

FEMINIST GUIDING PRINCIPLES AND COMMITMENTS FOR GLOBAL AI GOVERNANCE

An Advocacy Paper by the Gender in Digital Coalition

As AI governance discussions move forward, an important question remains:

What guiding principles, indicators and accountability mechanisms are needed to make AI governance commitments meaningful in practice?

AI systems are not neutral. They are designed, trained, deployed, evaluated and governed within social, economic, and political structures that reflect existing inequalities. As a result, they reproduce and amplify discrimination, exclusion, surveillance, exploitative labour practices, oppression, human rights violations and war crimes. Research has demonstrated that the impacts of AI are not experienced equally. Women and girls, gender-diverse people, and other historically marginalised groups are distinctly and disproportionately affected by these harms while remaining underrepresented in the institutions, processes, leadership and industries that shape AI development and governance.

The Global Digital Compact and WSIS+20 have already affirmed gender equality within their digital governance frameworks. The first-ever Global Dialogue on AI Governance is an opportunity for governments and decision-makers to embed that feminist commitment into AI governance from the start.

Principles with consequence: Feminist commitments for AI governance

Principle 1: Safety, dignity and interoperability by design

Commitment: AI systems will be safe, secure, and private by design, equitable in deployment, with local, gender-transformative redress mechanisms fully operational before any high-risk tools are rolled out, and guaranteeing that risk classification processes are participatory, accessible, and grounded in International Human Rights Law. This will also include interoperability requirements that prevent monopolistic control over digital infrastructure, enabling the right to repair and establishing shared, human rights-based standards that are adaptable to diverse local contexts for transparent reporting, evidence preservation, and identity protection that places user safety and dignity at the centre.

Recommendation:

Establish an independent mechanism for:

- recognising, documenting and addressing digital neocolonial, authoritarian and extractive digital practices,
- mandating pre- and post-deployment human rights impact and risk assessments that
- incorporate gendered and diversity considerations across the AI lifecycle, and ensuring corporate and public accountability across the AI supply chain through periodic post-deployment audits.

Principle 2: Precaution before deployment

Commitment: Precautionary principle as the baseline standard will be applied for AI development and deployment with transparency, accountability and human oversight, at all stages of the AI lifecycle and across the entire supply chain, including its environmental, labour, and resource dimensions, in line with international human rights law, international labour standards and international environmental law.

Recommendation:

- Require annual public audits and binding due diligence obligations.
- Assess environmental, labour, and socio-economic impacts across the full AI value chain, from an intersectional perspective.
- Establish clear thresholds for risk before deployment.
- Require ongoing monitoring and review after deployment.

Principle 3: Advance feminist data governance and sovereignty

Commitment: All people, offline and online, have the right to dignity, autonomy and safety. Data governance approaches will uphold gender-transformative methodologies, informed consent, collective rights, meaningful community participation, data transparency and data sovereignty. In this way, cultural, knowledge and linguistic diversity, indigenous knowledge, community-owned and feminist data, as well as public-interest information and data commons resources will be safeguarded from extractive neocolonial practices.

Recommendation:

- Align AI governance with existing international commitments on data justice, sovereignty, and collective rights.
- Build formal institutional links between AI and data governance processes (e.g. the UNCSTD, GDC, the ILO, the HRC).
- Support participatory and community-led data governance models to strengthen collective agency over data.
- Invest in local languages, Indigenous knowledge systems, and locally governed digital public resources.
- Protect public-interest and community-owned data from extractive practices.

Principle 4: No AI above human rights

Commitment: Development, deployment, and use of AI systems will be fully compliant with international human rights law and humanitarian law throughout their life cycle, and no AI system or platform may be developed, deployed, procured, or profited from, where it demonstrably facilitates violations of human dignity, autonomy, or human

Recommendation:

- Operationalise existing UN recommendations moratoria on the development, sale, procurement, deployment, export, and cross-border transfer of AI systems that are incompatible with international human rights or humanitarian law, including systems that facilitate, enable or profit from technology-facilitated gender-based violence, discriminatory surveillance, and other gendered harms and human rights violations.
- Require mandatory public transparency and meaningful penalties for non-compliance.
- Ensure accessible remedies for technology-related harms.
- Strengthen monitoring through existing human rights accountability mechanisms.

Principle 5: Ensure gender representation and diversity in algorithmic design lifecycle

Commitment: AI systems will be designed and trained to ensure equal participation. In order to mitigate algorithmic bias, the design of AI systems must account for and ensure that the perspectives and lived experiences of various stakeholders, including those most impacted by its deployment, are meaningfully represented.

Recommendation:

- Establish representativeness as a scientific standard for algorithmic design and development, defined before deployment and assessed across relevant demographic, social, economic, geographic, linguistic, and disability dimensions.
- Measure performance where systems perform worst—not through aggregate averages that obscure unequal outcomes. Performance should be disaggregated and thresholds met for populations most at risk of exclusion, discrimination, or harm.
- Require a single integrated Human Rights Impact Assessment across the algorithmic lifecycle in which representativeness, gender-disaggregated performance, environmental impacts, labour impacts, and other human rights considerations are assessed.
- Require meaningful representative participation throughout AI algorithmic lifecycle, including in problem definition, data collection, training, model development, testing, deployment, monitoring, and governance.
- Increase the representation of women, LGBTQIA+ groups, and historically excluded communities, particularly from the Global South, in technical, governance, and decision-making roles.

Principle 6: Invest in feminist, inclusive, equitable and social AI ecosystems

Commitment: Public and multilateral financing will be mobilised for AI development that is feminist, inclusive, equitable, and oriented toward public good. This financing will be geared towards feminist, community-based, and Global South-led AI initiatives that are severely underfunded to reduce the existing structural concentration of AI investment in a small number of dominant tech corporations.

Recommendation:

- Establish safe and sustainable international financing mechanisms for public-interest digital infrastructure.
- Support feminist, community-based, and open-source AI initiatives.
- Create long-term funding streams for women-led and community-led AI development.
- Prioritise locally governed AI ecosystems and public digital resources.
- Condition public funding on human rights compliance and accountability.

Principle 7: Ensure feminist AI governance mechanisms

Commitment: Governance frameworks will be grounded in intersectional feminist analysis to ensure that "human-centred AI" accounts for the diverse realities of women, persons with diverse sexual orientation and gender identities, and marginalized groups. Global AI governance processes will commit to intersectionality and expand traditional definitions of expertise beyond data scientists and software engineers to include technology-facilitated gender-based violence survivors, feminist technologists, labour advocates, human rights defenders, advocates in prevention of and response to gender-based violence, indigenous leaders, persons with disabilities, anti-racist advocates, public policy experts, and community organisations. Governance processes will prioritise agency, authorship, and decision-making power over mere representation.

In this regard, AI governance will recognise care, wellbeing, equality and social justice as objectives of AI innovation and evaluation, moving beyond the focus on economic growth and technological advancement and innovation.

Recommendation:

- Ensure meaningful participation, including through travel funding support, of feminist advocates, civil society, Indigenous peoples, persons with disabilities, labour organisations, and historically excluded groups.
- Expand definitions of expertise to include lived experience and community knowledge.
- Publicly track both participation and influence in global, regional and local AI governance outcomes.
- Require governing bodies to explain how recommendations from marginalised and underrepresented groups were incorporated—or why they were not.
- Measure outcomes using intersectional and gender-transformative indicators.

A call to action

Governments have a critical opportunity to truly shape AI governance to advance human rights, gender justice, equality, and sustainable development. The choices made today will determine whether AI reinforces existing systemic harm, patterns of exclusion, and the imbalanced concentration of power or contributes to a more equitable, gender-just, diverse and inclusive future.

The Gender in Digital Coalition calls on governments to place gender equality, diversity, and human rights grounded in International Human Rights and Humanitarian Law, feminist political economy, justice and the guiding principles outlined in this paper, at the centre of AI governance frameworks, institutions, and decision-making processes during the Global Dialogue on AI Governance and future discussions on global digital governance frameworks.

About the Gender in Digital Coalition

The Gender in Digital Coalition (GiDC) is a global feminist coalition that works to promote and ensure intersectionality and gender mainstreaming in global digital governance frameworks. Initiated in 2023 by Alliance for Universal Digital Rights, Association for Progressive Communications, Derechos Digitales, Equality Now, IT for Change, Pollicy, UNFPA, UN Women, World Wide Web Foundation (and now represented by Open Data Collaboratives), and Women at the Table to advocate for inclusion of gender across the international policy landscape, including the Global Digital Compact and Pact for the Future.

THESE PRINCIPLES WERE DEVELOPED BY

