# Towards Substantive Equality in Artificial Intelligence:

## Transformative AI Policy for Gender Equality and Diversity

### November 2024

**GPAI** / THE GLOBAL PARTNERSHIP
ON ARTIFICIAL INTELLIGENCE

# Acknowledgements

## Citation

## Project Supported by

# Table of Contents

# Foreword

AI offers tremendous opportunities for social good, but this wealth of possibility can and must be harnessed equitably. With this report, we seek to provide a path towards substantive equality in AI ecosystems, considering actions to enable the participation of all voices. We aim to transform the paradigms of who can participate in AI, and where and how, democratising its development, deployment, and governance.

The evidence is clear: AI systems are not neutral. They reproduce the world models, cultural values, knowledge, and languages of the contexts in which they are conceived, thereby replicating or amplifying systemic inequalities based on gender, race, ethnicity, abilities, social class, and education, among others. Policies must prevent AI systems from having different impacts on diverse groups and from widening gaps within and between countries. However, there is no single perspective for understanding how these inequalities are reproduced, nor a single approach to address their causes and offer solutions throughout the AI lifecycle. Efforts are needed at all levels and in all areas to articulate responses that drive transformative action.

In this report, we recognise the fundamental causes of global inequality and focus on the current harms AI is causing in the lives of millions of people. We take a systemic, participatory, and socio-technical approach centred on human rights and social justice to advance transformative change. Inequalities in AI ecosystems include biases in data or algorithmic models, but they go beyond them. This is why we nuance and emphasise meaningful inclusion in the development and decision-making processes for AI systems. We also recognise that the benefits of technological development, wealth, power, knowledge, and infrastructure are currently concentrated in the hands of a few, while the global majority bears the costs through resource exploitation in their territories, precarious work, and data extractivism.

Most importantly, we focus on identifying concrete ways to move forward differently. Therefore, alongside valuable initiatives driven by various actors, this report offers a roadmap for decision-makers, with recommendations and examples of promising practices that can be taken as benchmarks to enact transformative change. Additionally, it offers guidance to align the design, development and deployment of AI with the principles of the common good, and the well-being of local and global populations.

In 2023, the GPAI Responsible AI Working Group trusted us to co-lead the project *Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity* – an honour we are grateful for. With this report and other project activities, we hope to contribute to the building of truly inclusive, equitable and just AI ecosystems.

**Paola Ricaurte Quijano,**

Full Professor, Tecnológico de Monterrey and Faculty Associate at the Berkman Klein Center for Internet & Society at Harvard University

**Benjamin Prud'homme,**

Vice President, Policy, Safety and Global Affairs, Leadership Team, Mila

# Executive Summary

The rapid advancement of artificial intelligence (AI) is transforming societies and driving economic growth, holding great potential to improve lives and livelihoods globally. However, it risks exacerbating existing inequalities by mirroring and magnifying societal biases, particularly those affecting historically marginalised groups. Challenges such as discrimination, unfairness, bias and harmful stereotypes persist throughout the AI lifecycle, impacting many aspects of human life. Robust regulatory frameworks are urgently needed to mitigate these disparities, prevent harm and work towards substantive equality and diversity in AI ecosystems.

*Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity* aims to strengthen the capacity of States and other stakeholders to foster inclusive, equitable and just AI ecosystems. It examines promising practices, provides policy insights and offers actionable recommendations to enhance gender equality and diversity in AI and related policy making. The *Policy Guide for Implementing Transformative AI Policy Recommendations* provides additional guidance in implementation.

**Key Recommendations:**

1. **Inclusive Design and Democratic Innovation:** Integrate affirmative action and measures for institutional inclusion, and support inclusive technology design.

2. **Meaningful Participation in AI Governance:** Foster and ensure the active involvement of marginalised groups in AI governance to ensure better AI policy for all.

3. **Transparency and Accountability for Harm Prevention:** Establish ex ante safeguards and mechanisms for accountability among all AI actors to prevent harm and ensure fairness.

4. **Effective Access to Justice:** Implement measures to ensure that marginalised groups have access to legal recourse against AI-driven discrimination and bias.

The insights and recommendations are based on regional and group-specific consultations and a conceptual framework anchored in a human rights–based approach to AI. Through these consultations, the report explores the current state of gender equality and diversity within the AI sector, highlights persistent disparities and outlines pathways and promising practices towards substantive (actual) equality in AI. The framework assumes that inequality is structural and can be addressed and remedied through appropriate, transformative measures.

Transformative AI policy tackles the root causes of inequality to achieve substantive equality in AI and beyond. Achieving substantive equality through transformative change in AI will advance human rights and drive economic and social development. Through transformative AI policies, we can enhance the quality, usability and effectiveness of AI systems, contributing to a more equitable, sustainable and prosperous future for all.

African women discussing about a command on a computer. Photo by Iwaria Inc. on Unsplash.

# Summary of Key Recommendations for Transformative AI Policy

The summary of key recommendations provides a brief overview of concrete measures that policy makers can take to effectively integrate gender equality and diversity principles throughout AI policy frameworks, laws, regulations and practices. The recommendations for transformative AI policy are grouped into the following four categories:

- Inclusive Design and Democratic Innovation
- Meaningful Participation in AI Governance
- Transparency and Accountability for Harm Prevention
- Effective Access to Justice

## Inclusive Design and Democratic Innovation

1. **Involve Marginalised Groups in Technical and Non-Technical Roles Throughout the AI Ecosystem**

   Implement affirmative action across the AI ecosystem to involve women and other historically marginalised groups in technical and non-technical roles throughout the AI ecosystem to increase diversity in perspectives. Allocate resources to identify and remove barriers to diverse representation. This includes ensuring accessible, inclusive education beyond AI ecosystems.

2. **Invest in Capacity Building for Institutional Inclusion**

   Invest in capacity development and awareness raising, within public and private institutions and teams, on the experiences and rights of historically marginalised groups. Ensure regular dialogue with representatives of marginalised groups to understand and eliminate the specific barriers they face.

3. **Permit Processing of Special Categories of Data**

   Permit the processing of special categories of data under certain exceptional circumstances, based on substantial public interest, to achieve equality and non-discrimination. To prevent discriminatory outputs, AI system providers must test for systemic bias and ensure the representation of diverse datasets. This should be done without contravening personal data protection rights.

4. **Fund Transformative Technology Research and Design Approaches in AI Innovation**

   Fund research and provide grants and public recognition to incentivise the application of inclusive and transformative techno-design approaches in AI, such as those anchored in feminist technology design principles. These approaches address the gaps between technical and political fairness. Supporting AI system innovations that align with these principles advances more equitable and just applications, practices and processes.

# Meaningful Participation in AI Governance

**5. Promote Effective Public Engagement and Community Participation**

Employ various public engagement methodologies on national and international levels. Include marginalised voices in national AI governance discussions and amplify the Global Majority in international AI governance forums. Enable participation of representatives of marginalised groups by allocating budgets for participation costs ensuring that information and consultation processes are accessible, free, and comprehensible.

**6. Invest in Capacity Development Among Marginalised Groups**

Fund and support educational programmes, networking structures, and other resources that seek to develop the skills and confidence among marginalised groups to participate meaningfully or to actively lead the processes that serve their needs. Work with marginalised communities and representative organisations of marginalised groups to hold their own awareness sessions and consultations on AI-related issues.

**7. Legislate for Ex Ante Public Participation Rights**

Ground AI decision-making processes in ex ante public participation rights such as those established through the UNECE Aarhus Convention. Applying these principles to AI decision-making processes enables affected parties, as well as civil society organisations and the general public, to contest algorithmic decision-making consequences through public reasoning and deliberation.

**8. Protect Collective Data and AI Rights**

Revise rights frameworks that are impacted by AI systems and processes, such as intellectual property and data rights frameworks to 1) safeguard the data and knowledge sovereignty of Indigenous people and marginalised groups, including linguistic, religious and ethnic minorities; and 2) ensure the right to benefit from scientific progress.



Female delegate in hijab and suit speaking in microphone before tribune stand at political conference.

## Transparency and Accountability for Harm Prevention

**9. Establish the Right to Information in AI Systems and Enhance Algorithmic Transparency**

Establish the right to information in AI. This right should grant individuals the right to access clear, accessible details on when AI is employed, what algorithms are used, what data are used for input, and what criteria are used in decision-making processes. Requiring enhanced algorithmic transparency allows individuals negatively impacted by AI systems to challenge their outcomes. It also encourages technological innovation to confront limitations, such as behavioural opacity, and enhance interpretability and explainability.

**10. Enable and Conduct Obligatory Human Rights Impact Assessments (HRIAs)**

Enable and conduct impact assessments by providing policy guidance on how to conduct them. The assessments should evaluate whether risks of harm are acceptable under fundamental rights law and include clear duties to eliminate or prevent such risks. The assessments must also consider and compare possible non-technological approaches to identify the least intrusive measures to human rights.

**11. Develop Accountability Measures for Public-Sector Algorithmic Systems and Processes**

Develop AI-specific public procurement guidelines to protect human rights and due process, addressing complexities and risks introduced by algorithmic and AI systems and processes. Promote open data initiatives to build open libraries of algorithms used in public-sector systems. Ensure that policy makers undergo capacity-building so they can effectively conduct due diligence in AI procurement.

## Effective Access to Justice

**12. Strengthen Contextual Liability for Non-Discrimination in AI Systems**

Strengthen contextual liability for non-discrimination in AI systems in proportion to other accountability measures such as level of transparency, interpretability, and explainability. Product and fault liability regulations require revision to accurately reflect the complexities of AI systems and data-driven decision-making. Effective accountability in AI development and deployment takes into account specific characteristics such as opacity, explainability, autonomous behaviour, continuous adaptation and limited predictability. Chart a path towards liability in AI to ensure appropriate accountability among public and private providers and deployers.

**13. Empower Equality Bodies to Initiate Action**

Empower equality bodies, including national human rights institutions and other public interest organisations, to take action in the public interest. Allow these bodies to submit complaints to supervisory authorities even without identifiable complainants. Ease the burden of proof and equip these bodies with the legal authority and necessary training to effectively address discrimination and harms caused by AI systems and related processes.

**14. Ease the Burden of Proof for Claimants**

Review and revise evidence rules to ease the burden of proof for claimants (World Commission on the Ethics of Scientific Knowledge and Technology, 2005). Existing product liability rules often require harmed parties to demonstrate the causal link between product faults and specific damages. Consider adjusting these rules to make it easier for claimants to prove their cases and claim compensation.

# Introduction

The rapid advancement of artificial intelligence (AI) is transforming industries and driving economic growth; it holds great potential to improve lives and livelihoods across the world. It also risks exacerbating existing inequalities. AI systems can mirror and magnify societal biases and reproduce societal inequalities, particularly those affecting historically marginalised groups. As AI systems and processes are increasingly employed in critical areas such as criminal justice, education, health care and employment, addressing growing inequalities and a lack of diversity in AI ecosystems and related policy making becomes imperative.

Recognising gender equality and diversity as a priority in AI decision-making processes is central to achieving the Sustainable Development Goals (SDGs). Growing inequalities and a lack of diversity in AI pose a significant risk to achieving the SDGs by overlooking the specific needs, local expertise and cultural contexts of historically marginalised and excluded groups. Overlooking Global Majority perspectives in worldwide policy debates risks further exacerbating existing inequalities on a global scale (UNCTAD, 2021). There is an urgent need for robust regulatory frameworks to mitigate these disparities, to prevent harm and to ensure a beneficial development of AI systems and processes for all (Chauhan and Kshetri, 2022; Dankwa-Mullan et al., 2021; Dignum, 2023; Joyce et al., 2021; UNESCO, 2022; Vinuesa et al., 2020).

*Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity* provides AI ecosystems, and particularly States in their role as duty bearers, with practical policy recommendations and insights from promising practices on how to effectively integrate gender equality and diversity approaches throughout the AI lifecycle and related policy making.

---

## Artificial Intelligence (AI) Systems

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

(OECD, 2024b)

---

The report stems from an initiative of the Responsible AI Working Group in the Global Partnership on AI and engages localised experience and expertise from five regions worldwide. Regional and group-specific consultations included approximately 200 participants representing over 50 countries and a diverse array of communities and identities. Participants included representatives from academia, civil society, industry and government. By prioritising consultations, the project sought to hear and understand a multiplicity of voices, perspectives and experiences in their specific contexts. The insights and recommendations in this report build upon results from the consultations and extensive desk and literature reviews. These results include identified critical challenges to gender equality and diversity in AI ecosystems and processes that illustrate profound inequalities embedded within AI ecosystems.

*"The starting point should be the harms that people experience and will likely experience. This requires listening to those who are affected, as well as to those who have already spent many years identifying and responding to harms. Women, minority groups, marginalised people, in particular, are disproportionately affected by bias in AI. We must make serious efforts to bring them to the table for any discussion on governance."*

**Volker Türk,**
**UN High Commissioner for Human Rights, 2023**

The critical challenges highlight how new forms of individual and community harms and discrimination that arise alongside AI systems and processes disproportionately affect historically marginalised and excluded groups. AI systems can, for instance, disproportionately affect individuals who experience barriers based on sex, gender identity, gender expression and/or sexual orientation, including women (Costanza-Chock, 2018; Donnelly and Stapleton, 2022; Hamidi et al., 2018). AI systems and processes have further facilitated technology-facilitated gender-based violence (TFGBV), such as non-consensual image distribution and online harassment and stalking. The anonymity and scale of AI-driven platforms exacerbate these abuses. Gender feature extraction in biometric tasks also raises privacy and ethical concerns, particularly for LGBTIQI+ communities (Ovalle et al., 2023a). AI systems such as body scanning technologies that assess deviance and risk for example, are often based on cis- and hetero-normative assumptions, leading to discrimination against transgender individuals (Costanza-Chock, 2018).

## Historically marginalised groups

Historically marginalised and excluded groups include but are not limited to women; Indigenous Peoples; racialised people; people with disabilities; people who experience barriers on the basis of sexual orientation, sex, gender identity and/or gender expression and sex characteristics (SOGIESC)[1]; people on the move;[2] and individuals with low income or from low-income backgrounds. There is no consensus regarding the definition or use of the terms "marginalisation" or "historically marginalised groups". Marginalisation generally refers to a form of exclusion, either through economic, political, cultural or social elements, and is "a form of acute and persistent disadvantage rooted in underlying social inequalities" (UNESCO, 2009). The term "historically marginalised" highlights the fact that the marginalisation of certain groups has expanded over several decades or even centuries, and that systems that perpetuate such marginalisation and its effects are intergenerational.

(Adapted from UNESCO, 2009)

Another significant challenge is the lack of diversity and inclusion throughout the AI ecosystem (Cheong et al., 2021; Suresh et al., 2022). Exclusion in both technical and non-technical roles limits diverse perspectives, leading to AI systems that perpetuate knowledge exclusion and reinforce harmful stereotypes and discrimination (e.g. Badaloni and Rodm, 2022, Buolamwini and Gebru, 2018; Hull, 2023; Ricaurte, 2019; UNESCO, 2024a). Biased social benefit decision-making systems may, for example, allocate fewer resources to individuals with disabilities and restrict employment opportunities for marginalised groups.

The intersectionality of the identities of individuals who are part of historically marginalised groups can result in compounded disadvantages, as negative impacts from AI systems intersect with broader social inequalities (Buolamwini and Gebru, 2018; Forster, 2022; Kong, 2022; Ricaurte, 2022; Ricaurte and Zasso, 2023; UNCTAD, 2021; UNHRC, 2020). AI systems have, for instance, mistakenly identified crutches as weapons, exposing people with disabilities to increased risk of harm in situations where they are already at risk (UNHRC, 2021). Similarly, the deployment of surveillance technologies and predictive analytics in migration contexts often targets marginalised communities, exacerbating existing tensions and contributing to human rights violations (Dumbrava, 2021; McGregor and Molnar, 2023). These examples highlight deep-rooted inequalities and critical challenges to gender equality and diversity in AI ecosystems and processes that require concerted policy efforts.

## Algorithmic discrimination

Algorithmic discrimination occurs when AI systems unjustifiably treat or impact people differently based on race, colour, ethnicity, sex, religion, age, disability or other protected classifications (White House, 2022). This complex issue arises from the interplay of technology deployment, social practices and political objectives. It extends beyond data biases and includes all instances where AI perpetuates inequality through social interactions.

(Adapted from the Global Index on AI, 2024)



Woman listening to her colleague's presentation. Photo by Christina @ wocintechchat.com on Unsplash.

In light of these challenges, this report offers a comprehensive approach to address gender inequality and the lack of diversity and inclusion in AI through the transformative AI policy recommendations. The Transformative AI Policy Recommendations in this report build upon and advance the implementation of the OECD AI principles (OECD, 2024a), the UNESCO Recommendation on the Ethics on AI (UNESCO, 2022), the Sustainable Development Goals (SDGs) (United Nations, 2024), the Global Digital Compact (United Nations, 2024b), and other work calling for the alignment of AI systems with human rights frameworks to further gender equality and diversity in AI ecosystems and related policymaking (World Economic Forum, 2024).[3] International governance of AI systems and processes is a key enabler for achieving the SDGs. The Transformative AI Policy Framework advanced here is based on the assumption that inequality is structural and therefore possible to address and remedy through appropriate – transformative – measures.

The Transformative AI policy recommendations towards substantive equality are anchored in a human rights–based approach to AI systems and processes. This approach integrates human rights into the core of AI eco-systems by removing constraints and enhancing capabilities to ensure the enjoyment of human rights within and throughout AI ecosystems and related policy making. The recommendations effectively work towards substantive equality by addressing structural barriers preventing historically marginalised and excluded groups from benefiting from the gains of AI development and actively seek to empower women and other marginalised groups so they can fully enjoy their rights.

## Substantive equality

Substantive equality means de facto equality (equality in fact or actual equality) or equality of results (Fredman, 2016; UN Women, 2019). This includes making sure that everybody has equal opportunities and equal access to opportunities, as well as an enabling environment to achieve equal results. Substantive equality, as advanced in this report, incorporates an understanding of equity,[4] gender equality [5] and diversity that calls for equality of opportunity and of results through measures that accommodate difference, enhance voice and participation, and eliminate discrimination, marginalisation and the unequal distribution of power and resources. Substantive equality is an objective in itself, a transformational agenda and a means for achieving other sustainable development goals, including health, economic growth and education.

(Buckup, 2009; OECD, 2015; Ostry et al., 2018)

Achieving substantive equality through transformative change in AI will not only advance human rights but also drive economic and social development. Incorporating gender equality and diversity principles guided by human rights frameworks is essential for building AI innovation ecosystems that are inclusive and aligned with societal needs, and represents a formal commitment for the OECD and all UNESCO Member States. Through inclusive AI policies, we can enhance the quality, usability and effectiveness of AI systems and processes, contributing to a more equitable, sustainable and prosperous future for all.



Group of LGBTQI+ Activists protesting against the Religious Discrimination Bill in Sydney, Australia.
Photo by Nikolas Gannon on Unsplash.

# Report Outline

## Part I. Consultations, Critical Challenges and Potential Pathways

The report starts by describing the regional and group-specific consultations performed in collaboration with delivery partners to better understand unmet needs, systemic barriers and proposed strategic solutions. This section is followed by an overview of the results from the consultations, identifying critical challenges to and potential pathways for gender equality and diversity in AI ecosystems and related policy making. The consultations serve as a foundation for the content of this report.

## Part II. Conceptual Framework

Following the critical challenges and potential pathways, the report presents the conceptual framework underpinning the analysis of promising practices and development of key recommendations for policy makers.

The conceptual framework is divided into three parts. First, it explains the concept of substantive equality and transformative change as an approach. Second, it adopts a socio-technical approach to AI systems and the AI lifecycle, and presents opportunities for intervention throughout the stages of the AI pipeline. Third, it presents the Transformative AI Policy Framework and its three key dimensions, namely 1) Remedying systemic disadvantage (the right to inclusion); 2) Redressing the democratic deficit (the right to participation); and 3) Reversing misrecognition (the right to dignity).



Man in hoodie standing out in a crowd of people. Photo by Jake Weirick on Unsplash.

## Part III. Promising Practices and Recommendations

The next section showcases promising practices that either include, are led by or work to promote the rights and meaningful participation or inclusion of marginalised groups. This set is illustrative and meant to give policy makers and other interested parties concrete examples of actions that may be replicated or adapted to their contexts.

Finally, the report provides an in-depth overview of key recommendations for transformative AI policy. The recommendations for transformative AI policy are grouped in the following four categories reflecting key dimensions of the framework:

- Inclusive Design and Democratic Innovation
- Meaningful Participation in AI Governance
- Transparency and Accountability for Harm Prevention
- Effective Access to Justice

With these categories, the key recommendations for transformative AI policy take a holistic approach towards substantive equality.

# PART I.
# Consultations, Critical Challenges and Potential Pathways



Three women sitting from different generations in a line against a blue wall, dressed in vibrant traditional attire. Photo by Srimathi Jayaprakash on Unsplash.

# Regional and Group-Specific Consultations

The project engaged localised expertise from five regions worldwide, with approximately 200 participants representing over 50 countries and a diverse array of communities and identities. Participants included representatives from academia, civil society, industry and government who shared their perspectives through regional or group-specific consultations. This section describes these consultations alongside considerations and limitations.

The project team collaborated with regional research and civil society organisations (delivery partners) to conduct regional consultations. The project delivery partners included Derechos Digitales, Research ICT Africa and Data Pop Alliance, as well as a number of distributed Research Associates. Delivery partners mapped key stakeholders and organisations for each region, streamlining the organisation of the regional consultations. Emphasising a global, international, intersectional and multi-stakeholder reach, the collaborators ensured that they worked inclusively with a wide range of participants.
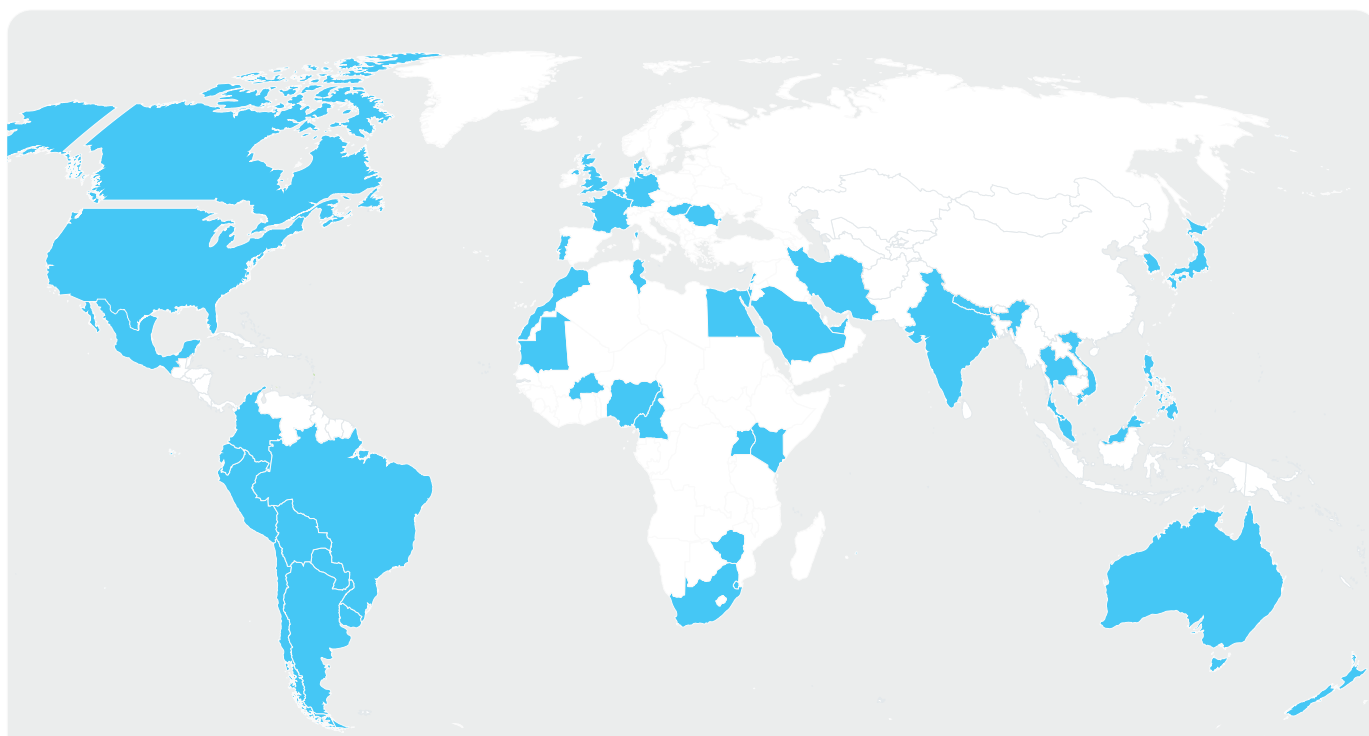


World map representing the geographic regions where the consultation occurred. Visual by Leonardo Studio Design. [6]

Regional consultations took place across five regions:

- Sub-Saharan Africa
- Middle East and North Africa (MENA)
- Latin America
- North America and Europe
- Asia and the Pacific

**50+ COUNTRIES**

**200+ PARTICIPANTS (180 REGIONAL + 20+ GROUP SPECIFIC)**

In addition to regional consultations, the project aimed to channel the perspectives of marginalised groups and persons self-identifying as members of such, including but not limited to women, Indigenous Peoples, racialised people, people with disabilities, people who experience barriers on the basis of various SOGIESC, people on the move, individuals with low income or from low-income backgrounds, and speakers and/or revitalisers of Indigenous and minoritised languages, among others, situated in various regions and contexts representing different key groups.

The consultations took various forms, such as individual interviews, round-table discussions and written contributions, with the overarching objective of hearing a wide spectrum of perspectives and voices. Participants represented a diverse array of communities and identities and included members from different key groups. In addition, a Consultations Expert Group (CEG) and a Project Advisory Group (PAG) contributed input to the project. During the formative stages of the project, the CEG played an expert advisory role for the project team. It included persons with disabilities, those with diverse SOGIESC, Indigenous peoples, racialised groups and representatives of people on the move, from different regions of the world. CEG members contributed key diverse and intersectional perspectives. To validate the project's methodological steps, a PAG was formed for regular consultation. The members of this group included GPAI Experts and External Experts/Specialists. The project team analysed the insights and discussions generated during the interviews thematically as qualitative data to identify emerging themes, convergence and divergence of opinions, consensus among participants, promising practices, lessons learned and knowledge sharing.

## Considerations and Limitations

### Cultural and legal sensitivities

Topics that can be considered controversial, taboo or illegal in specific areas, such as those related to sexual orientation and gender expression, affected the number of participants willing or able to participate in interviews. Local attitudes, levels of acceptance and laws can result in limited acceptance and protection for these individuals.

### Language barriers

The project was conducted primarily in English, and offered participation in Spanish and French. This limited equal opportunities for participation.

### Access and time

Consultations were virtual, requiring access to an Internet-connected device and consequently excluded the participation of individuals who would already be considered digitally marginalised. Technical questions in questionnaires meant limited accessibility depending on the level of prior AI knowledge. Short timelines also represented a limitation for participation and quality of results.

### Unequal representation

Civil society and academic representatives participated to a higher extent than representatives of government or industry. Most participants also had a high level of education and do not provide adequate representation of voices of less formally educated or marginalised groups. Representation of non-binary participants and groups was less prominent than that of other identities.

### Regional and Country Representation Limitations

Some countries and regions were not included in the report due to time, accessibility, and resource constraints. In particular, Caribbean countries were not represented in the consultative process.

# Results from Regional and Group-Specific Consultations

This section outlines the core findings from the in-depth, empirical mapping of opinions, perspectives, experiences and best practices of multiple stakeholders in respect of AI, gender equality and diversity as they interact with systemic barriers and strategic solutions for reaching gender equality and diversity in AI ecosystems. In the synthesis of findings, special care is taken to represent diverse voices and perspectives from multiple constituencies and regions. The following summary does not represent the nuanced views of all participants, but provides an overview of stated experiences of critical challenges and potential pathways to address gender equality and diversity in the full AI lifecycle and related policy making. The results are categorised as follows. The first part of this section outlines and describes the three major challenges to implementing gender equality and diversity in AI ecosystems, as identified by participants. The second part of this section outlines four pathways to address gender equality and diversity challenges in AI. Lastly follows a brief summary of the challenges and pathways.

## Critical Challenges to Gender Equality and Diversity in AI

The regional and group-specific consultations revealed three major challenges to implementing gender equality and diversity in AI ecosystems.

- **Conflation of access with inclusion**

  A recurrent theme in the regional consultations, particularly in Sub-Saharan Africa, was the need to differentiate between mere access to AI products and services and meaningful inclusion in AI ecosystems. Respondents were unequivocal about what comprised meaningful inclusion: systemic transformation in AI design and development that puts historically marginalised groups in the driver's seat.

- **Knowledge exclusion and invisibilisation**

  AI models lack diverse representation and perpetuate knowledge and other types of exclusion, resulting in epistemic injustice. Consultations in Latin America, Sub-Saharan Africa, and Asia, and with Indigenous communities in North America described the incompatibility of dominant modes of AI development with their knowledge ontologies.

- **Unbalanced distribution of resources**

  The distribution of resources in the international AI economic ecosystem is unbalanced. All regional consultations identified the systemic exclusion of historically marginalised groups, particularly in the Global Majority, from development gains in the AI economy as a critical challenge. Labour exploitation and data extractivism continue to perpetuate this inequality and the AI value chain entrenches environmental injustice.



Group of women in Sierra Leone. Photo by Annie Spratt on Unsplash.

## Inclusion

Inclusion refers to ensuring the full and meaningful participation of marginalized groups in all sectors, at all levels, resulting in a positive and measurable impact in their quality of life and full access to their human rights on equal basis with others. Inclusion requires a) measures to empower and support marginalized groups, and b) structural, policy and institutional measures to redress systemic injustice at all levels. An inclusive initiative is necessarily diverse and creates an environment of mutual respect, belonging, trust, support, and engagement.

(DESA-UN, 2009)

**"Inclusion in AI is not just about adding or giving access. It is actually about access that leads to flourishing or access that leads to benefiting. So if your access does not result in benefiting, then that access becomes exclusionary."**

*Angella Ndaka,*
*Centre for African Epistemic Justice, Kenya*

**"[Gender equality and diversity] is largely [reduced to] about recruiting more women. But the environment to which we are bringing them in is not welcoming."**

*Luisa Olaya,*
*GIZ (FAIR Forward), Germany*

**"We should be careful of creating generic gender and diversity solutions because each country or region has its own social differentiations and priorities for inclusion."**

*Anonymous,*
*Asia*

A recurrent theme in the regional consultations, particularly in Sub-Saharan Africa, was the need to differentiate between mere access to AI products and services and meaningful inclusion in AI ecosystems. Focusing solely on ensuring access disincentivises meaningful inclusion and reduces gender transformative programming to isolated initiatives that are not comprehensive, sustainable or transformational. Respondents in the regional consultations were unequivocal about what, on the other hand, comprises meaningful inclusion: systemic transformation in AI governance, design and development that puts historically marginalised groups in the driver's seat. Participants view context-specific measures to enhance the self-determination of marginalised groups in AI ecosystems as essential to addressing social difference and structural injustice, moving beyond deracinated approaches to gender equality and diversity in AI governance.

## Knowledge Exclusion and Invisibilisation

> "A lack of diverse representation can lead to AI development teams ignoring the needs and perspectives of women, people of colour and other marginalised groups."
>
> *Anonymous,*
> *Asia*

> "The most urgent issue is the representation of non-binary, trans and Indigenous people, as developers and as leaders in technology."
>
> *Marcos Cornelio Sánchez Ramírez,*
> *PIT Policy Lab, Mexico*

> "Aligning our technologies with our history and culture is really important and sovereignty and reciprocity are key values for AI."
>
> *Florian Lebret,*
> *Indigenous communities, Canada*

AI models lack diverse representation and perpetuate knowledge and other types of exclusion, resulting in epistemic injustice: the exclusion of knowledge, cultures, languages. The regional consultations in Latin America, Sub-Saharan Africa and Asia, along with discussions with Indigenous communities in North America, emphasised that dominant modes of AI development were incompatible with their knowledge ontologies. Participants highlighted concerns about the reinforcement of intersectional hierarchies of social difference in AI ecosystems. AI systems embed the cultural values and practices of the countries and spaces in which they are conceived and developed. This creates incongruencies when these technologies are applied in diverse global societies and cultures. The datasets underpinning such systems often lack representational diversity, leading to the invisibilisation and lack of representation of historically marginalised groups. This may also lead to the amplification of existing social biases. Additionally, AI developers' biases can impose cultural erasures or violate human rights. Dialogue with Indigenous communities revealed the critical concern of incompatibility between dominant AI development modes and their knowledge ontologies. In the regional consultations in Asia, resource constraints were named as a major obstacle to addressing such epistemic injustices in AI systems.

## Epistemic Injustice

Epistemic injustice refers to the unfair treatment of individuals or groups in their capacity as knowers or contributors to knowledge (Fricker, 2007; Hull, 2023). This concept highlights how biases and prejudices can undermine or disregard the knowledge, experiences, and perspectives of marginalised individuals or groups.

- Hermeneutical injustice happens when there is a gap in collective understanding that prevents marginalised individuals from making sense of their experiences. This occurs because their social experiences are not well understood or acknowledged by society at large, often due to a lack of shared language or concepts.

- Testimonial injustice occurs when a speaker's credibility is unfairly judged based on prejudice. For example, a person from a marginalised group might be disbelieved or ignored simply because of their identity, regardless of the validity of their information.

- Contributive injustice arises as a result of testimonial injustice, when marginalised persons are excluded from participating in the creation and validation of knowledge. Their insights and contributions are not valued or recognised, limiting their impact on decision-making processes and broader societal understanding



Group of people activists with megaphone protesting on streets, strike and demonstration concept. Photo by Halfpoint on iStock.

## Unbalanced Distribution of Resources

> "When we're talking about the AI lifecycle, we're also talking about the extraction of minerals from cobalt mines in Congo. To what extent are we thinking about the gendered implications of AI development in that regard?"
>
> *Kelly Stone, AI Observatory,*
> *South Africa*

> "The nice words about start-ups and economic liberation masks the reality that [...] we are in another form of colonisation [...]."
>
> *Maha Jouini,*
> *African Center for Artificial Intelligence*
> *and Digital Technology, Mauritania*

The unbalanced distribution of resources in the international AI economic ecosystem, and the systemic exclusion of women and other historically marginalised groups, particularly in the Global Majority, from development gains in the AI economy was identified as a major challenge in all regional consultations. Labour exploitation and data extractivism continue to perpetuate this inequality and the AI value chain entrenches environmental injustice.

Pervasive labour exploitation of data subjects (persons) from the Global Majority emerged as a critical concern in the Asia regional consultation. Labour-heavy tasks in AI value chains such data labelling, data and image annotation, transcription, translation and content moderation are disproportionately performed by contract workers hired from and located in the Global Majority. The Latin American consultations also highlighted that some forms of work in the AI industry are valued less than others, and that these labour-heavy tasks are commonly reserved for marginalised persons also within the Global Majority. This tends to render these contributions – and those of women and other marginalised groups more broadly – invisible to the overall AI ecosystem.

Data extractivism was also identified as a concern, where people in the Global Majority, as well as marginalised groups and persons in the minority world, are reduced to being mere providers of raw data and end consumers of AI products and services. The Latin American regional consultation cast a spotlight on how the free and informed consent standard that is commonly adopted in personal data protection legislative frameworks may not be adequate to challenge data extractivism. Such a contractarian approach fails to account for the distortions that arise because of the imbalance of power between AI system providers and data subjects. As transnational digital corporations control critical digital infrastructures underpinning major parts of economic and social life, individuals often find themselves co-opted into unfair terms of service, especially data surveillance–based business models.

The regional consultation in the Middle East and North Africa surfaced concerns about AI value chains perpetuating neo-colonial economic dependencies. The potential for global majority countries to develop AI on their own terms, for their own benefit, according to their own priorities, is severely limited. Participants expressed the view that, in the long term, the current distribution of AI resources may intensify global inequality. Lastly, as highlighted in the Sub-Saharan Africa regional consultations and by Indigenous participants, the AI value chain entrenches environmental injustice through ongoing resource extraction and other activities that disproportionately affect local communities of the global majority.

## Pathways to Address Gender Equality and Diversity Challenges in AI Ecosystems

The regional consultations identified four main pathways to address substantive equality challenges:

- **Develop the capacity of the public sector and the general public**
  Develop and strengthen the public sector's and the general public's overall knowledge and understanding of, and ability to engage with, AI ecosystems and development.

- **Incentivise inclusive design in the private sector**
  Incentivise the private sector with various measures to integrate inclusive design processes throughout the full AI lifecycle.

- **Ensure the accountability of developers, providers and operators**
  Institute legally binding regulatory frameworks to ensure accountability for harms and rights violations by public- and private-sector entities that design, own, deploy and operate AI systems.

- **Ensure meaningful participation by marginalised groups**
  Amplify voices and expand spaces for meaningful participation in AI development and AI-related decision-making processes.



Diverse group of five people sitting at a table and talking. Photo by Christina Morillo on Unsplash.

## Develop the Capacity of the Public Sector and the General Public

> **"AI or algorithmic systems are seen as something that's meant for a small group of people who are experts. There's hardly any cognizance of how it impacts all of us and is embedded in our everyday lives. "**
>
> *Dr Preeti Raghunath,*
> *University of Sheffield, Malaysia*
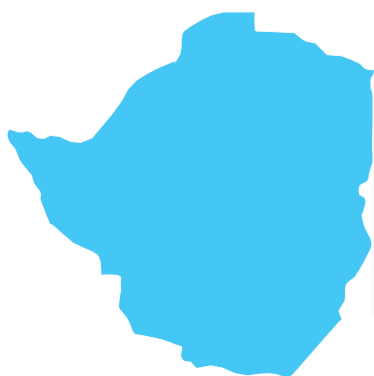
> **"The greatest resource that we can provide policy makers is awareness. Educating them and making them understand what AI all is about, what it can bring to the society."**
>
> *Kudakwashe Dangajena,*
> *African Union, Zimbabwe*

The regional consultations emphasised that strengthening policy makers' and the general public's capacity to participate in decision-making processes related to AI is crucial to address existing inequalities. Capacity building of public officials and civic education will develop everyone's knowledge and understanding of, and ability to engage with, AI ecosystems.

Developing the capacity of the public sector was identified as a major area for ensuring the effective incorporation of measures to ensure substantive equality in AI governance. According to several participants, civic education programmes should focus on promoting public understanding of critical debates in AI ethics and normative principles for human rights–based AI, while actively working to dispel techno-pessimism. This involves creating more inclusive and participatory mechanisms for persons to be involved in shaping AI trajectories.



Woman and man sitting in front of monitor. Photo by Desola Lanre-Ologun on Unsplash.

## Incentivise Inclusive Design and Democratise Innovation

> "AI should be open source. In the communities, we should be able to open the code, see the code, audit the code and see what it is doing."
>
> *Gerardo López Gómez,*
> *Derechos Digitales, Ecuador*

> "You can draw inspiration from a concept, but it's got to be rooted, contextualised and ultimately designed, developed, trained, deployed, learned and refined In the country that is meant to benefit. I don't think there are shortcuts to that."
>
> *Kelly Stone,*
> *AI Observatory, South Africa*

> "Diversity and gender equality integration necessitates the adoption of an intersectional, education-driven, collaboration-focused and policy-supported strategy that places a premium on the creation of ethical, inclusive and equitable AI. "
>
> *Anonymous,*
> *Asia*

The consultations identified a need to incentivise the integration of inclusive design processes throughout the full AI lifecycle. The regional consultations in Asia, Sub-Saharan Africa and Latin America stressed the importance of inclusive technological design and development to achieve effective gender equality and diversity outcomes in the AI ecosystem. This requires incentivising all actors to prioritise ethical, inclusive and equitable AI development. This also means providing incentives for AI developers to customise their models to local needs and conditions, contextualising models to ensure their sensitivity to the unique social, cultural and environmental contexts of their deployment. Linguistic diversity is one aspect of this type of contextualisation, and emerged as a critical dimension in the regional consultation in Asia.

Inclusive design practices can be promoted through public policies that democratise innovation. These policies should provide diverse communities with access to resources, training, and infrastructure, enabling them to develop technologies that are tailored to their needs, on their own terms.

**Ensure the Accountability of Developers, Providers and Operators**

> "I don't think you can have tech companies policing themselves and holding themselves accountable. They need to be held accountable to our democratic institutions."
>
> *Anonymous,*
> *Thailand*

> "In the contracting and procurement of technologies, states need to establish minimum standards of transparency, auditability and social participation."
>
> *José Alfredo Hau Caamal,*
> *Mozilla Foundation e Ação Educativa, Brazil*

Holding developers, providers and operators of AI systems accountable emerged as a strong priority for all participants. The regional consultations revealed a consensus among respondents on the need to move towards legally binding regulatory frameworks for accountability. Public- and private-sector entities that design, own, deploy and operate AI systems must be accountable for harms and rights violations.

External audits and human rights impact assessments were considered important tools for incorporating gender diversity and equality into the AI sector, but they are not a silver bullet. The risks of letting the private sector alone be in charge of them were mentioned, along with lessons learned from the environmental and financial sectors regarding corruption and capture. State and civil society actors should be prepared and have the capacities to develop public interest auditing and assessments.

As respondents in the regional consultation in Latin America emphasised, transparency and auditability of the source code are key to ensuring accountability. The Middle East and North African and Latin American regional consultations highlighted the need to address the increasing risk of a public-private AI surveillance nexus as a crucial priority. Respondents in the Middle East and North African consultation stressed the role of States in setting human rights due diligence requirements and accountability and reparation mechanisms when it comes to the development, use, vetting and procurement of digital technologies.

**Ensure Meaningful Participation by Marginalised Groups**

> "One thing that makes it very difficult is still the prevailing inequality in Latin America as to who can, and who cannot, talk about technology, since to a great extent it is still presumed that the only ones who are allowed to make criticisms about AI are engineers."
>
> *Marcos Cornelio Sánchez Ramírez,*
> *PIT Policy Lab, Mexico*

> "Conversations on AI governance are convened by people who are in power, and therefore the conversations remain in those circles, and then they feed off each other and create opportunities for each other – narrowing the stakes instead of broadening public interest."
>
> *Dr Preeti Raghunath,*
> *University of Sheffield, Malaysia*

The consultations emphasised the need to democratise AI development and decision-making processes through meaningful participation. Respondents supported transitioning from a narrow, expert-led approach to a broad-based, society-wide process rooted in extensive citizen engagement. This shift is essential for addressing potential injustices and biases and the opacity of AI technologies.

Participants from Asia, Sub-Saharan Africa and Latin America highlighted the benefits of co-designing AI systems with end users and promoting interdisciplinary collaborations. Engaging the wider public in these debates is crucial for developing AI systems that are sensitive to real-world contexts and diverse perspectives, particularly focusing on the rights of those most at risk of harm.

Ensuring effective multiparty mechanisms with inclusive and meaningful participation is vital according to respondents all over. All parties, including governments, the private sector, civil society, the technical and academic communities, and users, must be involved in their respective roles in accordance with the principles for global digital co-operation enshrined in the WSIS Tunis Agenda. Currently, civil society and marginalised groups are under-represented. Specific measures, such as allocating funding, preparing budgets and initiating partnerships, are necessary for marginalised groups to engage effectively in policy making. Additionally, the consultation in Asia highlighted the need to prevent elite capture, where dominant groups misuse public funds intended for the broader population for their own interests. The United Nations Guiding Principles on Business and Human Rights (UNHRC, 2011) were mentioned as key to assure that due diligence processes are implemented within all the lifecycle of AI and to guarantee that meaningful participation – including by people with disabilities – and impact evaluations are in place.

# A Brief Summary of Challenges and Pathways

The regional and group-specific consultations revealed three major challenges to implementing gender equality and diversity in AI ecosystems: 1) conflation of access with inclusion; 2) knowledge exclusion and invisibilisation; and 3) unequal distribution of resources.

The regional consultations identified four main pathways to addressing substantive equality challenges: 1) develop the capacity of the public sector and the general public; 2) incentivise inclusive design in the private sector; 3) ensure the accountability of developers, providers and operators; and 4) ensure meaningful participation by marginalised groups.
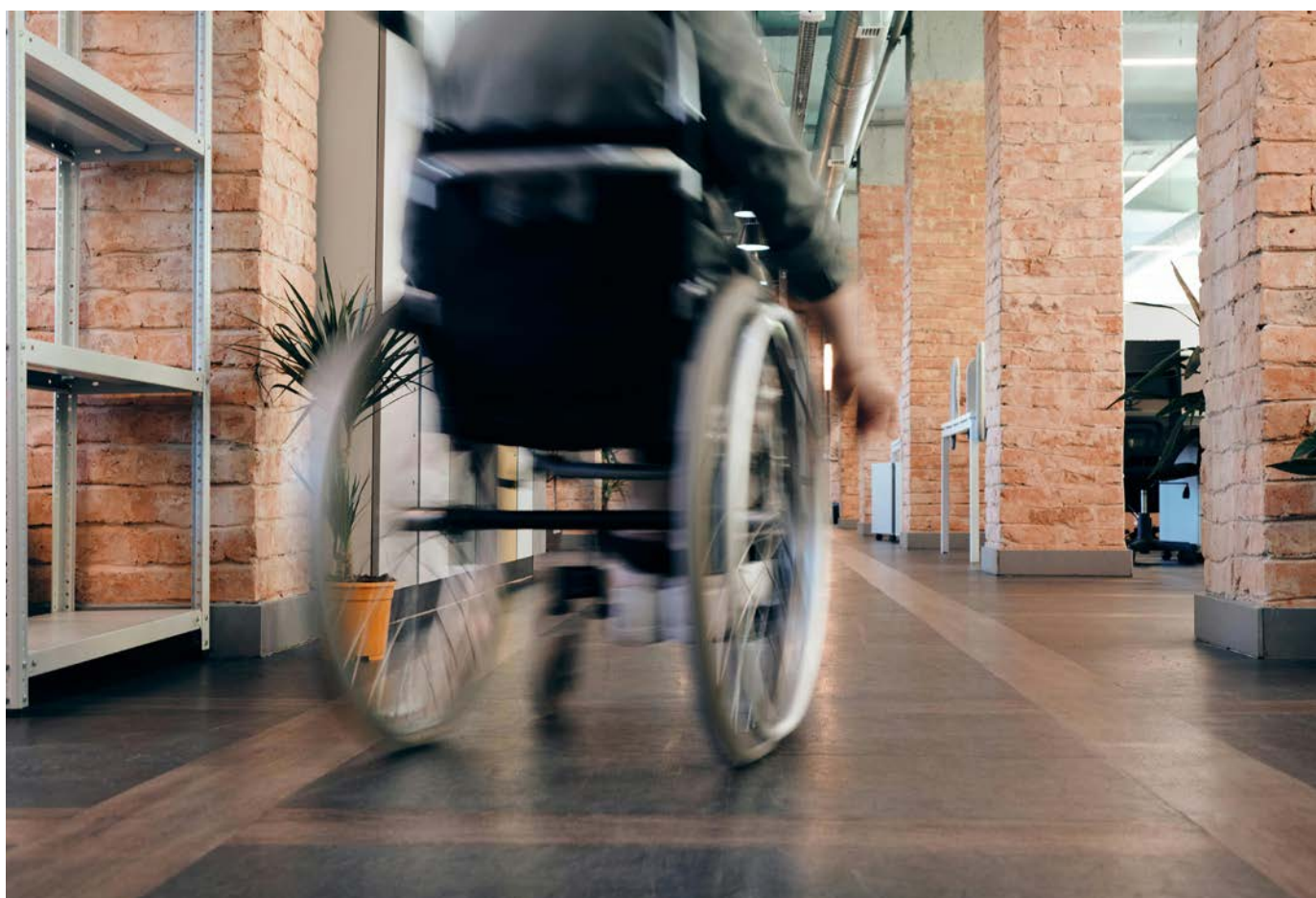


Woman in a wheelchair rolling through a corridor of a brick building. Photo by Marcus Aurelius on Unsplash.

# PART II.
## Conceptual Framework

Close-up of a hand, vividly covered in a spectrum of mixed paint colors against a blue background. Photo by Alexander Grey on Unsplash.

# A Transformative Approach Towards Substantive Equality

Drawing on gender-transformative approaches, this report advances a transformative approach to AI policy to achieve *substantive equality*.

## Substantive Equality

Substantive equality means de facto equality (equality in fact or actual equality) or equality of results (Fredman, 2016; UN Women, 2019). This includes making sure that everybody has equal opportunities, equal access to opportunities and an enabling environment to achieve equal results. Substantive equality, as advanced in this report, incorporates an understanding of equity, [7] gender equality [8] and diversity that calls for equality of opportunity and of results through measures that accommodate difference, enhance voice and participation, and eliminate discrimination, marginalisation and the unequal distribution of power and resources.

Substantive equality is anchored in the human rights framework and is reflected through several mutually reinforcing rights. Gender equality, meaningful participation, and inclusion work to ensure that all individuals, regardless of sex or SOGIESC, age, disability, ethnicity, religion, or economic or other social status can fully enjoy their rights. Gender equality is recognised as a core principle of human rights and is reflected throughout the SDGs. It ensures that individuals, regardless of their gender, have equal opportunities and are free from discrimination. [9] Meaningful participation and inclusion are human rights established in international conventions and involve individuals having a genuine and impactful role in decision-making processes that affect their lives. Achieving substantive equality is essential for the realisation of gender equality, meaningful participation and inclusion for all.



Group of seven Indian women sitting in circle and weaving. Photo by AROYBARMAN on iStock.
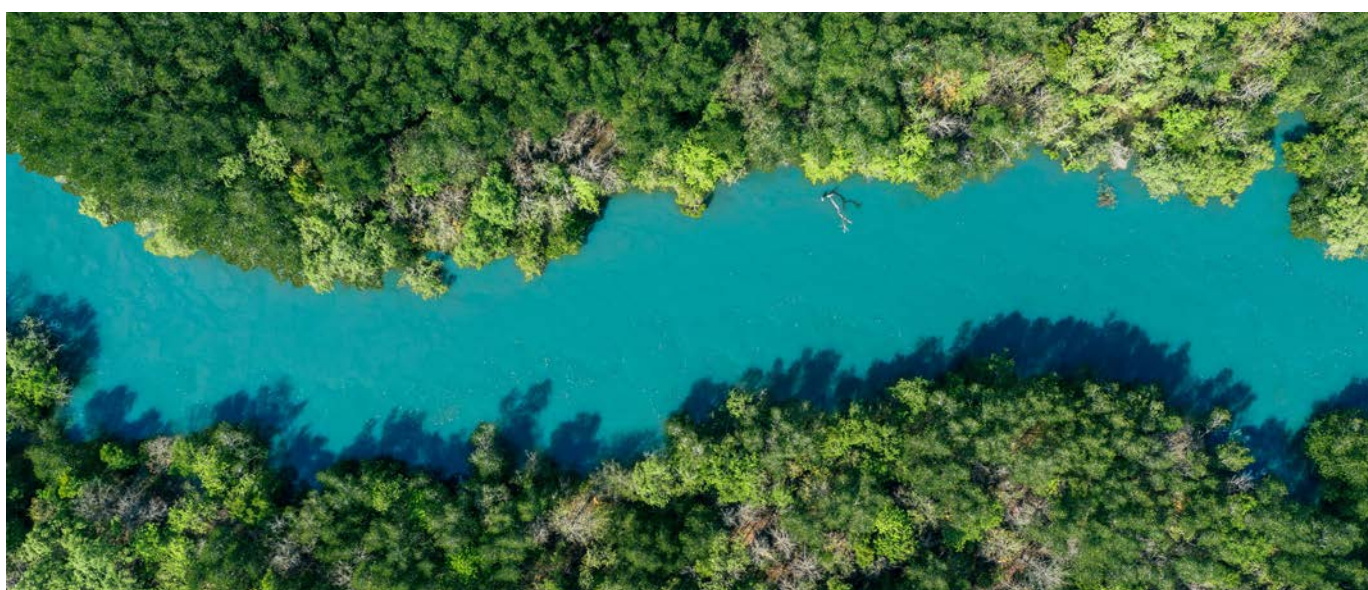


River passing through mangrove forest at Krabi, Thailand. Photo by Jitti Narksompong on iStock.

## Substantive equality in international human rights law

There are nine core international human rights instruments (Office of the High Commissioner for Human Rights, 2024b). Some instruments that reflect principles of substantive equality are presented below.

**The Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)**

The CEDAW shifts the baseline for gender equality from de jure equality in law and policy to de facto equality by effectively addressing systemic barriers to the equal participation of women and girls in all their diversity in all domains of life (IWRAW, n.d.; Shapiro, n.d.). The CEDAW Committee acknowledges that gender-based discrimination is inextricably linked with other factors such as race, ethnicity, religion or belief, health, status, age, class, caste, sexual orientation and gender identity (Campbell, 2015).

**The International Covenant on Economic, Social and Cultural Rights (ICESCR)**

The ICESCR requires substantive non-discrimination in the enjoyment of social and economic rights without discrimination of any kind as to race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status (Office of the High Commissioner for Human Rights, 1966).

**The International Convention on the Elimination of All Forms of Racial Discrimination (ICERD)**

The ICERD imposes obligations on State Parties to adopt specific policy measures that guarantee full and equal enjoyment of human rights and fundamental freedoms for groups marginalised due to race, colour, descent, or national or ethnic origin and individuals belonging to such groups (Office of the High Commissioner for Human Rights, 1965).

**The Convention on the Rights of Persons with Disabilities (UNCRPD)**

The UNCRPD shifts the approach to disability from a medical model to a human rights model that includes the systemic legal, policy, social and other changes that are required to ensure inclusive societies. The Convention provides specific measures for States, as duty bearers, to remove barriers and enable persons with disabilities to access and enjoy their rights fully (Goldschmidt, 2017).

Substantive equality is based on the principle that the human right to equality should be responsive to people who are disadvantaged, demeaned, excluded or ignored. It is the responsibility of all states as duty bearers to protect, promote and ensure the realisation of human rights. Human rights based approaches to policies, projects, programmes and systems generally include the following key activities:

- Recognition of the human rights and human rights instruments that the initiative impacts or is impacted by.

- Empowerment of the rights holders and access to justice to enable them to claim their rights individually and collectively.

- Capacity development for duty bearers to protect and promote human rights.

- Meaningful participation of marginalised groups.

Specific obligations for States include eliminating systemic barriers to gender equality and the many forms of discrimination that contribute to the persistent marginalisation. They also include actively promoting inclusive practices, to create an environment where everyone can fully participate and benefit from societal resources and opportunities. [10]



Diverse group of people, mainly women and First Nations from America with face paint, protesting. Photo by Pascal Bernardon on Unsplash.

From the substantive equality standpoint, policy must tackle the root causes to redress gender inequalities, remove structural barriers, including geopolitical entrenchment and power imbalances, and empower marginalised populations. This requires addressing intersecting social factors such as age, gender identity, sexual orientation, disability, ethnicity, religion, migration or economic or other status to ensure that all individuals, regardless of identity, are included in all aspects of society and have the power to influence outcomes. In other words, it must be transformative.

Actual equality can be achieved through the use of specific measures for transformative change (UN Women, 2015). Effectively inclusive policies, programmes and budgets for gender equality and diversity are the tools that enable the strategic and meaningful expansion of substantive equality and ensure access to human rights for marginalised groups. The concept of "substantive equality" provides a robust foundation for pursuing transformative change that addresses the challenges of inequality and exclusion throughout the full AI lifecycle and related policy making. The next section describes the socio-technical approach to AI systems and the AI lifecycle necessary for a successful implementation of transformative AI policy.

## Transformative Change

Working towards transformative change or taking a gender-transformative approach means tackling the root causes of inequality to make social institutions and relations more inclusive and equitable by removing structural or legal barriers and empowering marginalised groups (UNRISD, 2017). Actions include making changes in law and policy, adapting and improving systems and services, redistributing resources and changing negative or harmful norms and beliefs (including stereotypes), behaviours and practices.

(UNICEF, 2021)

# A Socio-Technical Approach to AI

Working towards gender equality and diversity in AI requires one to consider the full AI lifecycle from a socio-technical perspective. A socio-technical approach to AI systems and lifecycle recognises that the performance and impacts of AI systems stem from the interaction between technical design, infrastructure, and social dynamics and incentives. AI systems are commonly referred to in terms of machine-based systems or a set of computer techniques that may have an impact on their environments. Anchored definitions of AI systems in such terms describe them as "machine-based systems that can influence their environments (OECD, 2024b). These descriptions recognise the crucial role that AI systems can or will have in impacting the environment, societal practices, norms, rights and structures (UNESCO, 2024c). Model-based approaches to AI policy typically adopt this understanding of AI systems and lifecycle.

## Artificial Intelligence (AI) Systems

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

(OECD, 2024b)


People crossing a pathway. Photo by bee32 on iStock.

AI policy often focuses on model-centred approaches to achieve diversity, equity and inclusion goals. These approaches, such as technical de-biasing, aim to correct data inaccuracies caused by social biases and improve AI systems for fair representation, equitable treatment and non-discrimination (Cachat-Rosset and Klarsfeld, 2023). For instance, de-biasing data points that undervalue women can correct biases in performance review scores, while choosing better proxy variables can reduce racial discrimination in recidivism risk algorithms or models seeking to predict reoffending (the "target trait"). Using "arrest records" as a proxy variable for predicting reoffending, for example, reflects already biased racial profiling by the police force (Benjamin, 2019). Incorporating mathematical adjustments, conscientious design, diverse and representative data collection, and collaborative testing and monitoring can address gender inequality and social discrimination (Hellman, 2023; Jora et al., 2022; Solaiman et al., forthcoming). However, data accuracy alone cannot ensure fairness when systemic biases are reflected in the target trait (Joyce et al., 2021; Kong, 2022; Ovalle et al., 2023a). While model-centric approaches correct data inaccuracies, a socio-technical approach addresses the root causes of inequality embedded throughout the socio-technical dimensions of the AI system lifecycle (Iason, 2022). This approach emphasises the mutual impact of social and technical elements on each other.

## The AI Lifecycle

A socio-technical understanding of AI views society and technology together as one coherent system composed of a multiplicity of systems – and an iterative process. This includes the norms, practices, materials and infrastructures in which – and through which – technological systems operate. It also encompasses narratives, imaginaries, ecosystems, geopolitical and economic interests, and AI as a subject matter (Ricaurte and Zasso, 2021). All of these socio-technical elements interconnect throughout the AI system lifecycle (Chen and Metcalf, 2024). The AI system lifecycle described in this report draws on existing conceptualisations and integrates these with additional socio-technical systems key to policy considerations for gender equality and diversity.[11]
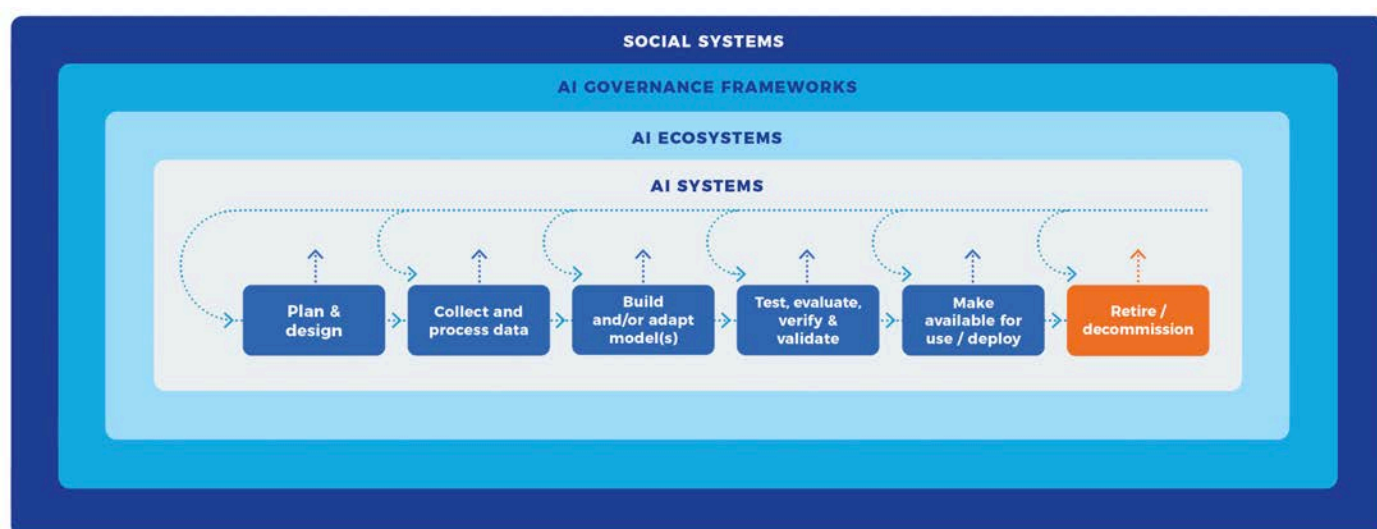


(Image adapted from OECD, 2024a; Ricaurte, 2024)

## The AI System Lifecycle Stages

### Planning and design

This stage includes problem formulation. The actors involved define the AI system's goals, objectives and scope as well as the key problems or opportunities the AI system should address.

### Data collection and (pre)processing

At this stage, researchers and developers identify data sources necessary for training and validating AI models. Sources can include databases, APIs, sensors and manual input. Determining the most relevant data collection practices for the task is important to ensure that all necessary items and processes are included. This is followed by data preprocessing. Preprocessing includes cleaning the data, handling missing values and transforming them into a suitable format for AI modelling. Other tasks involve normalisation (scaling data to a standard range), feature scaling (adjusting the scale of data) and feature engineering (creating new features from existing data).

### Model building

This stage entails modelling the acquired data using appropriate AI models based on the problem requirements. For example, use vision-based models for image data and language-based models for textual data.

### Tests, verification and validation

Testing, evaluation, verifying and validating are key activities during the AI system lifecycle. This includes using evaluation benchmarks to identify errors, biases or limitations in the AI models. Other activities involve validating the models with techniques such as A/B testing (a method of comparison) to ensure accuracy and reliability

### Deployment, operation and monitoring

During deployment, AI system deployers determine a deployment schedule to make the system available for use. At this stage, collaboration with key IT teams and system administrators is key to integrating the AI system with existing systems, databases and user interfaces. The implementation of monitoring and logging mechanisms to track the AI system's performance and identify any issues will also assist in the following stage. During the operation and monitoring stage, deployers maintain and continuously monitor the AI system's performance.

### Retirement or decommissioning

The decision to retire or decommission an AI system from operation may occur at any point during the system's lifecycle.



Overhead perspective of an intricate highway network, featuring a tangle of roads and bridges with numerous vehicles. Photo by Denys Nevozhai on Unsplash.

## Opportunities for Intervention

Viewing the AI system lifecycle through a socio-technical lens helps identify opportunities for intervention to create more inclusive and equitable AI systems. Inequality, inequity and different types of negative impacts or harms occur throughout the stages of AI system lifecycles. Many of these challenges stem from epistemic injustices, which occur when individuals' or communities' research insights and experiences are disregarded or undermined in key stages of the AI system lifecycle due to limited representation and recognition of diverse types of knowledge and expertise (Cheong et al., 2021; Nihei, 2022). Understanding and navigating the social-technical elements at play in any given context relating to AI systems is therefore key to developing successful interventions.

## Transparency, interpretability and explainability in AI systems and processes

Transparency involves system developers documenting the selection process for datasets, variables and models used in AI development, training, validation and testing. They should detail the measures ensuring data and output quality and indicate the confidence level of system outputs, with human intervention required for low-confidence outputs. Transparency also includes documenting methods for detecting potential biases. Independent verification and validation should be mandatory for systems with significant impacts on life and well-being.

Interpretability refers to providing clear information on the procedures followed by algorithms. This helps stakeholders understand the internal workings of AI systems.

Explainability involves detailing the specific decisions made by algorithms. It is crucial for public policy and other contexts where algorithmic decisions may benefit certain groups unfairly. Ensuring that marginalised groups have the right to explanations prevents the perpetuation of inequality. Distinguishing between genuine explanations and after-the-fact rationalisations is essential to maintain accountability and trust in AI systems.

(Adapted from ACM, 2022)

The socio-technical systems interacting with the AI system lifecycle include AI ecosystems, AI governance frameworks, and social systems tied to the economic context and financing, social relations, norms, values, practices, institutions, infrastructure and the environment. AI ecosystems include all the actors, or stakeholders, involved in or affected by the AI system lifecycle. Ensuring the engagement of key stakeholders with an awareness of positionality is key for meaningful participation. Involving relevant multi-stakeholder and interdisciplinary parties throughout decision-making processes helps prevent models from inheriting or amplifying harmful biases and discriminatory practices, for example. AI governance frameworks include frameworks, policies, laws, and regulations on the national, international or global level that shape the use, development, experience and impact of AI systems. Identifying and developing applicable frameworks, strategies and mechanisms at all levels and addressing inadequacies to ensure gender equality and diversity are important tasks to ensure transparency, accountability, and compliance in AI systems and processes. Social systems such as economic context and financing strongly impact the possibilities for change. Identifying and addressing barriers to gender equality and diversity by ensuring equitable access to financial resources, knowledge and networks is crucial for meaningful inclusion. Similarly, working to transform institutional and societal norms and practices through awareness, education, institutional policies and other tools is important to ensure impactful AI strategies. The next section introduces the Transformative AI Policy Framework for substantive equality, providing a holistic foundation for AI policy–related interventions.

# The Transformative AI Policy Framework

The Transformative AI Policy Framework takes a transformative approach to achieve substantive equality in the full AI lifecycle and related policy making. [12] Transformative AI policy integrates human rights into the core of AI ecosystems by removing constraints and enhancing capabilities to ensure the enjoyment of human rights within and throughout technological environments.

The framework seeks to actively empower women and other marginalised groups to fully benefit from their rights. This includes eliminating both direct and indirect discrimination and supporting affirmative actions to promote substantive equality (Gurumurthy and Deepti, 2023). The framework proposes measures to ensure the enactment of any changes required to eliminate institutional barriers that perpetuate the exclusion of women and other marginalised groups from AI development, decision-making, and governance. Recognising structural exclusion and the unequal distribution of power and resources makes it possible to intervene in AI-driven decision-making models and other aspects of AI design, development and deployment. This increases the chances of achieving equitable outcomes, leading to fairer results.

The following section outlines the key dimensions of how a transformative AI ecosystem can promote gender equality and diversity.

## Key Dimensions

The Transformative Policy Framework for Gender Equality and Diversity in AI is based on three core dimensions of substantive equality. The framework takes a systemic approach to integrating human rights (Bartoletti and Xenidis, 2023) and elaborates on these dimensions so that policies and laws for gender equality and diversity in AI innovation can be transformative. This section describes the following three core dimensions:

1. **Remedying systemic disadvantage:**
   **The right to inclusion**
2. **Redressing the democratic deficit:**
   **The right to participation**
3. **Reversing misrecognition:**
   **The right to dignity**

## 1. Remedying Systemic Disadvantage: The Right to Inclusion

A transformative policy framework for gender equality and diversity can address systemic disadvantage in AI ecosystems by actively furthering a right to inclusion. Systemic disadvantage may be understood as "a deprivation of genuine opportunities to pursue one's own valued choices" due to the unfair allocation of resources, goods and opportunity in a society (Fredman, 2016). AI ecosystems can perpetuate systemic disadvantage through distributive injustice, e.g. exclusionary allocation of opportunity in markets, such as racial discrimination in housing ad-serving algorithms, or exclusions from public services, such as discrimination in welfare automation (Stauffer, 2023). They can also perpetuate redistributive injustice, such as the seizure of what could be data commons by large corporations to consolidate intellectual monopolies (Rikap, 2022).

A transformative policy can actively further the right to inclusion by taking action on the following fronts:

- Eliminating harms of direct and indirect discrimination in AI ecosystems.

- Instituting obligations on AI ecosystem developers and providers to promote equality inclusive design processes and affirmative action for women and other marginalised groups (Bartoletti and Xenidis, 2023).

- Steering public and private innovation to generate benefits for marginalised groups. This includes, for example, dedicated budgets to support assistive AI innovations for accessibility and social security for racialised populations.

- Protecting the data commons and open AI innovation from private capture through appropriate IP frameworks grounded in the principle of open science to democratise the benefits of scientific progress (Krishna, 2020).

## 2. Redressing the Democratic Deficit: The Right to Participation

A transformative policy framework for gender equality and diversity needs to address the democratic deficit in AI ecosystems through the right to participation and to create space for the voices of marginalised groups. This enables duty bearers to uphold the basic premise of representative democracy: that every person subject to a policy has a voice in its making (Pande, 2002).

The bias, discrimination and unfairness of AI models represent a democratic deficit in AI ecosystems and is a product of the intersectional exclusion and inequality perpetuated in the spaces where AI is created (West et al., 2019). Considering that spaces of AI innovation are mostly owned by transnational corporations based in global minority countries, diversity and equity initiatives tend to be reduced to a shallow technical fix of adding more data to the neural network or enhancing the visibility of additional identities (Molnar, 2024; Niesen, 2016; Office of the High Commissioner for Human Rights, 2023). Furthermore, the highest-risk AI systems (biometric identification, emotion recognition and automated surveillance) tend to be deployed on the bodies of groups that are already marginalised by our systems – people on the move (migrants, refugees and other displaced persons) or individuals in conflict with law, for instance – who tend to be excluded from the state protections guaranteed to the national political community of citizens (#ProtectNotSurveil, 2024).

Two kinds of policy measures become critical:

- Instituting societal mechanisms for public scrutiny and feedback in private AI standards-setting processes
- Guaranteeing the rights of access to information and justice and the right to public participation in decision-making about AI laws and policies (Smuha, 2021).

## 3. Reversing Misrecognition: The Right to Dignity

Misrecognition – denigration, humiliation and failure to value individuals – is a violation of the principle that all human beings are equal (Fredman, 2016). Stigma, stereotyping, humiliation and violence "on grounds of gender, race, disability, sexual orientation, or other social locational factors" (Fredman, 2016) is an attack on personhood: an individual's claim to being a "social being, who possess(es) rationality, and [is] capable of autonomous action" (Martineau et al., 2012).



Group of people, mainly women protesting for their rights in Heroica Puebla de Zaragoza, México. Photo by Domingo Alvarez E on Unsplash.

Machine-human loops in AI development and deployment formalise forms of discrimination and act to reify and amplify existing forms of social inequality. This is recognised as algorithmic oppression by feminist scholars (Benjamin, 2019; Eubanks, 2017; Noble and Roberts, 2019; West, 2020). Such algorithmic oppression manifests not only in the unjust denial of opportunities for marginalised groups (discussed in Dimension 1) but also as representational injustice (such as AI image generators sexualising racialised women or social media algorithms amplifying sexism and misogyny).

A transformative policy framework for gender equality and diversity needs to reverse such misrecognition and acknowledge the moral struggle of marginalised groups to have their personhood recognised by protecting and promoting the right to dignity. Dignity is foundational to any society that truly embraces diversity. The right to dignity in human rights jurisprudence requires that states recognise the need and pave the way for concerted action against any discrediting process that stigmatises an individual or a group on any kind of ground and the institutional mechanisms that perpetuate it.

This calls for two-pronged policy action on the following fronts:

- Instituting a positive obligation for AI system providers and users to prevent algorithmic discrimination (directly and by proxy).

- Putting in place radical transparency and accountability guardrails in AI ecosystems through ex ante algorithmic audit obligations, and ensuring that trade secrets and patent law do not work at cross purposes with accountability measures to eliminate misrecognition in AI.



Mining excavation on a mountain. Photo by Vlad Chețan on Pexels.

# PART III.
# Promising Practices and Recommendations



A woman standing in a field of tall, dry grass with arms raised and crossed above the head, wearing a colorful patterned garment and braided hair, against an overcast sky. Photo by Ian Kiragu on Unsplash.

# Promising Practices Towards Substantive Equality In AI

**This section brings together a set of promising practices from around the globe that demonstrate how the key dimensions of substantive equality as outlined in the Transformative AI Policy Framework – addressing systemic disadvantage, redressing the democratic deficit and reversing misrecognition – can be effectively furthered in AI ecosystems. Learnings from each practice are summarised as policy insights.**

The practices encompass a range of actions that effectively foster gender equality and diversity objectives in AI ecosystems created by various stakeholders: multilateral agencies, civil society organisations, public agencies at the national level and network organisations. These practices were identified during either the regional or the group-specific consultations. The initiatives either include, are led by or work to promote the rights and meaningful participation of various marginalised groups, either locally, regionally or globally. This set is illustrative and not exhaustive and is meant to provide policy makers and other stakeholders with concrete examples of actions that may be replicated, supported or adapted to other contexts.

The promising practices are categorised as examples of (but are not restricted to) the following categories:

- **Resources for capacity development and public education**

  Expert guidance and resources for policy makers and the general public.

- **Inclusive technology design and democratic innovation practices**

  Technology models that effectively integrate gender equality and diversity objectives at appropriate stages in the AI lifecycle, developed by non-profit organisations and public agencies.

- **Accountability measures**

  Indexes, monitoring initiatives and awareness-building measures that aim to promote the accountability of AI providers and deployers to respect, protect and promote the human rights of all.

- **Meaningful inclusion and participation initiatives**

  Initiatives that promote active and meaningful participation and co-governance in data and AI ecosystems.



Photo of a woman holding a sign that reads "We are equal". Photo by Jacob Wackerhausen on iStock.

# Resources for Capacity Development and Public Education

**UNESCO's Global Dialogue**

**Initiative:**
Artificial Intelligence and Gender Equality:
Key Findings of UNESCO's Global Dialogue

**Organisation:**
United Nations Educational, Scientific and
Cultural Organization (UNESCO)

**Country/Region:**
Global

## Description

UNESCO's Global Dialogue focuses on integrating gender equality into AI policy and practice. The report reviews and identifies gaps in the integration of gender equality in existing AI principles. It provides recommendations and an action plan which provides guidance for each stakeholder group on how to operationalise the recommendations. The action plan includes a focus on five areas: 1) awareness raising; 2) ensuring that gender equality remains a priority; 3) coalition building; 4) capacity building and funding; and 5) research, monitoring and funding. This resource was created in dialogue with global AI and gender equality experts from various stakeholder groups and funded through a partnership with the German Agency for International Cooperation (GIZ).

## How the initiative advances substantive equality

The initiative promotes interventions for transformative gender equality. By advocating for the inclusion of a gender equality lens in AI development and governance, it works to eliminate barriers that perpetuate gender inequality. By recommending ways to enable and empower women and their participation, it tackles the root causes of systemic disadvantage. By enabling gender equality advocates to identify opportunities and build skills for emancipatory purposes through a focus on increasing capacity, awareness and AI literacy, it works to redress the democratic deficit. It seeks to reverse misrecognition by addressing biases and promoting a more inclusive narrative in AI development and deployment.

## Policy insights

- **Shift the narrative from "personal behaviours" to system change**

  Emphasising the imperatives of gender equality, the report recommends that the AI industry and policy makers should shift the focus from individuals to system change. This will enable them to successfully embrace gender transformative principles, frameworks and recommendations.

- **Provide guidance on how to operationalise gender transformative principles, frameworks and recommendations**

  The report highlights the lack of guidelines for operationalising principles such as fairness and transparency. Policy makers should build upon this resource and others to reflect principles in action. They should also develop guidance and ensure that effective gender transformative AI ecosystems are developed.

**Data Justice Policy Brief**

**Initiative:**
Data Justice Policy Brief

**Organisation:**
Global Partnership on Artificial Intelligence (GPAI)

**Country/Region:**
Global

## Description

The *Data Justice Policy Brief* aims to embed principles of data justice into AI policy and practice. It provides a comprehensive framework for policy makers to ensure that AI development and deployment are fair, transparent and inclusive. The policy brief emphasises the importance of addressing both social and economic challenges associated with the proliferation of AI systems and processes. It offers practical recommendations for creating policies that mitigate harm, promote accountability, and ensure that AI benefits are distributed equitably across all segments of society. The ultimate goal is to support a more just and ethical AI landscape.

## How the initiative advances substantive equality

GPAI's *Data Justice Policy Brief* addresses systemic disadvantage by promoting fair and transparent AI policies that are based on principles of inclusion, equity and economic justice, and redress of harms. It bridges the democratic deficit by advocating for inclusive policy-making processes that involve diverse stakeholders, thereby enhancing representation and accountability. In addition, by recognising and ensuring that marginalised communities are involved by design and by promoting transparency and accountability in data and AI practices and systems, it works to uphold the right to dignity, reversing misrecognition.

## Policy insights

■ **Democratic participation of affected communities – by design**

Data justice requires policy makers to identify the full set of stakeholders who might be impacted by data collection and use, and data-driven activities. Individual and collective data subjects, as well as primary data generators, are essential stakeholders. Ensure that their participation is built in democratically in the design, development and deployment of data-intensive systems, including AI systems and processes.

■ **Promote transparency in data and AI practices and systems**

The brief recommends that people who have power in the collection and use of data and data-driven innovation should be obliged to make information publicly available about what data are collected and how they are used, including information about AI inputs and algorithms. It also recommends that they provide this information directly to impacted individuals and communities.

**AI & Equality Human Rights Toolbox**

**Initiative:**
AI & Equality Human Rights Toolbox

**Country/Region:**
Global

**Organisation:**
Women at the Table and École Polytechnique Fédérale de Lausanne (EPFL), in consultation with OHCHR

## Description

The AI & Equality Toolbox is an educational platform designed to catalyse informed debate and collaboration in creating a human rights–based approach to AI. It provides methodology, workshops and online courses curating resources aimed at technologists, AI practitioners, social scientists, activists, policy makers and the public. The initiative focuses on context and purpose, multidisciplinary collaboration and community engagement to achieve equity and inclusion in AI innovation. It is an initiative of Women at the Table in collaboration with the UN Office of the High Commissioner for Human Rights (OHCHR) and EPFL. Its online course sits on the Sorbonne Center for AI (SCAI) learning portal in collaboration with the Sorbonne University.

## How the initiative advances substantive equality

In promoting a human rights–based approach to AI, the Toolbox addresses systemic disadvantage by ensuring marginalised groups are included in AI development, reducing biases and promoting equity. It fosters informed debates and collaboration between technologists, policy makers and the public, ensuring marginalised communities are included in AI policy making. This redresses the democratic deficit. Finally, through education and interdisciplinary collaboration, the Toolbox challenges the myth of AI neutrality, working to create an environment where technologists, policy makers and people impacted by AI systems share a common vocabulary that recognises and integrates a range of social perspectives.

## Policy insights

■ **The adoption of a human rights–based approach**

Adopting a human rights–based approach throughout the AI lifecycle ensures that AI systems are developed, deployed and audited with respect for human rights, fostering informed debates and shared vocabulary on the ethical use of data. The human rights baseline emphasises AI developers' responsibility to adhere to the duties of transparency and explainability, and the "right to know" of people impacted by AI systems.

■ **A platform for multidisciplinary collaboration and community engagement**

The AI & Equality Human Rights Toolbox sets out to create a global community of AI researchers, social scientists, data scientists and activists who collaborate and exchange on how AI systems can serve the public good and prevent harm. The initiative fosters and promotes community engagement and capacity development through a platform with resources such as online courses, and possibilities for discussion. By supporting spaces for multidisciplinary collaboration and community engagement anchored in human rights, policy makers can contribute to building greater AI literacy, capacity and awareness.

## Indigenous Pathfinders in AI

**Initiative:**
Indigenous Pathfinders in AI

**Country/Region:**
Canada

**Organisation:**
Mila - Quebec AI Institute and Indspire

### Description

Indigenous Pathfinders in AI is a transformative career pathway programme that empowers Indigenous talent to learn, develop and lead the evolution of AI. Rooted in community and Indigenous world-views, the programme bridges Indigenous perspectives with AI systems and processes. Led by Mila – Québec AI Institute in partnership with Indspire, the programme engages participants in both technical and non-technical aspects of AI, including topics like responsible AI and ethical data governance. Participants are provided with stipends, and fully covered travel, accommodations, and meals – ensuring they can fully engage in the programme without financial constraints. The Pathfinders initiative is creating new opportunities for Indigenous innovators to shape the future of AI and drive impactful, community-centred solutions.

### How the initiative advances substantive equality

Systemic barriers, including financial challenges, limited access to training opportunities and inadequate digital infrastructure, restrict both skill development and awareness of AI career opportunities. Indigenous Pathfinders in AI advances substantive equality as a targeted initiative that dismantles barriers to democratic innovation and meaningful participation in AI governance. The programme aims to foster early and meaningful inclusion in AI ecosystems, empowering Indigenous talent and perspectives to shape AI systems and processes from their inception, addressing epistemic exclusions. By addressing these gaps proactively, the programme ensures a more equitable and diverse future. This approach addresses systemic disadvantages rooted in historical inequalities and strengthens capacity for meaningful participation, redressing the democratic deficit.

### Policy insights

- **Community-driven AI solutions**

  In addition to creating career pathways that boost AI literacy among Indigenous talent, the Pathfinders programme actively engages Indigenous voices in developing AI solutions tailored to the unique needs of their communities. Pathfinders sets a new standard for integrating Indigenous perspectives throughout the AI development process, from design to ethical data governance. This initiative introduces a transformative approach to AI education while empowering Indigenous communities to shape the future of AI in ways that reflect their values and priorities.

- **A holistic approach that prioritises Indigenous knowledge**

  The Pathfinders programme's learning environment is shaped by Indigenous perspectives, with its core founded on the four R's of Indigenous wisdom: responsibility, respect, reciprocity and relevance. Holistic in nature, Pathfinders emphasises participants' well-being, providing access to Elders and knowledge keepers, as well as sessions dedicated to cultural teachings. By empowering Indigenous talent to tackle challenges they are uniquely positioned to understand, the programme fosters the development of AI solutions that address social issues important to and defined by their communities.

# Inclusive Technology Design and Democratic Innovation Practices

**Design from the Margins (DFM) Methodology**

**Initiative:**
Design From the Margins (DFM) Methodology

**Organisation:**
The De|Center

**Country/Region:**
Global

## Description

Design from the Margins (DFM) is a methodology developed to create equitable technology by centring the lived experiences of marginalised, or decentred, groups throughout the entire AI lifecycle. DFM resides at The De|Center, an interdisciplinary centre for community-based research, corporate accountability and technology design intervention to reduce the harms of existing and emerging technologies. The methodology incorporates community-focused research, harm reduction, and deep engagement with marginalised groups (decentred users) – people who are most at risk of experiencing harmful impacts of technology. Examples of the impact of this approach include improvements made in platforms like Grindr, Signal and WhatsApp. The initiative is funded by grants and donations.

## How the initiative advances substantive equality

DFM focuses on harm reduction and aims to transform technology tools to prevent their use in perpetuating systemic oppression. In so doing, the approach addresses systemic disadvantage. The methodology involves deep community-based research, collaboration and interventions embedding the principles of safety, dignity and rights of marginalised groups. DFM redresses democratic deficit and helps ensure that tech advances serve everyone, particularly those most impacted by injustice. It recognises that, when technology is designed for those most affected by systemic injustices, it becomes better for everyone. The approach also reverses misrecognition, as marginalised voices are amplified and centred throughout the methodology and related work.

## Policy insights

- **Create or support a Mutual Aid Fund**

  Through a Mutual Aid Fund, marginalised groups can have access to no-strings-attached financial support in the shape of cash or in-kind assistance. Companies who benefit from DFM are encouraged to contribute funds. By creating or supporting similar funds, policy makers can enable marginalised groups to lead the development they seek.

- **Translate identified harms and wants into implementable changes**

  The DFM enables technology interventions by translating documented harms and injustices into implementable changes to AI-related technology and practice. This also includes wants expressed by marginalised groups. Policy makers can also support similar practices that achieve concrete changes directed by the experiences of those who are most impacted.

## The Feminist AI Research Network (f<A+i>r)

**Initiative:**
The f<A+i>r Network

**Organisation:**
<A+> Alliance for Inclusive Algorithms

**Country/Region:**
Global

### Description

The Feminist AI Research Network, f<A+i>r, initiative aims to create inclusive AI technologies by addressing systemic gender, racial and intersectional biases. The invitation-only Global Directory includes approximately 100 feminist AI academics, activists and practitioners from different fields that share multidisciplinary research and feminist AI innovation to advance transformative change. f<A+i>r fosters South-South-North-South knowledge sharing and highlights feminist innovations globally. The f<A+i>r network was created in 2020 by the <A+> Alliance for Inclusive Algorithms, supported by the International Development Research Centre of Canada (IDRC), the Oak Foundation and the Swiss Federal Department of Foreign Affairs (FDFA) Human Rights and Women's Rights Divisions.

### How the initiative advances substantive equality

The f<A+i>r initiative furthers the goals of addressing systemic disadvantage by actively combating gender, racial and intersectional biases in AI technologies. It seeks to correct future harms with new data, algorithms, models, policies and systems that can be researched and piloted for transformative change. It redresses the democratic deficit by fostering the inclusion of and supporting the capacity development of feminist innovation teams from the Global Majority. The initiative reverses misrecognition by producing effective, innovative, interdisciplinary models that harness emerging technologies, correcting for real-life bias and barriers to women's and other marginalised groups' rights, representation and equality.

### Policy insights

- **The promotion of transformative AI systems**

  By funding and supporting in other ways the development and implementation of AI systems and processes that actively integrate techno-feminist principles and correct systemic gender and intersectional biases (creating new datasets, algorithms and models), policy makers support gender equality and diversity in AI.

- **Collaboration and knowledge exchange**

  The network of feminist technologists from the Global Majority contributes to the development of context-specific AI solutions that address regional needs and challenges. This leads to more inclusive and equitable AI policies globally.

**Fixing the bAIs Initiative**

**Initiative:**
Fixing the bAIs: Using AI to Correct Gender Bias in AI

**Organisation:**
MullenLowe MENA for Aurora 50

**Country/Region:**
United Arab Emirates/MENA

## Description

The Fixing the bAIs campaign aims to address and mitigate gender biases in AI systems by training AI datasets using AI. Women are vastly underrepresented in the datasets used to train AI. Using image-generating AI tools such as Midjourney, Dall-e and Stable Diffusion, and employing gender-intentional text prompts, the campaign corrects representational biases in training data pools. The images are royalty- and rights-free to encourage dissemination that will further inform various datasets that will be used to train AI. It also fosters collaboration and dialogue among all parties to promote best practices and innovative solutions for bias reduction in AI.

## How the initiative advances substantive equality

The Fixing the bAIs campaign addresses systemic disadvantage by seeking to reduce discrimination and promoting fair treatment across demographics through bias correction in AI systems. It redresses the democratic deficit by enhancing transparency and accountability in AI development, ensuring participation by marginalised communities. The campaign also reverses misrecognition by intentionally designing gender-inclusive AI systems: providing open-licensed, gender-diverse training datasets; challenging dominant gender scripts; and ensuring that marginalised groups are better represented and their contributions recognised.

## Policy insights

■ **The mandating of diverse datasets and gender-intentional training**

Policy makers can reduce biases in AI systems by mandating the use of diverse datasets and gender-intentional training processes. By incorporating diverse and inclusive data practices, ensuring that training datasets are representative of various demographics, AI systems can better reflect and serve the needs of all societal groups.

■ **Disclosure requirements**

Policy makers can ensure continuous improvement and public trust in AI systems by adopting regulations that require organisations to disclose their efforts and progress in mitigating AI biases. Regular audits and public reporting of AI systems' biases and corrective measures promote transparency and accountability in AI development.

**Indigenous Jobs Map**

**Initiative:**
Indigenous Jobs Map

**Country/Region:**
Australia

**Organisation:**
Australia National Research Agency
(Commonwealth Scientific and Industrial
Research Organisation; CSIRO)

## Description

The Indigenous Jobs Map, developed by CSIRO, provides a comprehensive overview of job opportunities in Indigenous communities across Australia. It aims to support Indigenous employment and economic development by mapping and visualising current job data. The Indigenous Jobs Map is a data analytics tool developed with the intent to aid Australia's national research agency in effectively implementing its Reconciliation Action Plan to improve employment outcomes for Aboriginal and Torres Strait Islander workers in the national job market. The initiative seeks to enhance transparency and access to employment information, enabling better decision-making and planning for both job seekers and employers.

## How the initiative advances substantive equality

The Indigenous Jobs Map enhances access to employment opportunities for Indigenous communities by mapping current job data, promoting economic development and reducing employment disparities. In so doing, it addresses systemic disadvantages. By providing transparent and accessible employment information, the initiative ensures Indigenous voices and needs are included in decision-making processes, supporting more inclusive planning and policies. This redresses the democratic deficit. The map highlights the contributions and potential of Indigenous communities, countering stereotypes and ensuring that their skills and opportunities are recognised and valued, reversing misrecognition.

## Policy insights

- **Data-driven targeted affirmative action for employment strategies**

  The Indigenous Jobs Map uses detailed job data to create tailored employment strategies for Indigenous communities, helping policy makers address employment disparities and promote economic growth. This initiative highlights the importance of a substantive equality approach in AI, showing how data analytics can guide targeted affirmative actions for policy makers.

- **Enhancing transparency and access**

  The initiative provides a transparent, accessible platform for visualising employment opportunities. Policy makers can use similar tools to ensure that job information is easily accessible to Indigenous communities, facilitating informed decision-making and planning. This inclusive technology design helps bridge the information gap and empowers Indigenous individuals in the job market.

# Accountability Measures

## The Global Index on Responsible AI

**Initiative:**
The Global Index on Responsible AI

**Organisation:**
Global Center on AI Governance (GCG)

**Country/Region:**
Global

### Description

The Global Index on Responsible AI is an initiative aimed at assessing and promoting responsible AI practices worldwide. It provides a comprehensive evaluation of countries' AI policies and practices, focusing on inclusivity, accountability and ethics. The index serves as a multidimensional tool and benchmark with globally representative data for policy makers, researchers and journalists to track and measure progress on steps countries are taking to ensure the enjoyment and realisation of human rights in AI. The index is led by the GCG with support from the International Development Research Centre of Canada (IDRC), the Government of Canada and USAID.

### How the initiative advances substantive equality

The index encourages countries to adopt policies that reduce biases and discrimination in AI systems, ensuring that marginalised groups benefit from AI technologies, addressing systemic disadvantage. It measures country-level commitments and practices ensuring that AI promotes, rather than harms, gender equality and women's empowerment. Redressing the democratic deficit, the index was developed in consultation with affected parties and is based on principles of accessibility and openness, inclusion and participation, and the ambition to fairly reflect local contexts and realities. Reversing misrecognition, the index highlights best practices that ensure fair representation of all societal groups in AI development and deployment, combating stereotypes and promoting inclusion.

## Responsible AI

Responsible AI is defined by the Global Index on Responsible AI as the design, development, deployment and governance of AI in a way that respects and protects all human rights and upholds the principles of AI ethics through every stage of the AI lifecycle and value chain. It requires all actors involved in the AI ecosystem to take responsibility for the human, social and environmental impacts of their decisions. The responsible design, deployment and governance of AI are proportionate to the purpose of its use and meet the technological needs of the individuals and societies it seeks to serve.

(Adapted from the Global Index on Responsible AI, 2024)

**Policy insights**

- **Benchmarks for regulations**

  The initiative provides benchmarks that take into account actors' responsibilities across the entire AI lifecycle and ecosystem. Policy makers can turn to these to create regulations that mandate transparency in AI operations and clear accountability mechanisms to ensure a responsible use of AI systems that does not perpetuate biases.

- **International co-operation through collaborative monitoring**

  The initiative promotes international co-operation through collaborative monitoring. Policy makers can further strengthen and support the role of global communities in collaborative monitoring and accountability of AI actors. International co-operation on responsible AI is an area of shared commitment between countries around the world.



A close-up of two hands holding, with one hand featuring a ring.. Photo by Nathan Dumlao on Unsplash.

**The Migration and Technology Monitor**

**Initiative:**
The Migration and Technology Monitor

**Country/Region:**
Global

**Organisation:**
Refugee Law Lab

## Description

The Migration and Technology Monitor (MTM) is a collective of journalists, filmmakers, academics and international communities working from the ground up to share the realities of migrants affected by evolving border technologies. It is a community built by and for people on the move. It monitors surveillance technologies, automation and the use of AI to screen, track and make decisions about people on the move. The MTM supports migrant communities by fostering a fellowship for storytelling on border surveillance impacts. The initiative aims to share knowledge and promote participatory work, connecting rigorous analysis with policy and advocacy to engage the public, media and policy makers.

## How the initiative advances substantive equality

MTM addresses systemic disadvantage by including people on the move, such as refugees and displaced persons, in the design, development and deployment of AI systems. It mitigates systemic harms by foregrounding lived experiences and highlighting discriminatory practices at borders and in humanitarian settings. By involving migrants, who are often excluded from democratic processes, in policy conversations about technology, MTM addresses the democratic deficit and embodies "nothing about us without us".[13] Additionally, MTM reverses misrecognition by emphasising migrants' experiences, ensuring fair representation in policy discussions and challenging exclusionary narratives.

## Policy insights

■ **Transparent reporting and oversight**

Documenting the discriminatory practices and harmful impacts experienced by people on the move promotes accountability. Policy makers can promote accountability by requiring detailed public reporting and independent oversight of AI tools used at borders and in humanitarian settings to ensure they uphold human rights standards.

■ **Community-led monitoring**

Supporting community-led monitoring of AI systems affecting migrants ensures that affected individuals have a direct role in assessing and holding accountable the technologies that impact their lives. Policy makers can enhance accountability and responsiveness in AI deployment by establishing mechanisms for community-led audits and feedback loops.

**The Algorithmic Justice League**

**Initiative:**
The Algorithmic Justice League

**Organisation:**
The Algorithmic Justice League

**Country/Region:**
USA/North America

**Description**

The Algorithmic Justice League (AJL) combines art, research and advocacy to raise awareness of and develop capacity to mitigate harms and biases in AI. Through public awareness campaigns, educational programmes, and capacity development, they strengthen the people most impacted by AI systems and equip policy makers, researchers and industry professionals to ensure that AI systems are developed and deployed in ways that protect human rights and promote social justice. One of its critical focus areas is tackling the intersectional injustice of sexism and racism in the specific ecosystem of facial recognition technologies. The AJL is funded by the Ford Foundation, the McArthur Foundation and individual contributions.

**How the initiative advances substantive equality**

By advocating for fair and equitable AI, the AJL works to reduce systemic discrimination against marginalised communities. It addresses systemic disadvantage and reverses misrecognition by uncovering and mitigating biases in AI technologies, ensuring that these systems do not perpetuate existing inequalities. Through audits and advocacy campaigns that highlight racism, sexism, ableism and other forms of discrimination, it brings to the fore core concerns about the violations of the right to autonomy and dignity. It promotes transparency and accountability in AI development, fostering public dialogue and participation. This helps to democratise AI policy making, ensuring that diverse voices are included in discussions about technology's role in society.

**Policy insights**

■ **Public awareness raising through art and storytelling**

In addition to participating in talks and producing educational materials, the AJL raises public awareness about the impacts of AI systems and processes through art and storytelling amplifying personal stories of people whose lives have been directly impacted by unjust algorithms. By supporting and enabling initiatives working to raise public awareness about the harmful impacts of AI systems and processes, policy makers can contribute to more effective and inclusive AI governance processes.

■ **Advocacy through collective documentation**

The AJL invites anyone to submit and share their lived experiences of AI systems and processes, such as facial recognition programmes at TSA checkpoints. These accounts allow for a better understanding of marginalised groups' experience of documented issues such as racial discrimination, privacy concerns or the consequences of choosing to opt out. Policy makers can invite or support initiatives that invite people to share their lived experiences to document harms for increased understanding and possibilities for justice.

# Meaningful Inclusion and Participation Initiatives

**Māori Data Governance Model**

**Initiative:**
Māori Data Governance Model

**Organisation:**
Te Kāhui Raraunga Charitable Trust

**Country/Region:**
New Zealand

## Description

The Māori Data Governance Model initiative by Te Kāhui Raraunga focuses on ensuring that Māori data are governed by Māori principles and values. Māori data represent a taonga (treasured possession) that requires culturally grounded models of protection and care. The Model advocates for data sovereignty, emphasising the control, ownership and application of data by Māori communities. The Model provides guidance for the systemwide governance of Māori data, consistent with the Government's responsibilities under te Tiriti o Waitangi (Treaty of Waitangi). The initiative is the result of a co-design process between Māori leaders, representatives of Māori organisations with data interests, individual Māori data experts and Crown agencies.

## How the initiative advances substantive equality

The initiative promotes data sovereignty, ensuring that Māori communities control and benefit from their data. By aligning data practices with Māori principles, it supports development and reduces systemic disadvantages rooted in historical inequities. The co-design process involves Māori leaders, organisations and experts, ensuring that Māori voices are integral to data governance decisions. This inclusive approach strengthens democratic participation and aligns with Treaty of Waitangi obligations. By centring and amplifying Māori voices, the initiative ensures that Māori perspectives and rights are recognised and respected in data governance, correcting historical misrepresentation and exclusion or epistemic injustice.

## Policy insights

- **"Nothing about us without us" treaty-based co-design process**

  The initiative emphasises a treaty-based co-design process, where outcomes are defined by Māori and decision-making authority rests with them. By recognising their right to steward their data commons as integral to their right to self-determination and as structurally enabling, it ensures meaningful participation and inclusion of Māori communities in data governance.

- **Amplification of Indigenous voices**

  Equipping Indigenous communities to have a greater say in the management of their collective data resources is essential to prevent representational injustice and cultural appropriation. By addressing inadequacies in past censuses and prioritising Māori voices in the co-design process, the initiative ensures that data governance reflects the Māori communities' perspectives and rights. This practice enhances representation and supports the accurate and respectful use of Māori data.

**The RIADIS Workshop**

**Initiative:**
RIADIS Workshop to Raise Awareness on Artificial Intelligence (AI) and Assistive Technologies among Persons with Disabilities in Latin America (in Spanish)

**Country/Region:**
Latin America

**Organisation:**
RIADIS with funding from Christian Blind Mission

## Description

The Latin American Network of Non-Governmental Organizations of Persons with Disabilities and Their Families (RIADIS) is a regional network dedicated to promoting and defending the human rights and social inclusion of people with disabilities. The network focuses on advocacy, capacity building and collaboration to influence policy and raise awareness about the rights and needs of people with disabilities. The workshop sought to present and share good practices, actions and research in AI and assistive technologies for people with disabilities. The event featured tools like speech-to-text converters and discussed the risks of AI systems in the labour market for people with disabilities if they are not consulted throughout the AI lifecycle.[14]

## How the initiative advances substantive equality

The RIADIS initiative demonstrates practical strategies to centre the needs and concerns of persons with disabilities in AI innovation. By ensuring their voices are included at every stage of the AI lifecycle, RIADIS helps to address systemic exclusion. Targeted workshops give persons with disabilities the knowledge and tools needed to engage in discussions and decision-making, empowering them to advocate for inclusive design considerations and hold AI developers accountable, redressing the democratic deficit. By centring the needs and voices of persons with disabilities, RIADIS works to reverse their historical misrecognition. The initiative promotes their visibility and ensures that their contributions are valued and acknowledged.

## Policy insights

■ **Empowerment through education**

Targeted workshops that empower persons with disabilities to engage in decision-making processes related to AI enable participants to understand the issues, advocate for their needs and hold AI developers accountable for inclusive design. Policy makers can replicate this approach by funding and supporting educational programmes that seek to develop the skills and confidence among marginalised groups to participate meaningfully in policy discussions and technology design processes.

■ **Adoption of disability-specific ethical frameworks**

Adopting disability-specific ethical frameworks in AI design and development ensures that the unique needs and rights of persons with disabilities are considered throughout the AI lifecycle. Policy makers should mandate the integration of these frameworks into AI development processes to promote ethical considerations that specifically address disability issues.[15]

**The Gender and Responsible Artificial Intelligence Network (GRAIN)**

**Initiative:**
The Gender and Responsible Artificial Intelligence Network (GRAIN)

**Country/Region:**
Senegal, Nigeria, Uganda

**Organisation:**
Initiative Prospective Agricole et Rurale (IPAR), Centre for the Study of the Economy in Africa (CSEA) and Sunbird AI

## Description

GRAIN focuses on integrating gender equality and responsible AI practices in sub-Saharan Africa. The network connects organisations, universities and hubs to collaborate on AI and gender inclusion. It aims to learn, share, advocate and propose solutions to link AI development with gender equality, supporting inclusive development in the region. The initiative is a consortium of think tanks and non-governmental organisations and is part of the Artificial Intelligence for Development in Africa (AI4D Africa) programme. It is conducted with the financial support of Canada's International Development Research Centre (IDRC) and the Swedish International Development Cooperation Agency (Sida).

## How the initiative advances substantive equality

GRAIN reduces systemic disadvantages by connecting organisations, universities and hubs, ensuring that AI development in sub-Saharan Africa includes diverse perspectives. The initiative tackles challenges such as under-representation of women, lack of attention to African languages and gender-inclusive development solutions. It promotes inclusive, multidisciplinary collaboration and decision-making by involving various parties in AI and gender-related discussions. This approach helps to redress the democratic deficit. The initiative advocates for AI systems that recognise and respect the identities and contributions of all individuals, fostering a more inclusive and context-attentive technological environment and reversing the historical misrecognition of marginalised communities.

## Policy insights

■ **Inclusive AI policy development**

GRAIN highlights the need for AI policies that address intersectional discrimination and under-representation by ensuring that AI development includes diverse perspectives, especially from women in sub-Saharan Africa, and attention to African languages. Policy makers should adopt frameworks that mandate the consideration of intersectionality and local context in AI development to reduce systemic disadvantages.

■ **Multidisciplinary and participatory collaboration**

GRAIN emphasises inclusive and multidisciplinary collaboration involving various parties in AI and gender-related discussions. Policy makers should encourage engagement by affected parties in AI policy making to create more representative, context-attentive AI systems.

# Key Recommendations for Transformative AI Policy

**The report, *Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity*, offers a comprehensive framework and key recommendations for achieving substantive equality within AI ecosystems and related policy making. The following set of recommendations outlines concrete measures that policy makers can take to effectively integrate gender equality and diversity principles throughout AI policy frameworks, laws, regulations and practices.**

The transformative AI policy recommendations build upon the results of regional and group-specific consultations, the policy insights from promising practices; they reflect key dimensions of the Transformative AI Policy Framework. The challenges identified through consultations—including conflating access with meaningful inclusion, knowledge exclusion and the unbalanced distribution of resources—demonstrate the profound inequalities embedded within the AI ecosystem. These challenges require concerted policy efforts that 1) effectively address structural barriers preventing historically marginalised and excluded groups from benefiting from the gains of AI development; and 2) actively seek to empower these groups to fully enjoy their rights. The recommendations for transformative AI policy are grouped in the following four categories addressing these challenges:

- **Inclusive Design and Democratic Innovation**
- **Meaningful Participation in AI Governance**
- **Transparency and Accountability for Harm Prevention**
- **Effective Access to Justice**

With these categories, the key recommendations for transformative AI policy take a holistic approach towards substantive equality. First, **inclusive design and democratic innovation** are key to addressing systemic disadvantages faced by women and other marginalised groups and upholding the right to inclusion. Policy makers can actively contribute to this through affirmative action: investing in capacity building for institutional inclusion, permitting processing of special categories of data and funding transformative technology research and design approaches. Second, ensuring **meaningful participation in AI governance** redresses the democratic deficit and upholds the right to participation. This includes promoting effective public engagement and community participation, investing in capacity development among marginalised groups, legislating for public participation rights and safeguarding collective rights relating to data and AI. Third, **transparency and accountability are essential for harm prevention**, redressing the democratic deficit and reversing misrecognition. They empower advocates of substantive equality and other stakeholders to scrutinise AI systems and processes, detect biases, and hold private and public providers and deployers accountable for harmful or discriminatory impacts. Recommendations include ensuring the right to information, enhancing algorithmic transparency and establishing clear accountability among all actors involved in the AI ecosystems through human rights impacts assessments and public procurement guidelines. Lastly, **ensuring effective access to justice** reverses misrecognition and upholds the right to dignity. By strengthening contextual liability, empowering equality bodies and easing the burden of proof, policy makers can improve access to justice for individuals and groups facing discrimination and harms in relation to AI systems and processes. Together, the categories address all dimensions of the Transformative AI Policy Framework to effectively achieve actual equality.

The recommendations are meant to be considered and applied in context. To successfully implement these recommendations, policy makers must therefore take into account contextual considerations – geographical, social, cultural, historical, economic, legal and political – that shape AI ecosystems in different regions and adapt the recommendations accordingly. This also includes considering variations in legal and regulatory frameworks, technical capacities, resource availability and stakeholder co-operation, and identifying and addressing obstacles for implementation. For guidance in implementation, please see the Policy Guide for Implementing Transformative AI Policy Recommendations. By committing to transformative AI policies for gender equality and diversity, we can move towards substantive equality in AI ecosystems and related policy making.

## Inclusive Design and Democratic Innovation

Inclusive design and democratic innovation are key to addressing systemic disadvantages faced by women and other marginalised groups and upholding the right to inclusion. The lack of diversity and representation, awareness and understanding of diverse experiences and perspectives, and processes that enable meaningful inclusion throughout AI ecosystems represents a critical challenge to implementing actual equality. Through affirmative action, investing in capacity building for institutional inclusion, permitting processing of special categories of data and funding transformative technology research and design approaches, policy makers can actively contribute to inclusive design and democratic innovation processes.

### 1. Involve Marginalised Groups in Technical and Non-Technical Roles Throughout the AI Ecosystem

Implement affirmative action across the AI ecosystem to involve women and other historically marginalised groups in technical and non-technical roles throughout the AI ecosystem to increase diversity in perspectives. Allocate resources to identify and remove barriers to diverse representation. This includes ensuring accessible, inclusive education beyond AI ecosystems. Diverse teams and interdisciplinary approaches bring wider perspectives to AI development, making systems more inclusive, equitable and transformative (Buhl, 2023). To further ensure compliance with gender equality and diversity standards, mandate that AI design processes obtain approval from ethics advisory boards and certification programs.



Native Maori woman discussing business matters with another woman, looking at a mobile tablet device in a workplace in New Zealand. Photo by corners74 on iStock.

## 2. Invest in Capacity Building for Institutional Inclusion

Invest in capacity development and awareness raising, within public and private institutions and teams, on the experiences and rights of historically marginalised groups such as people experiencing barriers on the basis of sex, sexual orientation, gender identity or expression, racialised people, people with disabilities or Indigenous peoples (Bartoletti and Xenidis, 2023). Ensure regular dialogue with representatives of marginalised groups to understand and eliminate the specific barriers they face. Capacity development and awareness raising for institutional inclusion will strengthen institutions' ability to be more inclusive of marginalised groups, both in their role as employers and in their role as service providers. Develop and regularly review guidelines for communications infrastructure and written and verbal communications to ensure they do not pose accessibility barriers, discriminate by use of certain language/expressions, or fail to represent the diversity in society. This includes carrying out accessibility assessments and establishing accessibility and inclusion checklists for all forums and online activities. These measures contribute to meaningful inclusion by promoting inclusive practices and an increased understanding within institutions of how to accommodate diverse experiences and perspectives. Every researcher or staff member should understand what they need to change specifically in their practices, protocols and guidelines to contribute to achieving the institution's objective to be more inclusive and diverse.

## 3. Permit Processing of Special Categories of Data

Permit the processing of special categories of data under certain exceptional circumstances, based on substantial public interest, to achieve equality and non-discrimination. This should be done without contravening personal data protection rights, similar to the provisions in the EU AI Act's Recital 44c. To prevent discriminatory outputs, AI system providers must test for systemic bias and ensure the representation of diverse datasets. This may require the processing of sensitive personal data to assess how protected attributes such as race and gender might correspond to proxy variables in the model and perpetuate discrimination (Deck et al., 2024; Dwork et al., 2012). Such processing can be restricted under data protection laws. Assessing substantial public interest in order to achieve equality and non-discrimination while respecting personal data protection rights is essential to protect and promote human rights. Robust data protection frameworks are unevenly distributed globally but required to prevent discrimination and exploitation in the context of AI. Support research to expand collective perspectives on data commons and public interest approaches to inform the implementation of this recommendation (Cofone, 2023).

## 4. Fund Transformative Technology Research and Design Approaches in AI Innovation

Fund research and provide grants and public recognition to incentivise the application of inclusive and transformative techno-design approaches in AI, such as those anchored in feminist technology design principles. These approaches address the gaps between technical and political fairness. Supporting AI system innovations that align with these principles advances more equitable and just applications, practices and processes.



Australian woman and two younger women looking at Aboriginal flag. Townsville, Australia during NAIDOC Week. Photo by Ann Smith on iStock.

**Feminist technology design principles in AI innovation include:**

**Integrating embodied knowledge reflecting diverse identities**

Respect the autonomy and dignity of all individuals, especially marginalised communities, as a core design principle in AI development.

- Apply the principle "Nothing about us without us" as a gold standard for full participation and equalisation of opportunities for, by and with historically excluded groups. Recognise this as part of the "do no harm" principle.
- Co-design AI systems with people who experience barriers on the basis of sex, sexual orientation, gender identity and/or gender expression to ensure gender-inclusive innovation.

**Adopting intersectional fairness metrics in AI models**

Implement intersectional fairness metrics to equalise AI systems' performance across intersectional subgroups. These metrics should:

- Consider multiple protected attributes.
- Protect all intersecting values of these attributes, e.g. racialised women.
- Continue protection for individual attribute values, e.g. women.
- Ensure protection for marginalised groups heavily affected by societal discrimination. Rectify systematic differences due to structural oppression, rather than codifying them (Buolamwini, 2023; Buolamwini and Gebru, 2018; Ovalle et al., 2023b).

**Developing and promoting Indigenous and minoritised languages and cultures in AI development**

- Encourage both large-scale public initiatives and smaller community-driven projects to foster active community participation in creating AI datasets, algorithmic schemas and use-case formulations. Establish incentives and infrastructures at national and sub-national levels in local languages.[16]



Flat lay with black cubes arranged in Feminist word on brown wooden surface. Photo by Jen Theodore on Unsplash.

## Meaningful Participation in AI Governance

Ensuring meaningful participation in AI governance redresses the democratic deficit and upholds the right to participation – leading to better policies for everyone. Access is often mistaken for inclusion, but access alone does not guarantee meaningful participation or inclusion. When policy makers listen to and engage with the people most affected by AI systems and processes and those experienced in identifying and addressing its harmful impacts, policies, laws and regulations can more effectively prevent and mitigate harm. Recommendations include promoting effective public engagement and community participation, investing in capacity development among marginalised groups, legislating for public participation rights and safeguarding collective rights relating to data and AI.

### 5. Promote Effective Public Engagement and Community Participation

Foster equity and inclusion in the design, development, deployment and governance of AI systems by employing various public engagement methodologies on national and international levels. This means including marginalised voices in national AI governance discussions, as well as amplifying the Global Majority in international AI governance forums. Utilise deliberative polling, community juries, citizen assemblies, consensus conferences, deliberative mini-publics, online deliberation, participatory budgeting, iterative participatory research and design practices, community assemblies and community reference panels (OECD, 2021).[17] Engage with marginalised communities and organisations representing marginalised groups to understand and eliminate specific barriers they face (United Nations, 2019, 2021). Allocate budgets for participation costs for representatives of marginalised groups such as compensation for their time and expertise, and costs for reasonable accommodations to ensure accessibility (e.g. sign language interpreters, Indigenous language interpreters). These efforts must also ensure that information and consultation processes are accessible, free, and comprehensible to marginalised groups, using appropriate language and communication channels and formats.

### 6. Invest in Capacity Development Among Marginalised Groups

Fund and support educational programmes, networking structures, and other resources that seek to develop the skills and confidence among marginalised groups to participate meaningfully or to actively lead the processes that serve their needs. Work with marginalised communities and representative organisations of marginalised groups to hold their own awareness sessions and consultations on AI-related issues, thereby enhancing their understanding and engagement. Invest in public awareness sessions about AI, individual rights, relevant institutions, and mechanisms for seeking redress and reparation.

### 7. Legislate for Ex Ante Public Participation Rights

Ground AI decision-making processes in ex ante public participation rights such as those established through the UNECE Aarhus Convention. The Aarhus Convention guarantees societal rights regarding access to information, public participation in decision-making and access to justice in environmental matters (UNECE, 1998). Applying these principles to AI decision-making processes enables affected parties, as well as civil society organisations and the general public, to contest algorithmic decision-making consequences through public reasoning and deliberation (UNECE, n.d.). This recommendation affirms the crucial role of the public and allows for wider scrutiny and accountability in AI-related decision-making processes.

## 8. Protect Collective Data and AI Rights

Revise rights frameworks that are impacted by AI systems and processes, such as intellectual property and data rights frameworks to 1) safeguard the data and knowledge sovereignty of Indigenous people and marginalised groups, including linguistic, religious and ethnic minorities; and 2) ensure the right to benefit from scientific progress. This recommendation contributes to a more equitable distribution of gains. Ensure strong institutional safeguards to protect social sector datasets, especially where there is a risk of proprietisation of core development functions through AI models (e.g. in health, education and welfare). Ensure that data and AI governance frameworks protect public innovation, with IP law reforms encouraging non-commercial AI ecosystems that contribute to social good and collective rights. The right of all individuals and communities to benefit from scientific progress must also be extended to AI ecosystems. This acquires special significance in view of the marginalisation of local cultures and knowledge systems in dominant data and AI models. Access and use regimes governing public datasets should further the goals of democracy, development and human rights, without eroding societal and environmental well-being (Gurumurthy and Deepti, 2023).

Possible pathways towards protection of collective data and AI rights are presented in the following paragraphs:

■ **Implement conditional access to public data**

Access to public domain and open government data should be conditional, with purpose limitations and clear sunset clauses on use. Establish robust institutional safeguards for social sector datasets, such as health, education and welfare, to ensure AI models uphold public service principles and protect marginalised communities (Tomasso Fia, 2021).

■ **Enforce fair use limitations**

Impose fair use limitations on how AI models learn from and utilise training data to prevent profiteering through reuse.

■ **Propose collective licensing**

Develop collective licensing proposals that balance the moral rights of creators with the value systems that consider the intellectual commons as public heritage.

■ **Ensure reciprocity in common data pools**

Establish reciprocity guarantees in common data pools, ensuring that private model developers who use public data layers are obligated to share back and enrich the commons.



People from Native American communities at the Chumash Day Pow Wow and Inter-tribal Gathering. Photo by Hanna Tor on iStock.

# Transparency and Accountability for Harm Prevention

Transparency and accountability are essential for harm prevention, redressing the democratic deficit and reversing misrecognition. They allow stakeholders to scrutinise AI systems and processes, detect biases, and hold private and public providers and deployers accountable for harmful or discriminatory impacts. They empower marginalised groups to challenge discriminatory practices and foster public trust by ensuring that decision-making processes are open and understandable. The following recommendations ensure the right to information, enhance algorithmic transparency, and establish clear accountability among all actors involved in the AI ecosystems through human rights impacts assessments and public procurement guidelines.

### 9. Establish the Right to Information in AI Systems and Enhance Algorithmic Transparency

Establish the right to information in AI. This right should grant individuals the right to access clear, accessible details on when AI is employed, what algorithms are used, what data are used for input, and what criteria are used in decision-making processes. This recommendation enhances requirements for algorithmic transparency and allows individuals negatively impacted by AI systems to challenge their outcomes (Grochowski et al., 2021; Kossow et al., 2021). The information provided must be made understandable for the general public. To prevent exploitation of intellectual property or trade secrets as a means to avoid accountability, transparency obligations should supersede such claims (Fink and Finck, 2022). Requiring enhanced algorithmic transparency also encourages further technological innovation to confront limitations such as behavioural opacity and enhance interpretability and explainability (Bommasani et al., 2024; Winfield et al., 2021).

Possible additional pathways towards enhancing the right to information in AI and algorithmic transparency include the following:

- Review and consider revising trade secret laws to establish clear limits and public interest exceptions to balance proprietary rights with transparency (Kilic, 2024).

- Consider establishing a new standard of post facto adequation for algorithmic transparency. Such a standard would require inferences from AI systems and processes in public functioning to provide clear and accessible information about the data, model and criteria used in decision-making processes. This standard ensures decisions are supported by recorded justifications, enhancing explainability and enabling public audits to mitigate harmful impacts (Mathews and Sinha, 2020).

- Consider a combination of model explanation (i.e. the coupling of an opaque AI system with an interpretable, transparent model that fully captures the logic of the opaque system), model inspection (i.e. a representation that makes it possible to understand some specific properties of an opaque model or its predictions) and outcome explanation (i.e. an account of the outcomes of an opaque AI in a particular instance).

- Consider prohibiting black-box models, especially in high-risk cases and public decision-making processes and supporting use of interpretable models (Rudin and Radin, 2019).

## 10. Enable and Conduct Obligatory Human Rights Impact Assessments (HRIAs)

Enable and conduct impact assessments by providing policy guidance on how to conduct them. Private- and public-sector providers and deployers are required to respect international human rights and principles, including through the application of human rights due diligence and impact assessments throughout the AI lifecycle (Global Digital Compact, 2024). The assessments should evaluate whether risks of harm are acceptable under fundamental rights law and include clear duties to eliminate or prevent such risks (Jakubowska et al., 2024). The assessments must also consider and compare possible non-technological approaches to identify the least intrusive measures to human rights. If the assessment reveals serious concerns, deployers should refrain from use.

Possible pathways towards obligatory HRIAs for AI systems include the following:

- Consider requiring comprehensive ex ante impact assessments with an explicit focus on gender and diversity.

- Consider mandating that the results of such impact assessments must be submitted to a regulatory body or oversight agency for review and approval before the AI system in question can be deployed. Review processes should follow appropriate standards of representation to effectively address substantive equality considerations.

- Consider requiring the prohibition of certain AI systems and processes if potential or actual impacts are not justified under international human rights law (McGregor and Molnar, 2023).

- Consider establishing penalties or sanctions for organisations that fail to conduct the required impact assessments.

- Consider establishing a legal presumption that the deployment of an AI system without a comprehensive impact assessment constitutes prima facie evidence of discrimination.

- Consider establishing mechanisms for periodic review and re-evaluation of risks.

## 11. Develop Accountability Measures for Public-Sector Algorithmic Systems and Processes

Develop AI-specific public procurement guidelines to protect human rights and due process, addressing complexities and risks introduced by algorithmic and AI systems and processes (Hickok, 2022). Promote open data initiatives to build open libraries of algorithms used in public-sector systems. Ensure that policy makers undergo capacity-building so they can effectively conduct due diligence in AI procurement. Consider establishing mechanisms for moratoriums or bans on Automated Information Systems (AIS) and AI systems and processes that may impact human rights.



Millennial Indian woman explains to colleague details of collaborative project sit together at desk share ideas, talking, making joint task. Photo by fizkes on iStock.

## Effective Access to Justice

Ensuring effective access to justice is seminal for achieving substantive equality. It requires accountability, reverses misrecognition and upholds the right to dignity. By strengthening contextual liability, empowering equality bodies and easing the burden of proof, policy makers can improve access to justice for individuals and groups facing discrimination and harms in relation to AI systems and processes.

### 12. Strengthen Contextual Liability for Non-Discrimination in AI Systems

Strengthen contextual liability for non-discrimination in AI systems in proportion to other accountability measures such as level of transparency, interpretability, and explainability. Product and fault liability regulations require revision to accurately reflect the complexities of AI systems and data-driven decision-making. Effective accountability in AI development and deployment takes into account specific characteristics such as opacity, explainability, autonomous behaviour, continuous adaptation and limited predictability. Chart a path towards liability in AI to ensure appropriate accountability among public and private providers and deployers.

Potential pathways include:

- Introducing Terms of Use and encouraging public regulatory sandboxes for AI systems in uncharted territories. This will establish clear boundaries for safe use and reduce harm while enabling innovation and iterative testing.

- Considering a mixed liability framework for AI system providers and deployers that balances fault-based and strict liability in relation to other accountability measures.

- Evaluating software-focused liability regulations (EPRS, 2024).

- Requiring public and private providers and deployers to remove any discovered harm and discriminatory impact stemming from processes associated with the development and deployment of the systems in given uses.

- Compensating for identified (material and non-material) damages suffered due to algorithmic discrimination.

### 13. Empower Equality Bodies to Initiate Action

Empower equality bodies, including national human rights institutions and other public interest organisations, to take action in the public interest. Allow these bodies to submit complaints to supervisory authorities even without identifiable complainants (Equinet, 2023).[18] Ease the burden of proof and equip these bodies with the legal authority and necessary training to effectively address discrimination and harms caused by AI systems and related processes.

### 14. Ease the Burden of Proof for Claimants

Review and revise evidence rules to ease the burden of proof for claimants (World Commission on the Ethics of Scientific Knowledge and Technology, 2005). Existing product liability rules often require harmed parties to demonstrate the causal link between product faults and specific damages. Given the complexity and limited predictability of AI systems, this requirement is challenging. Consider adjusting these rules to make it easier for claimants to prove their cases and claim compensation.

**Easing the burden of proof for discrimination and damages in AI – Insights from the proposed AI Liability Directive, EU**

- **Introduction of a "presumption of causality"**

  The proposed EU AI Liability Directive introduces a presumption of causality, assuming a causal link between non-compliance with a duty of care under EU or national law (the fault) and the output or lack of output from the AI system that caused the damage.

  This principle balances the interests of claimants and defendants in product liability disputes without increasing liability risks that could hinder innovation and the adoption of AI-enabled products and services. The onus is on AI system providers, operators and users to demonstrate the validity and fairness of their systems, reducing the burden on harmed parties to prove the connection between the AI system's operations and harmful outcomes (European Commission, 2022; European Parliament, 2023).

- **Granting national courts authority to order evidence disclosure (concerning high-risk AI systems suspected of causing harm)**

  Individuals harmed by AI systems face challenges in meeting liability claim requirements due to the complexity of these systems. Granting national courts the authority to order the disclosure of evidence concerning high-risk AI systems suspected of causing harm can enhance accountability (World Commission on the Ethics of Scientific Knowledge and Technology, 2005). This aligns with the precautionary principle, provided courts follow necessity and proportionality principles and avoid blanket requests.



Silhouette of a person with a raised fist, set against a vibrant sunset gradient. Photo by Miguel Bruna on Unsplash.

## Conclusions and Exploring Paths Forward

The key recommendations for transformative AI policy presented in this report provide a comprehensive pathway for achieving substantive equality within AI ecosystems and related policy making. Categorised to reflect key dimensions of the Transformative AI Policy Framework, the recommendations focus on four categories: inclusive design and democratic innovation, meaningful participation in AI governance, transparency and accountability for harm prevention, and effective access to justice. Together, these pillars create a robust foundation for AI ecosystems and related policy making that advances substantive equality.

The successful implementation of these recommendations will require intentional investments, budget allocations, capacity building, stakeholder collaboration and regulatory innovations to navigate the obstacles and complexities of rapid AI development. Most importantly, it will require an ongoing commitment to address technological, legal, financial or other barriers and to actively elevate and empower typically excluded voices to achieve transformative change for the benefit of all.

Exploring paths forward, continued investment in interdisciplinary, intersectoral and multistakeholder research, and collaboration will contribute to bringing about more inclusive, equitable and transformative AI systems and processes. Advancing our understanding of how AI systems and processes impact other rights frameworks, expanding collective perspectives on, for example, privacy rights, the right to freedom of expression, education and health, or environmental and labour rights, and continued research on data commons and public interest approaches will also inform future applications of transformative policies. Ongoing iterative testing and innovation in safe, secure regulatory environments will allow for increased learning and development of effective policy frameworks. Moreover, national and international collaboration across sectors and regions will be necessary to adapt policies to diverse local contexts, ensuring that AI systems and governance are not only globally equitable but also sensitive to regional variations in legal, cultural and economic environments. Investing in policy research to further evaluate the impact of policies, laws and regulations in the rapidly evolving AI environment will allow policy makers to further improve and attune transformative policies and practices.



Detailed black and white image of interlocking hands, highlighting a bracelet on one and a ring on the other. Photo by Aarón Blanco Tejedor on Unsplash.

Ultimately, if these recommendations are implemented with a commitment to context-sensitive approaches and ongoing innovation, they will lead us towards a future where the benefits of AI systems and processes are equitably shared, ensuring a more gender equal, diverse, and inclusive world for all.

# Endnotes

1    In this report, we employ the term SOGIESC as inclusive of all sexual orientations, gender identities, gender expressions and sex characteristics, including intersex traits. Some also use SOGI (sexual orientation, gender identity) or SOGIE (sexual orientation, gender identity and gender expression). The report recognises that the acronym refers to all humans with sexual orientations and gender identities, including cisgender and straight people. The term LGBTQI+, however, specifically emphasises specific people with marginalised identities (e.g. transgender, nonbinary, esbian, etc.). For more information, see Office of the High Commissioner for Human Rights (2024a) and Smith (2023).

2    People on the move include refugees, asylum seekers, migrants and internally displaced persons (IDPs).

3    See, for example, the Women for Ethical AI Action Request calling on all countries to join UNESCO's activities to support the mission to drive systematic actions for gender-inclusive AI (UNESCO, 2024b).

4    Equity recognises that each person has different circumstances, and allocates the exact resources and opportunities needed to reach an equal outcome. See International Women's Day (2023).

5    The report adopts a non-binary approach to gender equality and refers to barriers faced not only by women, but also by persons with diverse sexual orientation, gender identities and expressions, and sexual characteristics (SOGIESC).

6    Representation from ten countries in Africa (Burkina Faso, Cameroon, Kenya, Mauritius, Nigeria, South Africa, Swaziland, Tunisia, Uganda and Zimbabwe); ten countries in Latin America (Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Mexico, Paraguay, Peru and Uruguay); ten countries in North America and Europe (Belgium, Canada, Denmark, France, Germany, Hungary, Portugal, Romania, United Kingdom and USA); nine countries in the MENA region (Egypt, Iran, Israel, Lebanon, Mauritania, Morocco, Saudi Arabia, Tunisia and the United Arab Emirates); eleven countries in the Asia-Pacific region (Australia, India, Japan, Malaysia, Nepal, New Zealand, Philippines, Singapore, South Korea, Thailand, and Vietnam).

7    Equity recognises that each person has different circumstances, and allocates the exact resources and opportunities needed to reach an equal outcome.

8    The report adopts a non-binary approach to gender equality and refers to barriers faced not only by women, but by women and persons with diverse SOGIESC.

9    This principle is enshrined in various international human rights instruments, such as the Universal Declaration of Human Rights and the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW) (Office of the High Commissioner for Human Rights, 1979).

10   See CEDAW (UN Women, 2009), Art. 2(3) and 7, and Office of the High Commissioner for Human Rights (2006) for examples.

11  The descriptions in this section draw on the OECD's visualisation of the AI system lifecycle and are further informed by UNESCO's description of the AI lifecycle and its various stages, from research and inception to disassembly and termination (UNESCO, 2024c, p. 10). However, it places additional emphasis on data collection, data preprocessing and statistical modelling (Ovalle et al., 2023b). The key dimensions of the OECD Framework for the Classification of AI Systems further inform the description of social systems (OECD, 2022).

12  The framework builds on existing AI frameworks and goes a step further to assert the state's duty to actively eliminate discrimination and address structural injustice.

13  The motto "Nothing About Us Without Us" relies on this principle of participation, and it has been used by Disabled Peoples Organizations throughout the years as part of the global movement to achieve the full participation and equalization of opportunities for, by and with persons with disabilities. https://www.un.org/esa/socdev/enable/iddp2004.htm.

14  In accordance with the recommendations issued by the Special Rapporteur on the Rights of Persons with Disabilities.

15  In accordance with the recommendations issued by the Special Rapporteur on the Rights of Persons with Disabilities.

16  See, for example, Spain's initiative to develop a foundational AI model trained in Spanish and co-official languages, in collaboration with institutions such as the Barcelona Supercomputing Center and the Spanish Academy of Language, which exemplifies transparency and accessibility (Computerworld Spain, 2024).

17  OECD Principle 1.3 on Transparency and Explainability underscores the need to address transparency concerns by promoting public discourse and establishing dedicated entities, if necessary, to enhance awareness and understanding of AI systems. This fosters acceptance of and trust in AI technologies.

18  The legal power to take notice of a matter without receiving a formal complaint.

# References

ACM (2022), Statement on Principles for Responsible Algorithmic Systems, ACM Technology Policy Office, Washington, DC, https://www.acm.org/binaries/content/assets/public-policy/final-joint-ai-statement-update.pdf.

Badaloni, S. and A. Rodà (2022), "Gender knowledge and Artificial Intelligence", paper presented at the 1st Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming, BEWARE-22, co-located with AIxIA 2022, University of Udine, Udine, Italy.

Bartoletti, I. and R. Xenidis (2023), Study on the Impact of Artificial Intelligence Systems, their Potential for Promoting Equality, Including Gender Equality, and the Risks They May Cause in Relation to Non-discrimination, Gender Equality Commission (GEC) and the Steering Committee on Anti-Discrimination, Diversity and Inclusion (CDADI), Council of Europe, Strasbourg, https://rm.coe.int/prems-112923-gbr-2530-etude-sur-l-impact-de-ai-web-a5-1-2788-3289-7544/1680ac7936.

Benjamin, R. (ed.) (2019), *Captivating technology: Race, carceral technoscience, and liberatory imagination in everyday life*, Duke University Press, Durham, NC.

Bommasani, R. et al. (2024), "The Foundation Model Transparency Index v1.1", https://crfm.stanford.edu/fmti/paper.pdf.

Buckup, S. (2009), *The price of exclusion: The economic consequences of excluding people with disabilities from the world of work*, Employment Working Paper No. 43, ILO, https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcm s_119305.pdf.

Buhl, N. (2023), Human-in-the-Loop Machine Learning (HITL) Explained, Encord Blog, https://encord.com/blog/human-in-the-loop-ai/.

Buolamwini, J. (2023), *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*, Random House, New York.

Buolamwini, J. and T. Gebru (2018), "Gender shades: Intersectional accuracy disparities in commercial gender classification", in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Association for Computing Machinery, New York, https://proceedings.mlr.press/v81/buolamwini18a.html.

Cachat-Rosset, G. and A. Klarsfeld (2023), "Diversity, equity, and inclusion in Artificial Intelligence: An evaluation of guidelines", *Applied Artificial Intelligence*, Vol. 37(1), article 2176618, https://doi.org/10.1080/08839514.2023.2176618.

Campbell, M. (2015), "CEDAW and women's intersecting identities: A pioneering new approach to intersectional discrimination", *Revista Direito GV*, Vol. 11(2), pp. 479–504, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2726534.

Chauhan, P. and N. Kshetri (2022), "The role of data and artificial intelligence in driving diversity, equity, and inclusion", *Computer*, Vol. 55, pp. 88–93, https://ieeexplore.ieee.org/abstract/document/9755238.

Chen, B.J. and J. Metcalf (2024), *Explainer: A Sociotechnical Approach to AI Policy*, Data & Society Policy Brief, https://datasociety.net/wp-content/uploads/2024/05/DS_Sociotechnical-Approach_to_AI_Policy.pdf.

Cheong, M.C.Y. K. Leins and S. Coghlan (2021), "Computer science communities: Who is speaking, and who is listening to the women? Using an ethics of care to promote diverse voices", in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, https://api.semanticscholar.org/CorpusID:231639365.

Cofone, I. (2023), The Privacy Fallacy: Harm and Power in the Information Economy, *Cambridge University Press, Cambridge*, UK, https://www.cambridge.org/core/books/privacy-fallacy/547578F2A1AE0C40963105CE-066B412E.

Computerworld Spain (2024), "Spain will create foundational AI model in local languages", *Computerworld*, https://www.computerworld.com/article/1612218/spain-will-create-foundational-ai-model-in-local-languages.html

Costanza-Chock, S. (2018), "Design justice, A.I., and escape from the matrix of domination", *Journal of Design and Science*, https://doi.org/10.21428/96c8d426.

Dankwa-Mullan, I. et al. (2021), "A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle", *Journal of Health Care for the Poor and Underserved*, Vol. 32, pp. 300–317, https://muse.jhu.edu/article/789672.

Deck, L. et al. (2024), "Implications of the AI Act for Non-Discrimination Law and Algorithmic Fairness", https://arxiv.org/html/2403.20089v1.

Dignum, V. (2023)." Responsible artificial intelligence: Recommendations and lessons learned", in D. O. Eke, K. Wakunuma, and S. Akintoye (eds.), *Responsible AI in Africa: Challenges and Opportunities*, Springer International Publishing, Cham, Switzerland, https://doi.org/10.1007/978-3-031-08215-3_9.

Donnelly, N. and L. Stapleton (2022), "The social impact of data processing: The case of gender mapped to sex", *IFAC-PapersOnLine*, Vol. 55(39), pp. 117–122, https://doi.org/10.1016/j.ifacol.2022.12.021.

Dumbrava, C. (2021), "Artificial intelligence at EU borders: Overview of applications and key issues", In-Depth Analysis, Think Tank, European Community, https://www.europarl.europa.eu/thinktank/en/document/EPRS_IDA(2021)690706.

Dwork, C. et al. (2012), "Fairness through awareness", in *ITCS '12*: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, https://doi.org/10.1145/2090236.2090255.

Equinet (2023), Ensuring European AI that Protects and Promotes Equality for All, https://equineteurope.org/wp-content/uploads/2023/07/Recommendations-amendments.pdf.

Eubanks, V. (2017), Automating Inequality: *How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York.

European Commission (2022), *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, European Commission, Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0496.

European Parliament (2023), *Artificial intelligence liability directive*, Briefing, EU Legislation in Progress, https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf.

Fink, M. and M. Finck (2022), "Reasoned A(I)dministration: explanation requirements in EU law and the automation of public administration", *European Law Review*, Vol. 47(3), pp. 376–392.

Forster, M. (2022), *Refugee Protection in the Artificial Intelligence Era. A Test Case for Rights*, Chatham House, London.

Fredman, S. (2016), "Substantive equality revisited", *International Journal of Constitutional Law*, Vol. 14(3), pp. 712–738, https://doi.org/10.1093/icon/mow043.

Fricker, M. (2007), *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford Academic, Oxford.

Goldschmidt, J.E. (2017), "New perspectives on equality: Towards transformative justice through the Disability Convention?", *Nordic Journal of Human Rights*, Vol. 35(1), pp. 1–14, https://doi.org/10.1080/18918131.2017.1286131.

Global Center on AI Governance (2024), Methodology: Our Multidimensional Framework, *Global Index on AI*, https://www.global-index.ai/methodology

Grobelnik, M. et al. (2024), "What is AI? Can you make a clear distinction between AI and non-AI systems?", *The AI Wonk*, OECD.AI Policy Observatory, https://oecd.ai/en/wonk/definition.

Grochowski, M. et al. (2021), "Algorithmic transparency and explainability for EU consumer protection: Unwrapping the regulatory premises", *Critical Analysis of Law*, Vol. 8(1), pp. 43–63, https://doi.org/10.33137/cal.v8i1.36279.

Gurumurthy, A. and B. Deepti (2023), "Reframing AI governance through a political economy lens", IT for Change, https://itforchange.net/sites/default/files/add/IT%20for%20Change_Reframing%20AI%20Governance%20through%20a%20Political%20Economy%20Lens.pdf.

Hacker, P. (2024), Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence: Complementary Impact Assessment, *European Parliamentary Research Service*, https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf.

Hellman, D. (2023), "Big data and compounding injustice", *Journal of Moral Philosophy*, Vol. 21(1-2), pp. 62–83, https://doi.org/10.1163/17455243-20234373.

Hickok, M. (2024), "Public procurement of artificial intelligence systems: New risks and future proofing", *AI & Society*, Vol. 39, pp. 1213–1227, https://doi.org/10.1007/s00146-022-01572-2.

Hull, G. (2023), "Dirty data labeled dirt cheap: Epistemic injustice in machine learning systems", *Ethics and Information Technology*, Vol. 25(3), article 38, https://doi.org/10.1007/s10676-023-09712-y.

Iason, G. (2022), "Toward a theory of justice for artificial intelligence", *Daedalus*, Vol. 151(2), pp. 218–231, https://doi.org/10.1162/daed_a_01911

International Women's Day (2023), "#EmbraceEquity for IWD 2023 and beyond?", https://www.internationalwomensday.com/Missions/18707/Equality-versus-Equity-What-s-the-difference-as-we-EmbraceEquity-for-IWD-2023-and-beyond.

IWRAW (n.d.), Substantive equality, https://cedaw.iwraw-ap.org/cedaw/cedaw-principles/cedaw-principles-overview/substantive-equality/.

Jakubowska, E. et al. (2024), "EU's AI Act fails to set gold standard for human rights", Joint analysis, https://www.amnesty.eu/wp-content/uploads/2024/04/EUs-AI-Act-fails-to-set-gold-standard-for-human-rights.pdf.

Jora, R.B. et al. (2022), "Role of Artificial Intelligence (AI) in meeting Diversity, Equality and Inclusion (DEI) goals", in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1.

Joyce, K. et al. (2021), "Toward a sociology of artificial intelligence: A call for research on inequalities and structural change", *Socius*, Vol. 7, article 2378023121999581, https://doi.org/10.1177/2378023121999581.

Kilic, B. (2024), Into Uncharted Waters: Trade Secrets Law in the AI Era, CIGI Papers No. 295, *Centre for International Governance Innovation*, https://www.cigionline.org/static/documents/no.295.pdf.

Kong, Y. (2022), "Are 'intersectionally fair' AI algorithms really fair to women of color? A philosophical analysis", in *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, https://doi.org/10.1145/3531146.3533114.

Kossow, N., S. Windwehr and M. Jenkins (2021), "Algorithmic transparency and accountability", Transparency International Anti-Corruption Helpdesk Answer, Transparency International, https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf.

Krishna, V.K. (2020), "Open science and its enemies: Challenges for a sustainable science–society social contract", *Journal of Open Innovation: Technology, Market, and Complexity*, Vol. 6(3), article 61, https://doi.org/10.3390/joitmc6030061.

Martineau, W., N. Meer and S. Thompson (2012), "Theory and practice in the politics of recognition and misrecognition", *Res Publica*, Vol. 18, pp. 1–9, https://doi.org/10.1007/s11158-012-9181-7.

Mathews, H.V. and A. Sinha (2020), "Use of Algorithmic Techniques for Law Enforcement", *Economic and Political Weekly*, Vol. 55(23), https://www.epw.in/journal/2020/23/special-articles/use-algorithmic-techniques-law-enforcement.html.

McGregor, M. and P. Molnar (2023), *Digital Border Governance: A Human Rights Based Approach*, University of Essex and UN High Commissioner for Human Rights, https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf.

Molnar, P. (2024), "The deadly digital frontiers at the border", *Time*, 21 May, https://time.com/6979557/unregulated-border-technology-migration-essay/.

Nihei, M. (2022), "Epistemic injustice as a philosophical conception for considering fairness and diversity in human-centered AI principles", *Interdisciplinary Information Sciences*, Vol. 28(1), pp. 35–43, https://doi.org/10.4036/iis.2022.A.01.

Noble, S. and S. Roberts (2019), "Technological elites, the meritocracy, and postracial myths in Silicon Valley", in *RACISM POSTRACE,* R. Mukherjee, S. Banet-Weiser and H. Gray (eds.), Racism Postrace, Duke University Press, Durham, NC. Report #: 6. Retrieved from https://escholarship.org/uc/item/7z3629nh.

OECD (2021), Open government and citizen participation, https://www.oecd.org/en/topics/open-government-and-citizen-participation.html.

OECD (2022), The OECD Framework for the Classification of AI Systems, OECD.AI Policy Observatory, https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/02/oecd-framework-for-the-classification-of-ai-systems_336a8b57/cb6d9eca-en.pdf.

OECD (2024a), OECD AI Principles overview, OECD.AI Policy Observatory, https://oecd.ai/en/ai-principles.

OECD (2024b), "Explanatory memorandum on the updated OECD definition of an AI system", OECD *Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en.

Office of the High Commissioner for Human Rights (1965), I*nternational Convention on the Elimination of All Forms of Racial Discrimination*, https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-elimination-all-forms-racial.

Office of the High Commissioner for Human Rights (1966), *International Covenant on Economic, Social and Cultural Rights*, https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights.

Office of the High Commissioner for Human Rights (1979), *Convention on the Elimination of All Forms of Discrimination against Women New York*, 18 December 1979, https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women.

Office of the High Commissioner for Human Rights (2023), *Digital Border Governance: A Human Rights Based Approach*, https://www.ohchr.org/sites/default/files/2023-09/Digital-Border-Governance-A-Human-Rights-Based-Approach.pdf.

Office of the High Commissioner for Human Rights (2024a), LGBTI people, https://www.ohchr.org/en/topic/lgbti-people?gad_source=1.

Office of the High Commissioner for Human Rights (2024b), The Core International Human Rights Instruments and their monitoring bodies, https://www.ohchr.org/en/core-international-human-rights-instruments-and-their-monitoring-bodies.

Ostry, J.D. et al. (2018), "Economic gains from gender inclusion: New mechanisms, new evidence", IMF Staff Discussion Note, IMF, https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2018/10/09/Economic-Gains-From-Gender-Inclusion-New-Mechanisms-New-Evidence-45543

Ovalle, A. D. Liang and A. Boyd (2023a), "Should they? Mobile biometrics and technopolicy meet queer community considerations", in *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms,* and with https://dl.acm.org/doi/abs/10.1145/3617694.3623255.

Ovalle, A. et al. (2023b), "Factoring the matrix of domination: A critical review and reimagination of intersectionality in AI fairness", in *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, https://doi.org/10.1145/3600211.36047.

Pande, R (2003). "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India.", *American Economic Review*, Vol. 93 (4), https://www.aeaweb.org/articles?id=10.1257/000282803769206232.

#ProtectNotSurveil coalition (2024), Joint statement – A dangerous precedent: How the EU AI Act fails migrants and people on the move, https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/.

Ricaurte, P. (2024), "A sociotechnical approach to the AI Lifecycle", *Feminist AI Research Network*, Forthcoming.

Ricaurte, P. (2019), "Data epistemologies, the coloniality of power, and resistance", *Television & New Media*, Vol. 20(4), pp. 350–365, https://doi.org/10.1177/1527476419831640.

Ricaurte, P. (2022). Ethics for the majority world: AI and the question of violence at scale. Media, Culture & Society, 44(4), 726-745. https://doi.org/10.1177/01634437221099612.

Ricaurte, P. and M. Zasso (2023), "AI, ethics, and coloniality: A feminist critique", in M. Cebral-Loureda, E.G. Rincón-Flores and G. Sanchez-Ante (eds.), *What AI Can Do*, Chapman and Hall/CRC, New York.

Rikap, C. (2022), "The expansionary strategies of intellectual monopolies: Google and the digitalization of healthcare", *Economy and Society*, Vol. 52(1), pp. 110–136, https://doi.org/10.1080/03085147.2022.2131271.

Rudin, C. and J. Radin (2019), "Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition", *Harvard Data Science Review*, Vol. 1, https://doi.org/10.1162/99608f92.5a8a3a3d.

Shapiro, J. (n.d.), "CEDAW as a tool for promoting substantive gender equality," http://www.cedaw.org.tw/en/upload/media/Capacity%20Building/1-1The%20Mechanism%20of%20CEDAW%20Committee.pdf.

Smith, R.A. (2023), "From LGBTQIA+ to SOGIESC: Reframing sexuality, gender, and human rights", Open Global Rights, New York.

Smuha, N.A. (2021), "Beyond the individual: governing AI's societal harm", *Internet Policy Review*, Vol. 10(3), https://doi.org/10.14763/2021.3.1574.

Solaiman, I. et al. (forthcoming), "Evaluating the social impact of generative AI systems in systems and society", in P. Hacker, A. Engel, S. Hammer and B. Mittelstadt (eds.), *Oxford Handbook on the Foundations and Regulation of Generative AI*, Oxford University Press, Oxford.

Stauffer, B. (2023), Automated Neglect: How The World Bank's Push to Allocate Cash Assistance Using Algorithms Threatens Rights, Human Rights Watch, https://www.hrw.org/report/2023/06/13/automated-neglect/how-world-banks-push-allocate-cash-assistance-using-algorithms.

Suresh, H. et al. (2022), "Towards intersectional feminist and participatory ML: A case study in supporting feminicide counterdata collection", in *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, https://api.semanticscholar.org/CorpusID:249872586.

Türk, V. (2023), "Artificial intelligence must be grounded in human rights, says High Commissioner", Office of the High Commissioner for Human Rights, Geneva.

UNCTAD (2021), *Technology and Innovation Report 2021*, https://unctad.org/page/technology-and-innovation-report-2021.

UNECE (n.d.), Introduction, https://unece.org/environment-policy/public-participation/aarhus-convention/introduction.

UNECE (1998), *UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters*, https://unece.org/environment-policy/public-participation/aarhus-convention/text.

UNESCO (2009), "Concept paper on marginalization", Paper presented at the Tenth meeting of the Working Group on Education for All (EFA), Paris.

UNESCO (2022), *Recommendation on the Ethics of Artificial Intelligence*, UNESCO, Paris, https://unesdoc.unesco.org/ark:/48223/pf0000381137.

UNESCO (2024a), "Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes", https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes.

UNESCO (2024b), Statement W4EAI, Women for Ethical AI Action Request, SHS/2024/PI/H/1, https://unesdoc.unesco.org/ark:/48223/pf0000388754/PDF/388754eng.pdf.multi.

UNESCO (2024c), Ethics of Artificial Intelligence: The Recommendation, https://unesdoc.unesco.org/ark:/48223/pf0000381137.

UNESCO (2024d), Ethical Impact Assessment, Global AI Ethics and Governance Observatory, https://www.unesco.org/ethics-ai/en/eia.

UNESCO, IDB and OECD (2022), *The Effects of AI on the Working Lives of Women*, UNESCO, Paris.

UNHRC (2011), *Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework,* A/HRC/17/31, https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.

UNHRC (2020), *Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*, A/HRC/44/57, http://undocs.org/en/A/HRC/44/57.

UNHRC (2021), *Report of the Special Rapporteur on the Rights of Persons with Disabilities*, A/HRC/49/52, http://docs.org/en/A/HRC/49/52.

UNICEF (2021), *UNICEF Gender Policy and Action Plan 2022-2025: Gender-Transformative Programming*, https://www.unicef.org/lac/en/media/43146/file.

United Nations, (2006), *Convention on the Rights of Persons with Disabilities and Optional Protocol*, https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf.

United Nations (2019), *United Nations Disability Inclusion Strategy*, United Nations, New York, https://www.un.org/en/content/disabilitystrategy/assets/documentation/UN_Disability_Inclusion_Strategy_english.pdf.

United Nations (2021), *Report of the Secretary-General's Task Force on Addressing Racism and Promoting Dignity for All in the United Nations Secretariat*, United Nations, New York, https://hr.un.org/sites/hr.un.org/files/sap_final_report_0.pdf.

United Nations (2024), The 17 Goals, https://sdgs.un.org/goals.

United Nations (2024b), *Global Digital Compact*, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf.

UNRISD (2017), Transformative Policies for Sustainable Development: What Does It Take?, https://www.unrisd.org/en/library/publications/transformative-policies-for-sustainable-development-what-does-it-take.

UN Women (2009), *Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)*, https://www.un.org/womenwatch/daw/cedaw/text/econvention.htm#intro.

UN Women (2015), "Chapter 1: Substantive equality for women: The challenge for public ;olicy", in *Progress of the World's Women 2015–2016*, UN Women, https://progress.unwomen.org/en/2015/pdf/ch1.pdf.

UN Women (2019), *CEDAW-based Legal Review: A Brief Guide*, UN Women in Ukraine, Kyiv, https://eca.un-women.org/sites/default/files/Field%20Office%20ECA/Attachments/Publications/2019/CEDAW%20Legal_ENG_compressed%20%281%29.pdf.

Vinuesa, R. et al. (2020), "The role of artificial intelligence in achieving the Sustainable Development Goals", *Nature Communications*, Vol. 11(1), article 1, https://doi.org/10.1038/s41467-019-14108-y.

West, S.M. (2020), "Redistribution and recognition: A feminist critique of algorithmic fairness", *Catalyst: Feminism, Theory, Technoscience*, Vol. 6(2), https://doi.org/10.28968/cftt.v6i2.33043.

West, S.M., M. Whitaker and K. Crawford (2019), *Discriminating Systems: Gender, Race, and Power in AI*, AI Now Institute, https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2.

White House (2022), *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*, White House Office of Science and Technology Policy, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

Winfield, A.F.T. et al. (2021), "IEEE P7001: A Proposed Standard on Transparency", *Frontiers in Robotics and AI*, Vol. 8, article 665729, https://doi.org/10.3389/frobt.2021.665729.

World Commission on the Ethics of Scientific Knowledge and Technology (2005), *The Precautionary Principle*, UNESCO, Paris, https://unesdoc.unesco.org/ark:/48223/pf0000139578.

World Economic Forum (2024), *AI Value Alignment: Guiding Artificial Intelligence Towards Shared Human Goals*, White Paper, World Economic Forum, https://www3.weforum.org/docs/WEF_AI_Value_Alignment_2024.pdf.

# Appendix: List of Regional Consultation Participants

**This is a list of all regional consultation participants who have given their consent to be named in this report.**

**Regional consultations for the following regions were led by Data Pop Alliance.**

## Asia

1.  **Apekchya Rana Khatri,**
    UNFPA Nepal

2.  **Gaurav Sharma,**
    GIZ India

3.  **Dr. Preeti Raghunath,**
    University of Sheffield

4.  **Philippe Doneys,**
    Asian Institute of Technology

5.  **Hazel T. Biana,**
    Research Fellow, Social Development
    Research Center, De La Salle University

6.  **Manila Hoang,**
    Vietnam Foreign Ministry

7.  **Eleonore Fournier-Tombs,**
    United Nations University

8.  **Omran Najjar,**
    Humanitarian OSM StreetMap Team

## Middle East and North Africa (MENA)

1.  **Mohammad Reza Bateni,**
    Amirkabir University of Technology

2.  **Hayfa Khalfaoui**

3.  **Maha Jouini,**
    African Center for AI and Digital Technology

4.  **Sukaina Al-Nasrawi, United Nations Economic**
    and Social Commission for Western Asia
    (UNESCWA)

5.  **Fouad Mrad**

6.  **Grace S. Thomson,**
    Center for AI and Digital Policy (CAIDP)

7.  **Ibrahim Helmy Emara,**
    Tanta University

8.  **Dr. Houda Chihi,**
    Tunisia Telecom

9.  **Limor Ziv,**
    Humane AI

10. **Prof. Hazem Abdelazim,**
    ESLSCA University

11. **Arwa ElGhadban,**
    Technical University of Munich

12. **Ihab Shoully**

## North America and Europe (NE)

1.  **Mariana Rozo Paz,**
    Datasphere Initiative

2.  **Yavar Baradaran-Khosrhoshahi**

3.  **Till Koebe,**
    Saarland Informatics Campus

4.  **H. Gijs van den Dool,**
    Freelancer / Independent Researcher

5.  **Luisa Olaya,**
    GIZ (Fair Forward)

6.  **Ruth Schmidt,**
    GIZ (Fair Forward)

**Regional consultations for the following region was led by Derechos Digiatales.**

## Latin America and Caribbean (LAC)

1.  Maria Encalada Córdova
2.  Tarcizio Roberto da Silva
3.  Luz Elena González
4.  Edson Prestes
5.  Marianela Milanes
6.  Silvana Bahia
7.  Gisele da Silva Craveiro
8.  Eduardo Carrillo
9.  Lutiana Valadares Fernandes Barbosa
10. Zinnya del Villar Islas
11. Cristina Pombo
12. Viviane Ceolin Dallasta Del Grossi
13. Danielle Costa Carrara Couto.
14. Luis Fernando Arias
15. Silvana Fumega
16. Valentina Rozo Ángel
17. Paz Peña
18. Daniela Camila de Araújo
19. Fernanda K. Martins
20. Virginia Pardo
21. Hugo Alvaro Miranda Colque
22. Virginia Brussa
23. Paola Ricaurte Quijano
24. Fernanda Campagnucci
25. Gustavo Macedo
26. Maximiliano Maneiro Vaz
27. Andres Lombana Bermudez
28. Ivana Feldfeber
29. Augusto Luciano Mathurin
30. Iván Terceros
31. Eliana Quiroz
32. María Lorena Flórez Rojas
33. Natali Viezna Figueroa Figueroa
34. Juan de Brigard
35. Juan Pablo Marín Díaz
36. Marcela Mattiuzzo
37. Thiane de Nazaré

**Regional consultations for the following region was led by Paola Ricaurte Quijano.**

## Latin America and Caribbean (LAC)

1. Gerardo López Gómez
2. José Alfredo Hau Caamal
3. Verónica Aguilar Martínez
4. Jhonnatan Rangel
5. Marcos Cornelio Sánchez Ramírez
6. Miriam Hernández
7. Kupijy Vargas
8. Manuela Yah Panti
9. Yamili Chan Canul
10. Andrés Ta Chikinib
11. Mariana Díaz

**Regional consultations for the following region was led by ICT Research Africa.**

## Sub-Saharan Africa (SSA)

1. Sandra Aceng
2. Adedeji Adeniran
3. Adebisi Adewusi
4. Shariffa Amolo
5. Houda ChiChi
6. Kudakwashe Dandajena
7. Laetitia Delaunay-Badolo
8. Khetsiwe Dlamini
9. Sinclair Ebimowei
10. Alison Gillwald
12. Rigobert Kenmogne
13. Jonas Kgomo
14. Slindile Khumalo
15. Cheruiyot Kiprono
16. Anja Kovacs
17. Michael Magoronga
18. Lorraine Matanda
19. Tholang Mathapo
20. Bhekani Mbuli
21. Katleho Mokoena
22. Margaret Muga
23. Judith Murungi
24. Ernest Mwebaze
25. Angella Ndaka
26. Nelson Nkari
27. Adaeze P. Nwoba
28. Dorcas Nyamwaya
29. Durnebi Okwasah
29. Titilola Helen Olojede
30. Eniola Olowu
31. Lavina Ramkisoon
32. Kelly Stone
33. Oluwatoyin Taiwo
34. Cynthia Tapera
35. Moses Thiga
36. Israel Olatunji Tijani
37. Simone Toussi
38. Josephine Zingani
39. Mpho Moyo
40. Pierrine Leukes