Advancing Global AI Policy Through the Lens of Gender Equity

Vaisnavi Nemala

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
vnemala@andrew.cmu.edu

Abstract

Current artificial intelligence (AI) regulatory frameworks have made significant progress in defining and developing risk-based measures to promote the transparency and accountability of AI systems at both national and global levels. However, the critical issue of gender-based bias and discrimination prevalent in AI requires special attention and mitigation strategies. This paper examines existing AI governance frameworks from a gender-transformative perspective, emphasizing the need for comprehensive global strategies to address these challenges and reshape the current AI ecosystem.

Introduction

As AI continues to evolve, it increasingly influences traditional societal decision-making processes. From hiring algorithms to policing decisions (Busuioc 2021), AI-driven systems have become integral to determining access to employment, allocation of public resources, and decisions with profound individual impacts. Harnessed responsibly, AI holds immense potential to transform systems and contribute to building a fairer, more equitable society. However, fulfilling this promise requires careful consideration of biases that can arise throughout the AI lifecycle and the potential to exacerbate systemic inequalities, especially against women and marginalized groups.

AI systems often reflect the data that they are trained on, mirroring societal biases and the perspectives of those who build them. A study by the Berkeley Haas Center for Equity, Gender and Leadership analyzed 133 biased systems across industries and found that 44.2 percent of these applications demonstrated gender bias, and 25.7 percent exhibited both gender and racial bias (Smith and Rustagi 2021). According to the World Economic Forum, it will take another 132 years to achieve gender equality on a global scale (World Economic Forum 2022). These harmful inequalities are reflected in technology, such as in depictions of women as less technologically skilled (Onta 2007) and the proliferation of other negative gender-based stereotypes (Sutko 2020). Emerging technologies such as AI can have staggering consequences and the potential to widen and perpetuate gender equality gaps, necessitating appropriate and direct regulation by governments to address gender-specific concerns.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

As a result of increasing concerns around the responsible use of AI, a growing body of research has emerged in the field of AI governance in recent years, discussing industry-specific guidelines and comprehensive global regulations to ensure the transparent and equitable use of AI. Despite this progress, current regulatory frameworks do not adequately address gender-specific biases. International cooperation on technology policy often prioritized technical, economic, and infrastructural issues, at the expense of recognizing how technological developments affect society's most vulnerable and historically excluded groups. There is a global deficit in governance for addressing the risks associated with AI and a growing need for the robust mainstreaming of gender considerations in the form of a global multi-stakeholder framework (UN Women Headquarters Office 2024).

This paper addresses the critical literature gap in AI governance research at the intersection of global AI policy and gender equity. It reassesses prominent global AI governance frameworks through the lens of gender equality and examines national policies based on the mitigation of gender-based discrimination in AI systems. Following an analysis of policy enforcement and effectiveness through two case studies, we propose a series of recommendations for advancing future gender-responsive AI policies.

The Global Landscape of AI and Gender Equality

The vast majority of existing global frameworks take a universal, broad approach to AI policy. The Bletchley Declaration signed at the 2023 AI Safety Summit held in the UK (Government of the United Kingdom 2023) recognized the importance of international collaboration for the management of unforeseen risks in the deployment of AI systems. It encouraged the building of risk-based policies for context-appropriate transparency and accountability. However, this directive serves as an example of a significant trend in current AI governance policies: the lack of resulting explicit guidelines on bias mitigation for vulnerable groups, including those based on gender or race. Similarly, the Hiroshima Process International Guiding Principles by the G7 (Ministry of Foreign Affairs of Japan 2023), promote responsible information sharing and reporting of incidents among organizations developing advanced AI systems; however, they

also fail to explicitly address gender-specific discrimination.

Additionally, Treaty No. 255 from the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Council of Europe 2024) mandates general measures for sustainable development and gender equality for signed parties, but provides very limited actionable guidance for addressing these biases. Thus, these examples underscore the urgency of developing global policies to close this gap of gender equity within AI governance frameworks.

The UNESCO Recommendation on the Ethics of Artificial Intelligence

One notable exception of global policy is the Recommendation on the Ethics of Artificial Intelligence (UNESCO 2022), UNESCO's first-ever global standard on AI ethics applicable to all 194 member states. This framework is remarkable in its specific guidelines to prevent gender-based bias in AI adoption by member states. Critical recommendations can be summarized as the following:

- Member states must maximize AI's potential to contribute to achieving gender equality while ensuring that the rights and freedoms of women and girls are not violated at any stage of the AI lifecycle
- Member states must have dedicated funds and adopt national policies that include a gender action plan and labor education to ensure girls and women are not excluded from the developing digital economy
- Member states should allocate resources to programs for increasing participation in STEM and information and communication technologies (ICT) disciplines
- Special care must be taken to ensure that already existing gaps are not exacerbated, including but not limited to: the gender wage gap, unequal representations in certain professions, lack of representation in leadership and senior management positions, the education gap, the digital access and adoption gap, and the unequal distribution of unpaid work and caring responsibilities in our societies
- Member states must make efforts to ensure that gender-based stereotypes and discriminatory beliefs are not propagated in AI systems, especially to avoid the magnifying effect of gender-based violence, harassment, and trafficking
- Female entrepreneurship and participation in AI development should be encouraged through economic initiatives and inclusive programs in the public and private sectors

This landmark piece of policy framework is distinctive in its specific guidelines for gender-based AI regulations. It represents a significant step forward by advocating for a holistic perspective of AI governance that prioritizes the needs of girls, women, and other vulnerable groups, while also providing specific protective measures to address gender biases and discrimination. While appreciable as an initial effort, it remains to see if future global regulations and treaties adopt this as a stepping stone for further exploration of gender equity in AI. Along these lines, recommendations from UN-Women during the sixty-seventh session of the Commission on the Status of Women discuss the importance of placing gender equality at the heart of the Global Digital Compact (UN Women Headquarters Office 2024). They underline the current gap in gender-transformative AI governance frameworks and advocate for the adoption of common global strategies to reshape the current AI ecosystem, in order to address the trust deficit for gender or racial bias in the governance model.

A comprehensive review and appraisal of current global AI governance practices reveals a critical gap in guidelines for preventing gender-based discrimination against women and girls. While certain frameworks begin the crucial work of identifying and addressing the needs of vulnerable communities in the advent of AI systems, the efforts remain preliminary and leave much to be desired in terms of global trends of AI governance for gender equity.

AI Governance Trends in National Policies

Several countries have introduced regulations on the responsible use of AI in industries and governments. For instance, the EU AI Act categorizes AI applications by risk based on adverse effects on human rights as protected by the Charter of Fundamental Rights of the EU (Article 28). It addresses particularly vulnerable AI applications, such as biometric identification for law enforcement (Article 18); access to education or training (Article 35); employment decisions (Article 36); and screening of applications for public benefits and healthcare services (Article 37). This approach exemplifies a dynamic risk assessment of AI systems that can potentially be affected by gender-based biases and discriminatory practices (The European Parliament and the Council of the European Union 2021; European Parliament, the Council of Ministers and the European Commission 2000).

In Canada, as part of Bill C-27, the Artificial Intelligence and Data Act (AIDA) designates AI bias as having an unjustified and adverse differential impact based on any of the protected characteristics outlined in the Canadian Human Rights Act (Innovation, Science and Economic Development Canada 2024). In South Korea, the proposed Act on Promotion of the AI Industry and Framework for Establishing Trustworthy AI (AI Act) similarly designates applications that are directly related to human life and safety as high-risk and are required to meet a certain level of trustworthiness (Roh and Nam 2023). In the United States, the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence in October 2023 affirms the necessity of holding AI accountable to standards that protect against unlawful discrimination and abuse, including in the justice system and the federal government (Biden 2023).

Among these national AI policies, a trend emerges of recognizing the need to keep high-risk AI systems accountable with respect to their countries' non-discrimination and human rights laws. However, there is still significant opportunity in defining exactly how those biases can be mitigated.

Analysis of Policy Effectiveness

Specific guidance unique to gender-based discrimination in global and national AI governance frameworks has been sparse. Many of the existing recommendations are still preliminary and do not necessarily go in-depth into recommendations for unique anti-discriminatory measures for protected characteristics such as gender or race. Nevertheless, efforts have been made to reduce overall discrimination and ensure algorithmic accountability in AI policies globally.

National policies vary in levels of enforcement and effectiveness - some have explicit legislation in place, whereas others have voluntary guidelines but few significant measures for enforcement of those policies. As a case study, we will review regulation from both the United States and the European Union, and discuss their respective policy efficacy.

Case Study of AI Policy Effectiveness for Gender Equity in the United States

In the United States, many proposed bills and existing frameworks have opted for a more cautionary approach of encouraging the development of ethical standards, and there has been limited comprehensive enforcement passed in terms of legislation for anti-discrimination in AI (White & Case LLP 2024). Policies such as the AI Risk Management Framework by the National Institute for Standards and Technology (NIST) are explicitly noted to be "for voluntary use" and organizations are invited to incorporate as many or as few suggestions into their AI systems (National Institute for Standards and Technology 2023).

However, this does not mean that these guidelines themselves have not had any impact on increasing algorithmic accountability. There have been several noteworthy actions taken to begin implementing these policies in federal agencies in the US (White House Office of Science and Technology Policy 2022), such as combating gender-based violence and AI-generated image-based sexual abuse. A White House Call to Action (Klein and Prabhakar 2024) called on leading AI developers and data providers to make voluntary commitments to curb the creation of AI-generated material containing gender-based violence and abuse. (The White House 2024). In addition, in accordance with Executive Order 14110, the Office of Federal Contract Compliance Programs developed guidelines addressing AI in the context of Equal Employment Opportunity (EEO) to mitigate the potential harmful impacts of AI in automating unlawful bias, such as on race, gender, or ethnicity, in hiring practices for all federal contractors (Office of Federal Contract Compliance Programs 2024). The Department of Education's Office for Civil Rights further expanded the domain of Title IX of the Education Amendments of 1972 to prevent discrimination on the basis of gender in AI used in educational environments. (United States Department of Education's Office for Civil Rights 2024).

Thus, the potential impact of AI policy for combating gender-based discrimination is promising. Current frameworks have demonstrated good initial progress for encouraging voluntary accountability in AI systems, however, there is still considerable potential for passing comprehensive legislation and enforcement processes for all agencies and stakeholders in the US.

Case Study of AI Policy Effectiveness for Gender Equity in the European Union

One excellent example for studying policy enforcement is the European Union's work with enforcing the EU AI Act. According to Article 113 in the EU AI Act, any AI systems posing unacceptable risks were banned within six months of the EU AI Act becoming enforceable, effective in February 2025 (The European Parliament and the Council of the European Union 2021). Rules on general-purpose AI systems were required within 12 months of the act becoming effective, and high-risk systems (such as those making critical decisions in recruitment and law enforcement) were given up to 36 months to comply with the requirements (European Parliament 2024).

The EU also established the European AI Office as part of the European Commission, which serves as the center for the enforcement and implementation of the EU AI Act. The AI Office is tasked with conducting evaluations of general-purpose AI models, ensuring accountability from model developers with necessary documentation requests, and investigating possible infringements of the guidelines (European Commission 2024). The EU has made significant strides in ensuring the necessary measures for the enforcement and implementation of its policies, from the creation of offices for the administration of the Act to penalties levied for failure to comply.

On the whole, regional and national policies have displayed early signs of success, however, there is still potential for improvement of enforcement and official legislation from all nations worldwide. Furthermore, while existing policies do highlight the necessity for anti-discrimination measures in the use of AI systems, there is a lack of precise guidelines and regulations for specific biases and discrimination against sensitive groups. In summary, we find that while an excellent starting point, there still remains substantial room for growth for ensuring national AI policies are directly encouraging and enforcing gender equity.

Proposed Recommendations

Based on an analysis of existing governance frameworks, we recommend the following considerations to be made when developing policies that reflect the intersection of AI and gender-based discrimination:

1. Addition of specific recommendations for combating gender-based discrimination in high-risk AI systems involved in decision-making processes of high impact, such as recruitment, employment, healthcare access, education access, etc.
2. Publication of a national gender action plan when developing AI policy and legislation for gender equity in AI governmental and corporate processes
3. Encouragement of representation and involvement of the voices of women and other marginalized groups at all

stages of the AI development and policy lifecycle, including leadership, senior management positions, development, and policy-making

4. Allocation of resources and funds for increasing participation of women and girls in STEM and information and communication technologies (ICT) disciplines, as outlined in The UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO 2022)
5. Creation of avenues for legal recourse and enforcement of the rights of girls and women (Global Partnership on Artificial Intelligence 2024)
6. Issuance of technical guidelines for continued transparency and accountability in AI systems for harm prevention against sensitive groups

Global and national policies should reflect the needs of specific sensitive demographics to ensure that existing systemic biases are neither exacerbated nor perpetuated in AI systems. Girls, women, and other vulnerable groups face significant obstacles in current societal processes, and global AI policies must address these unique challenges. In the context of global regulatory frameworks, we conclude that overall AI governance trends are moving in the right direction to begin addressing gender-specific inequalities, while still having scope for growth in developing comprehensive gender-responsive AI regulations and global strategies for the empowerment of all girls and women.

References

Biden, J. R. J. 2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. Accessed: 2024-11-24.

Busuioc, M. 2021. Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, 81(5): 825–836.

Council of Europe. 2024. Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. https://rm.coe.int/1680afae3c. Accessed: 2024-11-24.

European Commission. 2024. European AI Office. https://digital-strategy.ec.europa.eu/en/policies/ai-office. Accessed: 2025-01-18.

European Parliament. 2024. Artificial Intelligence Act: MEPs adopt landmark law. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law. Accessed: 2025-01-18.

European Parliament, the Council of Ministers and the European Commission. 2000. Charter of Fundamental Rights Of the European Union. https://www.europarl.europa.eu/charter/pdf/text_en.pdf. Accessed: 2024-11-24.

Global Partnership on Artificial Intelligence. 2024. Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity. https://gpai.ai/projects/responsible-ai/towardsrealdiversityandgenderequalityinai/towards-substantive-equality%20in-artificial-intelligence_Transformative-AI-policy-for-gender-equality-and-diversity.pdf. Accessed: 2025-01-11.

Government of the United Kingdom. 2023. The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023. Accessed: 2024-11-24.

Innovation, Science and Economic Development Canada. 2024. The Artificial Intelligence and Data Act (AIDA) – Companion document. https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document. Accessed: 2024-11-24.

Klein, J.; and Prabhakar, A. 2024. A Call to Action to Combat Image-Based Sexual Abuse. https://www.whitehouse.gov/ostp/news-updates/2024/05/23/a-call-to-action-to-combat-image-based-sexual-abuse/. Accessed: 2025-01-18.

Ministry of Foreign Affairs of Japan. 2023. Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems. https://www.mofa.go.jp/files/100573471.pdf. Accessed: 2024-11-24.

National Institute for Standards and Technology. 2023. AI RMF Playbook. https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook. Accessed: 2025-01-18.

Office of Federal Contract Compliance Programs. 2024. Artificial Intelligence and Equal Employment Opportunity for Federal Contractors. https://www.dol.gov/agencies/ofccp/ai/ai-eeo-guide. Accessed: 2025-01-20.

Onta, N. 2007. Recent Books on Gender and Technology. *Gender, Technology and Development*, 11(3): 401–404.

Roh, T.; and Nam, J. E. 2023. South Korea: Legislation on Artificial Intelligence to Make Significant Progress. https://www.kimchang.com/en/insights/detail.kc?sch_section=4&idx=26935#:~:text=To%20summarize%2C%20the%20AI%20Act,certification%20systems%20for%20AI%20trustworthiness. Accessed: 2024-11-24.

Smith, G.; and Rustagi, I. 2021. When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. *Stanford Social Innovation Review*.

Sutko, D. M. 2020. Theorizing femininity in artificial intelligence: a framework for undoing technology's gender troubles. *Cultural Studies*, 34(4): 567–592.

The European Parliament and the Council of the European Union. 2021. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. https://artificialintelligenceact.eu/ai-act-explorer/. Accessed: 2024-11-24.

The White House. 2024. Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order. https://www.whitehouse.gov/briefing-room/statements-releases/2024/10/30/fact-sheet-key-ai-accomplishments-in-the-year-since-the-biden-harris-administrations-landmark-executive-order/. Accessed: 2025-01-18.

UN Women Headquarters Office. 2024. Placing Gender Equality at the Heart of the Global Digital Compact. Technical report, UN Women Headquarters Office.

UNESCO. 2022. Recommendation on the Ethics of Artificial Intelligence. Technical report, UNESCO.

United States Department of Education's Office for Civil Rights. 2024. Avoiding the Discriminatory Use of Artificial Intelligence. https://www.ed.gov/media/document/avoiding-discriminatory-use-of-ai. Accessed: 2025-01-20.

White & Case LLP. 2024. AI Watch: Global regulatory tracker - United States. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states. Accessed: 2025-01-18.

White House Office of Science and Technology Policy. 2022. Blueprint for an AI Bill of Rights. https://www.whitehouse.gov/ostp/ai-bill-of-rights. Accessed: 2025-01-18.

World Economic Forum. 2022. Global Gender Gap Report 2022. Technical Report 978-2-940631-36-0, World Economic Forum, Geneva, Switzerland.