individual cases. Regular external reviews assess adherence to these limitations, while mandatory transparency reports detail its activities and decision-making processes. A robust whistleblower protection system encourages internal accountability, ensuring MOS remains true to its mission.

- **Collaborative Approach to Oversight**

Rather than working in isolation, MOS actively collaborates with existing institutions while respecting their distinct roles. It provides evidence and insights to courts and regulators without attempting to replicate their functions. Its assessments support informed policy-making and public debate without prescribing specific solutions. This collaborative approach enhances democratic oversight of AI systems while preserving the separation of powers essential to democratic governance.

- **Empowering Public Discourse and Understanding**

MOS's ultimate impact comes from its ability to shed light on complex issues for public understanding. Regular public meetings, community feedback sessions, and open data about its operations, help citizens to better understand and engage with questions about AI's role in society. Rather than making decisions for the public, it empowers informed public participation in crucial debates about how AI shapes our information environment and democratic processes.

This carefully balanced approach ensures that MOS enriches democratic oversight of AI systems without overstepping into censorship or judicial territory. Maintaining clear boundaries while providing crucial insights strengthens rather than supplants existing democratic institutions.

Closing remarks

The **AI Commons initiative** represents a vision for democratising AI governance through four interconnected pillars that address both immediate needs and long-term structural changes. By combining capacity building through the AI Equity Lab Network, community input via the Citizens' Design Council, policy experimentation in the AI Policy Innovation Lab, and transparent oversight through the Multi-stakeholder Oversight System, this framework creates multiple pathways for meaningful public participation in shaping AI's development. Importantly, it centers the perspectives of the Global South and addresses fundamental power imbalances in the current AI ecosystem, while building local expertise and decision-making capacity. This comprehensive approach recognises that effective AI governance requires both technical innovation and social transformation — from addressing racial and geographic inequalities to ensuring digital sovereignty and rebuilding shared social spaces. The success of the initiative will depend on a sustained commitment to inclusive governance, transparent processes and a genuine redistribution of power to ensure that AI systems serve, rather than undermine, democratic values and human rights.

Open for signatories until 10th February 2025 – the list of signatories will be published periodically

In order to sign on, organisations and individuals are invited to fill in this form.