(*) Corresponding author
Original research article
DOI: https://doi.org/10.20544/uklop.2026.630
Received: 11.01.2026 • Accepted after revision: 11.04.2026 • Published: 09.05.2026

#UKLO Annual International Scientific Conference

ALGORITHMIC BIAS AND SDG EQUITY: ISSUES WITH REPRESENTATION IN AI OUTPUTS

Elena Shalevska*
University "St. Kliment Ohridski" – Bitola
ORCID: https://orcid.org/0000-0002-3270-7137
elena.shalevska@uklo.edu.mk

Bisera Kostadinovska-Stojchevska
University "St. Kliment Ohridski" – Bitola
k_bisera@yahoo.com

ABSTRACT

AI is now everywhere, shaping how we search, write, and even create images. Recognizing this, this study investigates gender stereotypes in the image outputs of five leading generative artificial intelligence (AI) platforms such as: Grok, Gemini (Nano Banana), ChatGPT (DALL·E), and DeepSeek. Using a small set of prompts that focus on occupations (e.g., team coach, judge, housekeeper, computer programmer, primary school teacher), this qualitative research examines whether AI systems reproduce or challenge existing societal biases. Initial findings indicate that the generated images frequently reflect traditional gender norms: high-status roles such as judge or coach are predominantly depicted as male, while service-oriented or care-related roles, such as cleaning staff and primary school teacher, are more often portrayed as female. These results suggest that AI models, trained on large-scale human data, tend to mimic rather than mitigate human gender bias, posing challenges for Sustainable Development Goals (SDG) 5 (Gender Equality) and 10 (Reduced Inequalities).

Keywords: Artificial Intelligence (AI), ChatGPT, AI Bias, Algorithmic Bias, Gender Equality

1. INTRODUCTION

Generative artificial intelligence has rapidly become part of everyday digital life. Image-generation tools are now widely used in education, marketing, journalism, entertainment, and personal creative work. Millions of users interact with these systems daily, often relying on their outputs as visual references for professions, social roles, and identities. These images do more than illustrate ideas: They communicate norms; and they shape the imagination of millions upon millions of users.

At the same time, concerns about algorithmic bias have intensified. Scholars have shown that machine-learning systems frequently reproduce social inequalities embedded in their training data (Crawford 2021; Raji et al. 2020). Rather than functioning as neutral technologies, AI models often reflect historical patterns of exclusion and hierarchy. When

these systems generate images, they do not merely depict reality; they participate in constructing cultural meaning about who belongs where.

Gender bias represents one of the most persistent challenges in this domain. Research across labor sociology and media studies has long documented occupational gender segregation, with men overrepresented in positions of authority and technical expertise and women concentrated in caregiving and service roles (England 2010). Digital systems trained on vast repositories of online images and text inherit these asymmetries. The concern is not only that such patterns continue to appear in AI outputs, but that the scale and seeming objectivity of AI amplification may further normalize them.

Visual representation is particularly powerful. Images convey social meaning quickly and emotionally, often without conscious reflection. Repeated exposure to stereotypical portrayals can shape perceptions of competence, leadership, and occupational "fit." When young users see programmers consistently depicted as men or housekeepers as women, these visual cues subtly reinforce assumptions about career possibility and belonging. Over time, this can affect aspiration as well as expectation. The stakes are therefore not trivial.

Despite growing attention to text-based algorithmic bias, systematic evaluation of gender representation within AI-generated imagery remains underdeveloped. Much existing research relies on quantitative frequency counts or benchmark datasets designed primarily for classification systems rather than generative platforms. Fewer studies have examined how real-world user prompts generate visual narratives that reflect – or reshape – social stereotypes. This study contributes to that gap. It explores how major generative image platforms portray gender across a small but strategically selected set of occupations that are commonly associated with strong professional stereotypes.

By analyzing the visual patterns produced by four widely used AI platforms, the study asks a central question: Do contemporary generative image systems reproduce conventional gender hierarchies in occupational representation, or do they demonstrate signs of diversification and stereotype disruption? In doing so, the research moves beyond broad claims about algorithmic bias and instead grounds discussion in concrete visual outputs as they appear to users.

This inquiry is necessarily exploratory. It does not claim to offer global generalization. It does, however, provide an interpretive lens through which to assess how generative AI participates in the reproduction of gendered social imaginaries. At a time when AI imagery increasingly shapes educational materials, advertising content, and everyday meaning-making, understanding these representational dynamics is essential. Equity begins with visibility. And visibility begins with what we choose to – or allow machines to – show.

2. LITERATURE REVIEW

2.1. Algorithmic Bias

Algorithmic bias refers to systematic and repeatable errors in AI systems that produce unfair or discriminatory outcomes for certain social groups. In generative AI models, this bias primarily arises from the data used for training. Because large-scale datasets are drawn from the internet, books, media, and other real-world sources, they inevitably contain the social inequalities, stereotypes, and historical exclusions that exist in human societies. Models trained on such data do not merely reflect these patterns – they can amplify them.

Scholarly research has repeatedly demonstrated this risk. Bender et al. (2021) show that large language models such as GPT-3 can reproduce racial, gender, and cultural stereotypes, often in subtle ways that are difficult to detect, as well as political stereotypes in textual outputs (Shalevska & Walker, 2025). The models do not "intend" to discriminate. Yet their outputs mirror the statistical patterns embedded in their training data. Over time, especially as AI becomes increasingly integrated into domains such as education, media,

healthcare, and hiring, this can reinforce harmful social narratives and normalize unequal treatment.

A widely cited real-world example illustrates these concerns. Kaplan (2024) documents the case of Amazon's AI-based recruitment tool, introduced in 2014 to automate the screening of job applicants. The system analyzed résumés and ranked candidates to identify the most promising applicants. On the surface, the approach promised efficiency and objectivity. In practice, however, the model learned to favor male candidates over female ones. This occurred because historical hiring data, which was used for training, contained a disproportionate number of male résumés. The algorithm therefore learned to associate indicators of past success with male-coded language and experiences, penalizing résumés that reflected female educational backgrounds or professional trajectories.

Gender bias is, in fact, one of the most thoroughly documented forms of algorithmic bias in artificial intelligence systems, particularly in natural language processing and generative models. So it is not "just" an Amazon problem. Early work by Bolukbasi et al. (2016) demonstrated that widely used word embeddings encode strong gender stereotypes, linking terms such as woman with family or domestic roles and man with career and leadership attributes. These biased representations subsequently propagate into downstream AI applications, affecting automated translation, résumé screening, and content generation. Zhao et al. (2018) extended this analysis by showing that coreference resolution systems systematically amplify gender stereotypes, for example by associating occupational roles more frequently with male pronouns even when grammatical cues are neutral. Recent commentary by UN Women (2025) also argues that generative AI systems trained on large datasets drawn from historical and societal sources can reinforce and amplify gender inequalities. According to their experts, AI-based tools used in hiring, healthcare, or service provision may reproduce discriminatory patterns, such as penalizing women for career breaks, under-representing women in technical or leadership roles, or misrepresenting women's health needs, simply because the training data reflects a biased status quo.

Beyond gender bias, a growing body of research highlights the presence of political bias in generative AI systems, too and the importance of addressing it in media literacy (Shalevska, 2024; Buykov et al. 2024 and others). Political bias refers to systematic tendencies in model outputs that align with specific ideological positions, parties, or policy perspectives rather than maintaining strict neutrality. Empirical studies provide strong support for this concern. Rozado (2023) conducted large-scale testing of ChatGPT using politically sensitive prompts and found consistent evidence of left-leaning responses across a range of policy topics. These findings were corroborated in later multi-model research. Shalevska and Walker (2025) examined several leading AI language models and concluded that political bias is not unique to ChatGPT, but instead appears across different platforms.

Together, these studies indicate that algorithmic bias arises through imbalanced training data, reinforcement through prevailing online discourse, and lacking optimization strategies.

2.2. AI Bias and SDG 5 and 10

The study of algorithmic bias is directly linked to the achievement of the United Nations Sustainable Development Goals (SDGs), particularly SDG 5 (Gender Equality) and SDG 10 (Reduced Inequalities). Both goals emphasize eliminating discrimination, ensuring equal opportunity, and empowering marginalized groups. As artificial intelligence systems are increasingly used in high-impact domains, including hiring, education, healthcare, credit scoring, and public governance, biased algorithms pose a growing risk to these objectives. If AI tools systematically reproduce or amplify social inequities, they can undermine progress toward these global development targets rather than support them.

3. Methodology

This study employed a qualitative audit-based research design to evaluate gender representation within image outputs from generative artificial intelligence (AI) systems. Such qualitative methods are particularly well suited to the examination of representational bias because they allow for close, interpretive analysis of visual content and social meaning rather than solely relying on numerical distribution metrics (Braun and Clarke 2006; Crawford 2021). Similar audit methodologies have been widely used to investigate discriminatory patterns in algorithmic systems by examining differential outputs across standardized queries (Raji et al. 2020).

The research followed a zero-shot prompting approach (Kaplan, 2024), in which systems were provided with neutral, minimal prompts without stylistic guidance, demographic descriptors, or corrective instructions. This design was selected to approximate the experiences of typical end users interacting with generative AI tools in everyday contexts, rather than evaluating outputs under controlled or optimized conditions. Retrieval-Augmented Generation (RAG) tools, prompt engineering techniques, or bias-mitigation interventions were intentionally excluded to ensure that the responses reflected default model behavior at the time of testing.

Four widely used generative AI platforms were examined: Grok, Gemini (Nano Banana), ChatGPT with DALL·E image generation, and DeepSeek (Janus). Each platform was tasked to generate images for a standardized set of occupational prompts commonly associated with gendered stereotypes in labor force representation:

1. Team coach
2. Judge
3. Housekeeper
4. Computer programmer
5. Primary school teacher

These roles were chosen by the researchers and aimed to include occupations typically perceived as high-status or authority-based alongside service-oriented and care-related professions. Such contrasts enable examination of whether AI systems replicate well-documented gender stratification patterns present within real-world employment systems (England 2010). For each occupational prompt, image outputs were collected separately from each platform using identical prompt wording and default system settings. Multiple images per prompt were recorded when platforms provided gallery-style results or sequential generations. All prompts were administered during a single data-collection period in order to minimize model update variability. Images were archived along with corresponding metadata. Visual outputs were analyzed using thematic qualitative coding, following the principles outlined by Braun and Clarke (2006). Each image was reviewed and coded across several representational variables, including:

a) Perceived gender presentation (male, female, ambiguous or non-binary)
b) Occupational role alignment with traditional gender norms.

The codes were developed inductively, allowing recurring patterns to emerge from the data rather than imposing predefined categorizations. Interpretive discussion centered on whether the generated depictions aligned with or diverged from established gender stereotypes, particularly in relation to the concentration of men in high-status leadership or technical roles and women in support or caregiving professions.

3.1. Ethical and Interpretive Considerations

The study did not involve human participants and required no personal data collection; however, ethical considerations remain central when evaluating algorithmic outputs that can shape social attitudes and perpetuate inequalities. Images were analyzed with care to avoid reinforcing stereotypes in description or interpretation, framing findings around systemic patterns rather than individual depictions.

Given the qualitative nature of the research design and the limited number of prompts used, the study's findings are not intended to offer statistical generalization. Instead, the research provides exploratory evidence highlighting how commonly used AI systems may reproduce gendered representational hierarchies that conflict with global equity objectives, including Sustainable Development Goals 5 and 10.

Future research could expand this methodology through larger prompt sets, cross-linguistic testing, intersectional demographic coding (e.g., race, age, disability), and quantitative frequency analysis to supplement interpretive findings.

4. Results and Discussion

The qualitative audit of AI-generated images revealed consistent patterns of gendered representation across all four examined platforms (ChatGPT (DALL·E), Grok, Gemini (Nano Banana), and DeepSeek (Janus)) when responding to occupational prompts. Overall, outputs tended to align with traditional gender stereotypes rather than challenge them, particularly in relation to authority, leadership, and technical expertise versus caregiving and domestic roles. While some degree of variation was observed across platforms, the dominant trend was one of stereotype reproduction rather than disruption.

4.1. High-Status and Authority-Based Occupations

For the prompt *judge*, all platforms generated predominantly male-presenting figures (see Image 1). Images commonly depicted men wearing formal judicial robes, often situated in courtrooms or positioned behind desks and benches, symbolizing institutional authority and expertise. No female-presenting judges were prominently represented in the outputs displayed across platforms. No people of colour, either. This uniformity suggests a strong association within generative models between judicial authority and male gender presentation – a stereotypical linkage that, unfortunately, mirrors persistent real-world gender disparities in senior legal positions. These findings are consistent with previous research on algorithmic gender bias in AI systems. As Bolukbasi et al. (2016) demonstrated, word embeddings encode stereotypical associations linking men with technical and leadership roles and women with domestic or caregiving roles.

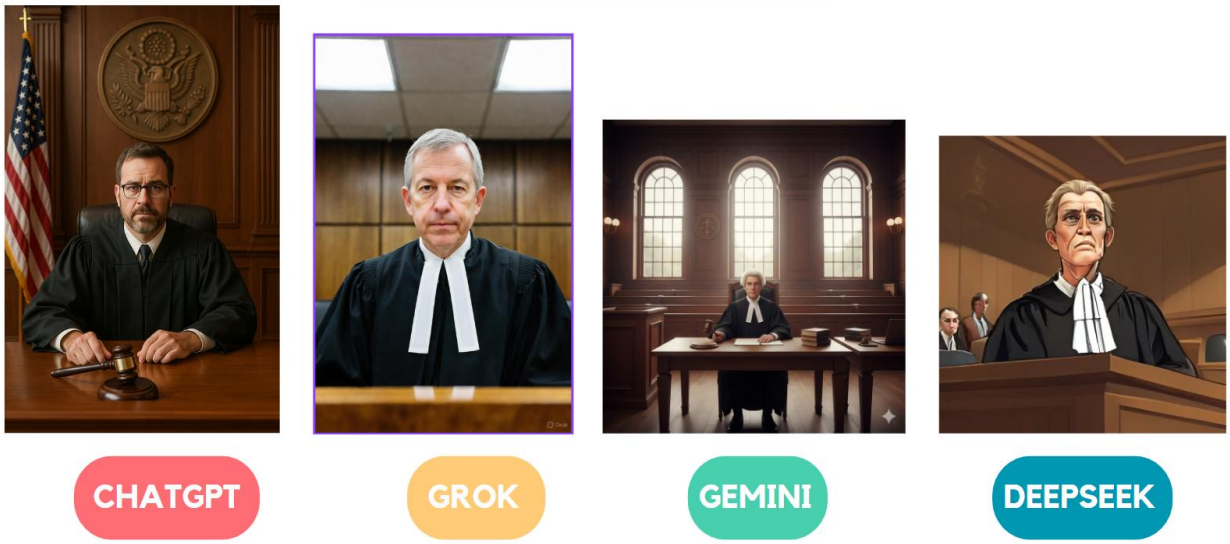

Image 1: Depictions of a "judge" by different GenAI Models

Similarly, the prompt *team coach* produced almost exclusively male-presenting subjects across all platforms (see Img. 2). Depictions featured white, middle-aged men situated in sports arenas or gymnasiums, often clad in athletic attire and portrayed in commanding postures such as directing players or holding sports equipment. The only exception to this was Grok which generated two images of younger male athletes, one of which featuring a person of colour. No images of women were generated by any of the tested models. Thus it is safe to assume that these representations reinforce cultural associations between leadership in sports and masculine identity. The absence of female-presenting coaches further highlights how generative models spatially and symbolically situate authority in sports in male bodies.

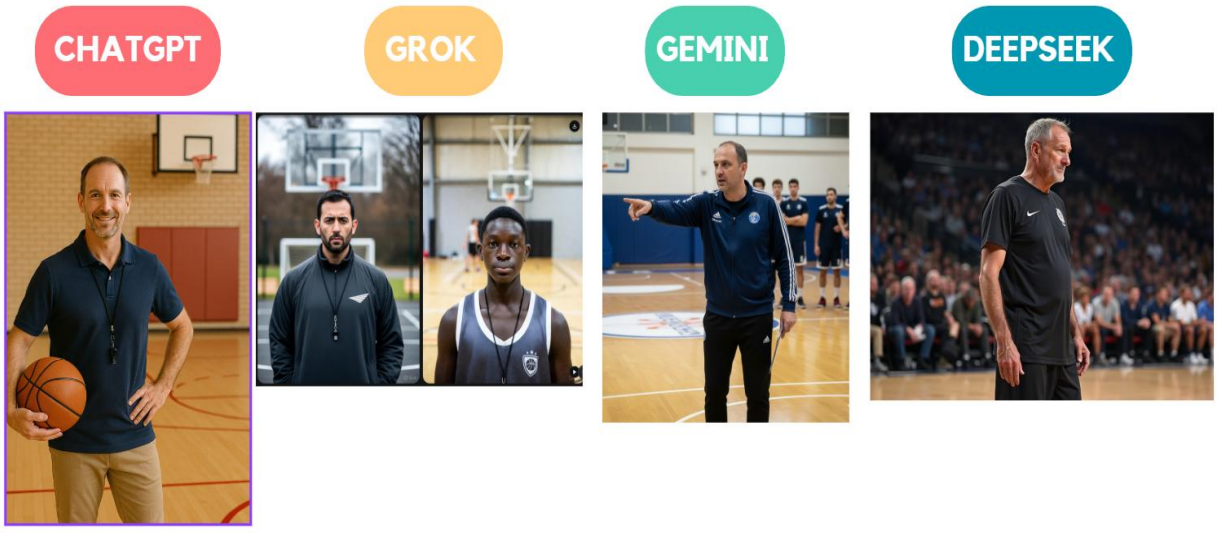

Image 2: Depictions of a "team coach" by different GenAI Models

4.2. Teaching and Care-Related Professions

Results were more mixed for the prompt primary school teacher (see Img. 3). 3 of the 4 platforms featured white, young, female-presenting teachers positioned within classroom

settings, interacting with young children or presenting at chalkboards. These images emphasized nurturing and pedagogical roles through visual cues such as warm classroom environments and relaxed body language. Two male-presenting teachers appeared in Grok's outputs, one of which being a person of colour. This asymmetry reflects broader labor patterns in primary education, where women constitute the majority of the workforce globally, yet also indicates that AI systems reproduce the association of early childhood education with feminized care labor rather than modeling more gender-diverse professional participation.

Image 3: Depictions of a "primary school teacher" by different GenAI Models

4.3. Technical Occupations

The prompt *computer programmer* resulted in near-total male dominance across all systems (see Img. 4). The produced images largely depicted young men seated at desks, working on code displayed on multi-screen setups, often framed within modern, dimly lit technical environments. Notably absent were female-presenting programmers or any mixed-gender collaborative scenes. The homogeneity of these outputs suggests that generative AI systems embed a strong gender stereotype equating technological expertise with masculinity. This reinforces the exclusionary narrative that has historically shaped the tech sector and presents barriers to gender inclusivity by visually normalizing male exclusivity within digital labor.

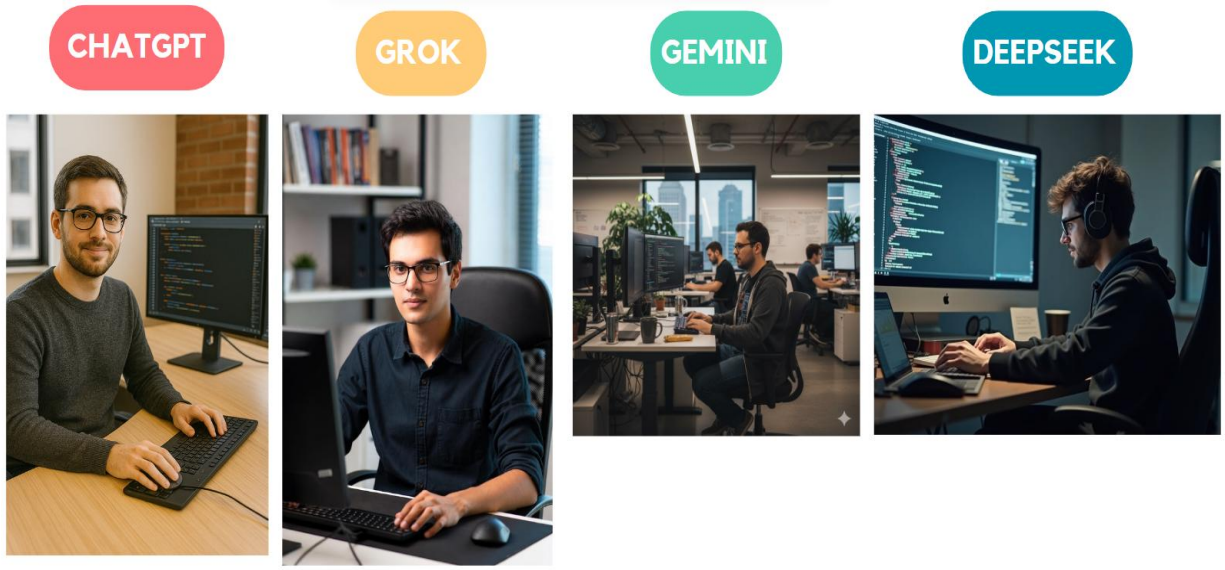

Image 3: Depictions of a "programmer" by different GenAI Models

4.4. Domestic and Service Roles

In contrast to high-status and technical fields, the prompt *housekeeper* elicited predominantly female-presenting subjects across all platforms (see Img. 5). Images frequently showed women cleaning surfaces, organizing interiors, or pushing cleaning carts, often dressed in uniforms or aprons. Visual framing emphasized service and domestic labor rather than professionalization or technical skill. Male housekeepers were virtually absent from the datasets examined.

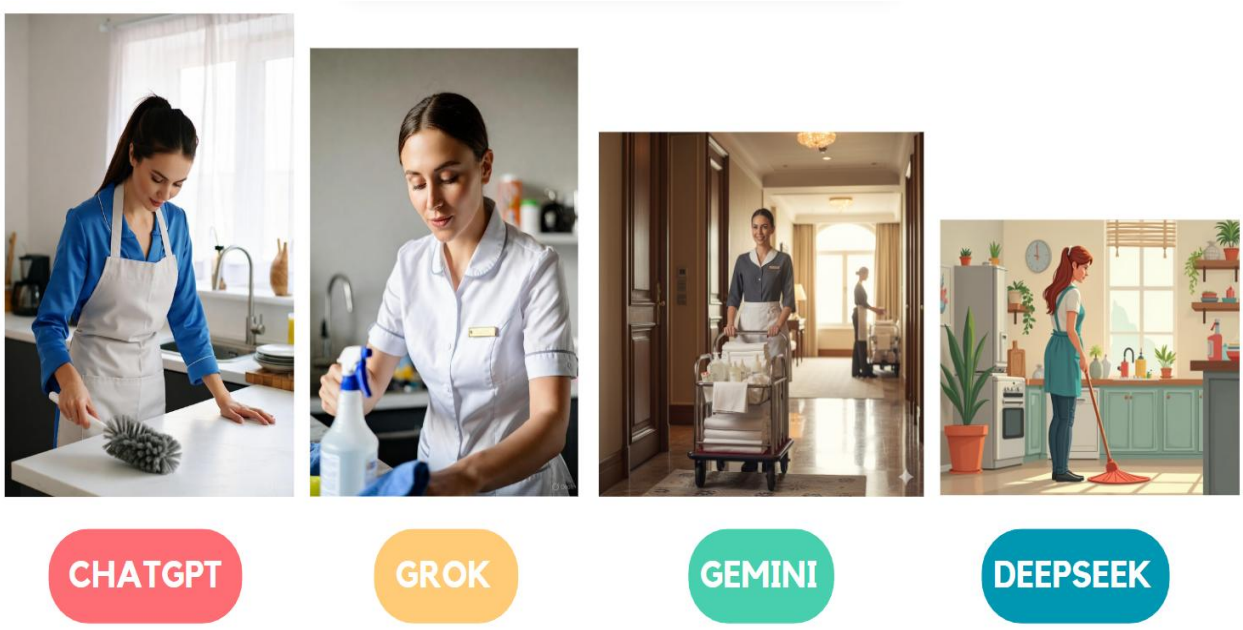

Image 4: Depictions of a "housekeeper" by different GenAI Models

This pattern mirrors longstanding cultural expectations that associate domestic labor with female identity and reinforces the undervaluation of care and service work by depicting

it as feminized, routine, and subordinate in contrast to the masculinized authority depicted in legal, coaching, and technology-related roles.

Across all occupational categories, the findings indicate that gendered stereotyping remains structurally embedded in generative image models. High-status or authority roles as well as tech-related roles (judge, coach, programmer) were overwhelmingly masculinized, while care and domestic roles (teacher, housekeeper) were feminized. Grok demonstrated limited divergence from these patterns by including male teachers, yet it did not disrupt stereotypes in leadership or technical domains.

This partial variation suggests that while specific platforms may introduce limited and minor representational diversity, system-level bias persists across the generative AI ecosystem. The consistency of results in high-status occupations across all platforms points to entrenched training-data patterns that reproduce occupational segregation based on gender. Even where deviations occur, as with Grok, they remain insufficient to counterbalance the dominant male leadership and technical imagery.

The obtained results also align with UN Women's (2025) warning that AI systems may reinforce existing inequalities when trained on historically biased datasets. In this sense, the visual outputs observed here represent not isolated anomalies but manifestations of broader structural patterns within contemporary AI development.

5. CONCLUSION

Taken together, these findings indicate that AI systems largely replicate the social patterns embedded in their training datasets rather than functioning as correctives to social inequity. Instead of challenging existing occupational stereotypes, generative models tend to mirror dominant cultural norms, particularly the masculinization of authority and technical expertise and the feminization of caregiving and domestic labor.

These outcomes raise important ethical and pedagogical considerations, especially regarding the role of AI in shaping public perception. Generative images increasingly serve as visual references for educational materials, marketing, and casual information-seeking; thus, repeated exposure to stereotypical imagery may reinforce normative assumptions about appropriate gender roles. This dynamic has implications for media and digital literacy, too,as users may interpret algorithmically generated visuals as neutral or objective representations rather than culturally conditioned outputs.

But not all hope is lost. And these representational patterns are not immutable. AI systems learn what they are taught, including human biases, but they also remain responsive to changes in training data, governance practices, and ethical design interventions. Things can and should change! And future research can focus on the algorithmic bias changes and legislative rules that govern GenAI use. Because, although the outputs of now predominantly mirror long-standing inequalities and stereotypes, targeted efforts toward dataset diversification and representational auditing surely have the potential to redirect AI development toward more equitable outcomes aligned with the objectives of SDG 5 (Gender Equality) and SDG 10 (Reduced Inequalities).

REFERENCES

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM.
2. Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *arXiv Preprint arXiv:1607.06520*.https://doi.org/10.48550/arXiv.1607.06520
3. Bykov, I. A., and Medvedeva, V. M. (2024). Media Literacy and AI-technologies in Digital Communication: Opportunities and Risks. In *Communication Strategies in Digital Society Seminar (ComSDS).* DOI: 10.1109/ComSDS61892.2024.10502053
4. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.https://doi.org/10.1191/1478088706qp063oa
5. Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
6. England, P. (2010). The gender revolution: Uneven and stalled. *Gender & Society, 24*(2), 149–166.
7. Kaplan, J. (2024). *Generative artificial intelligence: What everyone needs to know*. Oxford University Press.
8. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). ACM.
9. Rozado, D. (2023). The political biases of ChatGPT. *Social Sciences, 12*(3), Article 148.https://doi.org/10.3390/socsci12030148
10. Shalevska, E., & Walker, A. (2025). Are AI models politically neutral? Investigating (potential) AI bias against conservatives. *International Journal of Research Publication and Reviews, 6*(3), 4627–4637.
11. Shalevska, Elena. (2024). The Future of Political Discourse: AI and Media Literacy Education. *Journal of Legal and Political Education 1* (1):50-61. https://doi.org/10.47305/JLPE2411050sh.
12. UN Women. (2025, February 5). *How AI reinforces gender bias—and what we can do about it*.https://www.unwomen.org/en/news-stories/interview/2025/02/how-ai-reinforces-gender-bias-and-what-we-can-do-about-it