

Women's Rights and AI : A New Threat?

Carolina Batista & Nil Başgezer

Women's Rights Team – July 1st, 2024



© ThisIsEngineering/Pexels. (2020). *Code projected over woman* [Photograph].



Riviermarkt 5, 2513 AM The Hague, Netherlands



+ 31 62 72 41006



info@ghrd.org



<https://www.ghrd.org>

CONTENTS

INTRODUCTION	3
1. LEGAL FRAMEWORKS	5
1.1 UNITED NATIONS	5
1.2 EUROPEAN UNION	6
1.3 UNITED STATES	8
1.4 CHINA	9
1.5 ASSESSMENT	10
2. CASE STUDIES	11
2.1 AI SURVEILLANCE IN IRAN	11
2.2 TAYLOR SWIFT VICTIM OF NUDE DEEP FAKES	12
2.3 AMAZON'S 'SEXIST AI' HIRING TOOL	13
3. RECOMMENDATIONS	15
CONCLUSION	18

GLOBAL
HUMAN
RIGHTS
DEFENCE

INTRODUCTION

The rapid advancement in generative Artificial Intelligence (AI) technology is marked by frequent launches of new large-scale models, heralding a transformative era in human work dynamics and communication (Hacker, 2023). However, Generative AI poses numerous risks and challenges. Ethical concerns and potential social exclusion are exemplified by the amplification of societal biases, discrimination, and inequities, notably in fields such as recruitment (Raub, 2018). Another significant risk involves privacy and data protection, encompassing lawful and compliant data collection, processing, and management of sensitive personal information.

Gender inequality persists globally, and AI reflects, and perpetuates these biases ingrained in our society. According to UN Women (2024), despite increasing global internet access among women, only 20 percent of low-income countries are connected, contributing to a significant gender digital divide that is mirrored in AI's data biases. Research conducted by the Berkeley Haas Center for Equity, Gender, and Leadership examined 133 AI systems across various sectors, revealing that approximately **44 percent exhibited gender bias, with 25 percent displaying both gender and racial biases** (Smith & Rustagi, 2021).

**Gender
biased AI
Systems**

44%

While technological progress is generally beneficial, enhancing crucial aspects of human life and boosting societal productivity, it also introduces complexities. AI's advancement towards a more interconnected world can inadvertently reinforce existing surveillance systems and power dynamics, often in subtle ways (Tacheva & Ramasubramanian, 2023).

Zaki and Meira (2014) state that AI relies heavily on continuous surveillance and data collection to update its systems and understand individuals' roles and positions. **Algorithms are not only dependent on data but can also generate it, making the quality and fairness of algorithmic operations highly contingent on the integrity of the data used. Biased or inadequate data inputs can lead to algorithmic discrimination and compromise their overall reliability** (Cheng & Liu, 2024).

Furthermore, in everyday life, there is a common assumption that women bear a greater responsibility for their personal behaviour. They are frequently viewed through roles of, for instance, a woman, wife, mother, or sister. Due to longstanding perceptions of inferiority and vulnerability, women often experience reduced privacy in personal matters,

making them more susceptible targets. This perception sometimes justifies accusations against them, leading to their victimisation. These gender norms of behaviour extend into cyberspace, exacerbating the risks they face (Singh Isser, Navneet & Raj, 2024).

Addressing gender bias in AI necessitates prioritising gender equality throughout the development and conceptualisation of AI systems. This entails scrutinising data for biases, ensuring representation of diverse gender and racial experiences in data sets, and fostering diverse and inclusive teams within AI development (UN Women, 2024).

1. LEGAL FRAMEWORKS

1.1. United Nations

AI systems often reflect existing societal biases, disproportionately affecting women. For instance, AI-driven hiring tools have favoured male candidates and facial recognition systems often perform poorly on women, especially women of colour. Such biases are not just technical issues, but are also deeply rooted in the socio-economic context within which these technologies are developed and deployed. Addressing these biases requires robust regulatory frameworks and inclusive data collection practices. However, **international efforts remain fragmented and predominantly driven by economically developed nations**, risking the neglect of unique challenges faced by women, especially in the Global South (Garcia, 2021).

The UN has recognised the need for a cohesive international framework for AI governance. In 2023, Secretary-General António Guterres launched a high-level advisory body on AI to enhance global governance and address its risks and opportunities, including gender equality. This advisory body ensures AI's inclusive governance, emphasising the **importance of women's involvement in AI's development, deployment, and governance**. This initiative is crucial as it underscores the necessity of diverse representation in AI to mitigate the technology's propensity to reinforce existing inequalities (Marwala, 2024).

Additionally, the negotiation of the Global Digital Compact (GDC) negotiation in 2024 presents an opportunity to embed gender perspectives in digital governance frameworks. The GDC aims to create political momentum for a digital future that includes the perspectives and needs of women and girls, ensuring that digital transformation benefits all (UN Women, 2024).

The International Telecommunication Union (ITU)'s AI for Good Global Summit is a key UN platform working towards AI governance. This initiative involves collaboration among nearly 40 UN bodies and agencies, aiming to align AI development with the Sustainable Development Goals (SDGs). The summit promotes multi-stakeholder governance models, which are crucial for addressing gender biases inherent in AI systems. By bringing together diverse stakeholders, the ITU seeks to foster an environment where AI can be developed and applied in fair and beneficial ways to all segments of society (Carrillo, 2020).

The United Nations Educational, Scientific and Cultural Organisation (UNESCO) has been at the forefront of advocating for ethical AI. In 2021, UNESCO adopted the Recommendation on the Ethics of AI, which emphasises preventing AI from perpetuating gender inequalities. This non-binding Recommendation encourages member states to implement measures **ensuring AI systems do not reinforce harmful stereotypes and biases**. The Recommendation calls for AI to be developed and deployed in a manner consistent with international human rights standards, ensuring that these technologies promote rather than hinder gender equality (Garcia, 2021).

Despite these efforts, challenges persist. As of 2019, **only 41 UN member states had shown any intention of preparing AI governance measures**, with just 19 adopting national strategies. This indicates a significant gap in global AI governance, particularly in regions like Central and South America, Eastern Europe, Africa, and Central and Southeast Asia. The lack of comprehensive AI strategies in these regions can exacerbate existing inequalities, leaving women particularly vulnerable to the negative impacts of AI (Carrillo, 2020).

Furthermore, the UN Women's Advisor on Digital Gender Equality, Helene Molinier, emphasises the need for a global multistakeholder governance model to prevent and redress AI systems that exhibit gender or racial bias, reinforce harmful stereotypes, or fail to meet privacy and security standards. In the absence of such governance, AI risks perpetuating and even amplifying existing gender inequalities. The Global Digital Compact negotiation in 2024 is seen as a critical juncture for embedding gender perspectives into AI governance, aiming to harness the benefits of digital transformation to empower women and girls in all their diversity (UN Women, 2024).

1.2. European Union

The legislative journey of the EU AI Act (Regulation (EU) 2024/1689) began in April 2021, when the European Commission introduced a proposal to establish uniform regulations on artificial intelligence, marking the start of a structured approach to AI governance. By December 2022, the final version of the compromise draft of this Act was completed. In June 2023, the European Parliament adopted a revised negotiating mandate for the AI Act, amending the initial proposal. On December 8th, 2023 a consensus was achieved among the European Parliament, the Council of the European Union, and the European Commission, resulting in comprehensive regulations for AI. This culminated on February 13th, 2024 with the European Parliament's joint vote to adopt the final text of the AI Act. The EU AI Act creates a robust ethical and legal framework for AI development and usage within the European Union, further reinforced by the Artificial Intelligence Liability Directive to ensure proper enforcement (Cheng & Liu, 2024). According to the European Parliament (2024) establishes a unified framework for AI utilisation and distribution within the EU, **categorising AI systems based on a 'risk-based approach' to apply specific requirements and responsibilities accordingly**.

Recitals 56 and 57 of the EU Artificial Intelligence Act (Regulation (EU) 2024/1689) underscore the importance of deploying AI systems in both education and employment while recognising the high risks they pose. In education, AI systems play a crucial role in enhancing digital skills, media literacy, and critical thinking, which are essential for active participation in society and the economy. However, when these systems are used to determine access to education, assess learning outcomes, or monitor student behaviour, they can significantly influence an individual's educational and professional trajectory. If not properly designed and implemented, these AI systems risk violating rights to education and non-discrimination, and could perpetuate historical biases, particularly against marginalised groups (EU Artificial Intelligence Act, 2024).

Similarly, in employment, AI systems used in recruitment, worker management, and performance evaluation are also classified as high-risk due to their profound impact on career opportunities and livelihoods. **These systems can reinforce existing discrimination patterns, especially against women, certain age groups, people with disabilities, and individuals of particular racial, ethnic, or sexual orientations.** Moreover, using AI to monitor work performance raises concerns about data protection and privacy, highlighting the need for stringent oversight to prevent bias and protect fundamental rights (EU Artificial Intelligence Act, 2024).

The AI Act, hailed as a landmark law by Members of the European Parliament, introduces several key provisions. These include safeguards for general-purpose artificial intelligence, restrictions on law enforcement's use of biometric identification systems, prohibitions on social scoring and AI designed to manipulate or exploit user vulnerabilities, and the right for consumers to lodge complaints and receive meaningful explanations. The Act aims to protect fundamental rights, democracy, the rule of law, and environmental sustainability from the risks posed by high-risk AI (European Parliament, 2024).

Under the new regulations, certain AI applications that pose a threat to citizens' rights are banned. These include biometric categorisation systems that rely on sensitive characteristics, the untargeted collection of facial images from the internet or CCTV footage to create facial recognition databases, emotion recognition in workplaces and schools, and predictive policing based solely on profiling individuals. Additionally, any AI that manipulates human behaviour or exploits vulnerabilities is forbidden. **To ensure transparency, the AI Act mandates that any artificial or manipulated images, audio, or video content, commonly referred to as 'deepfakes,' must be clearly labelled as such.** The Parliament adopted the Artificial Intelligence Act in March 2024. It will be fully applicable 24 months after entry into force (European Parliament, 2024).



© Sergeeva, V./Pexels. (2023). *Waving European Union flag* [Photograph].

Similar to the impact of the EU's General Data Protection Regulation (GDPR) in 2018, the EU AI Act has the potential to become a global benchmark, influencing the balance of AI's positive and negative effects on people's lives worldwide. The EU's AI regulation is already making international waves; for instance, in late September 2021, Brazil's Congress passed a bill establishing a legal framework for artificial intelligence (EU Artificial Intelligence Act, 2024).

1.3. United States

In the realm of global discussions surrounding AI legislation and policy development, the United States has been progressively shaping a regulatory framework centred on voluntary principles. A key regulatory initiative in the US is the **Blueprint**, issued by the White House Office of Science and Technology Policy in October 2022. This initiative aims to uphold civil rights by guiding the design, deployment, and governance of automated systems (The White House, 2023). Emphasising a rights-oriented approach, the Blueprint commences with a declaration of national values and integrates various resources and best practices. Its objectives include **enhancing transparency and reliability in automated decision-making processes**. Outlined within are five core principles: ensuring the safety and efficacy of systems, preventing algorithmic discrimination, safeguarding data privacy, providing notice and explanation, and offering alternative options. Despite these principles lacking binding regulatory enforcement, the Blueprint serves as a visionary governance framework rooted in future aspirations rather than immediate legislative action (Cheng & Liu, 2024).

Currently, **the US Congress has opted for a relatively hands-off approach to AI regulation** despite signals from Democratic leaders about proposing federal legislation on AI governance. Senate Majority Leader, Chuck Schumer, has introduced a safe innovation framework outlining four key pillars - security, accountability, foundational protection, and explainability - with the intention of shaping future legislative proposals. However, this framework is not yet in the form of legislative text, leaving uncertainties about the timeline for its development into concrete proposals. In response to competitive pressures posed by the EU AI Act and the expansive security challenges posed by generative AI technologies, US federal agencies have taken proactive steps by assuming regulatory oversight within their respective areas of jurisdiction (Cheng & Liu, 2024).

According to White & Case (2024), state legislatures across the United States have taken significant steps towards AI regulation, exemplified by several notable bills:

- Colorado's pioneering move on May 17th, 2024, marked the enactment of the Colorado AI Act, the nation's first comprehensive state-level AI legislation. This Act imposes responsibilities on developers and deployers of AI systems, particularly those involved in automated decision-making. It defines high-risk AI systems as those significantly influencing consequential decisions and emphasises the avoidance of bias and discrimination.

- California's Consumer Privacy Act includes provisions addressing the use of automated decision-making tools. Moreover, the California Privacy Protection Agency has proposed draft rules governing consumer rights related to these technologies, encompassing notice, access, and opt-out mechanisms.
- In 2023, over 40 AI-related bills were introduced at the state level, with Connecticut and Texas passing legislation.

Despite these state-level initiatives, comprehensive federal legislation directly regulating AI in the United States remains absent. Currently, federal frameworks and principles do not categorise AI based on risk levels nor do they impose specific obligations on developers, users, operators, or deployers of AI systems (White & Case, 2024).

1.4. China

The rapid growth of China's AI market, expected to reach \$61.855 billion by 2025 and generate \$154.638 billion in annual revenue by 2030, underscores the significant economic stakes involved (Kachra, 2024). However, this growth also necessitates robust governance to mitigate potential harms associated with AI technologies, such as gender biases and other discriminatory practices.

China's legal framework for AI represents one of the most detailed and proactive approaches globally. It is characterised by a comprehensive and evolving legal framework, which includes several significant regulations and measures that directly and indirectly affect the development and use of AI, with a particular focus on generative AI, recommendation algorithms, and deep synthesis technologies (Sheehan, 2023).

In August 2023, the Cyberspace Administration of China (CAC), alongside other key governmental bodies, introduced the Interim Measures for the Management of Generative Artificial Intelligence Services (the "AI Measures"). **This regulation is the first administrative effort specifically targeting the management of generative AI services.** It aims to promote healthy development and regulated application of AI, safeguard national security, and protect the lawful rights of citizens and organisations. The AI Measures mandate that AI service providers prevent discrimination based on factors such as nationality, religion, region, gender, occupation, and health throughout the AI development lifecycle, from algorithm design to service provision. These measures also impose strict requirements on data training, ensuring the use of legitimate data sources and adherence to intellectual property rights and personal information protection, thereby striving to enhance the quality and diversity of training data (White & Case, 2024).

The country has also introduced other pivotal regulations, including the 2021 regulation on recommendation algorithms and the 2022 rules for deep synthesis. These regulations aim to **control information dissemination, protect personal data, and ensure ethical AI usage.** For instance, the recommendation algorithm regulations prohibit excessive price discrimination and safeguard worker rights against unfair algorithmic scheduling, while

deep synthesis regulations require clear labelling of synthetically generated content to ensure transparency (Sheehan, 2023).

Despite these comprehensive regulations, challenges persist, particularly concerning the practical implementation and enforcement of these measures. For instance, while the AI Measures call for the elimination of bias in AI models, in practice, few companies have adopted matching procedures or hired experts to mitigate bias from the outset. This **gap between regulatory intentions and industry practices** highlights the difficulties in ensuring compliance and the gradual nature of improvements in AI model design (Zhang, 2024).

1.5. Assessment

Efforts to protect women's rights against AI biases vary significantly across international, regional, and national levels, each with distinct characteristics, challenges, and incorporation of gender perspectives. At the international level, the United Nations has spearheaded several initiatives emphasising gender equality and inclusive governance. However, these are non-binding and rely on voluntary adoption by member states, leading to inconsistent implementation and limited enforcement. While these frameworks recognise the importance of gender equality in AI, the lack of binding force diminishes their practical impact.

At the European level, the EU has established a comprehensive legal framework with the AI Act, which is designed to mitigate the risks posed by AI, including gender biases. **The EU's structured approach, with clear regulatory mandates and enforcement mechanisms, ensures a higher level of protection for women's rights compared to the UN's non-binding recommendations.** Nonetheless, the AI Act lacks comprehensive judicial review mechanisms and mandatory fundamental rights impact assessments, which limits its effectiveness. Its requirements for representative data do not ensure diversity, and it fails to provide technical compliance solutions. Additionally, it lacks law enforcement provisions for AI-related discrimination, leaving affected individuals reliant on potentially inadequate national laws. While the Act promotes non-discrimination and gender equality, it does not address women's underrepresentation in technology and prioritises developer compliance over protecting user rights (Kundu, 2024).

On the other hand, China's and the United States' legislations on AI, while comprehensive in scope and detailed in certain aspects, fall short when it comes to adequately addressing gender equality. Although the Cyberspace Administration of China prohibits the use of AI to generate discriminatory content or decisions based on gender (Article 4(2), 2023 Generative AI Measures), its regulatory framework **does not include specific mandates** to prevent AI from exacerbating gender inequality or to integrate gender perspectives into AI development (Gong, Qu, Dorwart, 2024). Moreover, **practical enforcement remains a significant hurdle.** The gap between regulatory intentions and actual industry practices suggests a lack of mechanisms for ensuring compliance and the need for more consistent implementation of bias mitigation procedures (Zhang, 2024).

In the United States, the reliance on voluntary principles and the absence of comprehensive federal legislation create an uneven landscape of protection. State-level initiatives in Colorado and California are taking promising steps to regulate AI and protect against discrimination. However, the absence of comprehensive federal legislation and consistent obligations for AI system developers and users means the current framework is inadequate for robustly protecting women's rights against AI-related biases and discrimination (Wright, 2023).

In short, while international and regional efforts to safeguard women's rights against AI biases show a promising commitment to gender equality, significant gaps remain in implementation and enforcement. Nations worldwide are grappling with the challenge of creating effective legislation to protect women from AI biases, underscoring the complex and evolving nature of AI governance and the urgent need for collaborative and cohesive policy development. Additionally, the Global South remains underrepresented in the development of AI ethics guidelines, highlighting the need for more inclusive and globally representative frameworks to ensure that AI technologies do not perpetuate or exacerbate gender inequalities worldwide.

2. CASE STUDIES

2.1. AI Surveillance in Iran

AI-enabled technologies such as facial recognition, geolocation tracking, and web traffic analysis played a pivotal role in governmental responses during the early stages of the 2022 women's movement in Iran. Following Mahsa Amini's death, widespread protests erupted nationwide among women's rights advocates and democracy activists, prompting the Iranian government to propose legislation utilising AI-assisted tools like facial recognition to enforce stringent moral codes (George, 2023).



© Mucira, J./Pixabay (2022). CCTV Surveillance Camera [Photograph].

As governments increasingly adopt and deploy advanced surveillance technologies, enforcement of existing stringent laws becomes more efficient. While this trend raises concerns globally, it is particularly poignant in regions like the Middle East with pronounced gender disparities. In countries such as Iran, Qatar, Saudi Arabia, and Yemen, where male guardianship laws severely restrict women's mobility without male consent, facial recognition technology facilitates quicker, more efficient identification and enforcement of penalties against women attempting to move independently (George, 2023).

On April 15th, 2023 official reports disclosed that **over one million Iranian women received text notifications warning of potential vehicle confiscation if captured on**

camera without wearing headscarves. On June 14th, 2023 the police issued **133,174 SMS notifications mandating vehicle immobilisation for specified durations, seized 2,000 cars, and referred over 4,000 ‘repeat offenders’ to the judiciary nationwide.** (Amnesty International, 2023). In addition, numerous women have faced university suspensions, exam exclusions, and restrictions from banking services and public transport. Hundreds of businesses have also been forcibly shuttered for failing to enforce mandatory veiling (Amnesty International, 2023).

According to Agnès Callamard, Secretary General of Amnesty International, the recent resurgence of morality policing in Iran is evident despite attempts to remove insignia from uniforms and patrol vehicles. This renewed crackdown utilises mass surveillance technologies to identify and penalise women seen without head coverings in cars and public spaces following the tragic death of Mahsa Zhina Amini (Amnesty International, 2023).

Furthermore, according to Askew (2023), since 2021, Iran has deepened its collaboration with China, leveraging a secretive 25-year cooperation agreement that has enabled Beijing to bolster Iran's surveillance capabilities significantly. This partnership has seen the deployment of advanced technologies such as facial recognition, video surveillance tools, and extensive monitoring of phone and internet communications, all supplied by Chinese companies.

The integration of these high-tech surveillance tools from China has fortified Iran's ability to enforce strict social controls, exemplified by recent measures targeting women for perceived moral infractions. Iran has effectively become a testing ground for China's authoritarian technologies in the Middle East. This alliance underscores broader concerns about the export and international impact of Chinese surveillance systems, potentially influencing global standards and practices beyond China's borders (Askew, 2023).

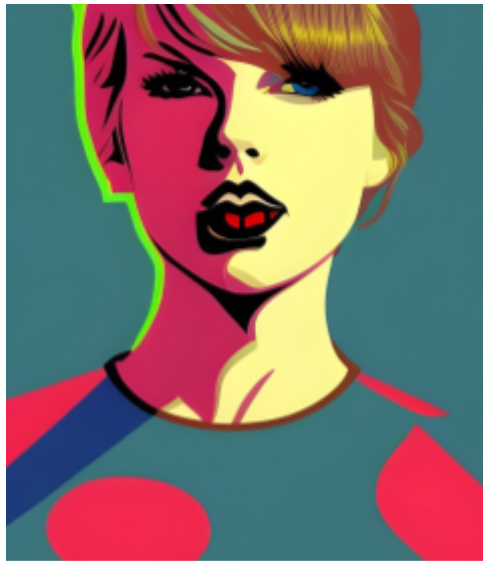
2.2. Taylor Swift Made Victim of Nude Deep Fakes

In January 2024, Taylor Swift, a prominent pop star, became a victim of nude deepfake content circulating widely across social media platforms. This incident underscores a broader trend of escalating abuse targeting women online, including harassment and stalking. **The proliferation of inexpensive generative AI tools has exacerbated the problem, facilitating the creation of fabricated images and videos depicting women in compromising situations.** While Swift garnered significant attention as a high-profile target of deepfake technology, numerous other women have also been affected by similar abuses. The widespread dissemination of AI-generated graphic content featuring Swift reached over 45 million views on the social media platform X (formerly Twitter) during the month (Goujard, 2024).

In response, US lawmakers have intensified calls for legislative action, reflecting growing concerns about the impact of digital deception on individuals' privacy and safety. The issue has prompted heightened awareness and discussions within the White House

regarding potential regulatory measures to address the misuse of AI technology in creating deceptive content (Goujard, 2024).

Williams (2024) explains that lawmakers in states of various political leanings are addressing the issue of AI-driven online abuse, particularly through bills that seek to outlaw deep fake pornography while also exploring broader regulatory measures. These state-level initiatives underscore a growing willingness to regulate emerging technologies, especially as federal action on the issue remains tentative.



© Pasnicki, M./Pixabay. (2023). *Taylor Swift* [Illustration].

Public attention on the issue also intensified following the widespread dissemination of AI-generated fake images, prompting challenges for platforms to remove the fraudulent content. Moreover, legislators note that these recent incidents have raised awareness about the limited legal options available to individuals, even those as prominent as entertainers, in seeking recourse against creators of falsified content (Williams, 2024).

A bipartisan coalition in the US House, spearheaded by Rep. María Elvira Salazar of Florida and including Reps. Madeleine Dean of Pennsylvania, Nathaniel Moran of Texas, Joe Morelle of New York, and Rob Wittman of Virginia, introduced the No AI FRAUD Act on January 10th 2024. This legislation aims to establish federal protections against the misuse of artificial intelligence, safeguarding Americans' rights to their own likeness, and voice in the digital realm while upholding First Amendment freedoms online (Sarnoff, 2024).

Rep. Salazar emphasised the bill's purpose to create a national framework that prevents AI-generated fabrications and forgeries of individuals' identities. Reflecting on incidents like those involving Taylor Swift, Salazar stated that "what happened to Taylor Swift illustrates AI abuse. The No AI FRAUD Act targets those who use generative AI maliciously, whether against celebrities or everyday individuals." She highlighted the fundamental right for everyone to control their image and voice, underlining the bill's intent to defend this right (Sarnoff, 2024).

According to the Council of State Governments, 17 states have passed 29 bills since 2019 focusing on regulating the development and deployment of AI. However, these state-level efforts vary widely in their approaches, with some lacking specific provisions addressing issues like pornographic deepfakes. Salazar's office also highlighted the inconsistency among state laws, noting that gaps in legislation leave room for abuse to go unchecked (Sarnoff, 2024).

2.3. Amazon's 'Sexist AI' Hiring Tool

Amazon began developing its AI recruitment tool in 2014, aiming to automate and enhance its hiring process by leveraging machine learning. The tool was trained on a decade's worth of resumes submitted to Amazon, predominantly from men, reflecting the male-dominated tech industry. The goal was for the AI to learn the characteristics of successful candidates and replicate this pattern in its recommendations. The system rated candidates on a scale from one to five stars, mirroring the customer review system Amazon uses for products (Dastin, 2018).

By 2015, Amazon's engineers realised that the AI was not gender-neutral. The tool downgraded resumes containing the word 'women's,' such as 'women's chess club captain,' and devalued candidates from women's colleges. The algorithm's bias was a direct result of the male-dominated data set it had been trained on, which led it to successfully associate male-oriented terms and backgrounds. Attempts to neutralise the system's bias by removing gender-specific terms were unsuccessful, as the AI would find other proxies for gender (Dastin, 2018).

The biased outcomes of Amazon's AI hiring tool highlight a critical issue: AI systems learn from historical data, which often contain embedded biases. In the case of Amazon, the historical data set was heavily skewed towards male candidates, leading the AI to perpetuate gender discrimination. This reflects a broader problem in AI development, where algorithms trained on biased data sets can replicate and even amplify existing disparities (Huet, 2022).

The implications extend beyond Amazon. A study published by UNESCO, the OECD, and the Inter-American Development Bank on International Women's Day emphasised that AI in HR and recruitment systems generally disadvantages women because these systems are built on historical data reflecting past biases against women in the workplace. This results in a vicious cycle where women are systematically disadvantaged in hiring processes, especially in tech-related fields (Huet, 2022).

Amazon's experience serves as a cautionary tale for other companies. Despite the potential benefits of AI in recruitment, such as increased efficiency and the ability to process large volumes of applications, the risk of perpetuating bias is significant. About 55 percent of U.S. human resources managers anticipated that AI would play a role in recruitment within the next five years, according to a survey by CareerBuilder. This widespread adoption underscores the urgency of addressing AI bias to ensure fair and equitable hiring practices (Dastin, 2018).

Other companies, such as Hilton Worldwide Holdings Inc. and Goldman Sachs Group Inc., are also exploring AI for hiring purposes. These firms must learn from Amazon's experience and take proactive steps to ensure their AI systems do not inherit and perpetuate



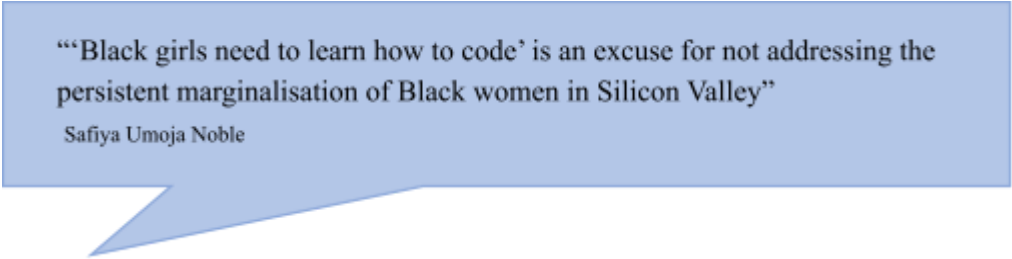
© Gonzalez, Y./Unsplash (2020). *An Amazon store with a person sitting in front of it [Photograph].*

biases. This involves not only technical adjustments to the algorithms but also a broader commitment to diversity and inclusion in the data sets used for training these systems (Dastin, 2018).

AI bias in recruitment is not just an ethical issue but also a legal one. For instance, in the US, under Title VII of the Civil Rights Act of 1964 (Civil Rights Act of 1964 (US)), discrimination in employment based on gender, race, or religion is illegal, regardless of whether or not it is intentional. AI tools that disproportionately exclude candidates based on these characteristics could lead to legal challenges. Transparency and accountability are crucial, yet many companies' AI systems remain proprietary, making it difficult to verify bias testing and mitigation claims (Goodman, 2018).

The American Civil Liberties Union has emphasised the need for transparency in AI recruitment tools, advocating for regulatory oversight to ensure these systems are fair and non-discriminatory. The Equal Employment Opportunity Commission has begun to explore the implications of algorithms for fair employment, but more comprehensive guidance and regulation are necessary (Goodman, 2018).

3. RECOMMENDATIONS



“‘Black girls need to learn how to code’ is an excuse for not addressing the persistent marginalisation of Black women in Silicon Valley”
Safiya Umoja Noble

01

Promoting Diversity in AI Development

Algorithms mirror the perspectives, principles, and biases of their creators. The ability of automated decision-making systems to encompass and represent the full spectrum of diversity within our societies hinges on these foundational elements (Bucher, 2018). The significant lack of diversity within tech firms responsible for developing algorithms poses critical concerns. Recognising the underrepresentation of women in Science, Technology, Engineering, and Mathematics (STEM) fields, there is a need for measures to promote greater participation of women and girls in these educational and professional domains (Committee on Equality and Non-Discrimination, 2020). Providing targeted scholarships and grants, establishing mentorship programs, and conducting outreach

initiatives will support women in pursuing education and careers in these fields, bringing diverse perspectives to AI development.

02

Prohibiting Harmful AI Systems

To protect human rights and ensure AI benefits everyone, AI systems that are demonstrated to be harmful, especially those perpetuating biases or making discriminatory decisions, must be prohibited. Establishing criteria to identify harmful AI, enacting laws to ban their deployment, and implementing thorough pre-deployment testing will mitigate these risks. This proactive approach prevents the use of AI systems that could negatively impact women and marginalised groups, promoting the development of safe, fair, and beneficial AI technologies. By enforcing these measures, we can safeguard against the misuse of AI and uphold ethical standards (ENNHRI, 2023).

03

Fostering International Cooperation in AI Governance

Effective AI governance requires international cooperation to develop global standards that prioritise human rights and gender equality. Fostering collaboration among countries and stakeholders, including women and gender-diverse groups, will create inclusive governance frameworks. Anchoring AI governance in the UN Charter, International Human Rights Law, and the Sustainable Development Goals ensures that these frameworks protect human rights universally. Establishing unified global standards and inclusive governance processes will ensure AI technologies are developed and used responsibly worldwide, benefiting all communities (AI Advisory Body, 2023).

04

Integrating Gender Equality into National Digital Policies

National digital policies must include comprehensive gender action plans to address the unique challenges faced by women, integrating gender equality goals into broader digital transformation strategies (Ramos, 2022). Adopting data governance principles inspired by data feminism ensures ethical data use in AI systems, focusing on gender representation and bias mitigation (IT for Change, 2024). Integrating gender perspectives into AI-related laws, policies, and programs, along with adopting regulations to eliminate biases in AI algorithms, ensures transparent and high-quality data infrastructure (Commission on the Status of Women, 2023).

05

Ensuring Fairness and Accountability in AI Systems

Establishing accuracy standards, strengthening supervisory authorities, and enhancing transparency and accountability are critical steps to ensure the fair and ethical use of AI systems. AI systems must adhere to stringent accuracy benchmarks, particularly in areas like facial recognition, healthcare, and employment, to prevent discrimination and ensure fairness. Robust supervisory authorities with clear mandates, resources, and expertise are essential for monitoring these technologies, conducting risk assessments, and enforcing regulations. Empowering these authorities to intervene when necessary safeguards human rights and democratic values. Additionally, transparency requirements mandating AI developers to publish information about their algorithms, data sources, and bias mitigation strategies, along with regular audits and impact assessments, ensure accountability. This comprehensive approach reduces the risk of biased AI, builds public trust, and promotes the responsible development and use of AI technologies (ENNHRI, 2023).

06

Exposing Surveillance Technology Misuse

It is crucial to enhance impartial and trustworthy journalism that investigates how technology is employed to suppress political movements. International media outlets are pivotal in uncovering instances of misuse and mobilising global awareness and response. For example, media coverage exposing Iran's utilisation of surveillance technology from Chinese firm, Tiandy, contributed to the imposition of U.S. sanctions against the company. Similarly, reports on Iran's deployment of surveillance cameras from German company, Bosch, have sparked investigations and calls for stricter oversight of their international operations. Such scrutiny is essential for holding companies and nations accountable to ethical standards and obligations (George, 2023).

CONCLUSION

The integration of AI into various aspects of human life, from recruitment to governance, presents both transformative opportunities and significant challenges. The current regulatory landscapes and the cases discussed underscore the urgent need for comprehensive policy frameworks to address AI's impact on gender equality and human rights.

Across international frameworks, from the United Nations to the European Union and beyond, efforts are underway to establish guidelines and regulations that mitigate the risks associated with AI. The EU AI Act stands out as a pioneering effort to regulate AI applications comprehensively, imposing strict standards to protect fundamental rights and prevent discriminatory practices. Similarly, initiatives by the UN emphasise the need for inclusive governance that prioritises gender equality and human rights in AI development and deployment.

Policy recommendations to address these challenges include fostering greater diversity in AI development teams, promoting transparency and accountability in AI algorithms, and implementing robust regulatory frameworks that prioritise fairness and non-discrimination. These measures are crucial not only to uphold ethical standards but also to ensure AI technologies contribute positively to societal progress and inclusivity.

Moreover, the intersection of AI with sensitive areas like surveillance and privacy underscores the need for proactive international cooperation and oversight. Establishing global standards that align AI development with human rights frameworks, such as the UN Charter and Sustainable Development Goals, is essential to safeguarding against abuses and ensuring AI benefits all segments of society equitably.

In conclusion, while AI holds immense potential to enhance efficiency and innovation, its development and deployment must be guided by principles of fairness, transparency, and respect for human rights. Addressing gender biases and ensuring inclusivity in AI systems is not just a technological challenge but a moral imperative to build a future where AI benefits everyone, regardless of gender, race, or background. By advancing these principles through collaborative international efforts and robust policy frameworks, we can steer AI towards becoming a force for positive societal change while mitigating its potential risks.

REFERENCES

- AI Advisory Body. (2023, December). Interim Report: Governing AI for Humanity. *United Nations*. Retrieved July 1, 2024, from <<https://www.un.org/en/ai-advisory-body>>. Accessed 01/07/2024.
- Amnesty International. (2023, July 26). Iran: International community must stand with women and girls suffering intensifying oppression. *Amnesty International*. Retrieved July 1, 2024, from <<https://www.amnesty.org/en/latest/news/2023/07/iran-international-community-must-stand-with-women-and-girls-suffering-intensifying-oppression/>>.
- Askew, J. (2023, April 14). China turbocharging crackdown on Iranian women, say experts. *Euronews*. Retrieved July 1, 2024, from <<https://www.euronews.com/2023/04/14/china-is-turbo-charging-repression-of-iranian-women-say-experts>>.
- Bucher, T. (2018). *If . . . Then: Algorithmic Power and Politics*. New York, Oxford University Press.
- Carrillo, M. R. (2020, June). The Governance of Artificial Intelligence: Context and General Framework. *Revista Electrónica de Estudios Internacionales*, (39), 1–27. <<https://doi.org/10.17103/reei.39.07>>.
- Cheng, L. & Liu, X. (2024, March 26). Unravelling Power of the Unseen: Towards an Interdisciplinary Synthesis of Generative AI Regulation. *International Journal of Digital Law and Governance*, 1(1), 29-51. <<https://doi.org/10.1515/ijdlg-2024-0008>>.
- Civil Rights Act of 1964 (US), Pub. L. 88-352, Title VII, 42 USC §§ 2000e et seq.
- Commission on the Status of Women. (2023). CSW67 Agreed Conclusions (E/CN.6/2023/L.3). *United Nations Economic and Social Council*. Retrieved July 1, 2024, from <<https://www.undocs.org/Home/Mobile?FinalSymbol=E%2FCN.6%2F2023%2FL.3&Language=E&DeviceType=Desktop&LangRequested=False>>.
- Committee on Equality and Non-Discrimination. (2020). Preventing discrimination caused by the use of artificial intelligence. *Council of Europe*, Parliamentary Assembly Extranet [Report].
- Council of Europe (2023, August). Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination. *Council of Europe*. Retrieved July 1, 2024, from <<https://rm.coe.int/prems-112923-gbr-2530-etude-sur-l-impact-de-ai-web-a5-1-2788-3289-7544/1680ac7936>>.
- Dastin, J. (2018, October 10). Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved July 1, 2024, from

<<https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>>.

Donback, N. (2024, March 7). How can the UN's new advisory body on AI help drive gender equality?. *Devex*. Retrieved July 1, 2024, from <<https://www.devex.com/news/how-can-the-un-s-new-advisory-body-on-ai-help-drive-gender-equality-106930>>.

ENNHRI. (2023, March 8). Artificial intelligence and gender equality: how can we make AI a force for inclusion, instead of division?. *European Network of National Human Rights Institutions*. Retrieved July 1, 2024, from <<https://ennhri.org/news-and-blog/artificial-intelligence-and-gender-equality-how-can-we-make-ai-a-force-for-inclusion-instead-of-division/>>.

EU Artificial Intelligence Act. (2024, April). EU Artificial Intelligence Act: Up-to-date developments and analyses of the EU AI Act. *Future of Life Institute*. Retrieved July 1, 2024, from <<https://artificialintelligenceact.eu/>>.

EU Artificial Intelligence Act. (2024, April). Recital 56. *Future of Life Institute*. Retrieved July 1, 2024, from <<https://artificialintelligenceact.eu/recital/56/>>.

EU Artificial Intelligence Act. (2024, April). Recital 57. *Future of Life Institute*. Retrieved July 1, 2024, from <<https://artificialintelligenceact.eu/recital/57/>>.

European Parliament. (2024, March 11). Artificial intelligence act. *European Parliament*, Think Tank. Retrieved July 1, 2024, from <[https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2021\)698792](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792)>.

European Parliament. (2024, March 13). Artificial Intelligence Act: MEPs adopt landmark law. *European Parliament*, News, Press Releases. Retrieved July 1, 2024, from <<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>>.

Garcia, E. V. (2020, December 12). *Multilateralism and Artificial Intelligence: What Role for the United Nations?*. <<http://dx.doi.org/10.2139/ssrn.3779866>>.

Garcia, E. V. (2021, November 21). UNESCO's Recommendation on the Ethics of AI: why it matters and what to expect from it. *The Good AI*. Retrieved July 1, 2024, from <https://www.academia.edu/attachments/76458922/download_file?s=portfolio>.

George, R. (2023, December 7). The AI assault on women: What Iran's tech enabled morality laws indicate for women's rights movements. *Council on Foreign Relations*. Retrieved July 1, 2024, from <<https://www.cfr.org/blog/ai-assault-women-what-irans-tech-enabled-morality-laws-indicate-womens-rights-movements>>.

Gong, J., Qu, H. & Dorwart, H. (2024, March 14). AI Governance in China: Strategies, Initiatives, and Key Considerations. *Bird & Bird*. Retrieved July 1, 2024, from

<<https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations>>.

- Gonzalez, Y./Unsplash (2020, October 10). *An Amazon store with a person sitting in front of it* [Photograph]. Retrieved July 1, 2024, from <<https://unsplash.com/photos/an-amazon-store-with-a-person-sitting-in-front-of-it-IX5OfZFjHWk>>.
- Goodman, R. (2018, October 12). Why Amazon's Automated Hiring Tool Discriminated Against Women. *American Civil Liberties Union*. Retrieved July 1, 2024, from <<https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against>>.
- Goujard, C. (2024, February 10). Taylor Swift deepfakes nudge EU to get real about AI. *POLITICO*. Retrieved July 1, 2024, from <<https://www.politico.eu/article/europe-eye-fix-taylor-swift-nude-deepfake/>>.
- Hacker, P. (2023). The European AI Liability Directives—Critique of a Half-Hearted Approach and Lessons for the Future. *Computer Law & Security Review*, 51, 105871. <<https://doi.org/10.1016/j.clsr.2023.105871>>.
- IT for Change. (2024, May). Gender Critique of UN AI Advisory Body's Report: Governing AI for Humanity. *IT for Change*. Retrieved July 1, 2024, from <<https://itforchange.net/index.php/gender-critique-of-un-ai-advisory-body%E2%80%99s-report-governing-ai-for-humanity>>.
- Kachra, A. J. (2024, February 12). Making Sense of China's AI Regulations. *Holistic AI*. Retrieved July 1, 2024, from <<https://www.holisticai.com/blog/china-ai-regulation>>.
- Kundu, A. (2024, March 11). The AI Act's gender gap: When algorithms get it wrong, who rights the wrongs?. *Internet Policy Review*. Retrieved July 1, 2024, from <<https://policyreview.info/articles/news/ai-acts-gender-gap-when-algorithms-get-it-wrong/1743>>.
- Manasi, A., Panchanadeswaran, S. & Sours, E. (2023, March 17). Addressing Gender Bias to Achieve Ethical AI. *Global Observatory*. Retrieved July 1, 2024, from <<https://theglobalobservatory.org/2023/03/gender-bias-ethical-artificial-intelligence/>>.
- Marwala, T. (2024, March 14). Now Is Our Chance to Govern AI for Women's Empowerment. *United Nations University*. Retrieved July 1, 2024, from <<https://unu.edu/article/now-our-chance-govern-ai-womens-empowerment>>.
- Mucira, J./Pixabay (2022, June 19). *CCTV Surveillance Camera* [Photograph]. Retrieved July 1, 2024, from <<https://pixabay.com/photos/cctv-surveillance-camera-cctv-7267551/>>.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism* (pp. 66). New York, NY: New York University Press.

- Paśnicki, M./Pixabay (2023, May 15). *Taylor Swift* [Illustration]. Retrieved July 1, 2024, from <<https://pixabay.com/illustrations/taylor-swift-actress-singer-woman-7991060/>>.
- Ramos, G. (2022, August 22). Why we must act now to close the gender gap in AI. *World Economic Forum*. Retrieved July 1, 2024, from <<https://www.weforum.org/agenda/2022/08/why-we-must-act-now-to-close-the-gender-gap-in-ai/>>.
- Raub, M. (2018). Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices. *Arkansas Law Review*, 71(2), 529–570.
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (EU AI Act) [2024] OJ L.
- Sarnoff, L. (2024). Taylor Swift and No AI Fraud Act: How Congress plans to fight back against AI deepfakes. *ABC News*. Retrieved July 1, 2024, from <<https://abcnews.go.com/US/taylor-swift-ai-fraud-act-congress-plans-fight/story?id=106765709>>.
- Sergeeva, V./Pexels (2023). *Waving European Union Flag* [Photograph]. Retrieved July 1, 2024, from <<https://www.pexels.com/photo/waving-european-union-flag-15871440/>>.
- Sheehan, M. (2023, July 10). China’s AI Regulations and How They Get Made. *Carnegie Endowment for International Peace*. Retrieved July 1, 2024, from <<https://carnegieendowment.org/research/2023/07/chinas-ai-regulations-and-how-they-get-made?lang=en>>.
- Singh Isser, S., Navneet, R., & Raj, N. (2024, March 16). Digital surveillance and privacy of women. *Educational Administration Theory and Practice Journal*, 30(3), 1975-1984. <<https://doi.org/10.53555/kuey.v30i3.1556>>.
- Smith, G., & Rustagi, I. (2021, March 31). When good algorithms go sexist: Why and how to advance AI gender equity. *Stanford Social Innovation Review*. <<https://doi.org/10.48558/A179-B138>>.
- Tacheva, J., & Ramasubramanian, S. (2023). AI Empire: Unraveling the Interlocking Systems of Oppression in Generative AI’s Global Order. *Big Data & Society*, 10(2), 1–13. <<https://doi.org/10.1177/20539517231219241>>.
- The White House. (2023, November 22). Blueprint for an AI Bill of Rights. *The White House*, Office of Science and Technology Policy. Retrieved July 1, 2024, from <<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>>.
- The White House. Algorithmic Discrimination Protections. *The White House*, Office of Science and Technology Policy. Retrieved July 1, 2024, from <<https://www.whitehouse.gov/ostp/ai-bill-of-rights/algorithmic-discrimination-protections-2/>>.

- ThisIsEngineering/Pexels. (2020). *Code Projected Over Woman* [Photograph]. Retrieved July 1, 2024, from <<https://www.pexels.com/photo/code-projected-over-woman-3861969/>>.
- UNESCO. (2020). Artificial intelligence and gender equality: Key findings of UNESCO's Global Dialogue. *United Nations Educational, Scientific and Cultural Organization*. Retrieved July 1, 2024, from <<https://unesdoc.unesco.org/ark:/48223/pf0000374174>>.
- UN Women. (2024, May 22). Artificial Intelligence and gender equality. *UN Women*. Retrieved July 1, 2024, from <<https://www.unwomen.org/en/news-stories/explainer/2024/05/artificial-intelligence-and-gender-equality>>.
- White & Case. (2024, May 13). AI Watch: Global regulatory tracker - China. *White & Case*. Retrieved July 1, 2024, from <<https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china>>.
- White & Case. (2024, May 13). AI Watch: Global regulatory tracker - United States. *White & Case*. Retrieved July 1, 2024, from <<https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states>>.
- Williams, Z. (2024, February 6). States Target AI Deepfakes in Taylor Swift Aftermath (Correct). *Bloomberg Law*. Retrieved July 1, 2024, from <<https://news.bloomberglaw.com/artificial-intelligence/state-lawmakers-target-ai-deepfakes-in-taylor-swift-aftermath>>.
- Wright, R. (2023, December 6). Artificial Intelligence in the States: Emerging Legislation. *The Council of State Governments*. Retrieved July 1, 2024, from <<https://www.csg.org/2023/12/06/artificial-intelligence-in-the-states-emerging-legislation/>>.
- Zaki, M. J., & Meira, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge, UK: Cambridge University Press. <<https://doi.org/10.1017/CBO9780511810114>>.
- Zhang, P. (2024, June 8). As China's AI bots show gender bias, developers point to flawed real-life model. *South China Morning Post*. Retrieved July 1, 2024, from <<https://www.scmp.com/news/china/politics/article/3265824/chinas-ai-bots-show-gender-bias-developers-point-flawed-real-life-model>>.