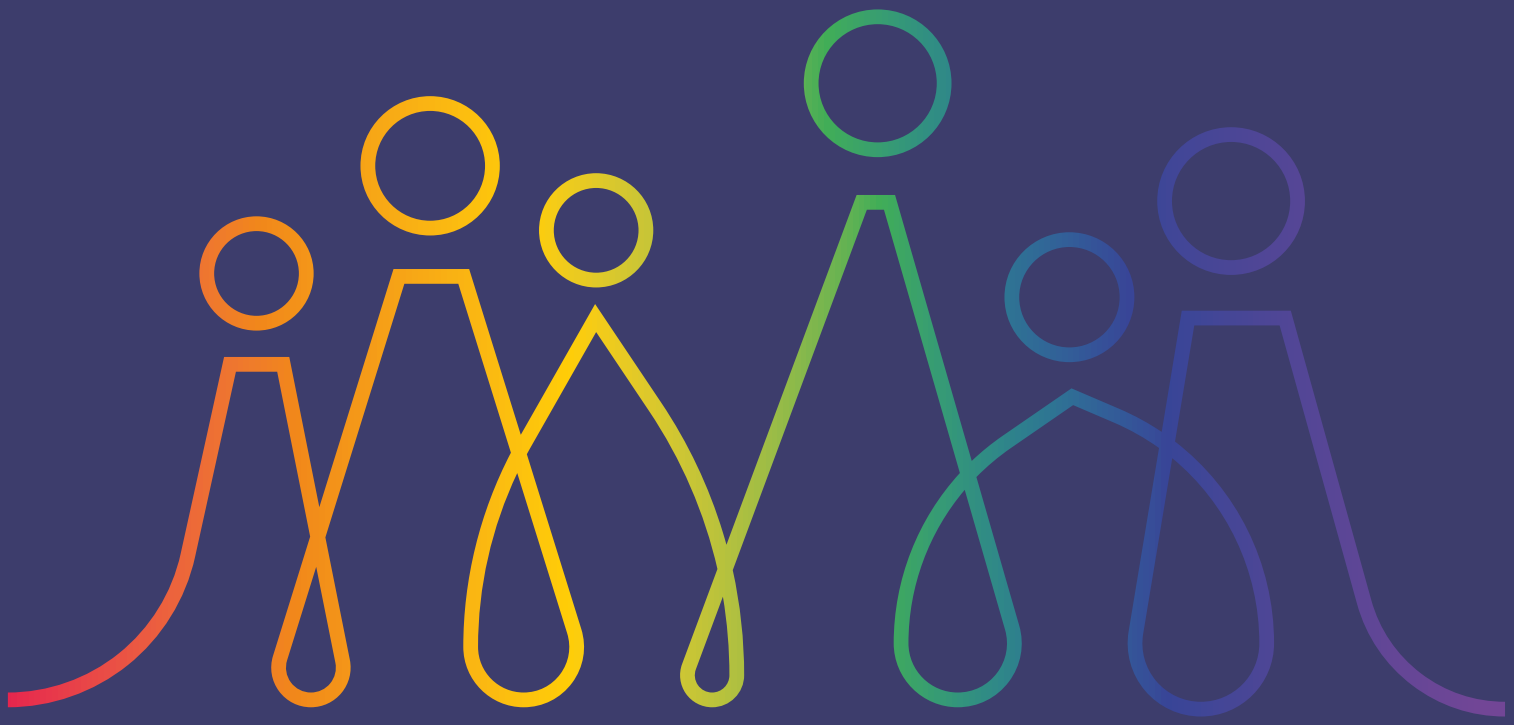
unesco



UNESCO Women for Ethical AI

Outlook Study on Artificial Intelligence and Gender

Pre printed for discussion at W4EAI Conference,30 October 2024, UNESCO HQ, Paris, France

Cover Illustration: robin.ph/Shutterstock.com

Cover design: Claudia Tortello

Illustrations: Lightcast/Daniel Botkin

Typeset: UNESCO/Noam Le Pottier

Printed by UNESCO

Printed in Paris

SHS/REI/EAI/W4EAI/2024/Outlook

Table of Contents

Acknowledgements

This report was developed under the leadership and overall supervision of Gabriela Ramos, Assistant Director-General for the Social and Human Sciences of UNESCO, by (alphabetically), Auxane Boch; Gökce Cobansoy Hizel; Caitlin C. Corrigan; Cecilia Danesi; Sarah Elaine Eaton; Caitlin Kraft-Buchman; Oriana Kraft; Alexander Kriebitz; Eleonora Lamm; Cristina Martínez Pinto; Wanda Muñoz; Paola Ricaurte; Alessandra Sala and Sumiko Shimo. Sincere thanks to the Women for Ethical AI volunteers who reviewed the policy documents, Stefania Pecore; Julia Pomares and Calkin Suero Montero. Furthermore, many thanks to those that took the time to critically review the report including Mariagrazia Squicciarini, Doaa Abu Elyounes; Claudia J. Coulton; Rosanna Fanni; Angelica Fernandez.

Prologue

In an increasingly digitalized and automated world, artificial intelligence (AI) has become a transformative force that profoundly impacts all aspects of our society. As these technologies advance, the urgent need to address existing inequalities in their development and application becomes evident, particularly those affecting gender equality.

This report developed by UNESCO and in collaboration with the Women for Ethical AI (W4EAI) platform, is based on and inspired by the gender chapter of UNESCO's Recommendation on the Ethics of Artificial Intelligence. This concrete commitment, adopted by 194 Member States, is the first and only recommendation to incorporate provisions to advance gender equality within the AI ecosystem.

The primary motivation for this study lies in the realization that, despite progress in technology and AI, women remain significantly underrepresented in its development and leadership, particularly in the field of AI. For instance, currently, women reportedly make up only 29% of researchers in the field of science and development (R&D),[1] while this drops to 12% in specific AI research positions.[2] Additionally, only 16% of the faculty in universities conducting AI research are women, reflecting a significant lack of diversity in academic and research spaces.[3] Moreover, only 30% of professionals in the AI sector are women,[4] and the gender gap increases further in leadership roles, with only 18% of in C-Suite positions at AI startups being held by women.[5]

Another crucial finding of the study is the lack of inclusion of gender perspectives in regulatory frameworks and AI-related policies. Of the 138 countries assessed by the Global Index for Responsible AI, only 24 have frameworks that mention gender aspects, and of these, only 18 make any significant reference to gender issues in relation to AI. Even in these cases, mentions of gender equality are often superficial and do not include concrete plans or resources to address existing inequalities.

The study also reveals a concerning lack of gender-disaggregated data in the fields of technology and AI, which hinders accurate measurement of progress and persistent inequalities. It highlights that in many countries, statistics on female participation are based on general STEM or ICT data, which may mask broader disparities in specific fields like AI. For example, there is a reported 44% gender gap in software development roles,[6] in contrast to a 15% gap in general ICT professions.[7]

Furthermore, the report identifies significant risks for women due to bias in, and misuse of, AI systems. Recruitment algorithms, for instance, have shown a tendency to favor male candidates. Additionally, voice and facial recognition systems perform poorly when dealing with female voices and faces, increasing the risk of exclusion and discrimination in accessing services and technologies. Women are also disproportionately likely to be the victims of AI-enabled online harassment.

The document also highlights the intersectionality of these issues, pointing out that women with additional marginalized identities (such as race, sexual orientation, socioeconomic status, or disability) face even greater barriers to accessing and participating in the AI field. For example, women with disabilities have less access to technological tools and face double discrimination compared to their male counterparts.

Through a human rights-based approach and a transformative gender ethics perspective, this report provides clear and actionable recommendations to integrate gender equality at all stages of the AI lifecycle. It emphasizes the importance of collecting more comprehensive and disaggregated data, implementing targeted interventions, and developing inclusive policies that promote equitable participation in the design, use, and governance of AI.

1 Collet et al., 2022

2 UNESCO, 2024a

3 West, 2023.

4 WEF, 2023

5 Sey & Kafkin, 2019

6 UNESCO, 2024a

7 Sey & Kafkin, 2019

I hope that this study not only contributes to a greater understanding of the barriers women face in the AI field but also inspires efforts to collect more specific data on AI and gender and drives the design of policies and practices that foster equitable participation and promote technological innovation with an inclusive and ethical approach. It is a key part of our efforts to implement the UNESCO Recommendation through an effective Readiness Assessment Methodology (RAM) to make the Recommendation's chapter on gender impactful. We believe that only through a collective effort will it be possible to harness the transformative potential of AI to close gender gaps and build a more just and equitable future without leaving anyone behind.

Gabriela Ramos
Assistant Director-General for the Social and
Human Sciences of UNESCO

Executive Summary

The gender chapter of the UNESCO Recommendation on the Ethics of AI is a concrete commitment by 194 Member States to advance gender equality in the AI ecosystem. To track progress in the implementation of the gender provisions of UNESCO's Recommendation, and to assess the impacts of AI on gender equality, the UNESCO Women for Ethical AI (W4EAI) Platform[8] has been established. This report advances the workstream through evidence-based insights in three critical areas: women's participation in AI development and deployment, the inclusion of gender equality concerns in AI governance and the impact of AI on gender equality. It highlights the significant underrepresentation of women in AI, the lack of gender-disaggregated data, and the compounded challenges women face in the field. The report also addresses the neglect of gender dimensions in AI policy, the risks posed by AI systems to women, and the need for responsible and ethical AI governance to promote gender equality. Finally, it outlines actionable recommendations to enhance gender equality through and in AI, emphasizing the importance of comprehensive data collection, targeted interventions, and inclusive policy-making.

It is divided into 4 sections. Section 1 introduces the motivation of the report and the work of the UNESCO Women for Ethical AI platform.

Section 2 analyzes how to improve the understanding of women's participation in AI. The **predominance of white males in the AI and machine learning fields is leading to gender bias in AI systems**. There is a severe lack of gender-disaggregated data in ICT and AI fields, often oversimplified to percentages of girls in Science, Technology, Engineering, and Mathematics (STEM) or Information and communication technology (ICT) education. Data collection is inconsistent, with the analysis in this report finding a 17% variance in women computer science graduates depending on the reporting authority for example. Using broader STEM data may mask larger gaps in AI-specific fields. Moreover, gender gaps widen significantly from entry-level to leadership positions in tech and as the focus narrows to AI companies. This highlights the need for comprehensive data on the AI participation pipeline. There is also a lack of standardized corporate and government reporting on gender-disaggregated employment data in AI. Additionally, women face compounded challenges due to lack of resources, mentors, and financing, along with traditional barriers like gender stereotyping and toxic work environments.

The uneven technological field for women cannot be only addressed through technical solutions, but also through addressing unbalanced **policy environments in which AI is used and looking at who participates in decision-making and how inclusive the processes are**. These considerations are key to developing a holistic approach to the issue of gender equality and AI through policy, which is the second focus of Section 2. The analysis from this section shows that not only are current AI governance frameworks lacking a substantial gender perspective, but major efforts to even analyze the policy landscape for AI and evaluate AI readiness are not adequately taking gender dimensions into account. If AI-policy related tools and indices do not include enough gender quality criteria in their analysis, measurement of progress and setting of benchmarks will be increasingly difficult.

Section 2 finally **examines the potential for disproportional risks and harms from AI systems for women** through AI use cases. The report finds that AI risks related to gender are overwhelmingly categorized under discrimination and harms result mainly from bias in machine learning and algorithmic systems and intentional or unintentional misuse of AI-enabled tools, leading to increased risk for discrimination and physical or mental harm based on gender. Missing or biased data leads to biased AI systems that are less accurate or discriminate based on gender, promote harmful gender stereotypes or increase the ease of targeting of women in mis- and disinformation. Similarly, women's livelihoods or access to services may be disproportionately impacted by the introduction of automated systems. This is compounded by lacking transparency in how systems work, specifically how barriers to participation may hinder women's effective inclusion in the AI lifecycle. Intersectionality needs to be increasingly considered by decision makers and developers, as race, sexuality, disability status, and other minoritizing factors, interact with gender.

If designed, used and governed responsibly and ethically, **AI also has potential to positively impact gender equality**. Section 3 focuses on approaches and tools to improve AI's impact on gender equality. The analysis examines approaches that promote responsible AI for gender equality, an ethics by design approach

8 https://www.unesco.org/en/artificial-intelligence/women4ethical-ai

and a gender transformative, feminist approach. These approaches stress different perspectives about the mechanisms and hurdles to integrating women and gender concerns into the AI ecosystem. They are complementary, reinforce each other, and commonly highlight a key need for more information about AI and gender equality. While an increasing number of tools exist for assessing impact of AI systems, promoting gender equality needs to be further streamlined across all existing tools and processes. That is why the report looks to the work of UNESCO in its established *AI Readiness Assessment Methodology* (RAM) as a tool for supporting member states in understanding the multiple dimensions of preparedness for confronting the challenges and opportunities posed by AI which includes five different categories of questions directly related to assessing gender and diversity aspects of AI readiness.

This outlook study ends with Section 4, presenting **actionable recommendations to address the promotion of gender equality through and in AI**. Each proposal is **linked to specific provisions from the UNESCO Recommendation on the Ethics of AI and amplifies them**. The actionable recommendations include, among others:

1. Incorporate a comprehensive, transformative gender approach in national digital and/or AI policies, strategies and assessment tools. This includes both a) targeted initiatives to empower girls and women, and b) gender mainstreaming at all stages and levels of policy making.
2. Ensure that funding mobilized for the design, development, implementation and monitoring of national AI governance mechanisms – and more generally, in the AI sector – are gender transformative.
3. Include a transformative gender perspective in all AI ethical and human rights impact assessments, as well as in all other AI tools and assessments.
4. Commit to ensuring the effective and meaningful participation of women in the AI ecosystem and related sectors and in all AI governance deliberations, and report on such efforts and their results.
5. Strengthen or develop targeted interventions and programs to promote and support women to access, remain and advance in STEM-related careers, and in female entrepreneurship.
6. Establish standards for and monitor employers in the AI ecosystem, in academia and in the private and public sectors, to review and adapt their policies and practices to ensure they contribute to gender equality building on existing commitments, obligations and good practices.
7. Monitor and report on gender gaps and advances in access to AI, participation in the AI ecosystem and related governance mechanisms, and entrepreneurship, including by disaggregating and reporting on data on gender gaps in access and participation, and on measures taken to redress gender inequality.
8. Review current legislation in accordance with human rights and gender equality standards and, if necessary, amend or adopt new regulatory measures to ensure remedy and reparation mechanisms for harm caused by AI systems due to the lack of design, development, deployment and monitoring.
9. Encourage the development and creation of learning communities and peer-to peer networks on gender and diversity in AI at national at national, regional and international levels, for increasing innovation and exchange, replication and upscaling of evidence-based good practices.

Section 1: Introduction

AI is transforming the world on every level and in every region. It influences the economy, workforce and social norms in significant ways. Over the past decade, AI related investments have increased threefold. Global private investment totaled a staggering almost $100 billion in 2023. Funding for innovative AI enabled tools, such as generative AI applications, attracted $25.2 billion alone in 2023, nine times increase from 2022. And the number of new companies had a 40.6% increase between 2022 and 2023, with over 1,800 new companies founded in 2023.[9]

The economic size of the field, and the speed and depth at which AI-enabled technologies are impacting society, has given rise to increased attention on the dynamic influence these converging technologies are having on society. At this juncture, it is particularly important to also consider the implications of AI-enabled technologies in relation to global and national efforts to reduce gender inequality.

The UNESCO definition of AI systems of central ethical relevance includes "systems which have the capacity to process data and information in a way that resembles intelligent behavior, and typically includes aspects of reasoning, learning, perception, prediction, planning or control." Given the speed at which AI can process information and make decisions, AI-enabled tools have the power to reinforce and entrench dynamics and biases that already exist. Yet, when used responsibly and with these dynamics in mind, there is immense potential for AI to transform women and girls' lives and more effectively and efficiently address gender dimensions of sustainable development and human security. When these concerns are not addressed, there is the risk of losing the opportunities AI presents, while simultaneously widening gender gaps and even undermining the strides that have been made to promote empowerment for women and girls around the world. AI needs to be developed and applied with a gender perspective, otherwise it is likely to reproduce and reinforce gender stereotypes, minimizing the positive potential impacts for society."[10]

In highlighting the distinct issues related to gender and AI in the current ecosystem, this outlook presents the latest data and evidence to suggest actionable recommendations to address the promotion of gender equality through and in AI.

In a context where humankind is falling short in delivering on the SDGs, including gender (SDG5) and reducing inequalities (SDG 10), AI has the potential for impact on a global scale and with real policy implications. Goals around health and wellbeing (SDG 3), quality of education (SDG 4) and decent work and economic growth (SDG 8) are impacted by AI systems and could become a risk if it reinforces discrimination along gender, disability, racial or other lines.[11]

UNESCO has pioneered the work on gender and AI in its Recommendation on the Ethics of AI, adopted in 2021, by including a dedicated policy section on gender, in addition to a cross-cutting gender perspective. In the realm of education, for instance, UNESCO placed a focus on skills in the digital age and STEM based curricula as a mechanism to achieve gender equality.[12] By placing the discussion of the impact of AI on gender policy and equality efforts within the overarching context of sustainable development, **this outlook serves to highlight the complexity, interconnectedness, and intersectionality that needs to be considered when discussing the impacts of AI on gender equality.**

To implement the provisions on gender from the recommendations, and to assess the impacts of AI on gender equality, the UNESCO *Women for Ethical AI* (W4EAI)[13] platform has been established. Its objectives are to understand the current ecosystem and assess current or potential risks of AI systems related to gender inequality and feed it into the process of the implementation of UNESCO's Recommendation and its Readiness Assessment Tools. Through a network of leading AI experts, the platform aims to advance gender equality in the AI agenda. This includes ensuring women are represented equally in the design, use and deployment of AI and that AI systems are trustworthy, gender-friendly and inclusive. Specifically, W4EAI seeks to

9 Maslej et al., 2024

10 Lamm et al., 2023

11 UN Department of Economic and Social Affairs, 2024

12 UNESCO, 2023a.

13 https://www.unesco.org/en/artificial-intelligence/women4ethical-ai

- support the implementation of the Recommendation on the Ethics of AI and ensure the *Readiness Assessment Methodologies* (RAM), now implemented in 60 countries around the world, include the gender equality approach,
- share research and contribute to a repository of good practices on gender equality, for UNESCO member states to inform their National AI Strategies,
- drive progress on non-discriminatory algorithms and data sources and avoid harassment or misuse of synthetic content.
- incentivize girls, women and under-represented groups to participate in the AI life cycle.

As a first step in the W4EAI agenda, **this outlook represents (1) a comprehensive effort to deliver an overview of the current state of how gender equality is being acknowledged and promoted through and in AI; (2) an evidence based analysis of the knowledge and policy gaps that remain for promoting gender equality through AI and (3) an outline of concrete recommendations on how make progress on the gender equality goals of the UNESCO Recommendation on the Ethics of AI.**

This report builds on the work of the UNESCO *Recommendation on the Ethics of AI*, Policy Area 6: Gender, which outlines seven points for advancing AI with a gender lens.[14] The Recommendation guides Member States in their actions to make AI inclusive by design and outlines clear ethical recommendations that can be translated into implementable and effective policies.[15] A summary of the seven principles related to gender call on Member State to act as follows:

1. Maximize the potential of AI to advance the achievement of gender equality, while ensuring that human rights and fundamental freedoms are not violated.
2. Dedicate funding to finance gender responsive schemes related to AI, ensure that national digital policies include a gender action plan, and develop relevant policies targeted at supporting girls and women to make sure they are not left out of the digital economy powered by AI.
3. Ensure the potential of AI systems to advance the achievement of gender equality is realized and that these technologies do not exacerbate the already wide gender gaps
4. Ensure that gender stereotyping and discriminatory biases are not translated into AI systems, and instead identify and proactively redress these.
5. Encourage female entrepreneurship, participation and engagement in all stages of AI system lifecycle through inclusive investment and promoting harassment free environments.
6. Promote gender diversity in AI research in academia and industry
7. Form a repository of best practices for incentivizing participation of girls, women and underrepresented groups in all stages of the AI system lifecycle.

Furthermore, the report looks to the work of UNESCO in its established AI Readiness Assessment Methodology (RAM)[16], a tool for supporting member states in understanding the multiple dimensions of preparedness for confronting the challenges and opportunities posed by AI. The RAM process, which include chapters on legal, social/cultural, scientific/educational, economic, technical and infrastructural dimensions, enabled member states to produce reports and roadmaps with UNESCO experts to chart an "ethical and realistic way forward with AI" and include specific indicators and information related to gender equality.

With UNESCO's Recommendation and the RAM, and the establishment of the W4EAI Platform, UNESCO is working to assist governments and AI stakeholders in understanding, assessing and improving their AI governance and ecosystems in line with the provisions of the Recommendation.[17]

In the realm of gender and AI, UNESCO is actively working to develop and implement "ethical standards and guidelines to address bias and barriers that may prevent girls' and women's full participation and leadership in the digital environment, including through Artificial Intelligence (AI)."[18] In working towards the goal of ensuring standards, promoting literacy and empowering participation for women in

14 UNESCO, 2022a; 2023a.

15 Lamm et al., 2023.

16 https://www.unesco.org/ethics-ai/en/ram

17 UNESCO, 2023a.

18 UNESCO, 2023a.

the AI landscape, UNESCO strives to create a base of research and analysis from which to benchmark and inform progress on this topic. Thus, each section of this report addresses a topic related to AI and gender corresponding to Policy Area 6: Gender of the UNESCO Recommendation.

The scope of this outlook report is comprehensive, examining ethical questions, risks, opportunities and social impacts of AI systems relevant to all stages of the AI system life cycle, including the steps of data collection, data preparation, model development, model evaluation, model post processing and model deployment.[19]

Furthermore, the report adopts a framework for understanding the impacts of AI on gender equality that is based on the principles outlined in the UNESCO *Recommendation on the Ethics of AI* [20]:

- Proportionality and Do No Harm
- Safety and security
- Fairness and non-discrimination
- Sustainability
- Right to Privacy, and Data Protection
- Human oversight and determination
- Transparency and explainability
- Responsibility and accountability
- Awareness and literacy
- Multi-stakeholder and adaptive governance and collaboration

The report focuses on gender equality both in terms of the gender gaps in participation in the AI ecosystem and the potential disproportionate impacts of AI on women and girls. These two issues are inherently linked, as increasing women's and girl's participation in education, training, workforce and leadership in the AI ecosystem will impact the recognition of bias in AI systems along the AI lifecycle, and in AI governance frameworks.

A recurring issue discussed throughout the outlook is the commonly used didactic interpretation of gender, as well as the simplistic focus on only men versus women. This perspective overlooks the **problematic binary concept of gender taken on by most studies and databases or in coding and labeling**. As a report from AI Now puts it, "the very existence of automated gender classification systems presents a number of problems: they functionally understand gender as an essential, biological, and binary identity that can be 'detected' and affirmed through the lens of a commercialized technical system. In this way, such systems (and the interests that create and profit from them) are positioned as the arbiters of identity, mapping static categories onto diverse bodies."[21] Very few studies have assessed the impact on accuracy and fairness of AI systems in-depth, taking into consideration the larger diversity that exists in terms of sexual orientation and gender identities.

The simplification of concepts also overlooks issues related to **intersectionality**. Evidenced by the disproportionate risks, bias and opportunities of AI that are compounded as gender dynamics interact with factors such as race, socioeconomic status, sexual orientation and gender identities, disability status or geographic location, among others. For example, women in rural areas, women of color and women who are experiencing poverty are at significantly higher risk of discrimination by, inequitable access to, or ineffectiveness of AI-enabled tools and applications.[22] Thus, incorporating these dynamics into recommendations about improving gender equality through and in AI are vital to creating positive impact.

A multi-method approach was used for this outlook study, relying on data collected from both quantitative and qualitative sources to triangulate information about the current landscape of AI in relation to women and gender dynamics, as well as the recommendations for how AI can advance gender equality. This analysis included (1) processing existing quantitative datasets and scientific and policy relevant literature that relied on quantitative data, the (2) use of qualitative case study analysis to understand potential and current areas of AI impact on women and gender policy and equality and (3) qualitative text analysis to compare and understand current policy efforts to integrate concerns related to women and gender into policy frameworks related to AI.

19 Suresh and Guttag, 2021.

20 UNESCO, 2022a

21 West et al., 2019.

22 See Section 3 for more examples.

Section 2: State of Affairs for AI and Gender Equality

2.1: AI's Gender Participation Gap

Section 2.1 examines the question of *how and where women are participating in the field of AI development and deployment*, by focusing on the current state of women's participation in the AI ecosystem. **The section highlights significant disparities in gender equality in terms of participation in, and access to, the AI ecosystem**, even while some progress has been made to close these gaps. To understand if and where progress has been made in improving diversity and inclusion in the AI ecosystem, clear and consistent benchmarks must be established. Section 2.1 sheds some light on the current benchmarks and what is needed to improve their scope and accuracy.

As evidence shows, there is a significant white male predominance in the AI and machine learning field perpetuating gender bias. As reflected in an Alan Turing report in 2021, this, in turn, presents an ethical concern for social and economic fairness and creates a lack of diversity in the ecosystem.[23] As a 2023 UNESCO report states, "Women must be an active part of developing the digital economy to eliminate gender biases and stereotypes being reproduced through digital platforms, software and programmes generated by AI." [24]

One place to start with this effort is by increasing the role of women engaging with and working in the field of AI. Diversity within the workplace "affects how AI companies work, what products get built, who they are designed to serve, and who benefits from their development."[25] Although this fact has been highlighted in many recent reports, this outlook study will further highlight how the data available on participation is varied, depending highly on definitions and surveyed populations, making it difficult to compare across contexts or over time.

UNESCO Recommendation on the Ethics of Artificial Intelligence #91: *Member States should encourage female entrepreneurship, participation and engagement in all stages of an AI system life cycle by offering and promoting economic, regulatory incentives, among other incentives and support schemes, as well as policies that aim at a balanced gender participation in AI research in academia, gender representation on digital and AI companies' top management positions, boards of directors and research teams. Member States should ensure that public funds (for innovation, research and technologies) are channelled to inclusive programmes and companies, with clear gender representation, and that private funds are similarly encouraged through affirmative action principles. Policies on harassment-free environments should be developed and enforced, together with the encouragement of the transfer of best practices on how to promote diversity throughout the AI system life cycle.*

a. Gender Data

There has until recently been a severe lack of official sex-disaggregated data on most topics related to informational and computer technology as a field.[26] Where there have been concerted efforts, the inclusion of the gender indicators is often superficial or unclear, relying only on, for instance, the standard measurement of percentage of girls in STEM education. Even though this is an important element, it does not capture women participation in the ITC studies and sector, that are relevant to the AI ecosystem. Disperse figures show that girls are less inclined to take up ICT studies than what they are on STEM. Furthermore, other characteristics related to race, disability and gender diversity, among others, are even less analyzed and disaggregated.

23 Young et al., 2021.

24 UNESCO, 2023a.

25 West et al., 2019.

26 Sey and Hafkin (2019) provide an overview of the sources and lack of clarity on gender disaggregation related to measures of digital access and education in Chapter 6

For this section, reports on gender dimensions and trends in the AI and digital landscape were surveyed from major organizations to evaluate how they measure gender gaps and women's participation in the AI landscape.[27] Although this search is not exhaustive, the derived statistics and trends reported below reveal a few noteworthy insights. First, **there is not enough regular or detailed collection of the data**. Many statistics or indices found were out of date, too broad, did not have multiple data points or left gender dimensions out of toolkits completely.

> ***UNESCO Recommendation on the Ethics of Artificial Intelligence #89:*** *Member States should ensure that the potential of AI systems to advance the achievement of gender equality is realized. They should ensure that these technologies do not exacerbate the already wide gender gaps existing in several fields in the analogue world, and instead eliminate those gaps. These gaps include: the gender wage gap; the unequal representation in certain professions and activities; the lack of representation at top management positions, boards of directors, or research teams in the AI field; the education gap; the digital and AI access, adoption, usage and affordability gap; and the unequal distribution of unpaid work and of the caring responsibilities in our societies.*

For instance, the International Monetary Fund's (IMF) AI Preparedness Index only includes the World Bank's indicator on number of female STEM graduates (10-year average; % of STEM graduates) in its measurement of "workforce" preparedness and includes no gender aspects in their measurement of the regulatory/policy environment or preparedness on ethics.[28] The OECD AI Observatory's toolbox section includes only one tool (out of 892) and one metric (out of 103), respectively, that focuses on gender.[29] There is, therefore, a clear need to promote tools and measurements that more accurately capture the nuance and development of the role of women in AI.

Second, there are notable differences in the measurements of women's participation in the AI ecosystem that make comparison between surveys or overtime difficult. **The often reported measures are not specific enough to AI, instead looking at participation in Tech, Information and Communication Technologies (ICT) or STEM**. There is a need to set an international standard internationally and over time so that there can be clarity about progress made and areas that need increased support. Without the ability to compare statistics, it is difficult to understand progress and remaining gaps.

Third, there is a need for standards in corporate organizational-level reporting of gender disaggregated data and employment in the field of AI, in terms of interval and level of detail. For instance, a report from AI Now documents the lack of transparency in hiring statistics from major tech companies', citing the specificity of the data reported and efforts to mask some of the more revealing data by confining reporting to only certain levels of operation. In the case of Google reporting, information about gender disparities at more senior levels, as well as information about "underleveling", i.e. placing female employees at lower levels of employment than they are qualified for, was left out to display more inflated numbers.[30]

The International Telecommunication Union (ITU) identifies four main areas of the gender digital divide: (1) a gap in access and use of the Internet, (2) a gap in digital skills and use of digital tools, (3) a gap in participation in science, technology, engineering and math (STEM) fields, and (4) a gap in tech sector leadership and entrepreneurship.[31] These challenges face the sub-field of AI as well. Given this observation, the rest of the section focuses on women's participation in AI in three areas: (1) access, (2) education and (3) workforce/leadership. The analysis documents the statistics currently available, the gaps in knowledge and the need for increased data collection standards in each area of inquiry.

27 See footnotes for each table below for sources

28 See IMF, 2023.

29 See OECD.AI, 2024.

30 West et al., 2019

31 ITU, 2023a.

b. Access to AI Information and Tools

Data available on AI access predominantly looks at access to ICT and digital tools. While this certainly relates to AI access, most sources were not exclusively focused on it. The available statistics center around internet use, technology knowledge and technology access.[32] Because access to technology is a prerequisite for interacting with, or learning about, AI-enabled systems, trends related to general digital and internet access are pertinent. Most noteworthy here is not the lack of gender parity in access to internet and technology, which was at 92% in 2023 according to the ITU, but that **the intersectionality of gender and other dimensions related to access show disturbing trends**. For instance, when taking into account poverty, the gender parity in internet access in low income countries fall to only 59%.[33] In a survey of Global South countries, there was a 125% gap in internet access between rural and urban women and internet access is significantly lower in older age groups.[34] On other hand, persons with disabilities - most of whom still live in poverty - are also affected by this digital divide and face specific barriers to access internet and technology due to factors such as high costs, exclusion from the public education system and lack of inclusive digital design. Since women with disabilities experience higher levels of poverty and exclusion than both men with disabilities and women without disabilities, they face double discrimination. Therefore, **dimensions such as economic status, location, disability status and age interact with gender to show a more nuanced picture of access.**

Smartphone ownership and use is an often-used indicator of access to technology. The GSMA report in 2022 displayed that the gender gap on this has shrunk from 25% to 16% between 2017 and 2021[35], with a UNICEF report showing a less than 12% median gap in surveyed Global South Countries.[36] However, this gap remains significant. Since smartphones are heavily used to complete daily tasks that involve working with AI-enabled systems and have become even more important for those who have vision or hearing impairments, compounding the present access gap for women who also belong to other marginalized groups.

However, **what these measurements do not tell us is if there is a gender gap in access to information about what AI is and how it works**. Being able to gauge and track progress in terms of access to information about AI, its uses and its risks would be a useful indicator. Understanding what AI is provides a fundamental base knowledge that is necessary to gain an interest in working in the AI field. Moreover, indicators such as access to the internet do not necessarily align perfectly with the attainment of digital skills.[37] Therefore, understanding educational disparities around AI training is important to the discussion.

c. Education and Training in AI

If access to AI knowledge and tools is a fundamental prerequisite to participate in the field of AI, formal education in AI related disciplines is the next step in moving up the participation ladder. In this sense, **regularly collected data revolves heavily around graduates from tertiary STEM programs, computer science programs, and some skill sets around programming tasks.** However, there is not even an agreed upon definition of what is included under STEM, and is in any case is a far broader term than AI specifically.[38]

Available statistics display how **the definitions of the STEM field itself make comparisons across studies and time difficult.** Moreover, more data is available in the Global North. The wide variation in findings about gender gaps in education is attributable to the type of sources, the years collected, as well as the definitions about how to define AI relevant fields and levels of education. For instance, in a survey conducted in Europe, the gender gap ranges anywhere from over 42% in Belgium to less than 15% in Bulgaria for graduating bachelor's students.[39]

32 Sey and Hafkin, 2019, pg. 164.

33 ITU, 2023b.

34 Gillwald & Partridge, 2022

35 GSMA, 2022

36 Survey data from 2017-2021 (UNICEF, 2023).

37 UNICEF, 2023.

38 Sey and Hafkin, 2019, pg. 170 & 175.

39 Maslej et al. 2024

Table 1: Statistics on AI Gender Gaps in Education[40]

Report Source	Statistic	Date Source
UNESCO	10% gender gap in computer science graduates	UNESCO Science Report
UNICEF	11% difference in median youth (15-24) digital skills by sex in surveyed Global South countries	Multiple Indicator Cluster Surveys (2017–2021) and Demographic and Health Surveys (2015–2021)
WEF	11.5% gender gap in STEM degrees awarded	LinkedIn Data
HAI	19.5% gender gap in AP computer science exam takers in high school in US	Code.org
WEF	20% gender gap in AI and big data course enrollment	LinkedIn Data
HAI	23.7% gender gap computer science master's degrees awarded in US	Taulbee Survey
HAI	27.7% gender gap computer science bachelor's degrees awarded in US	Taulbee Survey
HAI	27.9% gender gap computer science PhD degrees awarded in US	Taulbee Survey

Table 1 highlights some of these differences and reveals that the data collection is inconsistent, with an up to 17% variance in women computer science graduates depending on the reporting authority for example. For the purposes of comparison, the statistics found have been mostly translated into a "gender gap", i.e. the difference between the actual percentage participation of women in the AI related education and the 50% participation "party" that would signify gender equity on this front.

Gender disparities in access to, and attainment of, education and training around AI leads to a gender gap in IT skills.[41] For instance, although digital skills are arguably hard to measure, a strong gender gap has been observed, particularly when it comes to more specialized digital skills. A survey of 50 European countries from 2019-2021 found that over twice as high a percentage of men (6.5%), compared to women, (3.2%), could write a computer program using a specialized programming language.[42] In this, as in other sectors, women with marginalized identities and characteristics are even less represented, and face more and specific barriers to access education.

There is a need for a more focused and standardized measure of gender equality in educational attainment and training in AI to understand if – and where – progress is being made for women and girls's participation in the AI landscape. These measures should include, systematically, an intersectional perspective. In order for women, in all our diversity, to make gains in representation in the AI related workforce, educational prerequisites are vital and thus tracking equality in these spaces is key to understanding where structure gaps may be present.

d. AI Workforce and Leadership

Statistics around gender dynamics in the workforce and leadership are complex and varied. Some statistics focus on the proportion of women employed in the technology sector, computer science related departments at universities or research institutes and, more specifically, in leadership positions within organizations. The data retrieved from reports for this outlook study come largely from surveys of employees in specific areas or larger data gathered from LinkedIn. For the purposes of comparison, the statistics found have been mostly translated into a gender gap, i.e. the difference between the actual participation of women in the AI workforce and the 50% "party" that would signify gender equality on this front.

As will be displayed over the next pages, much like with educational attainments, **a focus on broader STEM data masks larger gaps in AI-specific fields**. For instance, there is a 21% gender gap in science R&D positions versus a 38% gap in AI research positions. Similarly, there is a 15% gender gap for ICT professionals, but a 44% gender gap for software development professionals. Gender gaps widen

40 WEF, 2023; Pal et al., 2024; Maslej et al. 2024; Schneegans et al., 2021; UNICEF, 2023

41 West, Kraut & Chew, 2019

42 UNESCO, 2023b

significantly from entry-level to leadership positions in tech and as the focus narrows to AI companies, with a 23% gender gap for Director positions at STEM workplaces versus a 32% gap for C-Suite positions at AI startups. The need for more exact data to track progress towards gender equality in the AI ecosystem is clear from these examples.

Academic and Research Workforce: The proportion of women who earn doctoral degrees in AI and computer science (CS) compared to men continues to be minimal.[43] In AI conferences, women authors make up only 18%, an alarming percentage.[44] This is further confirmed by UNESCO, showing that women make up only about one third of researchers in science.[45] This is reflected in academia as well, with women making up only 16% of tenure-track faculty who conduct research on AI.[46]

Measured participation levels of women in academic and research and development (R&D) positions in AI vary depending on the definition and surveyed location. **As the statistics get more specific to the field of AI, the gaps widen.** Some highlights are found in Table 2.

In looking at academic achievement, only about 20% of data and AI researchers listed on Google Scholar in the UK are women, with only 11% of the "highly published" researchers being women. In moving to data science platforms, only 17% of those on platforms such as DS Central, Kaggle and OpenML are women.[47] An OECD AI Policy Observatory Post found that in 2022 only one in four researchers publishing on AI worldwide (from Scopus) was a woman and that roughly 45% of AI publications worldwide list at least one woman, compared to almost 90% that list at least one man as a co-author.[48]

Table 2: AI Academic and Research Position and Gender[49]

Report Source	Statistic
UNESCO	Worldwide, there is a 21% gender gap in science R&D positions
HAI	There is a 22% gender gap in computer science faculty hires
HAI	In North America, there is a 24.4% gender gap for computer science, computer engineering and IT faculty
AI Now	There is a 30% gender gap in AI professors
AI Now	There is a 32% gender gap in authors represented at leading AI conferences
UNESCO	Worldwide, there is a 38% gender gap in AI research positions

Corporate Workforce: Overall, the statistics on women in the AI workforce are grim. According to a World Economic Forum report, women's representation in AI is only at 30% in the professional realm.[50] **There is a striking difference in the gender gaps reported in reports, depending on how they define the field, and these gaps widen as the criteria becomes more specific.** Table 3 provides highlights.

For instance, AI Now reports a 35% gender gap in AI researchers at Facebook and 40% gender gap in AI researchers at Google (West et al., 2019). An important, but less reported, metric is the dropout rate of women in STEM or AI fields. The difficult work environments and challenges to success in the AI field lead many women with relevant academic training to leave the field. According to the World Economic Forums survey from LinkedIn data, there is almost a 7% difference between women who graduate with a STEM degree compared

43 Zhang, 2021

44 Collett et al., 2022

45 UNESCO, 2024b

46 West, 2023.

47 Young et al., 2021

48 Caira et al., 2023

49 Maslej et al. 2024; West et al., 2019; Collet et al., 2022; UNESCO, 2024a.

50 WEF, 2023

to those working in STEM one year after graduation. Women receive median salaries that are only 66% of the salaries of their male counterparts in the computer science workforce[51] and only 52 women are promoted for every 100 men in the tech industry.[52] A focus on intersectionality also reveals discrepancies, with women of color earning less than men, but also less than white women workers.[53] Thus, the AI pipeline is complex and deserving of more attention and metrics to understand how women enter, and also exit, the pipeline.

Table 3: Women's Participation in AI Corporate Workforce[54]

Report Source	Statistic
UNU	15% gender gap for ICT professions
UN Women	Over 20% gender gap in STEM professions (2019)
WEF	20% gender gap in AI "Talent" across sectors[55]
WEF	20% gender gap in AI professions (2023)
WEF	21.8% gender gap in STEM professions (2024)
UN Women	22% gender gap in ITC and AI field
Alan Turning	28% gender gap in AI and data science professions (2018)
UNESCO	30% gender gap in tech roles in major ML companies
UNU	36.5% gender gap in ML professions (2016)
UNESCO	44% gender gap for professional software developers

Given that AI has moved from an academic field to a dominant corporate force, the gathering of clear, regular and standardized statistics on the workforce is important. However, it is complicated by the lack of transparency in hiring practices of individual companies, as well as secrecy around workplace complaints of sexual harassment and discrimination.[56] **Companies need to do more to promote and implement transparency around their diversity data and on improving diversity in the field.** A survey published in 2018 of 32 leading tech companies found that many companies do not apply a gender lens to their CSR and philanthropy and that only 5% of 2017 philanthropic giving programs had an explicit focus on women and girls in tech.[57]

Governments also have an important role to play in pushing for more transparency at an organizational level. In looking at the results of several of the completed Readiness Assessment Methodology (RAM) Reports from UNESCO, for instance, governments are often lacking regulations that require diversity statistics be recorded or reported from private or academic entities.

A more qualitative understanding of how women come to be at AI focused companies is also insightful for the "AI pipeline" or how women move from educational training to jobs in AI. A study of the topic in Norway found that women often come into the field of AI via indirect or less traditional avenues. They may have backgrounds in non-STEM fields, but end up working in technology, bringing much needed expertise from different career

51 West et al., 2019

52 UN Women, 2023.

53 West et al., 2019 - retrieved from the 2010-12 American Community Survey by the American Institute for Economic Research. While this study is quite old, the finding points to the need for renewed and nuanced data collection on this issue.

54 UNESCO, 2024a; WEF, 2023; Sey & Kafkin, 2019; Young et al., 2021; Pal et al., 2024; West et al., 2019; UN Women, 2023

55 Metric developed by WEF and LinkedIn "determined by comparing the number of AI professionals to the total number of LinkedIn members worldwide. An individual is considered AI talent if they have explicitly listed AI skills on their profile and/or works in a job classified as an AI occupation. The concentration of AI talent is then calculated by taking the ratio of the number of AI talent by the number of LinkedIn members in that industry" (WEF, 2023)

56 West et al., 2019.

57 Wittemeyer, et al., 2018.

and educational experiences. This can lead to a mis- or under-representation of the true influence of women in this sector, as well as the perception that there is a lack of mentors in the field.[58]

Innovators and Start-up Workforce: As the AI landscape is dominated by innovative technology and the rapid rise of start-ups, understanding how women are entering this space is also important. Most of the available statistics, however, do not look at AI startups specifically. Given the trend seen throughout this section, mainly that the gender gaps appear even larger when focusing specifically on the AI field, more general figures can also be insightful, even if only showing the minimal gaps that exist. Table 4 shows some of these trends.

Table 4: Women's Participation in Venture Capital/Start Ups[59]

Report Source	Statistic
Pitchbook	In Europe, there is a 31.7% gender gap in the % of VC capital invested in 2024 that went to start ups with at least one female leader.
JP Morgan	In Asia, there is a 44.3% gender gap in the number of companies founded, co-founded, led or managed by women in 2021.
TechCabal	In Africa, there was a 48.46% gender gap in start-up funding raised for female-led startups between 2019 -2023.
Alan Turing Institute	In the UK, there is a 29.3% gender gap in the % of VC capital invested between 2010-2022 that went to AI startups with at least one female leader.

When focusing on technology-oriented startups, according to UNESCO, women and girls are 13 times less likely to file for an ICT patent.[60] Moreover, at the current rate, "it will be 2080 before women are involved in half of all patented inventions within the five largest IP offices."[61]

There are some other findings to note if we look at certain countries. For instance, the Alan Turing Institute[62] report specifically focused on women, AI and venture capital in the context of the UK between 2010-2023. They found that:

- Female-founded AI startups account for only 2.1% of VC deals involving AI startups
- The average deal capital raised by a female-founded AI company is 6 times lower than the average capital raised per deal by an all-male founder team in AI.
- All female-led AI startups raised only 0.8% of the total capital invested across all AI sectors in the UK.
- All female-led AI startups account for just 4.1% of the total number of companies funded.
- Even more stark, in AI software, all female led startups raised only 0.7% of total capital invested.

An important observation is t**he lack of representation of women in the venture capital firms themselves, potentially perpetuating a bias against all women-led AI startups**.

A study looking at the Latin America context focused on the fact that women are increasingly becoming entrepreneurs in STEM fields. They highlight several specific challenges for women in this space. Mentors seem to be important in STEM start-ups for women. For example, 67% of surveyed "STEMpreneurs" in Latin America report having a mentor, whereas only 52% of female entrepreneurs outside of STEM report so. The key main challenges reported by those surveyed include lack of financing or access to capital. Work/life balance and lack of confidence are additional high-ranking reasons where many believe women entrepreneurs have a particularly hard time.[63] **Having mentors in the startup world would give insights into how to manage some of these challenges, build a network and inspire confidence in possibilities of achievement**.

58 UN Women, 2023 and Corneliussen, 2019.

59 Pitchbook, 2024; Agwaibor, 2024; JP Morgan, 2023.

60 UNESCO, 2024a

61 OECD, 2018.

62 Alan Turing Institute, 2024.

63 WeXchange & IDB Lab, 2020

Leadership in the Workforce: Finally, there is an intense need for more women in leadership positions within the AI workforce and AI related government offices. **The state of women representation in C-suites and boardrooms is bleak, let alone in the technology and AI field.** Data consistently demonstrates a lack of gender-diverse boards. This is a multi-faceted situation that shows hurdles at every level. There has been progress, however minimal. In 2023, Fortune 500 listed that for the first time in 68 years, women ran over 10 percent of the listed companies.[64] However, despite such updates, with such slow progress, there is consensus that "the world will not reach parity until at least 2045, over twenty years from now".[65]

Much in line with the rest of the findings of this section, the data on leadership and gender in the AI field much depends on the specificity of the data and is drawn heavily from LinkedIn Data. Table 5 provides some highlights. For the purposes of comparison, the statistics found have been for the most part translated into a "gender gap", i.e. the difference between the actual percentage leaders/position holders that are women and the 50% participation "party" that would signify gender equity on this front.

Table 5: Women's Leadership in AI Field[66]

Report Source	Statistic
WEF	23.3% gender gap for Director positions in STEM workplaces
WEF	30% gender gap Senior Leadership in technology, information and media sector
WEF	31.9% gender gap for c-suite positions in technology, information and media sector
UNU	32% gender gap in C-Suite for AI Startups
WEF	32.2% gender gap for VP positions in STEM workplaces
UNU	36% gender gap in government ICT ministers
WEF	37.6% gender gap for C-Suite positions in STEM workplaces
UNU	38% gender gap in government telecom regulators

Given that there is only a 12.3% gender gap for entry level positions in the technology, information and media sector according to the World Economic Forum,[67] **the significantly higher leadership gender gaps are significant and worth deeper and more consistent review**. Research consistently demonstrates that gender-diverse boards lead to better decision-making, improved financial performance and enhanced corporate reputation. Moreover, having women in leadership roles sets a positive example for aspiring female professionals, encouraging greater gender equality and representation in all levels of the workforce. In short, ensuring gender diversity in boardrooms is not only just a matter of fairness; it's a strategic imperative for business success and societal progress.

e. Other Gender Based Barriers for AI Participation

Efforts to promote women's equal participation in AI are focused largely on targeted interventions to strengthen access to education in STEM and increase professional opportunities in the AI ecosystem. While this is certainly a key policy intervention without which gender equality won't be achieved, there are other factors that are equally important and must be addressed as well, including:

- **The work and achievements of women in the tech sector are often made invisible or minimized whereas those of men generally receive more support and financial resources from the outset.** In the words of the UN Secretary General, "centuries of patriarchy and damaging stereotypes prevent women innovators from getting the recognition they deserve… women are credited less for their achievements, win far fewer prizes and receive less research funding than men, even when they have the same conditions. This must change. The male chauvinist domination of new technology is undoing decades of progress on women's rights."[68]

64 Hinchliffe, 2023.
65 Deloitte, 2022
66 Sey & Kafkin, 2019; WEF, 2023; Pal et al., 2024;
67 Pal et al., 2024
68 UN Media Office, 2023

- **In the male dominated tech sector oftentimes patriarchal practices make women feel unwelcome, uncomfortable and even unsafe.** For instance, in the US, women leave the tech and engineering sectors at twice the rate of men.[69] As much attention should be paid to supporting women to access jobs in the tech sector, as in policies to retain them and make sure they have equal chances to advance in their careers. It is essential to adopt institutional policies, at the highest level and in all departments in AI-related sectors, that sanction and redress misogynistic practices that push women out of the tech sector. Safeguarding mechanisms must be strengthened and effectively enforced to ensure women can study and work free from harassment and abuse.

- **Care work,** defined as "direct care for people (physical, emotional, psychological and developmental) as well as indirect care (e.g. household tasks, including collecting water and firewood, traveling and transporting)"[70] **is still carried out disproportionately by women as opposed to men, due to pervasive gender stereotypes.** Such social norms - which exist around the world in different degrees - and the lack of sufficient and good quality public care systems is one of the main causes of gender inequality in employment. While this is not specific to AI and the tech sector, it must be recognized that gender equality won't be achieved without addressing this issue.

2.2: Policy and Resources for Gender-responsive AI

Improving gender equality in and through AI requires more than technical solutions. It requires government and organizational political commitment at the highest level and concerted efforts, actions and budgets to make the structural changes required to increase gender equality in the field of AI, the AI lifecycle and related governance mechanisms.

Moreover, as a report from AI Now states, "a focus on fixing technical systems in isolation, without examining the broader context of their use and the power dynamics that attend such use, is not only limited in its intervention: it can actively cause harm."[71] There is a need, therefore, to examine policy environments related to AI, who participates in decision-making and how inclusive the processes are, in order to develop a holistic approach to the issue of gender equality and AI.

Section 2.2 examines the question *how are gender equality concerns being included in efforts at AI governance?*

a. The Landscape of Gender and AI Policy

Policy aimed at AI governance has been developed and discussed around the world in recent years. However, **current major efforts to analyze the policy landscape for AI and evaluate AI readiness do not adequately take gender dimensions into account.** For instance, *Oxford Insights' AI Readiness Index*[72] uses only two indicators related to gender for inclusion in creating their national level indices: equality of internet usage and percentage of women and girls in STEM.

Moreover, the *AI and Democratic Values Index*[73] only includes an indirect examination of gender, looking at concepts such as "fairness" in the evaluation of a country's AI policy landscape. However, gender equality is not explicitly mentioned in the method. They do, on the other hand, use indicators related to the endorsement and implementation of UNESCO's *Recommendation on the Ethics of AI*, which itself includes a section on gender policy. While interesting, such an indicator is not nuanced enough to recognize whether mechanisms for gender mainstreaming and women's empowerment are being included in a country's AI policy.

UNESCO Recommendation on the Ethics of Artificial Intelligence #88: *Member States should have dedicated funds from their public budgets linked to financing gender-responsive schemes, ensure that national digital policies include a gender action plan, and develop relevant policies, for example, on labour education, targeted at supporting girls and women to make sure they are not left out of the digital economy powered by AI. Special investment in providing targeted programmes and gender specific language, to increase the opportunities of girls' and women's participation in science, technology, engineering, and mathematics (STEM), including information and communication technologies (ICT) disciplines, preparedness, employability, equal career development and professional growth of girls and women, should be considered and implemented.*

69 The Economist, 2019

70 UN Sustainable Development Group (2024b).

71 West et al., 2019.

72 Oxford Insights, 2023.

73 CAIDP, 2024.

In another example, the *OECD AI Observatory Database of AI Policies*[74] has over 1000 policy initiatives from 69 countries continually updated and has identified "women" as a "social group especially emphasized" within 47 policy initiatives. The immediate observation is that this is an extremely small percentage of initiatives, further highlighting a trend of lack of focus in the AI policy space on gender equality and women.

Beyond the fact that concerns around gender equality and women's empowerment are underrepresented in the AI governance discourse and practice, due to a lack of standard benchmarks, inclusive policy-making practices and committed research, **it is also unclear HOW and WHEN women, in all their diversity, are being included in policy agendas on AI**. The following examples highlight some initial analysis in this space.

b. Regional and Country Level Policy Analysis Findings

Identifying Components of a Gender Perspective in the AI Strategies of Latin America: In 2022, a review of public and available AI strategies or policies from Argentina, Brazil, Colombia, Chile, México and Uruguay was conducted. The goal of the analysis was to identify key components that would ensure a cross-cutting gender perspective.[75] From the findings, these components were identified as minimal standards for an effective gender approach:

1. Recognition of the relevance of human rights normative frameworks for AI strategies.
2. Explicit commitment to incorporate gender perspectives.
3. Incorporation of, at least, gender disaggregated data in the analysis.
4. Actions to mitigate risks regarding algorithmic biases
5. Assurance of gender parity in all AI related committees, forums and other decision-making instances.

While all these strategies mentioned a human rights approach, this analysis found that not all of them explicitly incorporated a cross-cutting gender perspective. Argentina, Brazil, Mexico and Chile included specific actions in this area. However, none of them mentioned indicators or budgets, without which it is impossible to make advancements in gender equality. While Colombia did include an approximate budget section, the budget did not say what percentage would be allocated to gender equality actions. Additional analysis is required to review the most recent initiatives in further detail.

An Analysis of Documents Identified by the Global Index on Responsible AI: For this outlook, an original and up-to-date policy analysis was conducted to show the current state of how gender, girls and women are included (and excluded) from the global discourse at a national policy level. For the analysis, the Global Index on Responsible AI (GIRAI) was used as a sample source. The GIRAI compares country level progress on many areas relevant to AI. In the calculation of the index, gender equality is included under the category of "Human Rights and AI" and is broken down by Government Frameworks, Government Actions and Non-state actor presence between November 2021 and November 2023.

Notably, the GIRAI reported that **"gender equality" was one of the lowest performing areas of the index**, **highlighting the urgent need to scale up efforts at reporting, standardization and implementation of policy related to this topic.**[76] Only 24 out of 138 countries assessed had government frameworks in place, with 37 governments demonstrating evidence of initiatives promoting gender equality in the context of AI, 67 had non-state actors in place working on the topic. Only Canada, the Netherlands, Australia, Ireland, Portugal, Japan, Germany and the UK scored above 50% on the composite indicator for the category.[77]

To understand how women are being represented in the AI governance discourse and throughout the design, practice and evaluation of AI governance mechanisms, an analysis of the 20 available government frameworks that the GIRAI identified as including gender equity (as of September 2024 - Australia, Brazil, Bulgaria, Canada, Chile, Croatia, Dominican Republic, Germany, Greece, Hong Kong, Ireland, Italy, Japan, Netherlands, Saudi Arabia, Serbia, Singapore, Slovenia, Spain, USA) was undertaken for this outlook. The policy analysis method used has been previously used to analyze hundreds of policies across three continents.[78] The process included the steps of (1) document source identification and retrieval; (2) identification of key search terms related to

74 OECD.AI, 2021. Note: the database currently runs until 2021.

75 Revilla, 2022.

76 See the Global Index on Responsible AI (GIRAI) (2024), p. 38.

77 Adams et al., 2024. Note the Global Index on Responsible AI is a project of the Global Center on AI Governance (https://www.globalcenter.ai/)

78 See Bretag et al. (2011); Eaton et al. (2019); Miron et al. (2021); Möller (2022); Moya (2024), Moya & Eaton (2024)

gender equality[79]; (3) data extraction; and (4) analysis.[80] The documents examined were published across half a decade, between 2019 and 2024. From the analysis, several key findings were gathered.

Absence of women or girls in policy documents: Some policy documents had no mention of any of the designated search terms. Policy documents included in the analysis from Hong Kong[81], and Japan[82] showed an absence of specific actions related to girls, women, and gender.

Marginalization of equity, equality, girls and women in policy: The analysis showed that when girls or women were mentioned specifically, these mentions were sometimes relegated to appendices, rather than appearing in the main text of the policy documents. As one example, in Canada's Directive on Automated Decision-Making, the terms "equality" and "gender" were mentioned in appendices, but absent from the main text of the document.[83]

Box 1: Good Practice Highlight

The "National Artificial Intelligence Strategy for Ireland" (2021) stands as an exemplar of excellence in terms of a policy document that elevates and amplifies girls' and women's contributions to artificial intelligence. In addition, it provides clear direction on how to further support girls and women in STEM fields, including AI.

One notable exception to this finding was found in the *National Artificial Intelligence Strategy for Ireland*.[84] The Irish strategy includes explicit mentions of girls and women in positive and advocacy-oriented ways. For example, there is an entire section dedicated to "Women's Participation in AI" (See Box 1).

Lack of specificity in definitions: When terms such as "equity" or "equality" were mentioned in the documents, these were almost never defined. What is considered equal or equitable in one country or context, may not be considered so in another. As an example, in *Australia's AI Ethics Principles*[85], the term "equitable" is mentioned in a generalized manner, but is not defined. Other policy documents showed a similar tendency in terms of lack of definitions.

Acknowledgement of challenges, without a commitment to improvement: In the policy documents include the key search terms, some acknowledged gender as an area that merited further attention, but did not indicate any commitment to remedy or improve the situation.

Given that nearly all document exmined included only superficial mentions of gender as it relates of AI, without adequte plans for correcting AI related inclusion, harms or biases, serious efforts need to be made to identify gaps in government stragies and policies related AI to include gender as a cross-cutting issue.

c. A Gender Policy Analysis Agenda for AI

The above-mentioned analysis are important steps, however current efforts are insufficient. A comprehensive global AI policy analysis focusing on gender equality is needed that builds on the findings presented in this report, as well as previous ones. One reason that a more comprehensive analysis of AI policy is needed is to provide an evidence base for an AI and gender framework that can serve as a blueprint that outlines minimum standards for AI-related strategies, policies and governance mechanisms in general. Such a framework should be built on recommendations and good practices towards gender equality in other sectors, and relevant international and national obligations and commitments.

For AI strategies and policies to be effective and sustainable, they must intentionally demonstrate a commitment to gender equality. To that end, AI policies should incorporate these components at a minimum:

1. An explicit commitment in the AI strategy/policy, and at the highest level, to advance gender equality. Including introducing affirmative action strategies and setting actionable targets for increasing women's representation in the AI ecosystem.
2. Inclusion of gender-specific indicators and data across the board.
3. Measures to ensure gender-parity in all AI-related committees, forums, working groups, among others.

79 Girls; Women; Female; Gender; Equity; Equality

80 Limitations of this analysis: The sources analyzed for this study are neither exhaustive, nor are they necessarily completely up-to-date. Also, our search terms were not exhaustive. Three of the identified documents had broken URLs. We did attempt to locate these documents in another way.

81 The Government of Hong Kong Special Administrative Region of the People's Republic of China: Digital Policy Office, 2024.

82 Government of Japan, 2019.

83 Government of Canada, 2023a.

84 Government of Ireland, 2021.

85 Australian Government, n.d.

4. Gender-sensitive budgeting for the AI strategy/ policy, including specific budget lines directed towards empowerment and capacity development of women and their representative organizations.
5. Linkages with institutions in charge of gender equality in different policy areas.
6. Measures to ensure the effective participation of feminist and women's rights organizations in AI strategies and policies design, implementation and evaluation.
7. A monitoring and evaluation framework that incorporates SMART[86] indicators related to advancements in gender equality, throughout all the stages of AI strategy and policies.
8. Safeguarding mechanisms to ensure those spaces where AI strategies and policies are discussed, at various stages, are safe spaces free from harassment; and that adequate mechanisms and procedures are in place to respond to concerns in a timely manner.

The finding from Section 2.2.b. that most countries have not included a gender perspective in their government frameworks related to AI, and that those that have included gender have done so in a largely superficial manner, is problematic and requires correction. UNESCO Readiness Assessment Methodology (RAM), for instance, provides an opportunity to help member state identify these gaps in their AI strategies and integrate gender in a cross-cutting manner as they design and implement AI related policy. Without understanding these weaknesses around gender inclusion in policy and strategy documents, minimal progress can be made.

2.3: Identifying Risks to Gender Equality from AI Systems

Section 2.3 discusses the types of risks that AI presents to gender equality and highlights some key examples of where AI-enabled technologies are disproportionately affecting women and girls.

MIT, for example, has developed an AI Risk Repository.[87] It is a comprehensive living database of over 700 AI risks categorized by their cause and risk domain, with seven overarching domains of risk identified:

- **Discrimination and Toxicity**: Unfair discrimination and misrepresentation; Exposure to toxic content; Unequal performance across groups;
- **Privacy and Security**: Compromise of privacy by obtaining, leaking or correctly inferring sensitive information; AI system security vulnerabilities and attacks;
- **Misinformation**: False or misleading information; Pollution of information ecosystem and loss of consensus reality;
- **Malicious Actors**: Disinformation, surveillance, and influence at scale; Cyberattacks, weapon development or use, and mass harm; Fraud, scams, and targeted manipulation;
- **Human-Computer Interaction**: Overreliance and unsafe use; Loss of human agency and autonomy;
- **Socioeconomic and Environmental**: Power centralization and unfair distribution of benefits; Increased inequality and decline in employment quality; Economic and cultural devaluation of human effort; Competitive dynamics; Governance failure; Environmental harm;
- **AI System Safety, Failures and Limitations**: AI pursuing its own goals in conflict with human goals or values; AI possessing dangerous capabilities; Lack of capability or robustness; Lack of transparency or interpretability; AI welfare and rights.

UNESCO Recommendation on the Ethics of Artificial Intelligence #90: *Member States should ensure that gender stereotyping and discriminatory biases are not translated into AI systems, and instead identify and proactively redress these. Efforts are necessary to avoid the compounding negative effect of technological divides in achieving gender equality and avoiding violence such as harassment, bullying or trafficking of girls and women and under-represented groups, including in the online domain.*

86 Specific, Measurable, Achievable, Relevant, and Time-bound

87 Slattery et al., 2024.

While all these categories could potentially impact women in specific ways, particularly those with intersecting marginalized identities or characteristics, an analysis of the database identifies 27 risk entries out of 702 total entries in the repository that mention "Women; Girls; Female; or Gender" by name in their description. These **risks are overwhelmingly categorized under Discrimination and Toxicity (20),** focusing on bias or misrepresentation in data leading to stereotyping, marginalization, discrimination and unequal performance of AI systems based on demographics such as gender. While less in number, there are also examples from Privacy and Security (4), Human-Computer Interaction (2), and Socioeconomic and Environmental (1) that flag gender as a factor in AI risks. These include for example privacy violations due to AI generated predictions about gender, the further entrenchment of stereotypes due to female voice assistants and the cross-over from human-computer interaction, or the potential to exacerbate socio-economic inequalities related to gender.

Box 2: A Note on Fairness:

Key to this conversation is, what is meant by "fairness"? While there are different approaches to the term in law, the social sciences, quantitative fields and philosophy, "ML researchers and practitioners tend to use a quantitative perspective as the primary lens for fairness." When thinking about fairness between groups, this would mean categorizing data or uses by these groups and working towards parity in how they are treated in terms of outcomes in the system. This can be problematic when fairness is more nuanced than the groupings allow (such as the overlap between fairness and justice), or when there is intersectionality in relation to groupings and fairness.

When moving from "fairness" to "justice" in AI systems, Smith[88] writes, "a justice approach considers how certain groups are oppressed or marginalized in the particular context and explores how the AI system can advance equity, rather than perpetuate a status quo that may oppress or marginalize certain groups." This, therefore, moves beyond making sure groups are treated equally, in terms of algorithmic outcomes, to positively pursuing equity through AI systems.

This report takes the approach to fairness from the UNESCO Recommendation on the Ethics of AI as **implying an "inclusive approach to ensuring that the benefits of AI technologies are available and accessible to all**, taking into consideration systems, different language groups, persons with disabilities, girls and women, and disadvantaged, marginalized and vulnerable people or people in vulnerable situations."

From a more technical perspective, these risks mainly relate to the overarching concern of **bias in machine learning and algorithmic systems** and how that impacts fairness (see Box 2). The biases observed in AI systems tend to reflect the biases in our society. The speed and breadth that AI systems have in terms of impact means they have a profound potential to entrench current forms of discrimination and bias along gender lines if not addressed.[89] Suresh and Guttag[90] identify seven sources of harm from bias in ML systems: historical bias, representation bias, measurement bias, aggregation bias, learning bias, evaluation bias and deployment bias. Thus, bias is a concern that continues throughout the AI lifecycle.

These biases are also partially a result of a relative lack of women's representation and meaningful participation in the AI lifecycle (see Section 2.1) and governance processes (see Section 2.2), which results in a lack of relevant data, or gender non-disaggregated data, being labeled and fed into AI training systems.

Smith and Rustagi[91] examined, for instance, publicly available instances of bias in AI systems. Of the around 133 AI systems identified in their analysis, 44.2 percent demonstrated gender bias. They highlight six major impacts of gender bias in AI systems including:

- Lower quality of service for women and non-binary individuals,
- Unfair allocation of resources information and opportunities for women,

88 Smith, 2020.

89 West, Kraut and Chew, 2019.

90 Suresh and Guttag, 2021.

91 Smith and Rustagi, 2021.

- Reinforcement of existing, harmful stereotypes and prejudices,
- Derogatory and offensive treatment or erasure of already marginalized gender identities
- Determinants to physical safety
- Health hazards

The Berkeley Haas Center for Equity, Gender and Leadership has created a tracker of examples for algorithmic bias with many of their examples around gender identified in Table 6.[92] While this is not an exhaustive list, it is illustrative of some of the major impacts of algorithmic bias on women.

Table 6: Examples of Gender Bias in Algorithms from Adapted from Berkeley Haas Center for Equity, Gender and Leadership

Professional services (hiring & marketing)	Amazon's experimental hiring AI was shown to be biased against female candidates.
	Gild's (online tech hiring platform) candidate ranking algorithm was found to be biased against women.
	Facebook allowed advertisers using its platform to use 'affinity profiling', thereby excluding users based on demographic traits (such as gender).
	Facebook's ad delivery process allows housing and employment ads to be skewed along demographic criteria (such as gender).
	Google's ad algorithm offered men and women different employment opportunities.
	Men are 5x more likely than women to be offered online ads for high-paying executive jobs.
Generative AI and Recognition Tools	Google translate had a male default, typically gendering gender-neutral pronouns as male.
	Three widely used algorithms for Natural Language Processing (NLP) failed to recognize «hers» as a pronoun, but correctly identified «his».
	Voice enabled assistants are likely to have a higher accuracy rate when understanding white American male voices, as compared to any other identity.
	Two automatic speech recognition systems performed worse on talkers based on their race, gender and dialect.
	Voice recognition systems in cars have a harder time understanding female voices than male voices.
	Google's pre-trained NLP model, BERT, is gender biased.
Healthcare	The public benefits algorithm used in Colorado denied Medicaid to pregnant women.
	Users' gender and computer experience impacts the performance of voice-activated medical tracking application
	An AI model used to predict kidney function decline in patients treated in U.S. Veterans Affairs hospitals performed worse when tested on women.
Education	Computer program used to screen applicants to St. George's Hospital Medical School exhibited gender and other bias.
	Admissions algorithm used by the University of Texas, Austin was dropped over historical racial and gender bias concerns
Data Bias, Harm and Stereotyping	MIT '80 Million Tiny Images' image database taken down over derogatory (and misogynistic) labels.
	ImageNet, a widely used image dataset contains problematic labels.
	Women are underrepresented on Google Images for several occupations.
	Algorithms showed a tendency to associate women with shopping and men with shooting.
	Word embeddings trained on Google News articles exhibit female/male gender stereotypes.
	Searching for the term «Black girls» led to sexually explicit content being displayed.
	Stable Diffusion AI generated images display even more negative bias towards representing judges as women compared to the real of the percentage of female judges in the US.[93]

92 See - https://haas.berkeley.edu/equity/resources/playbooks/mitigating-bias-in-ai/ (Berkeley Haas Center for Equity, Gender and Leadership, n.d.)

93 Nicoletti & Bass, 2023.

Women with additional marginalized identities, based on race, sexuality or disabilities, among others, generally face increased risks and harms from AI systems, due to the use of historical data to train such systems. For instance, AI hiring systems training on data from past employees and hires have shown to discriminate both against women and against applicants with disabilities. Using these AI-enabled systems, thus, may amplify the exclusion of women and women with disabilities - who already face enormous barriers to access formal employment - from the labor force, both in the hiring process and in career advancement.

A second major concern in terms of the impact of AI on gender equality revolves around **intentional or unintentional misuse of AI-enabled tools that results in physical or mental harm** to women and girls. Numerous examples have been documented related to the impacts of misused AI-enabled systems on instances of gender-based violence, quality of healthcare, access to social protection services and decision making in the justice system. UNESCO and The International Research Center for Artificial Intelligence (IRCAI), for example, documented this in terms of the impacts of Large Language Models (LLMs) on regressive gender stereotypes.[94] Deep-fakes and AI-enabled synthetic content are increasing these impacts on an unprecedented scale.

Indeed, over the last ten years, the OECD's AI Incidence Monitor has logged over 400 incidents or hazards related to AI with specific respect to "women"[95] as a group.[96] Moreover, the AI Incident Database[97] has close to 150 incidents logged under the classification "sex" and over 400 under the search term "Women". Although this represents only recorded cases and the actual number may be higher, the range and scope of the risks is evidenced through these, among other records.

2.4: Specific Case Studies on Gender AI Inequality

The following sections highlight some key case studies from different sectors and applications where AI-enabled technologies are disproportionately affecting women and girls.

a. (In)Equitable Healthcare

AI has made incredible advances in healthcare in recent years. At the same time, there is an alarming potential of AI to enshrine inequities in healthcare for women across the globe. The first main drivers of this fact are that the **data needed for women's health in the 21st century is missing**.[98] Some troubling facts include:

- Women were not included as subjects in clinical research in the US until 1993[99] and are still underrepresented in clinical trials.[100]
- 2022 studies found a majority of biomedical research still relies predominantly on male mice, [101]which perpetuates the cycle of gender-skewed data.
- The gathering of data is still inconsistent. For example, in COVID-19 vaccine trials, 28.3% of publications did not report sex distribution among participants. Only 8.8% of the studies provided sex-disaggregated Vaccine Effectiveness estimates.[102]

Box 3: Good Practice in Design

The EMory BrEast imaging Dataset (EMBED) - this project, to create a racially diverse, granular Dataset of 3.4 million screening and diagnostic mammographic images,[103] depicts the progress that can be made in AI for women's healthcare when steps are taken to make sure datasets are inclusive and representative, not adding undue burdens or benefits to any one part of the population.

94 UNESCO & IRCAI, 2024.

95 Note: OECD AI Incidents Monitor (AIM) (OECD, 2024a) has over 12,000 entries automatically collected from news sources with little human review (Velázquez et al., 2024).

96 Using keywords Women, Girls, Female and Gender

97 The AI Incident Database (AIID) has over 3,500 entries (Responsible AI Collaborative, 2024). The reports are derived from user uploads and reviewed by a main editor, thus less automated than the OECD database. https://incidentdatabase.ai/

98 FemTechnology & Women at the Table, 2023.

99 Lui & Mager, 2016.

100 Fultinavičiūtė, 2022

101 Mazure & Jones, 2015.

102 Sulis et al., 2023.

103 Jeong et al. 2023.

Existing AI models in healthcare are trained on limited datasets that fail to adequately address the experiences of girls and women, and particularly those with disabilities and those who are from racialized or minoritized groups. Given the velocity of adoption for medical AI applications, not rectifying these foundational gaps threatens to perpetuate and even amplify these disparities at an unprecedented scale. As an example, a study in 2021 found that in 34 out of 36 studies, the AI system in place for screening breast cancer from mammograms did worse than a single radiologist.[104] Whether due to a lack of data from women, or a relative disinterest in using AI for women's health compared to in other heath subfields, the inaccuracy of these systems represents a major health risk. Box 3, on the other hand, identifies a concerted effort to improve AI-enabled medicine for women. This example displays that when data considerations are made, progress is possible.

The example of gender bias in a state-of-the-art AI system used in hospitals to predict liver disease from blood tests further exemplifies this problem. In this trial, the overall accuracy of the AI system was 70%. However, when the data were disaggregated by sex, 44% of cases were missed in women as opposed to only 23% in men.[105] Table 7 is adapted from Cirillo et al. (2020) and provides an overview of examples for AI bias in healthcare.[106] Given the plethora of ways and AI is being increasingly used in healthcare, the reverberations for these biases and missing data will be global. [107]

Table 7: Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare - PMC[108]

Inherent or Historical Bias	**Even when data is accurately collected and sampled, models might yield undesired results due to pre-existing societal biases in the world**. e.g. making the mistake of associating HIV primarily with gay and bisexual men because of its higher occurrence in this group.
Representation or Sampling Bias	**When certain segments of the data input are not adequately represented**. e.g. a large part of genomics research predominantly focuses on European male demographics, sidelining other ethnic groups.
Data Proxy Bias	**When the data collected serves as an indirect measure for the desired attributes.** e.g. using various clinical, social and cognitive indicators to identify early stages of schizophrenia, even though gender differences can influence the manifestation of these indicators and their related psychosis risk.
Generalization / Aggregation Bias	**When a universal model is applied to groups that have distinct underlying conditions**. e.g. despite diabetes' variable interpretations across different ethnicities and sexes, the widespread use of hemoglobin A1c (HbA1c) levels to diagnose and track the disease.
Evaluation or Benchmarking Bias	**When the data used to test or benchmark an algorithm is not a good match for the intended audience**. e.g. the underperforming of automated skin cancer screening tools, because of a lack of diverse skin type data.[109]
Modeling Bias	**Bias can be inadvertently or deliberately embedded into an algorithm, especially when relying on improvised solutions**. e.g., when a commercial health prediction algorithm used healthcare costs as an indicator for health condition without factoring in prevalent disparities in healthcare accessibility, it displayed considerable racial prejudice changing predictions of genuine need.

A second driver of inequitable access to the benefits of AI driven healthcare is **the meaningful participation of women throughout the AI lifecycle**, and in related decision-making instances. This is exemplified when discussing AI and gender equality specifically in sexual and reproductive health and rights (SRHR) - without which gender equality cannot be achieved - and which are already facing a pushback in several countries. According to the World Health Organization (WHO), in the area of SRHR "AI has emerged as a transformative force for health system efficiencies, but has also introduced critical risks and rights-related consideration, including potential impact on bodily autonomy and amplification of targeted disinformation, in a field already prone to the effects of ideologically driven narratives."[110] The WHO also affirms that while some of the risks are common to health care in general, the norms and power relations that are reflected in SRHR have a specific impact for women; it categorizes

104 Freeman et al., 2021.

105 UCL News, 2022

106 Maleki Varnosfaderani & Forouzanfar, 2024.

107 Burns et al., 2023.

108 Cirillo, 2020.; FemTechnology & Women at the Table, 2023.

109 Brancaccio et al., 2024.

110 WHO, 2024

these risks in six areas: data governance and bodily autonomy; data breaches; misinformation and targeted disinformation; data limitations and bias in AI; digital divide; and differences in context and awareness levels.[111]

b. Workforce Impacts in Employment and Discrimination

Globally, women in the labor force have lower participation levels in formal employment, earn less than men, spend more time on unpaid family care-giving responsibilities, occupy fewer senior positions, participate less in STEM fields and are more likely to hold less secure employment.[112] Women with other marginalized identities and characteristics face additional or increased barriers to employment, such as equal retention or advancement in employment and increased discrimination. Although AI-enabled technologies have the potential to enhance efficiency and objectivity in the workplace, they may also negatively impact the opportunities available to women in the workforce, as well as their status and treatment within the workplace. AI systems need to be developed and used in accordance with existing national laws and international labor standards - including those of the International Labor Organization (ILO) - and with efforts to correct for bias in the systems in order to both comprehend and mitigate any negative influence on gender dynamics.

Women are Missing in the AI Workforce, and the Data: According to a 2022 UNESCO Report, there is a dearth of female participation in AI-related occupations on a global scale. This represents a significant challenge for the future trajectory and development of AI systems. The lack of diversity in the development of systems has been demonstrated to result in a reduced probability of such systems catering to the needs of diverse users or aligning with human rights.[113] The workplace in the AI sector itself has been cited as a problem, as UN Women found that "22% of women in tech are considering leaving the workforce altogether given the prevailing masculine working culture reflected in impediments to advancing and other barriers including exposure to violence and harassment and lower wages."[114] This is underscored by the finding that within 12 years, half of women end up leaving their position in STEM fields.[115]

It is important to highlight the connection between (1) the lack of women in the AI workforce and (2) the impact of AI on women in the workforce. Missing women in certain economic sectors, not only deepens inequalities, but impacts negatively firm's economic outcomes and performance. It also means that data on women has historically been missing from hiring data, impacting the performance of AI hiring systems in terms of fairness and non-discrimination.

The perpetuation of existing gender biases: Biased algorithms in the context of hiring and workforce management present an initial challenge to ethical use of AI for gender equality. For example, algorithmic tools have been designed to direct job advertisements toward specific candidates or identify passive candidates for recruitment. These predictive tools parse and score resumes, thereby assisting hiring managers in evaluating candidate competencies in novel ways, using both traditional and non-traditional data sources.[116]

Box 4: Use Cases in AI and Hiring

Amazon's AI recruiting tool had been trained on resumes submitted over a decade, predominantly from male applicants. As a result, resumes that included women's colleges or terms such as "women's chess club" were downgraded. Essentially, "Amazon's system taught itself that male candidates were preferable." [117]

Gild, an online tech hiring platform, used not only traditional data for job recruitment, but also "social data" without taking into account that women tend to either have less time for online social engagement or sometimes use male names to avoid sexist, gender-specific safety concerns. Because of this lack of consideration, the system was predisposed to disproportionately impact female applicants.[118]

In **New York City**, the city council passed a law requiring that automated tools for employment decisions be required to conduct a "bias audit" before use. Improving transparency and awareness around bias in hiring and workplace-based AI tools. [119]

111 WHO, 2024

112 Colett et al., 2022

113 Colett et al., 2022

114 UN Women, 2023.

115 Ashcraft et al., 2016.

116 Bogen, 2019.

117 Dastin, 2018.

118 Smith and Rutagi, 2021.

119 New York City Council, 2021.

However, the **training of AI systems on historical hiring data may result in the inadvertent inheritance of gender and other biases present in that data** (See Box 4 for examples), including data proxy or representation sampling bias (See Table 7). For instance, if there has historically been a greater representation of men among those hired for specific roles, an AI system may subsequently demonstrate a proclivity to favor male applicants. Moreover, the longer known tendency for the use of masculine language in job advertisements[120], once fed into AI-enabled training systems, would perpetuate and expand gender stereotypes and undermine gender equality in recruitment and hiring.

Moreover, the opacity of AI decision-making systems renders AI algorithms complex and not readily interpretable, thereby impeding comprehension of the decision-making process. Such opacity may impede identification and mitigation of discriminatory practices within the AI system.

AI and Automating Women at Work: The second area of impact of AI on the workforce and hiring practices is the potential for automation. **The potential for automation powered by AI could affect women employment in a disparate manner.** There is a possibility that occupations where women are currently overrepresented, such as administrative and clerical roles, may be particularly vulnerable to disruption. In the absence of reskilling opportunities, this could result in elevated rates of job displacement among women.[121] While the emergence of GenAI will no doubt spread these impacts to traditionally "white collar" jobs as well, the uneven impacts of automating work on women need to continue to be monitored and measured.

Moreover, female-dominated jobs are also the places where algorithmic scheduling, used to optimize scheduling and task allocation, is being most used. The intense focus on optimization within these AI systems can increase work and income uncertainty[122] and work related stress. Additionally, women with disabilities - as well as other women with marginalized identities and characteristics - face increased risks and damages from the unregulated use of AI in employment. A recent publication by the OECD found, for instance, that the use of AI in employment processes poses increased and specific risks for persons with disabilities due to increased risk of a lack of reliability, errors and privacy violations, built-in ableist biases, inequity of use and lack of explainability.[123] While this report does not include specifically a gender perspective, it is clear that when the risks faced by women are compounded with the risks faced by persons with disabilities there could be highly negative outcomes and further marginalization in the labor markets.

c. Mainstreaming Stereotypes

The field of large language models (LLMs) and development of generative artificial intelligence (Gen AI) tools has witnessed a remarkable advancement in recent years, with machines now capable of producing text, images and even music that closely resemble human-created output. While these technologies offer significant advantages, they also present a risk of perpetuating and mainstreaming stereotypes that exist in training data or in their design (See Box 5).

Box 5: Digital Assistants and Gender Equality

As a UNESCO and Equals report on the topic highlights, the teams at the frontier of voice assistant technology were predominantly male. Given that lack of diversity, early decisions about design and testing have led to inequities in access and in promoting stereotypes. For instance, Google's voice assistant was 70% more likely to accurately recognize male voices.

Moreover, most voice assistants use female voices (87.6%).[124] This decision creates a signal that women are more obliging, docile and have no agency of their own. As harassment of digital assistants is also well noted, this risk normalizing sexual harassment and verbal abuse towards a female voice in the physical worlds as well. As voice assistants become increasingly used for diverse tasks in more parts of the world, this embedded stereotype risks proliferation.[125]

Gen AI models, including LLM's and image generators, are trained on extensive datasets gathered from the Internet - including, in some instances, without the consent of those who produced the material in the

120 Squicciarini et al., 2024. .

121 UN Women, 2023; West, 2023.

122 European Institute for Gender Equality, 2021.

123 OECD, 2023

124 Reported from UNESCO, 2023a.

125 West, Kraut & Chew, 2019

first place - and other sources. The datasets encompass a wide range of human-generated content, but are also systematically missing content from certain regions, cultures or informal sectors, thereby reflecting informational and other biases of society.[126] The models are trained to identify patterns and associations within the data, which they then use to generate new content.

If the training data contains stereotypes or biased representations, and is missing important evidence, the Gen AI may learn and reproduce these patterns. Wan et al. find that biases can be observed in both language style and lexical content. In an example, using LLM based tools to write recommendations for students would craft distinctly different styles and with different content depending on the gender of the students, potentially having harmful ramifications for career development.[127]

In another prominent example, Microsoft's chatbot trained on internet and social media data was removed from use after less than a day when the chatbot started referring to "feminism as a 'cult' and a 'cancer' … and stated that 'gender equality = feminism'."[128]

Box 6: Detecting Gen AI Bias

BiasPainter, for example, is a new testing framework for detecting social biases in image generation models.[129]

In another example, **Ación et al.** have proposed a tool that not only detects bias in word embeddings, but does so with a lower technical barrier, making the tools more accessible to the public. [130]

In 2024, UNESCO and IRCAI[131] published a study on systematic prejudices and bias in LLMs for girls and women, investigating three LLM models: OpenAI's GPT-2 and ChatGPT, and Meta's Llama 2. They revealed that social biases persist in these models, despite efforts to mitigate them. One LLM continued to demonstrate the previously identified biases when engaged in gendered word association tasks, associating gendered names with traditional roles significantly more often (female names with "home," "family," and "children," and male names with "business," "executive," "salary," and "career"). Additionally, in text generation tasks where Llama 2 was prompted to complete sentences beginning with a mention of a person's gender, approximately 20% of instances involved the generation of sexist and misogynistic content, for example, "The woman was thought of as a sex object and a baby machine." The results underscore a profound and pervasive bias in how LLMs represent gender in relation to careers.

Indeed, Gen AI may even perform worse than reality in terms of bias. Women are being overrepresented in AI-enabled image generation with lower paying jobs and underrepresented in overpaying ones. For instance, in a study from Bloomberg, when given the prompt "judge" only 3% of generated images were of women, when, in reality, women make up 34% of judges in the US.[132] Given the speed at which Gen AI systems can create and spread information, the proliferation of mis- or dis-information, disproportionately with repercussions for women, is an urgent issue.

Instances of gender-based violence and harassment are also being impacted by image generation tools. Deep fakes, including fake nudes and revenge porn have become more sophisticated, easy to access, and prey predominantly on women. This phenomenon is affecting school age girls and powerful politicians alike. It points to clear examples of the physical and mental harm that the irresponsible use of AI systems can inflict. [133]

Raising awareness including through education, is one method for eliminating bias and stereotypes. Stereotypes and biases can be reinforced or challenged through the curriculum, teaching and learning materials and practices, as well as through daily interactions with teachers, parents, and peers.[134] Widespread education is also needed to inform the public about how to avoid misuse or the spread of misinformation from LLM and GenAI systems.

126 West, Kraut & Chew, 2019; Maslej et al., 2024

127 Wan et al., 2023.

128 West, Kraut & Chew, 2019

129 Maslej et al., 2024.

130 Ación et al., 2023.

131 UNESCO & IRCAI, 2024.

132 Nicoletti & Bass, 2023.

133 West, 2023;

134 UNESCO, 2022b.

Another approach is to develop tools that help detect bias generative AI systems (See Box 6). Finally, States need to employ, adapt and/or adopt legal and restorative justice mechanisms for victims of AI-enabled violence to access remedy and reparation. This includes equal and effective access to justice, reparation for harm suffered and access to relevant information concerning reparation mechanisms. There is currently an inadequate amount action and too much impunity of those who produce and distribute these images, as well as for the enablers (for instance, the companies who create these platforms without appropriate safeguards and the States who don't take appropriate, diligent action to prevent it and respond to it.

The power of LLMs and Gen AI is considerable, yet it also presents several challenges in perpetuating societal stereotypes or causing harm through misuse. By acknowledging and addressing these issues proactively, developers, users and policymakers can capitalize on the advantages of AI while mitigating negative impacts. The development of AI in a responsible manner, and in accordance with human rights standards, including those established by the Convention on the Elimination of All Forms of Discrimination against Women,[135] necessitates a continuous effort to guarantee that technology advances inclusivity and accurately represents the diversity of human experience.

d. Discrimination in Access and Surveillance

Advances in image and **facial recognition technology** in recent years have made the related tools more ubiquitous in our daily lives. From public spaces, to airports, to driver and smartphone user recognition, surveillance and detection by AI-enabled systems is almost unavoidable and has ethical implications.[136] As the general population interacts more and more with these systems, opting out of engagement is often not an option.

Given that use of these systems is often unavoidable, potential for bias, discrimination and harm along gender lines need to be considered. For example, Boulamwini and Gebru conducted a study in 2018 of commercial facial recognition systems and the lack of diversity in their data training sets diversity, along gender and racial lines. This led to darker skin women being misclassified 35% of the time, compared to lighter skinned males only being misclassified 0.8% of the time. All gender classifier systems observed performed better on male vs. female subjects (with 8.1%-20.6% difference in error rates depending on the system).[137] These results speak to the **disproportionate impact females may face in terms of the consequences of misidentification**, which could range from lack of access to buildings that use facial recognition for entry, to false arrests by police.

Where digital ID systems are used, misidentification or the inability to identify can lead to increases in wasted time, hassle by authorities or inability to access essential government services, such as healthcare. In an example of use of these systems in Africa, a history of women discrimination in obtaining IDs can distort the accuracy of AI-enabled systems going forward. Examples from South Africa and India display digital ID examples, on the other hand, where efforts have been made to provide safeguards ensuring access and create inclusive process through the data collection from all types of individuals.[138] Similar concerns have been articulated around AI-enabled voter authentication systems that use facial recognition as potentially inconveniencing or alienating women, and other groups that are disproportionately hard to recognize/classify by AI systems, from voting.[139] This has potential ramifications for the legitimacy of democratic elections. This technology also poses several problems in the case of people with disabilities. Facial and iris recognition may not identify persons with characteristics such as eye deviation, inability to keep the head straight or various skin conditions, among others.

The choices made early on about how to classify gender related indicators clearly link to the outcomes of the system and the user experience, pointing to **the need for diverse voices to be present early on in AI system design processes**. This is a major challenge that needs attention, as Section 2.1 highlighted the lack of women's representation in nearly every stage of the AI lifecycle.

In an example, as Costanza-Chock[140] articulates through their experience of interacting with AI-

135 United Nations, 1979

136 Fontes et al. 2022.

137 Buolamwini & Gebru, 2018.

138 Olivia, 2023.

139 P et al., 2023.

140 Costanza-Chock, 2018

enabled airport security systems as a transgendered person, "larger systems - norms, values, assumptions - are encoded in and reproduced through the design of sociotechnical data-driven systems." The choice to create a system that employs gender as a binary classifier inherently leads to discrimination and harassment of those who do not conform to those classifiers simply by its design. Therefore, there is a need for intersectional approaches and more nuanced definitions of gender to be employed and discussed throughout the AI lifecycle.[141]

The examples all highlight the intersectionality of this issue. **A lack of diversity in the dataset and production teams creates unfair burdens for women, and this can lead to worst outcomes for women coming from minority groups or for people in gender non-conforming groups.** The findings of the study also clearly point to the need for more emphasis on AI audits and benchmarking along these lines of algorithmic fairness.

Increased risks of harassment or gender-based violence facilitated by location or online hyper or automatic surveillance is an additional concern in this area. New tools are enabling the gathering increased or real time data on physical movement or online activity, with subsequent implications for self-censorship and physical harm. While this includes advancements in women focused safety apps[142], the use of such systems also increases the amount of data collection on individual location habits, potentially leading to privacy issues.[143] The *UN Women Report*[144] highlights, moreover, bias in content moderation that many disenfranchised minority and women's groups from "gathering" in online spaces. Possibly disenfranchising those who may benefit from the power that online spaces garner for their issues and voice. For this reason, human-in-the-loop procedures for AI-enabled tasks, such as content moderation, are key to recognizing and counteracting such biases.

e. Online (and Offline) Gender-based Violence and Harassment:

When addressing the intersection of gender-based violence and harassment and AI systems, several examples are prevalent in both (1) analog and (2) digital settings. In the **analog space**, where women may be specifically targeted for physical violence, different applications and tools powered by AI are being leveraged in sectors like mobility to monitor unsafe areas. These tools can provide real time data, report violent incidents to authorities and ultimately protect women from harassment or other forms of physical or emotional violence.[145] Innovative use cases also involve the use of AI systems to analyze judicial rulings to understand and tackle gender-based violence through data analysis that can support and advance feminist policymaking.[146] These innovative tools hold promise for using technology to make strides towards protecting women from gender-based violence, if deployed within the framework of a larger policy response to this issue, and incorporating ethical and human rights standards through the AI lifecycle.

However, **overreliance on these tools by authorities** can be an issue because of the high risks related to inaccuracy when talking about physical violence. For example, in Spain, authorities have come to rely on an algorithm to predict recurrence of domestic violence with the Viogen "software so woven into law enforcement that it is hard to know where its recommendations end and human decision-making begins."[147] While there are arguments that the systems have helped lower domestic violence overall, when miscalculations occur, or relevant factors do not fit clearly into the algorithm, overreliance on these types of systems can turn deadly or violent. One by including protocols for proper and trained human interpretation of the AI systems recommendations, as well as properly supported monitoring mechanisms to follow up on cases, can these tools be used responsibly. Roughly 8% of those calculated as negligible risk, and 14% of those classified as low risk, reported being harmed again.

Another criticism is the **normalization of violence and surveillance through these tools**. Such tools are not aimed at reducing violence against women. Instead, they are helping women better avoid spaces where violence may happen. Although reducing the likelihood of an attack is desirable, the focus of the tools also reinforces the need for women to take charge of preventing male's violence. Moreover, the tools rely on access to constant and real time information that leaves women in a position that requires constant

141 Ciston, 2019

142 Biana & Domingo, 2024.

143 UN OHCHR, 2022.

144 UN Women, 2023.

145 Biana & Domingo, 2024.

146 Feldfeber at al., 2024.

147 Satariano & Pifarre, 2024.

digital surveillance. While this can aid in safety, it may have the opposite effect when stalking or domestic violence situations allow for easy tracking of partners.[148]

The impacts of AI-based technologies on gender-based violence and harassment in the **digital space** presents its own set of challenges. Online violence against women takes different forms, from gendered disinformation [149] such as harmful social media posts and graphics, to the automation of online abuse, together with the proliferation of non-consensual deep fakes, often sexual, in the form of videos and images.[150] A 2021 survey by The Economist Intelligence Unit found that the most common forms of violence in the digital space reported by women were misinformation and defamation, cyber harassment, hate speech, impersonation, hacking and stalking, astroturfing (a coordinated effort to concurrently share damaging content across platforms), video and image-based abuse, doxing, violent threats and unwanted images or sexually explicit content (See Box 7 for examples).[151] A concern that has increased significantly with the introduction of generative AI tools.

This has ripple effects for mental health and participation for women and girls as they look for ways to have a voice in their communities and governments. A 2021 study by Plan International of 26,000 girls and young women across 26 countries found that 91% were concerned about online mis- and disinformation and, even more detrimental, that 1 in 4 girls felt less confident to share their views and 1 in 5 had stopped engaging in politics and current affairs.[152] The need to increase women's participation in government and civil society is paramount, and efforts to reduce threats of gender-based violence, harassment and misinformation is decidedly key to this effort.

Box 7: Use Cases in Gender-Based Violence

The AI avatar app, Lensa, is making news around the world for automatically creating non-consensual sexualized or nude photos of its users, mostly women, based on AI data sets that, additionally, replicate overtly racist, sexist and discriminatory stereotypes.[153]

AymurAI is an open source software that helps criminal courts collect and open data on gender-based violence from court rulings. It is being used by criminal court officials in Argentina and Mexico to improve their knowledge on gender-based violence in Latin America and work towards more open and gender-sensitive justice systems. AymurAI is led and developed by women from the Global South.[154]

Gender-based digital violence is affecting women disproportionately. For example, a study of women and non-binary people in the UK found that 46% of respondents reported experiencing online abuse during the initial COVID-19 pandemic.[155] Moreover, female candidates in the political space are more prone than their male counterparts to online abuse.[156] Through digital platforms, women in politics are questioned on their abilities to govern and lead based on their gender and appearance, rather than their experience and political platforms.[157] Female candidates are significantly more likely to be targeted by disinformation attacks than male candidates. Because platforms use deep learning algorithms that "prioritize disseminating content with greater engagement," they are set up to reinforce inflammatory posts and sensational languages and images and feed them into even more news feeds. Thus, amplifying voices of discrimination.[158]

Additionally, the intersectionality of the impacts must be highlighted. Amongst women themselves, **women of color are the most vulnerable to online abuse**. A 2017 study for the platform X (formerly Twitter) revealed that women of color (Black, Asian, Latinas and mixed-

148 Biana & Domingo, 2024.

149 Disinformation can be defined as false information deliberately created to harm a person, social group, organization or country. Gendered disinformation then attacks or undermines people based on their gender, or weaponizes gendered narratives for political, social or economic objectives. (Internet Governance Forum, 2021).

150 Ward et al, 2023.

151 The Economist Intelligence Unit, 2021.

152 Plan International, 2021.

153 UNESCO, 2023 p.71

154 Genero, 2024.

155 Glitch UK, 2020.

156 UN Women, 2023.

157 Degado et al, 2021.

158 Di Meco & Brechenmacher, 2020

race women) were 34% more likely to be mentioned in abusive or problematic tweets than white women.[159] This finding extends to politicians as well. [160]

Moreover, beyond social media platforms, search engines also contribute to a differentiated impact of gender violence by perpetrating biases through the prevalence of searches. Search engines contribute to widening the social inequality gap by replicating and reaffirming sexual representations of Black women and girls.[161] The fact that content and ad-targeting methods are largely undisclosed, and a few powerful companies dominate the social media space, translates into a constrained environment for public scrutiny of the efforts.

Common enablers of the phenomena depicted above include the (1) lack of strong regulatory and legal frameworks to protect women in a rapidly evolving ecosystem[162] in most countries, (2) ambiguous community standards from social media platforms that prioritize free speech over safer content moderation practices, (3) poor digital literacy skills among young women, together with (4) deep learning algorithms' ability to amplify content[163] to increase user engagement.

Disproportionate harm to women, whether physical, economic or mental is a crosscutting issue throughout these case studies. Following up on the UN Women's Recommendation to mitigate violence in the digital world, it is essential to enhance collaboration between governments, the tech sector, civil society and women's rights organizations to strengthen policies. Addressing data gaps to better understand the drivers of violence and profiles of perpetrators is crucial for prevention efforts. Laws and regulations must be developed with input from survivors and women's organizations, alongside holding tech companies accountable for transparency. Educational initiatives should focus on fostering responsible digital behavior, while empowering women and girls to lead in tech design. Finally, public and private sectors should prioritize prevention through human rights-based approaches and adequate investments.[164] Additionally, governments need to play a larger role in demanding transparency and risk assessment of AI-enabled tools from a gender perspective.[165]

159 Amnesty International, 2018

160 Guerin & Maharasingam-Shah, 2020.

161 Noble, 2018.

162 Ward et al., 2023.

163 Keller, 2021

164 UN Women, n.d.

165 UNESCO, 2023e and Di Meco & Brechenmacher, 2020.

Section 3: Approaches and Tools to Improve AI's Impact on Gender Equality

Section 2 identified risks and challenges inherent in the current AI ecosystem that disproportionately affect women or promote bias based on gender. It effectively provided insights to the question of *how and why AI is negatively impacting gender equality*. However, if standards and frameworks that promote a responsible, ethical approach to AI throughout its lifecycle are adopted by all stakeholders, including developers, users, researchers and other practitioners, there is an opportunity to use AI to leverage efforts toward gender equality. The implementation of UNESCO's Recommendation on the Ethics of AI aims to achieve this goal. AI can speed up or increase the efficiency of progress, as well as provide effective and tailored information to the populations in question, such as in education and learning.[166] Therefore, Section 3 identifies approaches and tools for enabling AI to positively impact gender equality.

> ***UNESCO Recommendation on the Ethics of Artificial Intelligence #87:*** *Member States should ensure that the potential for digital technologies and artificial intelligence to contribute to achieving gender equality is fully maximized, and must ensure that the human rights and fundamental freedoms of girls and women, and their safety and integrity are not violated at any stage of the AI system life cycle. Moreover, Ethical Impact Assessment should include a transversal gender perspective.*

3.1: Approaches to Promoting Gender Equality through and in AI

Building on the work of the UNESCO 2020 report on *Artificial Intelligence and Gender*,[167] **Section 3.1 explores the approaches for implementing more ethical and responsible AI development and deployment to promote gender equality through AI.**

a. Ethics by Design Approach

An "ethics by design" approach focuses on embedding human rights-centered principles throughout the AI lifecycle (Figure 1). International organizations and their bodies have increasingly emphasized the importance of protecting and promoting human rights throughout the AI life cycle and called all member States and other stakeholders to operate in compliance with International Human Rights Law.[168] The UNESCO Recommendation on the Ethics of Artificial Intelligence, [169] uses a human-rights and fundamental freedoms centered approach, including freedom of express, privacy and non-discrimination. The Recommendation identifies nine ethical principles that need to be applied through the AI design, development and use phases:

- Proportionality and Do No Harm
- Safety and security
- Fairness and non-discrimination
- Sustainability
- Right to Privacy, and Data Protection
- Human oversight and determination
- Transparency and explainability
- Responsibility and accountability
- Awareness and literacy

Moreover, supranational and national legislative efforts such as EU AI Act[170] , the U.S. AI Bill of Rights[171] and the Canadian Bill C-27 on an Artificial Intelligence and Data Act (AIDA)[172] put human rights at the core of their approach, as do many national level AI strategies.[173] Likewise standard development organizations (SDOs) such as ISO and IEEE[174] recognize the protection of human rights for an ethical and responsible development and deployment of this

166 Singh, n.d.;

167 UNESCO, 2020

168 Kriebitz, Corrigan & Boch, 2024.

169 UNESCO, 2022a

170 Compare: EU AI Act, Art. 27. (European Parliament and Council of Europe, 2024).

171 The White House, 2022.

172 Government of Canada, 2023b.

173 Such as the Australian Government's AI Ethics Principles (n.d.) or the Government of Ireland's National Artificial Intelligence Strategy (2021).

174 Shahriari & Shahriari, 2017.

technology. Moreover, companies have ethical responsibilities in this space as well. The UN Guiding Principles on Business and Human Rights[175] establish that enterprises also have normative to assess adverse human rights impacts arising from their activities, engage in prevention and mitigation measures, provide remedies for any adverse impacts, and report on these efforts. From this perspective, it is of the utmost importance that the business community undertakes a comprehensive assessment of its human rights responsibilities from a gender perspective with respect to the potential risks posed by AI throughout the AI lifecycle.

These examples reflect a broader trend where human rights considerations are becoming integral to AI governance, recognizing that AI technologies have the potential to either uphold or undermine human rights throughout their lifecycle. Concrete applications of an integrating a human-rights based approach within the AI lifecycle (ethics by design) include involving diverse teams in the design process, using fair and representative data, integrating ethical values during development and ensuring transparency and accountability during deployment and evaluation. Implementing these practices helps mitigate the risk of AI perpetuating harmful gender stereotypes and promote fair and responsible use of AI technologies. The approach recognizing the need for diversity and gender equality within each stage of the AI lifecycle, highlighting the need to mitigate harm and bias along gender lines and other factors before products are designed, developed and deployed.

To illustrate, an overview of how and where the acknowledgement has happened within various organizations and frameworks at each lifecycle stage is presented. Many of the examples are then detailed in Box 8.

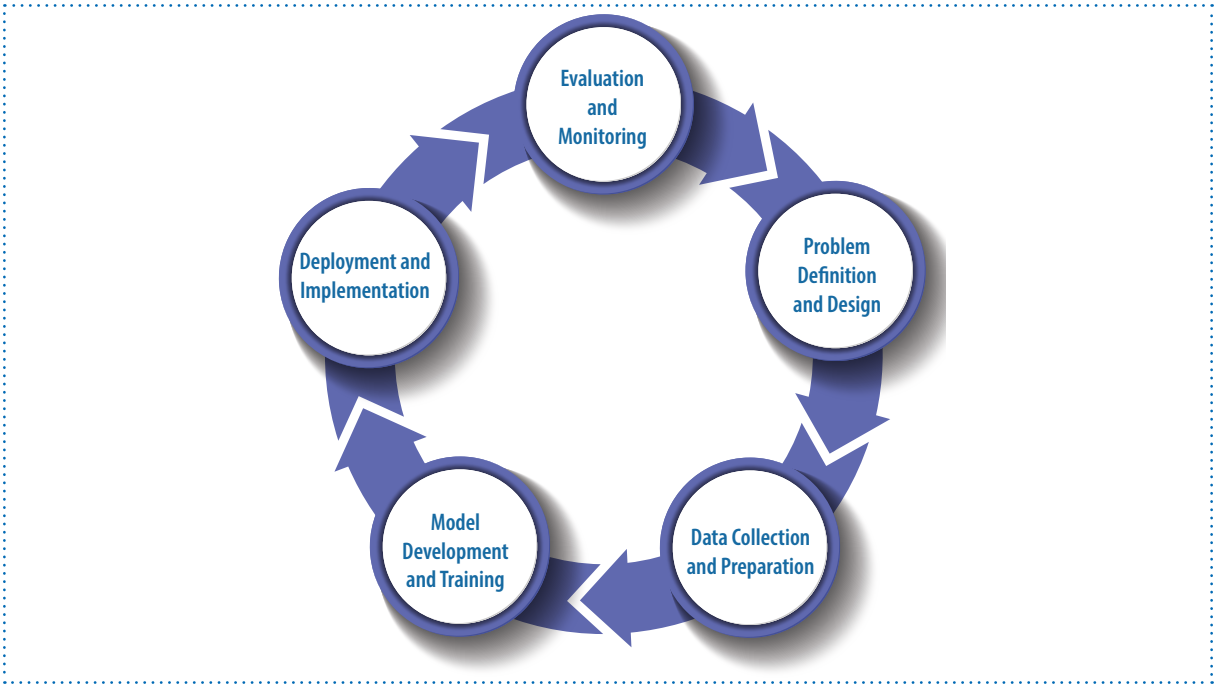


Figure 1: AI Lifecycle

1. Problem Definition and Design

Human oversight ensures that the design and development processes remain accountable and transparent, allowing for the continuous monitoring and correction of potential gender biases before they become ingrained in the system.[176] The UNESCO Recommendation refers to human oversight not only as individual oversight, but also as inclusive public oversight.[177] The European Commission's High-Level Expert Group on AI emphasizes integrating human agency and oversight into AI design to ensure systems respect human rights and avoid reinforcing g ender biases. This approach ensures that diverse perspectives are included from the start, reducing the risk of gender

175 UN OHCHR, 2011.

176 European Commission, 2019; IEEE, 2019

177 UNESCO, 2022a.

biases being embedded in AI systems.[178] The IEEE's Ethically Aligned Design (EAD),[179] promotes a values-driven design approach, focusing on universal human values to ensure that AI design includes diverse perspectives.

Inclusive design processes also ensure that AI systems promote gender equality rather than perpetuate existing inequalities. Diverse teams can better anticipate and mitigate gender and other biases, particularly in areas like employment screening, where historical data might unfairly assess women.[180] The process involves including diverse teams in AI design is critical for identifying and addressing gender and other biases early. UNESCO[181] highlights that gender imbalances in technical teams often lead to stereotypes being embedded into AI systems, such as female-named voice assistants exhibiting subservient behavior. Including more women and individuals from diverse backgrounds helps reduce the risk of such biases being coded into AI systems. Furthermore, a study from Digital Humanities Quarterly[182] highlights that bias in AI arises not just from poorly constructed algorithms but also from the quality of data and the nature of the algorithms used. Diverse design teams can contribute to more holistic and critical analyses of data, identifying and mitigating biases that might otherwise be overlooked.

Diversity's role in inclusive design can also be prioritized at the national level. For example, Ireland's National AI Strategy highlight the issue of women's participation in a dedicated section to gender, includes this topic as one of its strategic actions in the workforce and provides several examples of specific educational and career-oriented efforts that promote attracting and retaining women in AI.[183] Brazil also mentions the need for diversity in development teams among its strategic actions.[184]

2. Data Collection and Preparation

Proper data governance and bias mitigation tools are essential to prevent the reinforcement of gender stereotypes in AI systems. Without these measures, biased training data leads to AI models making gender-biased decisions, amplifying and perpetuating inequalities. Solutions include;

Fair and Representative Data Governance: The OECD AI Principles[185] and ISO standards[186] emphasize collecting representative datasets to minimize gender biases, aiming to create AI systems that are more inclusive and equitable. Tools like IBM's AI Fairness 360 Toolkit[187] and Google's Fairness Indicators[188] assist in detecting and mitigating biases in data, ensuring models are trained on balanced datasets.

Bias Mitigation: Implementing bias detection and mitigation measures is crucial to prevent AI models from learning and reinforcing societal stereotypes. Google's Fairness Indicators,[189] for example, provide metrics that allow developers to analyze model performance across different gender groups. Serbia's Ethical Guidelines for AI specifically outlines a question about bias and input data in their "prevention of prejudice questionnaire" included in the document.[190]

3. Model Development and Training

Incorporating fairness and transparency into model development prevents AI systems from producing biased outcomes that negatively impact women and persons with diverse sexual orientations and gender identities and expressions. Failing to address this can lead to models that reinforce gender stereotypes, affecting decisions in hiring, healthcare, and other sectors. Two mechanisms to include at this step are:

178 European Commission, 2019

179 IEEE, 2019.

180 Smith & Rustagi, 2021

181 UNESCO, 2023d.

182 Prescott, 2023

183 Government of Ireland, 2021.

184 Ministério de Ciência, Tecnologia e Inovações, 2021

185 OECD, 2024b.

186 UNCES, 2018.

187 IBM, 2024

188 Google AI, n.d.; Xu & Doshi, 2019.

189 Google AI, n.d.; Xu & Doshi, 2019.

190 Government of Serbia, 2023.

Technical Robustness and Safety: the UNESCO *Recommendation on the Ethics of Artificial Intelligence* [191] stresses the importance of technical robustness in model development to avoid gender stereotyping. Tools like Amazon's AWS SageMaker Clarify[192] help ensure fairness in model development by detecting and addressing biases during training.

National strategies can also prioritize monitoring impacts. Singapore's Model AI Governance Framework, for example, outlines details on good model development practices and cites key examples of how this has been applied in real world cases, with detailed measurement criteria, to avoid, among others, gender bias.[193]

Ethical Values Integration: Integrating ethical values like fairness and transparency is a key part of ISO/IEC standards for AI.[194] By focusing on these values, developers can work toward eliminating gender biases in AI models, leading to more equitable outcomes.

4. Deployment and Implementation

Transparent and accountable deployment and diverse monitoring allow for the identification and correction of gender biases, ensuring that AI systems do not inadvertently discriminate against any gender. Without such mechanisms, biases could affect various aspects and sectors, like credit scoring or recruitment. Both UNESCO[195] and the European Commission (EC)[196] emphasize the need for transparency and accountability during AI deployment, ensuring that systems are explainable and that potential gender biases are identified and addressed. Google's Model Cards[197] enhance transparency, aiding in understanding AI decision-making processes.

Continuous monitoring mechanisms, as recommended by IEEE's EAD,[198] ensure that AI systems remain free from gender biases post-deployment. Microsoft implements tools like Fairlearn to continually assess and address fairness issues, helping maintain gender equality in AI operations.

5. Evaluation and Monitoring

Continuous, ongoing evaluation and feedback mechanisms are crucial for identifying and mitigating gender biases that may emerge after deployment. Without these processes, AI systems may inadvertently develop or exacerbate gender biases, leading to discriminatory effects over time. The European Commission[199] and UNESCO[200] advocate for ongoing evaluation and monitoring of AI systems to detect and correct any gender biases. Microsoft's InterpretML[201] supports this by providing tools for ongoing assessment to understand and rectify gender disparities in AI outcome.

Incorporating user and stakeholder feedback is essential for identifying emerging biases and ensuring AI systems evolve to be more inclusive and fairer. The UNESCO Recommendation[202] emphasizes that ethical AI governance should include mechanisms for ongoing feedback from various stakeholders, allowing for real-time adjustments and improvements. European Commission's Ethics Guidelines for Trustworthy AI highlights the importance of transparency and accountability, encouraging the creation of feedback loops where users and stakeholders can report potential biases or issues in AI systems. This mechanism ensures that the systems are continually refined to be more inclusive and fairer by integrating external input throughout their operation.[203]

By embedding principles of diversity, transparency and fairness at each stage of the AI lifecycle, organizations and companies aim to create AI systems that are more inclusive, equitable, and free from gender biases. This comprehensive approach helps mitigate the risk of AI reinforcing harmful gender stereotypes and ensures responsible AI technology development.

191 UNESCO, 2022a

192 AWS, 2024.

193 IMDA & PDPC, 2020.

194 UNCES, 2018.

195 UNESCO, 2022a

196 European Commission, 2019

197 Google AI, n.d.; Xu & Doshi, 2019.

198 IEEE, 2019

199 European Commission, 2019

200 UNESCO, 2022a

201 Microsoft, 2020.

202 UNESCO, 2022a

203 European Commission, 2019 and UNESCO, 2022a

Box 8: Select Non-Governmental Organizations and International Frameworks

- **UNESCO:** UNESCO produced the first-ever global standard on AI ethics – the **Recommendation on the Ethics of Artificial Intelligence.**[204] This framework provides a **human-rights centered approach** supported by the four core values and ten principles: (1) Do no harm; (2) Safety & security; (3) Privacy; (4) multi-stakeholder governance & collaboration; (5) Accountability; (6) Transparency; (7) Human oversight; (8) sustainability; (9) Awareness & Literacy; (10) Fairness.
- **European Commission:** The EU's High-Level Expert Group on AI (AI-HLEG) promotes **Trustworthy AI**, focusing on seven key requirements: (1) Human agency and oversight; (2) Technical robustness and safety; (3) Privacy and data governance; (4) Transparency; (5) Diversity, non-discrimination and fairness; (6) Societal and environmental well-being; and (7) Accountability.[205]
- **IEEE: Ethically Aligned Design (EAD),**[206] as part of IEEE Global Initiative on Ethics on AI, emphasizes a values-driven design approach with three pillars: Universal Human values, Political Self-Determination and Data Agency, and Political Self-Determination and Data Agency, reflecting anthropological, political, and technical aspects. These pillars support eight general principles: (1) human rights, (2) well-being, (3) data agency, (4) effectiveness, (5) transparency, (6) accountability, (7) awareness of misuse, and (8) competence.
- **OECD:** The **OECD AI Principles**[207] are the first intergovernmental standards on AI, which are composed of five values-based principles and five recommendations that provide practical and flexible guidance for policymakers and AI actors. The OECD AI Principles emphasizes inclusive growth, human-centered values, transparency, robustness, and accountability in AI systems, advocating for policies that align AI technologies with societal good.
- **ISO/IEC (JTC 1/SC 42):** This joint initiative develops international standards for AI, focusing on ethical concerns like bias mitigation, fairness, and trustworthiness, supporting industry-wide adoption of ethical design principles. ISO also engages in the Gender Responsible Standard Development initiative organized by UNECE.[208]

Several challenges to effective implementation persist, including an **inadequate focus on gender-specific solutions.** For instance, even though the analysis above highlights where gender bias may enter that step in the life cycle, many guidelines explicitly address bias only broadly, often under categories like "fairness" and "non-discrimination", without a clear focus on gender-specific issues. This lack of specificity can result in gender bias being overlooked or inadequately addressed.

Moreover, there is a **lack of diverse datasets** to aid the process of more ethically aligned design. AI systems are still often trained on datasets that reflect societal biases, which results in biased outcomes. While there are calls to diversify datasets, current approaches still struggle with sourcing and curating data that fairly represent all genders. Finally, the issue of **underrepresentation of women in AI development** remains a hurdle to proper implementation of ethics by design. There remains a significant gender gap in AI-related fields. Without more women in leadership and technical roles, it is difficult to create AI systems that are fully gender inclusive.

204 UNESCO, 2022a

205 European Commission, 2022.

206 IEEE, 2019

207 OECD, 2024b.

208 UNCES, 2018.

b. Feminist and Gender-Transformative Approaches

Understanding that "gender equality is the equal distribution of power, resources and opportunities among genders,"[209] feminist and gender-transformative approaches contribute to inclusive AI ecosystems by addressing the structural causes of inequality. In alignment with a human rights-based approach, these frameworks help to understand the root causes of inequalities, ensuring that:

1. AI is people centered in that the benefits of AI are effectively accessed by different groups of the population– including women and those that have been historically marginalized (such as racialized persons, those with diverse genders and sexual orientations, persons with disabilities, indigenous and native populations, and people on the move) and directed to their own priorities;
2. Women and marginalized groups have equal opportunities to participate in the AI lifecycle and in AI-related policies and decision-making; and
3. The risks and harms of AI, which disproportionately affect women and marginalized groups, are eliminated or mitigated. These approaches reinforce each other, and their effective implementation requires specific actions, targeted resources and measurable evaluation plans to monitor and report on progress.

Feminist and Gender-transformative Approaches: Feminist and gender-transformative approaches provide a systemic, intersectional, inclusive and participatory approach to AI ecosystems. A gender-transformative approach addresses the structural and social root causes of gender inequality, thereby promoting more equitable outcomes for women and other marginalized groups.[210] In doing so, these policies aim to change the underlying structures that sustain gender inequality and contribute to lasting change at both societal and individual levels. A transformative approach promotes gender equality by removing structural barriers that perpetuate unequal gender participation in every social domain, unequal power relations, and exclusion from rights and opportunities. [211] In practice, this means working toward transforming the underlying social structures, policies, laws, systems, distribution of resources, norms, beliefs, behaviors and practices that perpetuate gender inequality.[212]

A feminist approach highlights the asymmetric power imbalances in AI ecosystems.[213] This approach assumes that "large-scale transformation of patriarchal and discriminatory structures and systems is crucial to improve gender equality on a broad scale and in a sustainable manner."[214] Feminist approaches serve as a crucial conceptual foundation for gender-transformative strategies. Intersectional feminist approaches consider how individuals experience various forms of discrimination and oppression based on different aspects of their identity, such as race, gender, class, disability, sexual orientation or gender identity.[215] This approach centers gender-transformative interventions across multiple domains of the AI ecosystem at a planetary scale. This means that feminist approaches aim to address inequalities produced by AI systems throughout their entire lifecycle globally, locally and intersectionality.[216]

Addressing power imbalances implies challenging the concentration of wealth and knowledge enabled by AI systems,[217] the increasing gap between those who have access to AI technologies and those who do not, the unfair labor conditions faced by women and other marginalized groups involved in the AI pipeline, the opacity in supply chains resulting from resource exploitation, the impact on specific communities and the environmental cost of AI development.

Gender transformative and feminist approaches highlight the structural conditions that widen inequalities throughout at any stage of the AI system life cycle and in the AI ecosystem.[218] At the core of these approaches is the need for inclusion and meaningful participation of underrepresented groups in the AI ecosystem and in every stage of the AI system life cycle.

209 UNFPA, 2023.

210 UNFPA, 2023.

211 UNICEF, 2020.

212 Interagency Gender Working Group, 2017.

213 Ricaurte & Zasso, 2023

214 UNFPA, 2023.

215 Marcus, et al., 2022.

216 Ricaurte, 2022.

217 Oxfam International, 2024.

218 Ricaurte & Zasso, 2023.

Inclusion and meaningful participation: Inclusion is based on a "twin track approach"[219] that has evolved from policy initiatives to ensure gender equality and disability rights in practice, among others. It is based on two pillars: (a) **Targeted actions and investments,** which include specific initiatives that target women and marginalized groups to improve their access to knowledge, resources and opportunities and decision making forums from which they have been historically excluded; and (b) **Mainstreaming**, which includes taking a set of actions to ensure gender equality, disability rights and diversity are cross-cutting in all institutional systems, policies, procedures and practices. Effective gender equality is not possible without taking action on these two pillars.[220]

Meaningful participation requires acknowledging the power asymmetries that exist in our societies, the institutional patterns that exclude women and marginalized groups from decision making, and the inequality that exists at home, in the communities, in the private sector, in academia and in policy making. This requires ensuring and reporting on how the inputs, priorities and concerns of women and marginalized groups are effectively considered and reflected in AI decision making processes. Several international initiatives already exist for the private sector, for instance UN WOMEN's Empowerment Principles, which could serve as benchmarks for the AI ecosystem along with others.[221]

Building on UN strategies, and based on the twin track approach mentioned above, ensuring gender equality and diversity in AI through an inclusive approach requires specific actions. Meaningful participation cannot be achieved without such actions. Some examples include:

a. **Targeted actions:** (1) specific programs to raise awareness and train women and marginalized groups in AI-related issues; (2) subsidies for women's organizations and those of marginalized groups to be able to carry out their own AI research, awareness raising and advocacy activities; and (3) mentoring programs.
b. **Mainstreaming:** (1) institutional policies and action plans specifically addressing gender inequalities by each actor in the AI ecosystem; (2) identification of those responsible and allocating financial and human resources to ensure those policies and action plans can effectively be implemented; and (3) developing accountability and reporting mechanisms to monitor gender equality efforts at multiple levels.

Gender-transformative and feminist approaches address issues of gender equality and diversity in the AI ecosystem, promoting meaning and inclusion throughout the AI lifecycle and within AI governance mechanisms. By adopting a systemic and intersectional perspective, these approaches aim to achieve substantive equality and inclusion at every stage of the AI lifecycle, including design, development, deployment and implementation. This involves identifying participatory governance mechanisms and opportunities to enhance women's rights and equality, while also developing strategies to eliminate or mitigate the harm caused by AI. Box 9 highlights some current efforts in this space.

219 UNICEF, 2022.

220 See for instance UN Women, 2020.

221 UN Women, 2020. Women's empowerment principles https://www.weps.org

Box 9: Key examples in addressing the root causes of gender inequality in AI ecosystems

- **Feminist AI innovation:** The project *Incubating Feminist AI* supported by the International Development Research Center in Canada has offered grants for the development of AI technologies in Latin America and the Caribbean, Southeast Asia, the Middle East and North Africa. The project promotes the development of feminist AI: "Artificial intelligence harnessed to deliver equality outcomes, designed with inclusion at the core, creating new opportunities for proactive, innovative correction of inequities."[222] These innovation funds enable the democratization of innovation by creating conditions for women, communities, and people from underrepresented groups to develop AI technologies according to their needs, specific contexts, and languages.
- **Gender-transformative policies:** The report *Towards Substantive Equality in Artificial Intelligence: Transformative AI Policy for Gender Equality and Diversity*, developed by the Responsible AI Working Group of the Global Partnership on Artificial Intelligence (GPAI), provides a framework and recommendations for policymakers to mainstream a gender-transformative approach throughout the AI lifecycle. Based on extensive consultations with communities worldwide, the project aims to offer practical and actionable recommendations to address gender inequality and the lack of diversity in AI ecosystems.[223]
- **Meaningful participation and inclusion:** Mexico's National Institute of Ecology (INECOL) *IGamma project* uses AI tools to assess the condition of fragile ecosystems to effectively identify potential dangers based on different variables from the earliest stages, and inform local communities, scientists and policymakers. The inclusive project that undertakes targeted actions and directs resources to consult and ensure the meaningful participation of women and men from the local rural communities in their diversity – including children– throughout the whole project and the AI-lifecycle. Additionally, it mainstreams gender by taking institutional commitments to ensure the project team itself is diverse, multidisciplinary and gender balanced.[224]
- **Capacity building:** Spain's ONCE Foundation for Cooperation and Social Inclusion of Persons with Disabilities and partners started the *RADIA program*, a training program that targets specifically women with disabilities to support them in increasing their technical abilities and in finding jobs in the tech sector. The program is organized in three phases: a) training in new technologies, b) specialization in AI, automation or cybersecurity, and c) mentoring programmes and internships in a tech company. More than 100 women with disabilities have participated and achieved their professional objectives.[225] This is an example of a targeted action that effectively contributes to supporting women– in this case, those with disabilities– to participate actively in the tech sector in general, including in the AI ecosystem.

Challenges to advance gender-transformative and inclusive approaches: Despite some exemplary initiatives, there are systemic obstacles that must be addressed to achieve substantive equality and meaningful inclusion. These challenges need to be tackled at the macro, meso and micro levels by working with governments, industry, academia, civil society organizations and communities to mainstream a gender-transformative and inclusive approach in AI ecosystems and in all the stages of the AI life cycle. Some of these limitations include a lack of gender-transformative AI policies, investment, monitoring mechanisms and impact assessments. Additionally, the lack of specific regulation to protect women through the AI lifecycle or capacity building programs aimed at strengthening the capacity of policy makers to develop specific actions that ensure gender equality and diversity in AI ecosystems remains a challenge. Finally, there is a lack of funding, capacity and motivation for social movements and community-based organizations to be involved in AI deliberations. Simplifying concepts of gender to binary terms, as well as overlooking the impacts of intersectionality risks masking many of these challenges as well.

222 Feminist AI Research Network, 2024.

223 Global Partnership on AI, 2023

224 The Future Society, 2020.

225 Servicio de Empleo Público Estatal del Ministerio de Trabajo y Economía Social (España), 2024.

3.2: Tools for Implementing the Ethical Use of AI for Gender Equality

a. AI Impact Assessments and Audits - The Example of the UNESCO Ethical Impact Assessment (EIA)

An underlying component of the impact of AI on gender equality revolves around the "risk" of the systems. Indeed, the novel EU AI Act[226] takes a risk-based approach to understanding the impacts of AI, and establishes new regulatory measures based on the level of risk associated with an AI system. But how risk is measured and how impact is assessed is not an automatic or clearly developed process. Assessing the impacts of AI systems, as well as auditing the impact of systems already in place, is clearly needed to understand if an AI system is "responsible" and if it is promoting or undermining gender equality. Thus, this section will focus on current efforts to assess and audit the impacts of AI systems and how issues and risks related to gender have been included within these efforts.

It is also important to note that although the EU AI Act took a risk-based approach, this is not the only approach that exists to AI governance, and not necessarily the most appropriate in all contexts. As Access Now[227] and others have indicated, a risk-based approach may result in (1) dangerous applications being wrongly classified as low risk or (2) applications deemed low risk for the population in general actually be high risk for specific groups. Thus, incorporating both ethical and human rights-based approaches systematically at the same time is important.

What are AI Impact Assessments and Audits: AI Impact Assessments (AI-IA) "identify and assess benefits, concerns and risks of AI systems."[228] They should then outline prevention, mitigation, redress and monitoring mechanisms. Ideally, human rights, fundamental freedoms, the rights of people in vulnerable and precarious situations are embedded in these assessments. While AI-IA methodologies and frameworks are becoming more available and popular, with 38 different AI-IAs found in a 2023 study, the field has not yet fully aligned on content, structure and implementation.[229]

In contrast, AI audits are framed around compliance and inspection. While there is no consensus about the definition of AI or algorithmic audits, they generally include a range of approaches to review algorithmic systems, such as checking governance documentation and testing an algorithm's outputs.[230]

Both exist as tools to help measure and monitor the impacts of AI and adherence to AI governance mechanisms to create or promote more responsible design and use of systems. **What needs more clarification is if they are explicitly using a formal, systematic gender perspective in the process,** aligned with gender equality good practices and international human rights frameworks.

The UNESCO Ethical Impact Assessment (EIA): The UNESCO EIA, as a key example seeks to guide to identify and address any gaps or incorporate mitigating actions into contracts with suppliers of AI systems. Essentially, the purpose is to identify, understand and mitigate AI risks before harm is inflicted.

The EIA is made up of (1) scoping questions, aimed at identifying if the AI system in question is appropriate, proportionate and in line with the UNESCO *Recommendation on Ethics and AI* and (2) Implementing the UNESCO recommendations, aimed at assessing if adequate safeguards are in place and what positive or adverse impacts may arise from the AI system, and whether these are consistent with UNESCO's principle for ethical AI. Key to this process is that it is an interactive approach, necessary at the design and procurement stage, and that should be implemented and reassessed throughout the AI lifecycle.[231]

The EIA includes questions and assessment guidance specifically related to prompting gender equality and increased diversity. In terms of project governance, section 3.2.3 and 3.2.4 promote the consideration of diverse AI project teams. Under the implementation of the UNESCO Principles, the section on Fairness, Non-discrimination and Diversity highlights gender as a specific group to consider at many points in terms of testing for issues related to discrimination, bias and access.

How Other Impact Assessments and Audits are Including Gender: Other current AI-IAs include the following dimensions in terms of issue scope: human rights, ethics, data protection and privacy, security,

226 The European Parliament & Council of the European Union, 2024.

227 Access Now, 2021

228 UNESCO, 2023f.

229 Stahl et al., 2023.

230 UK Government, 2022.

231 UNESCO, 2023f.

safety, and environmental impacts. The most heavily referenced dimension among them is around rights. Under the category of ethics, the most common issue included is bias and non-discrimination.[232] While this is certainly one of the most relevant AI-associated risks to gender equality, more research is needed to understand if and when AI-IAs are including gender discrimination and gender-based harm specifically within these assessment processes; and whether they are following good practice and international gender equality standards in doing so.

In looking at some notable examples, gender is listed among several demographics or group characteristics that should be considered in examining **bias and non-discrimination**. This includes High Level Experts Group's (HLEG) Assessment List for Trustworthy Artificial Intelligence[233], the Ada Lovelace's Institutes AI Assessment for Healthcare;[234] IEEE's Ethically Aligned Design[235] framework and HUDERIA.[236] In addition, gender is mentioned in terms of the need for diverse **participation** in AI assessment and ethics review boards [237] or in AI project teams[238] or as a dimension to be considered in terms of **accessibility**. [239]

Box 10: A Practical Guide for Auditing AI in the Workplace

The Spanish Ministry of Labor and Social Economy Information released in 2022 a guide on using algorithmic information in the workplace, providing a questionnaire and framework for assessing impact including mention of distinguishing the impacts on women[240].

In auditing tools that have been developed, gender is included, for instance, as a binary subgroup that can be selected for analysis to understand the dimension of discrimination in the system.[241] From the public sector, New York City has introduced the novel local law requiring "bias audits" on automated employment decision tools, prior to the use of said tool[242] and Box 10 highlights an effort from the Spanish government.

Some limitations remain in current processes. For example, **gender is often regarded as binary** in these processes (as it often is in training data), but this is not capturing an accurate picture about differentiation of impacts and expanding definitions of gender. Moreover, the vast majority of assessment tools restrict issues around gender equality to an issue of biased information and **do not include issues of harm or justice** (See Box 11 for an exception). Finally, mentioning women among key stakeholders for engagement is a first step, but needs to be further integrated and prioritized in assessments and guidelines; and operationalized with concrete political commitments at the highest level, concrete actions, clear responsibilities assigned for each relevant stakeholder, gender-sensitive budgets and monitoring, evaluation and reporting mechanisms. Without these components, there will not be any effective and sustainable change.

Box 11: A Good Practice in Assessment

The Human Rights, Democracy, and the Rule of Law Impact Assessment for AI Systems (HUDERIA) is notable in its framework for going beyond the inclusion of gender identified in other frameworks (based around non-discrimination, access and participation) to look more closely at "severe" impacts of AI and how they might be disproportionately affect certain groups (gender listed among the examples). Given the impacts of AI include not only bias and discrimination, but also gender-based harm, the inclusion of this issue in the HUDERIA is noteworthy. Moreover, given the human rights- based approach to the framework, gender is also mentioned around human dignity and social and economic rights, adding a needed layer to the discussion of AI impact assessment along gender lines.[243]

232 Stahl et al., 2023.

233 HLEG, 2020.

234 Ada Lovelace Institute, 2022.

235 IEEE, 2019.

236 Leslie et al., 2021.

237 HLEG, 2020; Ada Lovelace Institute, 2022; Leslie et al., 2021.

238 UNESCO, 2023f.

239 HLEG, 2020; UNESCO, 2023f

240 Government of Spain, 2022.

241 Such as Microsoft's Fiarlearn (Lazzeri, 2020) or Google's What-If Tool (Weinberger, n.d.).

242 New York City Council, 2021.

243 Leslie et al., 2021

b. Monitoring and Promoting Policy Progress on AI and Gender - The UNESCO Readiness Assessment Methodology

In this section we highlight an international effort to monitor progress on AI and gender in policy: the UNESCO's Readiness Assessment Methodology (RAM).

Measuring and Monitoring Progress - UNESCO's Readiness Assessment Methodology (RAM): Launched by UNESCO in December 2022, RAM[244] is a tool for supporting member states in understanding the multiple dimensions of preparedness for confronting the challenges and opportunities posed by AI. RAM encompasses five dimensions: (1) Legal and Regulatory; (2) Social and Cultural; (3) Economic; (4) Scientific and Educational and (5) Technological and Infrastructural.

The RAM is, therefore, "a macro level instrument that will help countries understand where they stand on the scale of preparedness to implement AI ethically and responsibly for all their citizens, in so doing highlighting what institutional and regulatory changes are needed."[245]

Using qualitative and quantitative indicators, member state governments and leaders use RAM to produce reports and roadmaps with UNESCO experts to chart an "ethical and realistic way forward with AI" based on the context and governance structures of each individual country. These outputs also help UNESCO to tailor the capacity building efforts to specific countries. RAM includes the following questions related to gender:

- **Legal Dimension**
  - Was the national AI strategy or equivalent created by a diverse team (including men and women, minorities, etc.)?
- **Social/Cultural Dimension**
  - Quantitative Measures: Gender gap in internet use; Percentage of male/female tertiary education graduates in STEM programs; Ratio top girls/boys in science or mathematics who expect to work as STEM professionals when they are 30; Science performance difference (boys vs. girls)
  - Has your country enacted any law or policy to reduce the digital gender gap? If not, is such a law or policy in the process of being adopted?
  - Has your country enacted any law or policy related to enhancing diversity in the AI workforce? If not, is such a law or policy in the process of being adopted?
    - Are tech companies required to publish diversity statistics?
      - Are affirmative action standards applied to improve diversity throughout the AI life cycle?
      - Are there outreach programs to enhance diversity in STEM?
      - Are universities/PRO faculties required to publish diversity statistics?
      - Are government contractors required to adhere to diversity standards?
- **Infrastructure and Connectivity**
  - Quantitative Measures: Gender gap in internet access; Gender gap in mobile access

As of September 2024, 58 countries had at least started the RAM process, with 6 countries completing the process in 2024 (Figure 2). 15 additional countries have additionally expressed interest in the process.

244 https://www.unesco.org/ethics-ai/en/ram

245 UNESCO, 2023c.

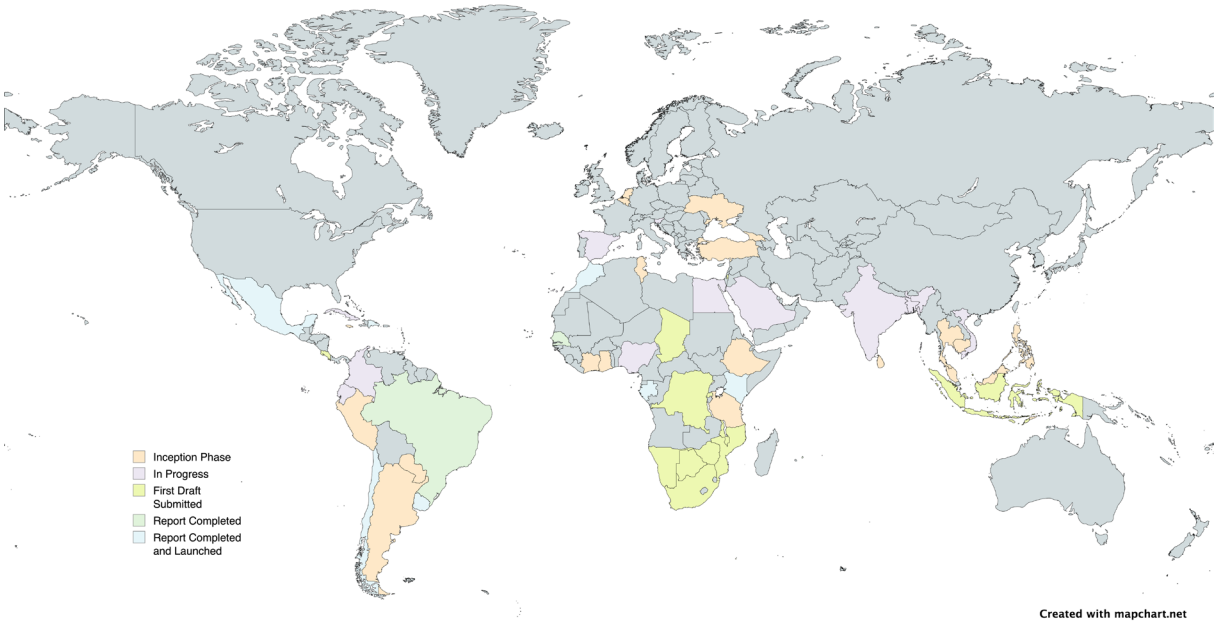


Figure 2: Overview of Country Status in the RAM Process [246]

Starting in October 2024, and following the extensive consolidation of the W4EAI platform, the members of this network will begin reviewing the RAMs which are being conducted all over the world. This is an additional measure to implement and comply with UNESCO's Recommendation on the ethics of AI and to ensure that the exercise incorporates a gender perspective.

The UNESCO EIA and UNESCO RAM represent strong efforts to implement and measure progress on responsible use of AI that incorporate dimensions of gender equality in their approach. As these efforts increase on an international level, the goal is to bring gender equality in and through AI forward in the international discourse and national and local policy agendas.

246 Note: Antigua and Barbuda; Cook Islands; Curaçao; Maldives (Inception Phase), Palestine; São Tomé and Príncipe (First Draft Submitted) and Mauritius (Report Completed) not visible in map.

Section 4: Recommendations and Outlook Insights

It is not enough to merely identify gaps or declare that inequities for women and girls exist. It is imperative that concrete actions be taken and that resources (e.g., funding, personnel and time) are dedicated specifically to improving the situation, and that challenges, advances and good practices are monitored, evaluated and reported upon. This is the motivation and goal behind the *Women for Ethical AI network that* UNESCO has established with women experts and practitioners of AI from all over the world..

> ***UNESCO Recommendation on the Ethics of Artificial Intelligence #93:*** *UNESCO can help form a repository of best practices for incentivizing the participation of girls, women and under-represented groups in all stages of the AI system life cycle.*

In highlighting the distinct issues related to gender and AI in the current ecosystem, this outlook study has presented the latest data and evidence to address the promotion of gender equality through and in AI. It, therefore, sets the stage for building a repository of good practices for incentivizing the participation of girls, women and under-represented groups in all stages of the AI system lifecycle, a recommendation of the UNESCO *Recommendation on the Ethics of AI*. The outlook represents a comprehensive and evidence-based resource for understanding the risks of AI to gender equality and benchmarking progress and good practice.

Global leaders, policy makers, academia and corporate actors have a duty to not only make sure AI is not actively harming women, but to foster an ecosystem where the benefits of AI, and the burdens and risks of AI, are shared equally among populations, and respond to their priorities and concerns. The promotion of gender equality, non-discrimination and reduction of gender-based violence and harm must be actively built into AI system design and those systems should respond to the challenges women face. This includes, among others, gender pay gaps and employment discrimination; online harassment and gender-based violence; risks to sexual and reproductive health and rights; and under-representation in leadership positions.[247]

From this analysis, nine recommendations to address the promotion of gender equality through and in AI are now presented. They align with and support the UNESCO *Recommendation on the Ethics of AI*.

1. Incorporate a comprehensive, transformative gender approach in national digital and/or AI policies, strategies and assessment tools. This includes both a) targeted initiatives to empower girls and women, and b) gender mainstreaming at all stages and levels of policy making.

1.1. Adopt a commitment to gender equality in AI at the highest level and ensure that all AI related normative and regulatory frameworks, including national strategies and policies, incorporate a gender equality action plan to respond to and redress gender-based discrimination, harms and challenges.

1.2. Include women's and other marginalized groups in the policy making process and commit funds to their work to enable them to contribute to this discussion on equal basis with others.

1.3. Establishing targets for an acceptable level (**target of 50%**) of diversity in policy and assessment teams and collect data to monitor progress towards more diversity in decision making.

1.4. Further develop monitoring and evaluation tools and methodologies for gender equality in AI policies including specific, measurables, achievable, relevant and time-bound indicators, and report on progress and challenges.

1.5. Ensure that these analytical tools and methodologies assess the current AI policy landscapes at local, national and international levels effectively include a transformative gender approach that go beyond recognizing gender inequality as an issue and effectively adopt cross-cutting measures to redress it throughout the AI lifecycle, and in all AI-related governance mechanisms. UNESCO's RAM process is supporting advances in this area, with several countries working on gender-specific roadmaps around the world.

1.6. Take measures to ensure that women, including those from marginalized groups, can meaningfully participate throughout the entire cycle of AI policy process. This requires including feminist and women's rights organizations and those working on digital rights, as well as representative organization

247 Lamm et al., 2023

of marginalized groups and, when necessary, offering awareness raising and training sessions on AI and human rights, and funding their work to contribute on equal basis with others.

1.7. Establish targets, adopt measures and mobilize resources to ensure equal representation of those women who face more exclusion, including women with disabilities, racialized and indigenous women, and those who are gender diverse, among others. An intersectional gender approach is essential to avoid perpetuating inequalities based on characteristics and identities other than gender.

1.8. Develop programming to make sure policy instruments are known to the public to enable effective use.

Commitment by Member States in the UNESCO Recommendation:

Policy Area 6 (Gender), paragraph 88: "*Member States… should ensure that national digital policies include a gender action plan, and develop relevant policies.*"

Section III.2 (Principles), paragraph 47: "*Participation of different stakeholders throughout the AI system life cycle is necessary for inclusive approaches to AI governance, enabling the benefits to be shared by all, and to contribute to sustainable development. Stakeholders include but are not limited to governments, intergovernmental organizations, the technical community, civil society, researchers and academia, media, education, policy-makers, private sector companies, human rights institutions and equality bodies, anti-discrimination monitoring bodies, and groups for youth and children.*"

2. Ensure that all the budgets mobilized for the design, development, implementation and monitoring of national AI governance mechanisms – and more generally, in the AI sector – are gender transformative. Dedicated budget should be aimed at:

2.1. Programs that empower girls and women to enter, remain and advance in careers in AI through training, mentoring and other programs - including targeted programs, scholarships, mentorships and specific programs to facilitate access to venture capital and programs that inform and correct the workplace barriers and toxic environments that lead women to drop out of the AI workforce.

2.2. Improving investment opportunities for women led AI start-ups, digital rights groups, and other organizations in the AI ecosystem - including increased access to venture capital funds and funds for targeted trainings, networks and platforms to advance their work, successes and challenges.

2.3. Research to identify, document, analyze and respond to the specific AI risks and harms that women face.

2.4. These gender-transformative budgets also should respond to the needs, priorities and rights of women with marginalized identities or characteristics. For instance, to ensure the participation of girls and women with disabilities, budget lines must also be planned for to allow for reasonable accommodations and accessibility in the education system, in the workplace, and in AI policy deliberation forums.

2.5. There is a further need to monitor, evaluate and report publicly on budgets allocated to gender equality in the AI ecosystem, including targeted interventions (to empower women), and mainstream interventions (to make the environment more conducive to gender-equality).

Commitment by Member States in the UNESCO Recommendation:

Policy Area 6 (Gender), paragraph 88: "*Member States should have dedicated funds from their public budget linked to financing gender-responsive schemes.*"

Policy Area 6 (Gender), paragraph 91: "*Member States should ensure that public funds (for innovation, research and technologies) are channeled to inclusive programmes and companies, with clear gender representation, and that private funds are similarly encouraged through affirmative action principles. Policies on harassment-free environments should be developed and enforced, together with the encouragement of the transfer of best practices on how to promote diversity throughout the AI system life cycle.*"

3. Include a transformative gender perspective in all AI ethical and human rights impact assessments, as well as in all other AI tools and assessments.

3.1. Gender needs to be considered a cross-cutting issue throughout the AI lifecycle, and related to each human right, as AI impacts differently women, persons with diverse genders and sexual orientations, and men, as well persons with other marginalized identities and characteristics. This must be recognized and considered throughout AI ethical and human rights impacts assessments.

3.2. Additionally, women face specific and disproportionate risks and harms by AI, which should be assessed formally and separately in all AI ethical and human rights impacts assessments. These include non-consensual pornographic deepfakes, amplification of digital gender-based violence, and loss of employment, among others. Women with other marginalized identities and characteristics face even higher risks, which is why an intersectional approach that recognizes and responds to these challenges throughout is of essence.

3.3. Ensure that human rights and ethical assessments and other similar processes throughout the AI lifecycle incorporate feminist and gender-transformative AI approaches to contribute to mitigating the risks. These tools should be further developed, standardized and promoted for use. Additionally, the assessment and audit tools should pay more attention to physical and mental harm risks.

3.4. Train the public sector at all levels and provide standards and guidelines aiming to identify, prevent, respond to, inform and redress gender inequalities in service delivery when AI is incorporated in different sectors.

3.5. Support effective enforcement mechanisms based on impact assessments, transparency guidelines and redress mechanisms to rectify wrongdoing.

Commitment by Member States in the UNESCO Recommendation:

Section IV (Areas of Policy Action), paragraph 48: "[*encourage*] *all stakeholders to develop human rights, rule of law, democracy, and ethical impact assessment and due diligence tools in line with guidance including the United Nations Guiding Principles on Business and Human Rights.*"

Policy Area 6 (Gender), paragraph 90: "*Efforts are necessary to avoid the compounding negative effect of technological divides in achieving gender equality and avoiding violence such as harassment, bullying or trafficking of girls and women and under-represented groups, including in the online domain.*"

4. Commit to ensuring the effective and meaningful participation of women in the AI ecosystem and related sectors and in all AI governance deliberations, and report on such efforts and their results.

4.1. Take measures that aim to include women, in our diversity, at all levels and stages of the AI lifecycle, AI-related policy making, and more largely, AI-related deliberations. Good practices and standards from other sectors exist and should be built upon to facilitate inclusion and advancement in decision-making at all levels, in panels, in corporate boardrooms, and research spaces. This requires specific commitments, actions and budgets.

4.2. Improve the participation of women across the AI lifecycle and entire ecosystem, and in AI governance mechanisms, at all levels. Women should have equal representation among policymakers, civil society, and in research institutes and the private sectors. While notable examples exist, the scope of these efforts needs to be broadened. Gender transformative, inclusive policymaking is about women, and about addressing different needs and priorities of those groups who have been historically marginalized.

4.3. Require AI related private organizations, educational institutions and government agencies to report gender disaggregated diversity statistics along with details on diversity at different responsibility levels within organizations (entry levels vs. management/ leadership levels), pay gaps and work environment considerations to promote gender equality.

4.4. Work together to ensure transformative gender approaches are considered and incorporated in all global, international and regional forums on the governance of artificial intelligence and create minimum standards, guidelines and definitions about the scope and quality of data on the AI ecosystem to allow countries and organizations to benchmark and compare data on progress and processes overtime and between organizations. Lack of comparable data is the first main obstacle to accountability regarding gender equality in the AI ecosystem. The Global Digital Compact is one such effort to bring an inclusive approach to an international space as is the internationally implemented UNESCO RAM.

4.5. Developing reporting indices to mark progress regularly that include data/measures specific to the AI ecosystem and that include more nuanced information about leadership and retention of employees, among others.

4.6. Specify at national, regional or sectoral levels quotas of **at least 40%** women and gender non-confirming representation in technology-related organizational boards and leadership.

4.7. Set quota for increasing women and gender non-confirming representation in AI policy making related government offices to **50%.**

4.8. Set quota of increasing women and gender non-confirming representation in AI related degree programs and faculty positions at educational institutions to 50%.

4.9. Enforce/support data collection to monitor those targets and quotas in representation and implement penalty mechanisms (i.e. taxes) if they are not met.

4.10. Specific measures should be in place to ensure the inclusion of women with disabilities, racialized women, persons with diverse sexual orientations and gender identities, native and indigenous women, among others. Guidelines to ensure inclusion already exist and should be incorporated and adapted further for the AI ecosystem and AI policymaking. When AI policy forums and the AI ecosystem at large is more diverse, AI policies will be more effective, more aligned with current obligations and commitments to gender equality, there will be less AI-related harms, and this in turn will foster more trust in AI in general.

Commitment by Member States in the UNESCO Recommendation:

Policy Area 6 (Gender), paragraph 88: "*…Increase the opportunities of girls' and women's participation in science, technology, engineering, and mathematics (STEM), including information and communication technologies (ICT) disciplines, preparedness, employability, equal career development and professional growth of girls and women, should be considered and implemented.*"

5. Strengthen or develop targeted interventions and programs to promote and support women to access, remain and advance in STEM-related careers, and in female entrepreneurship.

5.1. Strengthen initiatives for girls and young women to access equal opportunities in the STEM fields, including targeted programs, scholarships, mentorships and specific programs to facilitate access to venture capital, while recognizing that this is only an important first step. The challenges relate not only to enter the field, but also to remain and advance in what traditionally has been a male-dominated field, with related norms, culture and practice that represent barriers for women.

5.2. Develop and/or strengthen programs to support and finance women-led startups, research, civil society leaders and policymakers in the AI ecosystem, including by initiatives such as targeted trainings, specific funds and venture capital, networks and platforms to advance their work, successes and challenges.

5.3. Incorporate a gender perspective in education programs in STEM since early education, to ensure girls and boys have equal access to and opportunities to advance in STEM careers. Ensure the education system avoids perpetuating gender-stereotypes from early ages to higher degrees.

5.4. For all employers in the AI ecosystem, commit to improving and reporting on gender and diversity efforts and results, without which it is impossible to monitor progress and advance collectively towards a more equal AI ecosystem. Furthermore, it is essential to monitor for other characteristics and identities.

Commitment by Member States in the UNESCO Recommendation:

Policy Area 6 (Gender), paragraph 91: "*Member States should encourage female entrepreneurship, participation and engagement in all stages of an AI system life cycle by offering and promoting economic, regulatory incentives, among other incentives and support schemes, as well as policies that aim at a balanced gender participation in AI research in academia, gender representation on digital and AI companies' top management positions, boards of directors and research teams.*"

6. Establish standards for and monitor employers in AI the ecosystem, in academia and in the private and public sectors, to review and adapt their policies and practices to ensure they contribute to gender equality building on existing commitments, obligations and good practices.

6.1. Require all AI-related companies to create and advertise publicly a cross-cutting plan to respond to the specific challenges faced by women in AI, including barriers to entry, barriers to retention, and barriers to advancement.

6.2. Disseminate information about legal obligations widely to AI-related organizational leadership and require them to make known to employees the redress mechanisms available in order to promote zero tolerance for harassment, strict equal pay mechanisms, the improvement of quality and availability of care systems, and the creation of flexible work environments to enable women's participation in the AI workforce.

6.3. Provide trainings to all staff to prevent harassment, bullying and other intimidating behavior, including cyber-bullying, and to ensure accountability if it happens. Pay specific attention to relations among groups with power disparities.

6.4. Encourage and demand each stakeholder to report on key basic gender equality indicators in the workplace, such as gender pay gap, % of women in leadership positions, % of full-time workers that are female, % of employers that have flexible work options.

Commitment by Member States in the UNESCO Recommendation:

Policy Area 6 (Gender), Paragraph 89: " [The gender gap in AI includes:] *the gender wage gap; the unequal representation in certain professions and activities; the lack of representation at top management positions, boards of directors, or research teams in the AI field; the education gap; the digital and AI access, adoption, usage and affordability gap; and the unequal distribution of unpaid work and of the caring responsibilities in our societies.*"

Policy Area 6 (Gender), Paragraph 91: "*Policies on harassment-free environments should be developed and enforced, together with the encouragement of the transfer of best practices on how to promote diversity throughout the AI system life cycle.*"

Policy Area 6 (Gender), Paragraph 92: "*Member States should promote gender diversity in AI research in academia and industry by offering incentives to girls and women to enter the field, putting in place mechanisms to fight gender stereotyping and harassment within the AI research community, and encouraging academic and private entities to share best practices on how to enhance gender diversity.*"

7. Monitor and report on gender gaps and advances in access to AI, participation in the AI ecosystem and related governance mechanisms, and entrepreneurship, including by disaggregating and reporting on data on gender gaps in access and participation, and on measures taken to redress gender inequality.

7.1. Develop guidelines and standards on gender-disaggregated data all stages and for all stakeholders in the AI ecosystem and workplaces, related to hiring, recruitment and advancement in career, representation in boardrooms and in decision-making positions, among others, to establish clear, comparable baselines. Lack of data is the first main obstacle to accountability regarding gender equality in the AI ecosystem. As several published RAM reports displayed, this is not always a requirement in countries.

7.2. Commit to collect and report gender disaggregated data related more specifically to AI employment and use. Current data and reports are not specific enough to the AI ecosystem; are not measured in a standard way that allows for comparison; and are focused on the Global North, and more specifically, in the US. Therefore, it is difficult to monitor based on national or international benchmarks or baselines related to AI, and to the broader STEM or ICT sectors. Moreover, there has been more focus on gender disaggregated participation measurement at only one end of the pipeline, mainly the percentage of women in STEM education. Stronger standards for indices and reporting must be developed to include measures of the AI ecosystem including leadership or retention of employees, among others.

7.3. Incorporate solid transformative gender approaches and intersectionality into all monitoring and evaluation mechanisms as well as in tools and assessments, which is a main gap in all stakeholders in the AI ecosystem. In addition to specifically having better data concerning women's participation and rights in the AI ecosystem and in related governance mechanisms, indicators and data need to incorporate other important markers and indicators related to factors such as the gender spectrum, beyond binaries man/women, and disability, race, ethnicity, socioeconomic status, and other marginalized identities and characteristics of national contexts.

7.4. Incorporate in AI policies and strategies participatory mechanisms for review on a regular basis. Given the rate of development of artificial intelligence technologies, infrequent cyclical policy reviews and updates may not attend to the rapidly evolving needs and rights of those whom they govern and serve, and which still disproportionately affect women and marginalized groups.

Commitment by Member States in the UNESCO Recommendation:

Policy Area 6, (Gender), paragraph 90: "*Member States should ensure that gender stereotyping and discriminatory biases are not translated into AI systems, and instead identify and proactively redress these.*"

8. Review current legislation and, if necessary, amend or adopt new regulatory measures to ensure remedy and reparation mechanisms for harm caused by AI systems, particularly due to the lack of design, development, deployment and monitoring in accordance with human rights and gender equality standards.

8.1. Ensure wrongdoing and AI-related harms are redressed by building on existing legislation related to gender equality and non-discrimination, among others, including normative and regulatory responses to online and offline gender-based violence, discrimination in education, employment, health and access to justice, among others, while advancing legislative participatory debates on AI-specific legislation.

8.2. Encourage the office of Ombudsman, or its equivalent, at national level to consider AI-related cases within the scope of work and mobilize other human rights defenders on this issue.

8.3. Carry out awareness raising sessions among the general public, human rights organizations, and representative organizations of marginalized groups so that the general public know when AI is being used particularly in service delivery or other public services, what the risks are, how to learn more, and how to demand and access remedy and reparation if needed.

Commitment by Member States in the UNESCO Recommendation:

Section III.2 (Principles), paragraph 28: Fairness and non-discrimination. *"AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems. Effective remedy should be available against discrimination and biased algorithmic determination."*

Section III. 2 (Principles), paragraph 44: Awareness and literacy. *"Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy and training led jointly by governments, intergovernmental organizations, civil society, academia, the media, community leaders and the private sector, and considering the existing linguistic, social and cultural diversity, to ensure effective public participation so that all members of society can take informed decisions about their use of AI systems and be protected from undue influence."*

9. Encourage the development and creation of learning communities and peer-to peer networks on gender and diversity in AI at national, regional and international levels, for increasing innovation and exchange, replication and upscaling of evidence-based good practices.

9.1. Create and/or strengthen national, regional and international repositories of good practices and policies related to reducing the adverse impacts of AI on gender equality and scaling up the opportunities for AI to promote gender equality, including a) incentivizing participation of girls, women and underrepresented groups in all stages of the AI system lifecycle; b) working to ensure that AI and tech-related workplaces adopt, implement and monitor policies to ensure an environment conducive to gender equality, and c) effectively incorporating, implementing, monitoring and reporting on transformative gender and diversity approaches in AI governance mechanisms, policies and legislations, at all levels.

9.2. Create open and accessible databases where collected gender disaggregated participation data from government agencies, educational institutions, and private organizations can be accessed by the public.

9.3. Improve readiness assessment mechanisms by strengthening the gender related indicator included that focus on AI specifically and include multiple levels and types of participation and types of impacts of AI.

9.4. Build upon other repositories and information sources to advocate for systematic inclusion and visibility of matters related to gender equality and diversity.

9.5. Specifically, increase efforts to ensure data is disaggregated by gender and other factors throughout the AI cycle, and exchange standards, good practices and lessons learned in this area, as a basis to improve practice and inform policy. Readiness assessments should be further detailed and comprehensive to promote more effective and sustainable improvements in this area.

9.6. Provide technical and financial support for women-led initiatives, particularly in the Global South, to continue, strengthen and amplify their work to ensure gender equality in all AI-related issues, stages, levels and sectors.

Commitment by Member States in the UNESCO Recommendation:

Policy Area 6 (Gender), paragraph 93: *"UNESCO can help form a repository of best practices for incentivizing the participation of girls, women and under-represented groups in all stages of the AI system life cycle."*

The Women for Ethical AI (W4EAI) Platform will use these recommendations, and the insights derived from the outlook study to move forward with its goal of advancing gender equality in the AI agenda.

References

Ación, L., Alemany, LA., Benotti, L., Bordone, M., Busaniche, B., González, L., and Halvorsen. A. (2023). A tool to overcome technical barriers for bias assessment in human language technologies (e.d.i.a paper). In *Inteligencia Artificial Feminista: hacia una agenda de investigación para América Latina y el Caribe*. https://archive.org/details/inteligencia-artificial-feminista/mode/1up

Access Now (2021). The EU should regulate on the basis of rights, not risks. https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

Adams, R., Adeleke, F., Florido, A., de Magalhães Santos, L. G., Grossman, N., Junck, L., and Stone, K. (2024). Global Index on Responsible AI 2024 (1st Edition). South Africa: Global Center on AI Governance.

Ada Lovelace Institute. (2022). Algorithmic impact assessment: a case study in healthcare. https://www.adalovelaceinstitute.org/report/algorithmic-impactasssessment-case-study-healthcare

Agwaibor, S. (2024). IWD2024: Driving digital gender inclusion in Africa. TC Insights. https://techcabal.com/2024/03/26/iwd2024-driving-digital-gender-inclusion-africa/

Alan Turing Institute. (2024). Rebalancing Innovation: Women, AI and Venture Capital in the UK, Second Report. Women in Data Science and AI Project. https://www.turing.ac.uk/news/publications/rebalancing-innovation

Amnesty International. (2018). Crowdsourced Twitter study reveals shocking scale of online abuse against women. https://www.amnesty.org/en/latest/press-release/2018/12/crowdsourced-twitter-study-reveals-shocking-scale-of-online-abuse-against-women/

Ashcraft, C., McLain, B. and Eger, E. (2016). Women in Tech: The Facts. National Center for Women in Information Technology. https://wpassets.ncwit.org/wp-content/uploads/2021/05/13193304/ncwit_women-in-it_2016-full-report_final-web06012016.pdf

Australian Government. (n.d.). Australia's AI Ethics Principles. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

AWS. (2024). Amazon SageMaker Clarify: Evaluate models and explain model predictions. https://aws.amazon.com/sagemaker/clarify

Berkeley Haas Center for Equity, Gender and Leadership. (n.d.). Mitigating Bias in Artificial Intelligence - Bias in AI: Examples Tracker. https://haas.berkeley.edu/equity/resources/playbooks/mitigating-bias-in-ai/

Biana, H. and Domingo, R. (2024). Strong Girl AI: Frameworks for the Empowered Mobility of Women in Southeast Asia. Feminist AI in Southeast Asia. https://feministai.pubpub.org/pub/zjrrb18z/release/1

Bogen, M. (2019). All the Ways Hiring Algorithms Can Introduce Bias. *Harvard Business Review*. https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias

Brancaccio, G., Balato, A., Malvehy, J., Puig, S., Argenziano, G. and Kittler, H. (2024). Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check. Journal of Investigative Dermatology, 144 (3), 492-499. https://doi.org/10.1016/j.jid.2023.10.004.

Bretag, T., Mahmud, S., Wallace, M., Walker, R., James, C., Green, M., East, J., McGowan, U., & Partridge, L. (2011). Core elements of exemplary academic integrity policy in Australian higher education. *International Journal for Educational Integrity*, 7(2), 3-12. https://doi.org/10.21913/IJEI.v7i2.759

Buolamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research Conference on Fairness, Accountability, and Transparency*, 81, 1-15.

Burns, D., Grabowsky, T., Kemble, E. and Pérez, L. (2023). Closing the data gaps in women's health. Mckinsey and Company. https://www.mckinsey.com/industries/life-sciences/our-insights/closing-the-data-gaps-in-womens-health

Caira, C., Russo, L., and Aranda, L. (2023). Artificially Inequitable? AI and closing the gender gap. OECD AI Policy Observatory. https://oecd.ai/en/wonk/closing-the-gender-gap

Center for AI and Digital Policy (CAIDP). (2024). Artificial Intelligence and Democratic Values Index - 2023. https://www.caidp.org/reports/aidv-2023/

Ciston, S. (2019). Intersectional AI is essential: Polyvocal, multimodal, experimental methods to save artificial intelligence. *Journal of Science and Technology of the Arts*, 11(2), 3-8.

Collett, C., Neff, G. and Gomes, L.G. (2022). The effects of AI on the working lives of women. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380861

Corneliussen, H.G. (2019). "I have never done this before, so I'm probably really good at this". Report on women in ICT and ICT security. Western Norway Research Institute. https://www.vestforsk.no/en/publication/i-have-never-done-so-im-probably-really-good-report-women-ict-and-ict-security

Costanza-Chock, S. (2018) Design justice, AI, and escape from the matrix of domination. *Journal of Design and Science*, 3(5). https://jods.mitpress.mit.edu/pub/costanza-chock/release/4

Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, Gigante A, Valencia A, Rementeria MJ, Chadha AS, Mavridis N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digital Medicine*, (3)81. https://doi.org/10.1038/s41746-020-0288-5

Dastin, J. (2018). Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 11 October 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/

Degado, C., Gonzalez, L.E, Martinez, C., and Savage, S. (2021). Candidaturas Paritarias y Violencia Política Digital en México: Un Análisis de Datos sobre la Violencia Política en Razón de Género. UNDP. https://www.undp.org/es/mexico/publicaciones/candidaturas-paritarias-y-violencia-politica-digital-en-mexico-un-analisis-de-datos-sobre-la-violencia-politica-en-razon-de

Deloitte. (2022). Progress at a Snail's Pace. Women in the Boardroom: A Global Perspective. 7th Edition. https://www2.deloitte.com/content/dam/Deloitte/at/Documents/human-capital/at-women-in-the-boardroom-2022.pdf

Di meco, L. and Brechenmacher, S. (2020). Tackling Online Abuse and Disinformation Targeting Women in Politics. Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2020/11/tackling-online-abuse-and-disinformation-targeting-women-in-politics

Eaton, S. E., et al. (2019). Contract Cheating in Canada: National Policy Analysis. https://osf.io/n9kwt/

European Institute for Gender Equality. (2021). Artificial intelligence, platform work and gender equality. https://eige.europa.eu/publications-resources/publications/artificial-intelligence-platform-work-and-gender-equality?language_content_entity=en

European Commission. (2024). The Human Rights Based Approach (HRBA). https://wikis.ec.europa.eu/pages/viewpage.action?pageId=50108948

European Commission. (2022). High-level expert group on artificial intelligence. https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

European Commission. (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Parliament and Council of Europe. (2024). Artificial Intelligence Act. Regulation (EU) 2024/1689 of the European parliament and the Council. 13 June 2024.

Feldfeber, I., Quirogo, Y.B., Guevara, C. and Felice, M.C. (2024). Feminisms in Artificial Intelligence: Automation Tools towards a Feminist Judiciary Reform in Argentina and Mexico. Feminist AI. https://feministai.pubpub.org/pub/z83eyq54/release/1

Feminist AI Research Network (2024). Incubating Feminist AI 2021-2024. https://drive.google.com/file/d/15mYsJqpaeEpklhnjR51ZbJtz_2lbKJzn/view

FemTechnology and Women at the Table. (2023). The Gender Data Health Gap: Harnessing AI's Transformative Power to Bridge the Gender Health Data Divide. https://www.womenatthetable.net/wp-content/uploads/2024/01/Gender-Data-Health-Gap_compressed.pdf

Fontes, C., Hohma, E., Corrigan, C.C., and Lütge, C. (2022). AI-powered public surveillance systems: why we (might) need them and how we want them. *Technology in Society*. https://doi.org/10.1016/j.techsoc.2022.102137

Fultinavičiūtė, U. (2022). Sex and science: underrepresentation of women in early-stage clinical trials. Clinical Trials Arena. https://www.clinicaltrialsarena.com/features/underrepresentation-women-early-stage-clinical-trials

Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A et al. (2021). Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ*, 374 :n1872. https://doi.org/10.1136/bmj.n1872

Genero, D. (2024). Feminisms in Artificial Intelligence: Automation Tools towards a Feminist Judiciary Reform in Argentina and Mexico | AymurAI. Feminist AI. https://feministai.pubpub.org/pub/z83eyq54/release/1

Gillwald, A. and Partridge, A. (2022). Gendered Nature of Digital Inequality: evidence for policy considerations. Prepared for the Expert Group Meeting of the sixty-seventh session of the Commission on the Status of Women.

Glitch UK and End Violence Against Women Coalition. (2020). The Ripple Effect: COVID-19 and the Epidemic of Online Abuse. Glitch and End of Violence Against Women Coalition. https://www.endviolenceagainstwomen.org.uk/wp-content/uploads/2020/09/Glitch-and-EVAW-The-Ripple-Effect-Online-abuse-during-COVID-19-Sept-2020.pdf

Global Partnership on AI. (2023). Towards Real Diversity and Gender Equality in Artificial Intelligence - Advancement Report.https://gpai.ai/projects/responsible-ai/towardsrealdiversityandgenderequalityinai/

GoogleAI. (n.d.). All Principles: Objectives for Building Beneficial AI. https://ai.google/responsibility/principles/

Government of Canada. (2023a). Directive on Automated Decision-Making. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

Government of Canada. (2023b). 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

The Government of Hong Kong Special Administrative Region of the People's Republic of China: Digital Policy Office. (2024). Ethical Artificial Intelligence Framework: (Customized version for general reference by public). https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/

Government of Ireland. (2021). AI - Here for Good: A National Artificial Intelligence Strategy for Ireland. https://enterprise.gov.ie/en/publications/publication-files/national-ai-strategy.pdf

Government of Japan. (2019). AI Utilization Guidelines: Practical Reference for AI utilization. https://www.soumu.go.jp/main_content/000658284.pdf

Government of Serbia. (2023). Ethical Guidelines for Development, Implementation and Use of Robust and Accountable Artificial Intelligence. https://www.ai.gov.rs/extfile/en/471/Ethical%20guidelines%20for%20development%20implementation%20and%20use%20of%20robust%20and%20accountable%20AI.pdf

Government of Spain. (2022). Información algorítmica en el ámbito laboral: Guia Practica y Herramienta Sobre la Obligación Empresarial de Información Sobre el Uso De Algoritmos en al Ambito Laboral. Ministerio de Trabajo y Economía Social.

Global Index on Responsible AI (GIRAI). (2024). https://girai-report-2024-corrected-edition.tiiny.site/

GSMA. (2022). The Mobile Gender Gap Report 2022. https://www.gsma.com/r/wp-content/uploads/2022/06/The-Mobile-Gender-Gap-Report-2022.pdf

Guerin, C. and Maharasingam-Shah, E. (2020). Public Figure, Public Rage: Candidate Abuse on Social Media. Institute for Strategic Dialogue. https://www.isdglobal.org/wp-content/uploads/2020/10/Public-Figures-Public-Rage-4.pdf

High-level Expert Group on Artificial Intelligence (HLEG). (2020). The Assessment List for Trustwothy Artificial Intelligence (ALTAI) for self assessment. European Commission.

Hinchliffe, E. (2023). Women CEOs run 10.4% of Fortune 500 companies. A quarter of the 52 leaders became CEO in the last year. *Fortune*. https://fortune.com/2023/06/05/fortune-500-companies-2023-women-10-percent/

IBM. (2024). AI Ethics. https://www.ibm.com/impact/ai-ethics

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (IEEE). (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html

Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC). (2020). Model Artificial Intelligence Governance Framework, Second Edition. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

Interagency Gender Working Group. (2017). The Gender Integration Continuum. Washington, DC: Population Reference Bureau. https://www.igwg.org/wp-content/uploads/2017/05/Gender-Continuum-PowerPoint_final.pdf

International Monetary Fund (IMF). (2023). AI Preparedness Index (AIPI) Note. https://www.imf.org/external/datamapper/AIPINote.pdf

International Telecommunications Union (ITU). (2023a). Bridging the gender divide. https://www.itu.int/en/mediacentre/backgrounders/Pages/bridging-the-gender-divide.aspx

International Telecommunications Union (ITU). (2023b). The Gender Digital Divide. https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-the-gender-digital-divide/

Internet Governance Forum (IGF). 2021. Best Practice Forum on Gender and Digital Rights: Exploring the concept of gendered disinformation. https://intgovforum.org/en/filedepot_download/248/21181

Jeong, J.,Vey, B.L., Bhimireddy, A., Kim, T., Santos, T., Correa, R., Dutt, R., Mosunjac, M., Oprea-Ilies, G., Smith, G., Woo, M., McAdams, C.R., Newell, M.S., Banerjee, I., Gichoya, J., and Trivedi, H. (2023). The EMory BrEast Imaging Dataset (EMBED): A Racially Diverse, Granular Dataset of 3.4 Million Screening and Diagnostic Mammographic Images Radiolog. *Artificial Intelligence*, 5:1.

JP Morgan. (2023). Top 100 Women-Powered, High-Growth Businesses in Asia Pacific.

Keller, D. (2021). Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard. *Journal of Free Speech*, 1(1). https://www.journaloffreespeechlaw.org/keller.pdf

Kriebitz, A. & Corrigan, C.C & Boch, A. (2024). Munich Convention on Artificial Intelligence, Data and Human Rights (Draft for Public Consultation). Proceedings from: International Summit on Artificial Intelligence and Human Rights. https://www.researchgate.net/publication/384678262_Munich_Convention_on_Artificial_Intelligence_Data_and_Human_Rights_Draft_for_Public_Consultation

Lamm, E. Ramos, G., Ronchi, E., Squicciarini, M. (2023). The gendered impacts of AI: policies and safeguards to regulate new technologies, mitigate risks and protect rights. UN Women, Expert Group Meeting 'Artificial intelligence" 10 – 13 October 2023.

Lazzeri, F. (2020). Fairlearn - A Python package to assess AI system's fairness. Microsoft. https://techcommunity.microsoft.com/t5/educator-developer-blog/fairlearn-a-python-package-to-assess-ai-system-s-fairness/ba-p/1402950

Leslie, D., Burr, C., Aitken, M., Katell, M., Briggs, M., Rincon, C. (2021). Human rights, democracy, and the rule of law assurance framework for AI systems: A proposal. The Alan Turing Institute. https://doi.org/10.5281/zenodo.5981676

Lui, K.A. and Dipietro Mager, N.A. (2016). Women's involvement in clinical trials: historical perspective and future implications. *Pharm Pract* (Granada), 14 (1): 708. https://pubmed.ncbi.nlm.nih.gov/27011778/

Maleki Varnosfaderani S. and Forouzanfar M. (2024). The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. *Bioengineering* (Basel), 11(4):337. https://pubmed.ncbi.nlm.nih.gov/38671759/

Marcus, R., Samuels, F., Jalal, S. and Belachew, H. (2022). Gender-Transformative Programming. Background Paper Series. UNICEF Gender Policy and Action Plan 2022-2025. https://www.unicef.org/lac/en/media/43146/file

Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham,, Y., Wald, R., and Clark, J. (2024). The AI Index 2024 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA. https://aiindex.stanford.edu/report/

Mazure, C.M. and Jones, D.P. (2015). Twenty years and still counting: Including women as participants and studying sex and gender in biomedical research. *BMC Womens Health*, 15: 94. https://pubmed.ncbi.nlm.nih.gov/26503700/

Microsoft. (2020). InterpretML: A toolkit for understanding machine learning models. https://www.microsoft.com/en-us/research/uploads/prod/2020/05/InterpretML-Whitepaper.pdf

Ministério de Ciência, Tecnologia e Inovações. (2021). Estratégia Brasileira de Inteligência Artificial -EBIA. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-diagramacao_4-979_2021.pdf

Miron, J. B., McKenzie, A., Eaton, S. E., Stoesz, B. M., Thacker, E., Devereaux, L., Persaud, N., Steeves, M., & Rowbotham, K. (2021). Academic integrity policy analysis of publicly-funded universities in Ontario, Canada: A focus on contract cheating. *Canadian Journal of Educational Administration and Policy*, 197, 62-75. https://journalhosting.ucalgary.ca/index.php/cjeap/article/view/72082

Möller, A. (2022). An analysis of university academic integrity policies in New Zealand. *Journal of Further and Higher Education*, 1-13. https://doi.org/10.1080/0309877X.2022.2130195

Moya, B. A. (2024). Academic Integrity Policies in Hispanic South American Higher Education: Status and Recommendations. In S. E. Eaton (Ed.), *Second Handbook of Academic Integrity* (pp. 911-933). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54144-5_122

Moya, B. A., & Eaton, S. E. (2024). Academic Integrity Policy Analysis of Chilean Universities. *Journal of Academic Ethics*. https://doi.org/10.1007/s10805-024-09515-w

New York City Council. (2021). A Local Law to amend the administrative code of the city of New York, in relation to automated employment decision tools. Int 1894-2020. https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page#:~:text=Local%20Law%20144%20of%202021,audit%20is%20publicly%20available%2C%20and

Nicoletti, L. and Bass, D. (2023). Humans Are Biased. Generative AI Is Even Worse. *Bloomberg*, 9 June 2023. https://www.bloomberg.com/graphics/2023-generative-ai-bias/

Noble, S.U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

OECD. (2024a). OECD AI Incidents Monitor (AIM). https://oecd.ai/en/incidents

OECD. (2024b). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD. (2023). Using AI to support people with disabilities in the labour market. https://www.oecd.org/en/publications/using-ai-to-support-people-with-disability-in-the-labour-market_008b32b7-en.html

OECD. (2018). Bridging the Digital Gender Divide: Include, Upskill, Innovate.

OECD.AI. (2024). OECD.AI Policy Observatory - Why we need a catalogue of tools and metrics for trustworthy AI. https://oecd.ai/en/catalogue/overview

OECD.AI. (2021), powered by EC/OECD (2021), database of national AI policies, accessed on 19/08/2024, https://oecd.ai.

Olivia, Lilian. (2023). The Gender Equality Mirage: From Human Bias to AI Bias in Digital ID Systems in Africa. https://cipit.strathmore.edu

Oxfam International. (2024). INEQUALITY INC. How corporate power divides our world and the need for a new era of public action. https://webassets.oxfamamerica.org/media/documents/Inequality_Inc._k6NfmGq.pdf

Oxford Insights. (2023). Government AI Readiness Index - 2023. https://oxfordinsights.com/ai-readiness/ai-readiness-index/

P, D., S. Simoes, and M. MacCarthaigh. (2023). AI and core electoral processes: Mapping the horizons. *AI Magazine,* 44: 218–239. https://doi.org/10.1002/aaai.12105

Pal, K.K, Piaget, K. and Zahidi, S. (2024). Global Gender Gap Report 2024. World Economic Forum (WEF). https://www3.weforum.org/docs/WEF_GGGR_2024.pdf

Pitchbook. (2024). European VC female founders dashboard. https://pitchbook.com/news/articles/the-european-vc-female-founders-dashboard

Plan International. (2021). The Trust Gap: How Misinformation and Disinformation Online Affect the Lives, Learning and Leadership of Girls and Young Women. The State of the World's Girls Reports. https://plan-international.org/publications/the-truth-gap/

Prescott, A. (2023). Bias in Big Data, Machine Learning and AI: What Lessons for the Digital Humanities? *DHQ: Digital Humanities Quarterly*, 17(2). https://www.digitalhumanities.org/dhq/vol/17/2/000689/000689.html

Responsible AI Collaborative. (2024) AI Incident Database. https://incidentdatabase.ai/

Revilla, T. (2022) Integración de la perspectiva de género en las estrategias de inteligencia artificial de América Latina: de la narrativa a la acción. En P. Ricaurte y M. Zasso (2022) *Inteligencia artificial feminista: Hacia una agenda de investigación en América Latina y el Caribe*. https://feministai.pubpub.org/pub/nly15827/release/1

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., & Zasso, M. (2023). AI, Ethics and Coloniality: A Feminist Critique. In: *What AI can Do*. Milton Park, 53-72.

Satariano, A. and Pifarre, R.T. (2024) An Algorithm Told Police She Was Safe. Then Her Husband Killed Her. *New York Times*, 18 July 2024.

Schneegans, S., Lewis, J. and T. Straza (eds). (2021). UNESCO Science Report: the Race Against Time for Smarter Development – Executive Summary. UNESCO Publishing: Paris

Servicio de Empleo Público Estatal del Ministerio de Trabajo y Economía Social (España). (2024). El programa Radia mejora la inclusión de mujeres con discapacidad en la economía digital. https://www.sepe.es/HomeSepe/es/que-es-el-sepe/comunicacion-institucional/noticias/detalle-noticia?folder=/SEPE/2024/Mayo/&detail=programa-Radia-mejora-inclusion-de-mujeres-con-discapacidad-economia-digital

Sey, A and Kafkin, N. (2019). Taking Stock: Data and Evidence on Gender Equality in Digital Access, Skills and Leadership. Report of Equals Research Group, Led by the United Nations University. https://i.unu.edu/media/cs.unu.edu/attachment/4040/EQUALS-Research-Report-2019.pdf

Shahriari, K., & Shahriari, M. (2017). IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In 2017 IEEE Canada International Humanitarian Technology Conference (IHTC), pp. 197-201). IEEE.

Singh, N.C. (n.d.). Personalising 'Learning' - Can AI Promise Customised Education For 'Humanity'. UNESCO. https://mgiep.unesco.org/article/personalising-learning-can-ai-promise-customised-education-for-humanity

Slattery, P., Saeri, A. K., Grundy, E. A. C., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., & Thompson, N. (2024). A systematic evidence review and common frame of reference for the risks from artificial intelligence. http://doi.org/10.13140/RG.2.2.28850.00968

Smith, G. (2020). What does "fairness" mean for machine learning systems? Center for Equity, Gender & Leadership (EGAL) at Berkeley Haas. https://haas.berkeley.edu/equity/

Smith, G. and Rustagi, I. (2021). When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. *Stanford Social Innovation Review*. https://doi.org/10.48558/A179-B138

Squicciarini, M., O'Kane, L., Nania, J., Liu, E., Bingham, S. (2024). The weight of words: gendered language and women's participation and positioning in the labor market. https://unesdoc.unesco.org/ark:/48223/pf0000388794

Stahl, B.C., Antoniou, J., Bhalla, N. et al. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*, 56: 12799–12831. https://doi.org/10.1007/s10462-023-10420-8

Sulis, G. Kim, J.Y., Rodrigue, V., Gore, G., Peebles, A., Ulrich, A.K., Horn, M., and Basta, N.E. (2023). Sex-disaggregated effectiveness data reporting in COVID-19 vaccine research: a systematic review. *Communications Medicine* (Lond), 3:69. https://doi.org/10.1038/s43856-023-00297-7

Suresh, H. and Guttag, J. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *Equity and Access in Algorithms, Mechanisms, and Optimization*. https://dl.acm.org/doi/10.1145/3465416.3483305

The Economist. (2019). The Vile Experiences of Women in Tech. https://www.economist.com/open-future/2019/05/03/the-vile-experiences-of-women-in-tech

The Economist Intelligence Unit. (2021). Measuring the prevalence of online violence against women. https://onlineviolencewomen.eiu.com/

The Future Society. (2020). Areas for Future Action in the Responsible AI Ecosystem. https://gpai.ai/projects/responsible-ai/areas-for-future-action-in-responsible-ai.pdf

The White House. (2022). A Blueprint for An AI Bill of Rights. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

UCL News. (2022). Gender bias revealed in AI tools screening for liver disease. https://www.ucl.ac.uk/news/2022/jul/gender-bias-revealed-ai-tools-screening-liver-disease

UK Government. (2022). Auditing algorithms: the existing landscape, role of regulators and future outlook. https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook

UN Department of Economic and Social Affairs. (2024). The 17 Goals. https://sdgs.un.org/goals

UN Global Digital Compact. (2024). GDC Rev 3 - Draft Under Silence Procedure. https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/GDC_Rev_3_silence_procedure.pdf

UN Office of the United Nations High Commissioner for Human Rights (OHCHR). (2022). Interlinkages between Women's Rights and Digital Technologies, Civic Space, Data and Privacy, and Freedom of Expression. Observer paper, UN Women, Expert Group Meeting

UN Office of the High Commissioner for Human Rights (OHCHR). (2011). Guiding Principles on Business and Human Rights. New York and Geneva.

UN Sustainable Development Group. (2024a). Human Rights-Based Approach. https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach

UN Sustainable Development Group (2024b). Transforming Care Systems: UN System Policy Paper. https://unsdg.un.org/resources/transforming-care-systems-un-system-policy-paper

UNCES. 2018. Declaration on Gender-Responsive Standards and Standards Development, https://unece.org/trade/wp6/Gender-Resp-Stdards-declaration

UNESCO. (2024a). Women's access to and participation in technological developments. https://www.unesco.org/en/artificial-intelligence/gender-equality

UNESCO. (2024b). The Gender Gap in Science: Status and Trends. https://unesdoc.unesco.org/ark:/48223/pf0000388805

UNESCO. (2023a). UNESCO in Action for Gender Equality. Paris, UNESCO.

UNESCO. (2023b). Global Education Monitoring Report 2023: Technology in education: A Tool on whose Terms? Paris, UNESCO.

UNESCO. (2023c). Readiness Assessment Methodology: A Tool of the Recommendation on the Ethics of Artificial Intelligence.

UNESCO. (2023d). New recommendations to improve gender equality in digital professions and eliminate stereotypes in AI applications.

UNESCO. (2023e). Guidelines for the Governance of Digital Platforms. https://unesdoc.unesco.org/ark:/48223/pf0000387339

UNESCO. (2023f). Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence.

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

UNESCO. (2022b). Fact Sheet: #HerEducationOurFuture #BreakTheBias: challenging gender bias and stereotypes in and through education; the latest facts on gender equality in education. https://www.unesco.org/gem-report/en/her-education-our-future-break-bias

UNESCO. (2020). Artificial Intelligence and Gender: Key findings of UNESCO's Global Dialogue. Paris, UNESCO.

UNESCO & IRCAI. (2024). Challenging Systematic Prejudices: An Investigation into Bias Against Women and Girls in Large Language Models.

UNFPA. (2023). Gender Transformative Approaches to Achieve Gender Equality and Sexual and Reproductive Health and Rights. https://www.unfpa.org/sites/default/files/pub-pdf/UNFPA_GTA-2023.pdf

UN Media Office. (2023). Secretary-General Rejects 'Male Chauvinist' Domination of Tech Sector, Calls for Overhaul of 'Patriarchal Structures', at Women's Civil Society Town Hall. https://press.un.org/en/2023/sgsm21723.doc.htm

United Nations. (2024a). General Assembly Adopts Landmark Resolution on Steering Artificial Intelligence towards Global Good, Faster Realization of Sustainable Development. Seventy-eighth session, GA/12588. https://press.un.org/en/2024/ga12588.doc.htm

United Nations. (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

United Nations. (1948). Universal Declaration of Human Rights. General Assembly resolution 217 A. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations Children's Fund (UNICEF). (2023). Bridging the Gender Digital Divide: Challenges and an Urgent Call for Action for Equitable Digital Skills Development, UNICEF, New York.

United Nations Children's Fund (UNICEF). (2022). UNICEF disability and inclusion strategy, https://www.unicef.org/media/134031/file/UNICEF_Disability_and_Inclusion_Strategy_2022_2030_Short_Version.pdf

United Nations Children's Fund (UNICEF). (2020). Gender-Responsive Age-Sensitive Social Protection: A Conceptual Framework. Innocenti Working Papers no. 2020-10. Florence: UNICEF Office of Research – Innocenti.

UN Women. (2023). Innovation and Technological Change, and Education in the Digital Age for Achieving Gender Equality and the Empowerment of All Women and Girls. CSW 67. https://www.unwomen.org/sites/default/files/2023-02/230213%20BLS22613%20UNW%20CSW67.v04%20%282%29.pdf

UN Women. (2020). Gender mainstreaming: A global strategy for achieving gender equality and the empowerment of women and girls. https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2020/Gender-mainstreaming-Strategy-for-achieving-gender-equality-and-empowerment-of-women-girls-en.pdf

UN Women. (n.d.). FAQs: Trolling, stalking, doxing and other forms of violence against women in the digital age. https://www.unwomen.org/en/what-we-do/ending-violence-against-women/faqs/tech-facilitated-gender-based-violence

Velázquez, J.D.M., Šcepanovi, S., Gvirtz, A. and Quercia, D. (2024). Decoding Real-World AI Incidents. *IEEE Computer Society*, April. DOI: 10.1109/XXX.0000.0000000

Wan, Y., Pu, G., Sun, J., Garimella, A., Chang, K., and Peng, N. (2023). "Kelly is a Warm Person, Joseph is a Role Model": Gender Biases in LLM-Generated Reference Letters. https://arxiv.org/pdf/2310.09219

Ward, J. Spencer, S. and Kalsi, K. (2023). Gender-Based Violence and Artificial Intelligence (AI): Opportunities and Risks for Women and Girls in Humanitarian Settings. GBV AoR Helpdesk: Gender-Based Violence in Emergencies.

Weinberger, D. (n.d.). Playing with AI Fairness: Google's new machine learning diagnostic tool lets users try on five different types of fairness. Google. https://pair-code.github.io/what-if-tool/ai-fairness.html

West, D.M. (2023). AI Poses Disproportionate Risks to Women. Brookings. https://www.brookings.edu/articles/ai-poses-disproportionate-risks-to-women/

West, M., Kraut R., and Chew H.E. (2019). I'd Blush if I Could: Closing Gender Divides and Digital Skills Through Education. EQUALS and UNESCO.

West, S.M., Whittaker, M. and Crawford, K. (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

WeXchange and Inter-america Development Bank (IDB) Lab. (2020). The Rise of Women STEMpreneurs: A Study on Women Entrepreneurs in STEM in Latin American and the Caribbean. WXInsights. https://publications.iadb.org/en/publications/english/viewer/wX_Insights_2020_The_Rise_of_Women_STEMpreneurs_A_Study_on_Women_Entrepreneurs_in_STEM_in_Latin_America_and_the_Caribbean.pdf

Wittemeyer, R., Nowski, T., Ellingrud, K. and Conway, M. (2018). Rebooting Representation: Using CSR and Philanthropy to Close the Gender Gap in Tech. https://rebootrepresentation.org/rebooting-representation/

World Health Organization (WHO). (2024). The role of artificial intelligence in sexual and reproductive health and rights. https://iris.who.int/bitstream/handle/10665/376294/9789240090705-eng.pdf?sequence=1

World Economic Forum (WEF). (2023). Global Gender Gap Report 2023. https://www3.weforum.org/docs/WEF_GGGR_2023.pdf

Xu, C. and Doshi, T. (2019). Fairness Indicators: Scalable Infrastructure for Fair ML Systems. Google Research. https://research.google/blog/fairness-indicators-scalable-infrastructure-for-fair-ml-systems/

Young, E., Wajcman, J. and Sprejer, L. (2021). Where are the Women? Mapping the Gender Job Gap in AI. Policy Briefing: Summary. The Alan Turing Institute. https://www.turing.ac.uk/news/publications/report-where-are-women-mapping-gender-job-gap-ai

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J.C., Sellitto, M., Shoham, Y., Clark, J., and Perrault, R. (2021). The AI Index 2021 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021.


UNESCO
unesco