INTERNATIONAL JOURNAL OF SOCIAL SCIENCES BULLETIN
ISSN: (E) 3007-1917 (P) 3007-1909
Volume 4, Issue 1, 2026

INVISIBLE BIAS: GENDERED REPRESENTATIONS IN AI-GENERATED TEXT

Barkah[*1], Hareem Khattak[2], Laiba Javid[3]

[*1,3]*MPhil Scholar Department of English, Abdul Wali Khan University Mardan, Pakistan.*
[2]*MPhil Scholar Department of English, Northern University, Nowshera, Pakistan.*

[*1]barkhatilamuhammad@gmail.com, [2]hareemkhattak22@gmail.com, [3]laibajavidkhan567@gmail.com

Corresponding Author: *
Barkah
DOI: https://doi.org/10.5281/zenodo.18429773

Received	Accepted	Published
05 December 2025	15 January 2026	30 January 2026

ABSTRACT

This paper explores gender bias in AI-written texts using ChatGPT (GPT 5.2) and focus on how both male and female subjects are represented in the context of professional, social and personal life. Based on Feminist Critical Discourse Analysis (FCDA), the study discusses how language and discursive patterns depict the underlying gender ideologies and power relations. The AI generated descriptions of men and women were compared, using ten parallel prompts, in the role of CEOs, political leaders, teachers, parents, scientists and friends. The results show that although both genders are portrayed as competent, male subjects are portrayed as decisive, rational and action-oriented, and female subjects are portrayed as empathetic, relational and ethically conscious. These trends demonstrate gender conventions in society in AI output, which points to the ethical and social consequences of biased generative technologies. The research adds qualitative, FCDA-informed evidence to the discussion of AI bias, which can be used to detect bias, reduce it, and develop AI in a socially responsible way.

Keywords: *AI-generated text, large language models, Feminist Critical Discourse Analysis, gender bias, ChatGPT.*

1.INTRODUCTION

The rapid use of artificial intelligence, AI and its implementation into the daily communication, education, employment, and media have changed the manner in which human beings engage with technology. Large language models, LLMs like ChatGPT have become popular due to their capacity to produce coherent and contextually aware text in a broad variety of fields. As impressive as they may be in their natural language understanding and production, growing evidence indicates that those models are not neutral instruments and that they mirror and reproduce the cultural, historical, and social biases that are inherent in their training data (Khan, 2024; UNESCO, 2024). Specifically, the issue of gender bias in AI-generated texts has become a pressing issue, and such biases may support old stereotypes, develop the understanding of competence, and affect the expectations of the society concerning gender roles.

It has been found that AI systems tend to reflect human social biases, replicating gendered associations between men and public, authoritative, and professional occupations and women and private, relational, and nurturing occupations (UNESCO, 2024; Fang et al., 2024). These trends can be observed in various settings, such as leadership, education, parenting, scientific work, and social interactions. As an example, research on AI-based recommendation letter shows that the models often use achievement-oriented wording when addressing men and focus on relational and ethical attributes when addressing women, which demonstrates differences in framing that can influence the actual assessment and decision-making (Kaplan et al., 2024). Moreover,

the instances of bias to an extreme, like the creation of sexualized and violent images of women in AI results, present ethical and social issues on a serious level as the use of generative technologies has no proper management and protection (Wyer and Black, 2025). The results of these studies highlight the need to not only analyze the linguistic characteristics of AI-generated text but also the ideological frameworks underlying these outputs that reinforce ideologies.

Although the problem of AI gender bias and gender disparity is becoming more widely recognized, AI gender bias studies have tended to use quantitative indicators, including word frequency, co-occurrence, and statistical correlations between gendered words **(Mirza et al., 2025)**. Although these methods offer useful clues of prejudice, they are not the complete picture of the multifaceted ways language constitutes social meaning, expresses power relations, and recreates gender ideologies (Fatima, Yasmin & Irshad, 2025). Feminist Critical Discourse Analysis (FCDA) can be a strong theoretical approach to fill this gap, and it offers the discourse tools that reveal and construct social norms, hierarchy, and gendered power systems (Lazar, 2005). Through the application of FCDA, researchers can go beyond the surface patterns to observe how AI-generated stories encode implicit assumptions of authority, competency, emotional labor, and relations in their presentations of men and women. FCDA offers the means to connect linguistic patterns to larger social and ideological frameworks, which enables the subtle perception of the reproduction and/or defiance of gender norms by AI-generated narratives. Scholars, in turn, highlight the significance of bias detection, methods of mitigating bias, and ethical AI design in response to the reported biases. Some of the proposed solutions are the incorporation of fairness principles in the model architecture design, human-in-the-loop testing to assess outputs, and training data curation to be more gender-diverse (Khan, 2024; Wyer and Black, 2025, Dimgba et al. 2025). The current work fits the area of research on AI, gender studies, and discourse analysis with the purpose of revealing and analyzing gendered tendencies in AI-generated texts. Namely, it discusses the representation of male and female subjects in professional, social, and personal spheres as built by ChatGPT (GPT 5.2). Through a qualitative and in-depth analysis of AI-generated texts, this paper will be added to the current discussion on the ethical development of AI, bias reduction, and responsible use of technology. That way, it presents the significance of combining both quantitative and interpretive methods to gain a complete perspective on how generative AI mediates and reproduces gendered social norms.

Research Questions

1. How are male and female subjects represented differently in AI-generated texts across professional, social, and personal contexts?
2. What linguistic and discursive patterns in AI outputs reveal underlying gender ideologies and power relations?

Research Objectives

1. To examine the representation of gender in AI-generated texts and identify differences in depiction of male and female subjects.
2. To analyze linguistic and discursive patterns in AI-generated content using Feminist Critical Discourse Analysis (FCDA) to uncover embedded gender ideologies.

Literature Review

The study of gender bias in artificial intelligence (AI) and language models has grown quickly as generative technologies such as large language models (LLMs) are becoming part of daily communication, education, work, and media. A significant amount of empirical research demonstrates that AI systems tend to reproduce and propagate gender stereotypes common in society since they are trained on the corpora created by humans and have historical and cultural biases (Khan, 2024; UNESCO, 2024). Such biases are not limited to lexical associations but they also manifest in the framing of gender roles, agency, occupational identity and emotional behavior in AI-generated text.

A study by UNESCO that was mentioned extensively indicates that major LLMs, such as GPT-2 and GPT-3.5, are systematically gender-stereotyped in their generated texts. Female names were often connected with the words that referred to home, family and children, and the word male names were related to business, executive and career, which were stereotyped roles in division of labor (UNESCO, 2024). These results highlight how generative AI can also be used to strengthen

regressive gender ideals, particularly in the context of these models becoming more common in search engines, writing, and education.

Similar studies show that there is quantifiable word-level gender bias in several LLMs (**Mirza et al., 2025**). A comparative study of news text produced by a variety of models, such as ChatGPT, GPT-2, and GPT-3-based models, revealed that all of them demonstrated a high level of gendered variations in the distribution of words compared to professionally edited human news corpora. Even though more recent models like ChatGPT generated less biased content, the gender bias was not trivial, which means that there are still representational imbalances in AI text generation (Fang, Che, Mao, et al., 2024, Contreras, 2025). Such quantitative results are consistent with the general trends of bias in AI systems in different languages and cultural settings and indicate that gender bias is not a unique phenomenon of a particular model or linguistic setting (Khan, 2024).

Gender bias is also noted in certain functional applications of AI as pointed out by empirical work. As an example, recommendation letter analyses by ChatGPT showed that the model recreates much of the gender language patterns present in human-written recommendation letters, including the use of achievement and agentic language differently. It is not only indicative of biasness in the narrative creation but also a possible real-life implication in the situations where AI output is likely to be used in making a decision (Kaplan, Palitsky, Arconada Alvarez, et al., 2024). These biases are consequential since the use of language that portrays competence, agency and leadership will influence the judgments of qualification in education and professional environments.

In addition to the occupational and role representations, recent research pays attention to the extreme manifestations of bias. In one investigative study of GPT-3 text completions, women were portrayed with highly problematic depictions in situations that involved sexualized violence, and text completion prompts about women produced results supportive of harmful and violent stereotypes. The outputs bring up ethical issues regarding the internalization of sexualized bias in AI discourse and the social detriments that could occur through the implementation of biased systems without proper protective measures (Wyer and Black, 2025). The study broadens the scope of the knowledge on gender bias to encompass the concept of safety and harm in content created.

The cross-linguistic analysis is one more dimension of analyzing AI bias. Reports on AI products in various languages note that gender bias is observable in all languages, and models generate stereotypical correlations in job descriptions and expectations of behavior that differ across languages but follow gendered patterns (Khan, 2024).

These results indicate that model training data is influenced by structural characteristics of language and culture to reproduce gender stereotypes, which implies that culturally sensitive bias assessment models are required. Gendered representation does not appear only in AI-generated discourse it has been present in traditional media long before it as well. Barkah and Javid (2025) conducted a Critical Discourse Analysis of headlines in *Dawn* newspaper and discovered that Pakistani media treats working women in two ways, showing them as weak subjects with the burden of the lack of paid care and less support of the institution and as agents of societal and economic life. Their analysis shows the existence of patriarchal ideologies, gendered division of labor and social conventions incorporated in the media discourse. It means that gender bias in AI outputs could partially mirror the existing discursive tendencies in human-generated texts, which AI systems are trained on.

In the current paper, the theoretical and analytical approach of Feminist Critical Discourse Analysis (FCDA) is used to analyze AI-generated texts in terms of power relations, gendered ideologies, and discourse representation (Lazar, 2005). FCDA offers the means to connect linguistic patterns to larger social and ideological frameworks, which enables the subtle perception of the reproduction and/or defiance of gender norms by AI-generated narratives. Scholars, in turn, highlight the significance of bias detection, methods of mitigating bias, and ethical AI design in response to the reported biases. Some of the proposed solutions are the incorporation of fairness principles in the model architecture design, human-in-the-loop testing to assess outputs, and training data curation to be more gender-diverse (Khan, 2024; Wyer and Black, 2025, Dimgba et al. 2025). Nevertheless, even with the increased awareness and the efforts of intervention, research continues to detect bias in even advanced models

and demands more profound structural shifts in AI system development and deployment. Collectively, the available literature indicates an apparently stable empirical trend: AI systems are likely to reproduce gendered assumptions included in the training corpora, leading to biased associations, role modeling, and even damaging stereotyping. This literature contextualizes the purpose of the current research, which is to reveal and examine these biases in AI text in terms of Feminist Critical Discourse Analysis, giving a qualitative bias measure a deeper interpretive perspective and connecting it to a larger social and ideological meaning.

Methodology

Research Design

This research will have a qualitative research design that is based on Critical Discourse Analysis (CDA) and Feminist Critical Discourse Analysis (FCDA) as suggested by Lazar (2005). The method is chosen to reveal the process of discursive construction and reproduction of gender ideologies and power relations in AI-generated texts.

The data is AI-generated texts created by ChatGPT (GPT 5.2) with the help of a series of parallel prompts that were created to generate similar descriptions of both male and female characters. Ten prompts were employed, including professional, social, and personal roles such as CEO, political leader, teacher, parent, scientist, volunteer. Two texts were written in response to each prompt, one about a male subject and one about a female subject, and there were 20 texts in total.

The prompts were maintained the same way other than gender reference so as to be consistent and comparable. The texts generated were stored and divided into two sets male-referenced texts and female-referenced texts.

Analytical Framework

The Feminist Critical Discourse Analysis (FCDA) is used to guide the analysis and approach the language as a place where the gendered power relations are constructed and reproduced. The categories of analysis utilized were the following discursive categories:

Gender roles (professional, domestic, social) representation.

Agency and transitivity

The use of lexicon and adjectival patterns.

Emotionality/ rationality.

Leadership and authority

Construction of identity and naming.

Procedure of Analysis

Close reading of the texts was done through a manual qualitative analysis. All the texts were analyzed to find the recurring linguistic and discursive patterns concerning the categories of analysis. The relevant segments were coded and compared between male and female data to demonstrate asymmetries and differences in representation. The results were then explained with the prism of FCDA, in which linguistic patterns were associated with the larger gender ideologies and power dynamics that are embedded within the discourse generated by AI. The research utilizes texts generated by AI and available publicly, and it does not include human subjects. Thus, it does not use any personal or sensitive information and does not pose an ethical threat.

Data Analysis

The chapter is an in-depth critique of the AI-generated texts obtained on answering the prompts concerning the male and female subjects in professional, personal, and social settings. The paper looks at the presentation of gender in these texts in ten thematic issues, including CEOs, political leaders, teachers, parental care, friends, scientists, engineers, dealing with disappointment, community volunteering, and personal traits. It is analyzed through Feminist Critical Discourse Analysis (FCDA) (Lazar, 2005) that focuses on how discourse creates and reproduces gendered power relations and ideologies. Through a qualitative methodology that is manual, the study determines the patterns of language, syntactic structures, and the focus on themes and compares the depiction of men and women at similar situations.

The analysis is organized in two supplementary dimensions. To begin with, a descriptive analysis will be conducted where verbs and thematic focus, as well as explicit lexical choices, will be analyzed concerning male and female AI-generated texts. Second, the interpretive analysis based on the FCDA information studies the underlying ideologies, power relations, and gendered assumptions inherent in the language. In line with Lazar (2005), this paper views language as a social practice that reproduces and reinforces societal hierarchies, i.e. even purportedly neutral AI

outputs might reproduce gendered norms and expectations. The chapter through this dual lens understands the stories produced about each gender and makes a difference in terms of agency, emotionality, relational focus, and ethical framing. The computer-generated descriptions of CEOs show a uniform trend of gender representation. Male CEOs are described as decisive, strategic, and achievement-oriented, and verbs like oversaw, launched, modernized, etc. are used to highlight their agency and power in the work sphere. The male stories emphasize achievements such as market expansion, introducing new products, and adapting to economic unpredictability, and concentrate on quantifiable results and publicity. On the contrary, female CEOs are portrayed as equally competent in their profession but with added focus on ethical leadership, social responsibility, and mentorship. Although the AI models recognize female success, the presentation is framed with a relational and ethical aspect that is not as high as in male CEO stories. This disparity represents a slight prejudice of the way AI establishes leadership approaches, as men are depicted as aggressive performers and women as socially aware leaders (Kaplan et al., 2024; Fang et al., 2024).

In political leadership, male leaders are always portrayed as being strategic, pragmatic and decisive in their approach, which is crisis management and implementation of policies. In comparison, political leaders who are females are portrayed as cooperative, tolerant, and consensus-driven and focus on negotiating and risk-management rather than making decisions independently. According to this trend, the texts produced by AI absorb the gendered norms of power, according to which men are linked to decisiveness and social action, whereas women are placed at the role of mediators, who are more concerned with harmony and social stability (Wyer and Black, 2025; Khan, 2024).

Gendered tendencies are also present in educational roles. Male teachers are positioned in such a way that they are authoritative and supportive in that they focus on intellectual direction, critical thinking, and discipline. Women teachers, in their turn, are characterized as caring, motherly, and sensitive to emotional and academic needs of the students. Although both genders are qualified, the AI-generated texts focus on relational and emotional work on female teachers and cognitive and structural instructions on male teachers. This is in line with the existing literature on the subject of occupational gender stereotypes, placing women in the workforce as those who provide care and support and men as those who possess leadership and intellectual power (Kaplan et al., 2024; Lazar, 2005).

These tendencies are reflected in the description of the roles of parents. Male parents are said to be patient and supportive and tend to be systematic in their child care activities like helping in homework and routine organization. Female parents are always put in the perspective of an emotional caregiver and overworked nurturer, who places more emphasis on nurturing, paying attention, and being emotionally reassuring. The male counterpart is shown to be reliable, proactive and solution oriented, and as compared to the female counterpart, the female counterpart is shown to be empathetic, attentive and communicative, below which relational support over instrumental action is emphasized. These trends depict how AI recreates the classical gendered scripts, with men being the action-oriented problem-solvers and women being relational and emotionally sensitive characters.

As the description of scientific and technical jobs further shows, male scientists and engineers are defined by analytical thinking, persistence and objective problem solving, which is commonly attributed to innovation and personal success. Women scientists and engineers are portrayed as being competent but more focused on teamwork, mentorship, and ethics. This shows that, although AI acknowledges the technical potential in women, it still places the input of women in terms of relationships and cooperation, which implicitly legitimizes gendered expectations in even highly professionalized areas (Fang et al., 2024).

Speaking of emotional reactions and personal characteristics, male participants who manage disappointment are reflected as thoughtful, strong, and result-oriented, considering the ability to learn by failing and perceive in a rational way. Female subjects lay stress on emotional processing, mentoring, and adaptive strategies, which means that the method of coping is socially mediated. In the same vein, male characters are linked to analysis, discipline and emotional stability when listing personal strengths, whereas female characters are linked to competence and empathy, communication and relational intelligence. These trends represent the continuity of gendered

ideologies in AI products and correspond to men and women, respectively, as rational and relational, respectively. These trends are not unique to AI discourse but are similar to those that are represented in the traditional media. As an example, Barkah and Javid (2025) in their Critical Discourse Analysis of *Dawn* newspaper headlines discovered that working women in Pakistani media are described in two ways: as victims of lack of care, less institutional support and burden of not getting complete pay and as agents of social and economic input. This analogy implies AI generated discourse can be replicating existing media ideologies instead of creating bias by itself.

Themes that cut across all areas are unequal focus on agency, emotional labor, ethical responsibility, and relationality. Male AI characters are always depicted as action-oriented, leadership oriented and problem solving and the female characters are depicted in supportive, empathetic and socially responsible roles. These results suggest that AI-written texts recreate gender norms within the society, which is caused by bias in the training data as well as the cultural narratives, the models are trained on (Khan, 2024; Fang et al., 2024). Through the implementation of FCDA, one is able to interpret these textual patterns not only as descriptive productions but as an indication of ideological reproduction of gendered hierarchies. The discussion shows that AI-generated texts support gendered expectations, however subtle yet consistent, in the professional and personal and social sphere. Whereas male and female participants are equally effective in AI outputs, the framing and contextualization of these are different, with women being relational, empathetic and ethically oriented, and men decisive, rational and public agency. The results are in line with prior research that indicates gender bias in LLM in various situations, such as leadership, education, and interpersonal relations (Kaplan et al., 2024; Wyer and Black, 2025).

Conclusion

This paper examined gender bias in AI-generated texts using ChatGPT and examined how it represented men and women professionally, socially, and personally. The discussion shows that despite the apparent neutrality or objectivity of AI output, the choice of verbs and adjectives, the focus of description all have subtextual messages about agency, emotional labor, and authority. These results are supported by the previous empirical research, such as recommendation letter (Kaplan et al., 2024), news content (Fang et al., 2024), and extreme depiction of sexualized violence (Wyer and Black, 2025), which reinforces the conclusion that AI systems reproduce and sustain the culture and historical gender-related biases (Khan, 2024; UNESCO, 2024). The paper also highlights the social and moral concern of gender biasness in AI-generated content. Because of the increased use of LLMs in education, professional communication, and media and decision-making, biased outputs can become normalized, affect judgments of competence, and perpetuate structural inequalities. This highlights the role of bias detection and mitigation planning and ethical AI design that considers both social impact and linguistic representation (Khan, 2024; Wyer and Black, 2025).To sum up, the study proves that AI-generated text does not reflect reality in a neutral way but is a discursive space where gender ideologies are reproduced, enhanced, and naturalized.

REFERENCES

Barkah, & Javid, L. (2025). Representation of working women in Pakistani media: A critical discourse analysis of Dawn newspaper headlines. ***Global Sociological Review,*** ***10***(4), 69–79. https://doi.org/10.31703/gsr.2025(X-IV).06

Contreras, J. M. (2025). Automated evaluation of gender bias across 13 large multimodal models. *arXiv*. https://doi.org/10.48550/arXiv.2509.07050

Dimgba, M. O., Oba, S., Agrawal, A., & Giabbanelli, P. J. (2025). Mitigation of gender and ethnicity bias in AI-generated stories through model explanations. *arXiv*. https://doi.org/10.48550/arXiv.2509.04515

Fatima, W., Yasmin, M., & Irshad, I. (2025). Exploring gendered language and socio-cognitive impact in AI-generated texts: A critical discourse approach. *Pakistan Languages and Humanities Review,* 9(3), 504–524. https://doi.org/10.47205/plhr.2025(9-III)42

Fang, X., Che, S., Mao, M., Chen, Y., & Li, J. (2024). Bias of AI-generated content: An examination of news produced by large language models. *Scientific Reports, 14*, 5224. https://doi.org/10.1038/s41598-024-55686-2

Kaplan, D. M., Palitsky, R., Arconada Alvarez, S. J., Pozzo, N. S., Greenleaf, M. N., Atkinson, C. A., & Lam, W. A. (2024). What's in a name? Experimental evidence of gender bias in recommendation letters generated by ChatGPT. *Journal of Medical Internet Research, 26*, e51837. https://doi.org/10.2196/51837

Khan, V. (2024). Artificial intelligence and gender bias: Analyzing algorithmic discrimination in language models. *Journal of Gender, Power, and Social Transformation, 1*(2), 31-40. https://researchcorridor.org/index.php/jgpst/article/view/329

Lazar, M. M. (2005). *Feminist critical discourse analysis: Gender, power and ideology in discourse*. Palgrave Macmillan.

Mirza, I., Jafari, A. A., Ozcinar, C., & Anbarjafari, G. (2025). Quantifying gender bias in large language models using information-theoretic and statistical analysis. *Information, 16*(5), 358. https://doi.org/10.3390/info16050358

OpenAI. (2026). *ChatGPT (GPT-5 mini) [Large language model]*. https://openai.com/chatgpt

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

Wyer, S., & Black, S. (2025). Algorithmic bias: Sexualized violence against women in GPT-3 models. *AI and Ethics, 5*, 3293-3310. https://doi.org/10.1007/s43681-024-00641-0