Wah Academia Journal of Social Sciences
Volume 5, Issue 1, Mar 2026, pp. 532-558
ISSN – E 2958-8731 P 2958-8723
https://doi.org/10.63954/WAJSS.5.1.28.2026

Bots with Bias: Gender-Indexed Politeness in AI Chatbot Outputs

Maliha Kalsoom[1], Fatima Hasan Zai[2], Sassui Afzal[3] and Kaukab Saba[4]

Abstract

Human communication has also grown beyond face to face and computer-based interaction to include communication with machines with the fast development of conversational artificial intelligence. Such a change has brought up an issue regarding the way in which chatbots formulate politeness, interpersonal tone, and gendered linguistic behavior. The literature has discussed the bias against gender in AI systems, politeness in human-AI interaction, and how AI training data affect stereotypical responses, but there is a relative paucity of studies that explicitly compare the linguistic response of AI chatbots to male-coded and female-coded users. This paper fills that gap by examining lexical, pragmatic, and politeness-based differences in responses that were produced by various versions of ChatGPT. Based on the Politeness Theory by Brown and Levinson (1987) and the Gender and Language theory by Holmes (1995), the study will examine the hypothesis on whether AI replicates gender tendencies in positive/negative politeness, mitigation, hedging, and conversational style. Qualitative design was applied in which gender-coded prompts were then entered into ChatGPT and the results analyzed using word-frequency, contextual interpretation, and thematic comparison with politeness-marketing coded responses. The results demonstrate apparent differences: prompts marked as male were approached more directly, more concisely, and task-focused, whereas prompts marked as female tend to get much warmer and richer in detail, including more occurrences of positive politeness and indirectness. These tendencies imply that AI chatbots fail to assume an apolitical communicative position but rather reflect culturally transmit-ted gender standards that are represented in their training material. This research paper will add to sociolinguistic literature on non-human communicators and will indicate the necessity of more equitable, stereotype-free AI language systems.

Keywords: Artificial Intelligence, Politeness Theory, Gendered Language, Chatbots, Sociolinguistics, Pragmatics, Bias in AI

[1-2-3-4]Department of English, International Islamic University, Islamabad – Pakistan

Introduction

The fast-paced advancement of conversational artificial intelligence has marked a significant change in human interaction with computers. Chatbots, such as ChatGPT, Alexa, and Google Assistant, are more likely to determine the norms of communication. Different studies have explored chatbot discourse to determine how users perceive gendered AI voices (Kim et al., 2019; Guo et al., 2020). In addition, some have distinguished pragmatic dimensions in human-AI interaction (Nurmukhambetova, 2021; Lee et. Al., 2024)

Many recent studies on AI have addressed gender bias as a computational or ethical issue. Therefore, the current study focused on the gendered politeness depicted in AI chatbot responses. This is a sociolinguistic and pragmatic study of language used by AI in the generation of responses according to the situation or prompt owner. Sociolinguistic research has traditionally investigated the effects of gender on the politeness and linguistic behavior of human beings, but how these tendencies can be reflected in AI-produced discourse has not been widely addressed.

This study examines the presence of gendered tendencies in AI chatbots regarding politeness and mitigation in commercial as well as paid AI chatbots. It examines whether AI chatbots promote conventional gender communication rules or challenges, and dismantle them. In particular, this study aims to establish variations in politeness markers in chatbot responses under gendered prompts (e.g., hedges, modals, apologies, and indirectness). The current study draws attention to English language conversation in chatbots consequently analyzing the responses it generates under male and female coded prompts. Therefore, emphasis is placed on the lexical and pragmatic markers related to politeness.

The main research questions are as follows: Is there any exhibition of gender-specific behavior in politeness or mitigation during conversations in AI responses? How do these tendencies relate to the current sociolinguistic theories of gender and language? How does an AI chatbot use lexical and pragmatic markers towards a specific gender? The focus here is not on how gender bias is encoded by algorithms; rather, it investigates chatbot language choices during response generation. This is to examine whether chatbot language reflects, resists, or remains neutral toward gendered politeness norms.

This study is significant in that it expands sociolinguistic research on non-human communicators and, provides a solution to the problem of bias, fairness, and representation in AI systems. This study adopts gender and politeness theory (Brown and Lavinson, 1987; Holmes, 1995) when speaking about the AI language. These theoretical aspects help identify key insights into this research. This, in turn, contributes to the understanding of how conversational technologies can reinforce or transform social norms. In addition, it will determine how to create a more inclusive and respectful AI chatbot language that does not promote stereotypes.

Review of the Literature

The literature review contains three subsections. Each section discusses the key subtheme that is connected to the main topic of this study. Subthemes include; Politeness and AI, Gender Stereotypes and AI, Language and AI responses that are explained as following:

- **Politeness and AI**

In language use, politeness means respecting and considering feelings of other people. Morand and Ocker (2003) defines politeness as respecting and maintaining the social "face" of others. There are linguistic and relational cues that the speaker uses to maintain and respect social relation of both hearer and speaker. People use positive and negative politeness strategies to manage social distance and respect during interaction with people. Positive strategies create closeness and solidarity to make hearer feel valued and included whereas Negative strategies protect the hearer's negative face to express non-inference and distance. AI systems also exhibit these strategies in their algorithms.

As Ivkovic (2024) studied how GPT-3.5 answered to questions that are asked about five different topics using negative politeness strategies as well as positive politeness strategies. The findings showed that there is no any significant change in the structure, quality and coherence of argumentative topics when they are asked by using two different politeness types indicating that GPT-3.5 does not respond based in interpersonal cues, the way humans typically do. Instead responses typically vary on the basis of topic rather than type of politeness used. This finding indicates a shift in typical human interaction behavior and poses questions that whether AI politeness behavior genuinely adapt to interpersonal cues, particularly gendered

ones or not, which our study aims to investigate. Babaeva et al., (2020) in his study highlighted how chatbots use politeness markers, emotional expressions and personal pronouns to produce human like responses. It sets the framework of communicants' interaction - between a human and computer, in which the difference between human-computer interaction (HCI) and the interaction between humans (HHI) is clearly noticed.

Research indicates that politeness strategies do have an impact when interacting with robots. However, the context of the interaction also has a greater influence on how users perceive and interact with robots than politeness alone (Salem et al., 2013). For instance, Ndububa and Ugoala (2025) in their study provided an analysis about how LLM-based chatbots i.e. Claude, Gemini, Perplexity, ChatGpt and Capilot manage politeness strategies, style variation and prompt adaptation in human-AI interactions. They studied how phrasing patterns, formality, sentence length and idiomatic expressions vary across different chatbots (2023). Style variation, diplomatic language by AI and prompt engineering are the most significant features of this study showing how these features are uneven across different platforms.

These researches focused on the politeness strategies used by AI tools to produce human like responses but there remains a gap on how AI chatbots use politeness markers and other politeness strategies while responding to female users whereas employing directness to male users.

- **Gender Stereotypes and AI**

Gender stereotypes are reinforced in AI systems through biases in dataset training and algorithm development (Manasi et al., 2022). These AI systems replicate the societal norms as they are instructed based on historical data which carry underlying gender role biases. Voutyrakou et al. (2025) show in their re-search how gender bias is a complex and persistent problem that comes from both cultural influences and technical problems. They used two AI tools Chatgpt and Gemini to show how sociocultural norms influenced the outcomes of AI tools. The research highlighted that conversational Artificial Intelligence systems reinforce social stereotypes in their conversational and interactional style which exacerbates inequalities in their responses.

Studies also investigate users' behavior towards the responses of AI chatbots depending on the feminine and masculine traits they use. As Ahn et al. (2022), studied chatbots with "male" and "female" personalities. They revealed that users behave differently when chatbots act caring, warm and polite (feminine) or directive and confident (masculine). Also, people like chatbots when their response matches and reproduces social norms rather than subverting it which means AI influences perceptions of people. Not only this, responses of different AI systems respond differently to controversial topics. For example, Ghafouri et al. (2023), in their paper looked into how large language models handle debatable and controversial topics such as religion, gender and free speech. They compared responses generated by AI with human debated to measure biases and moderation while highlighting the need for more transparency to ensure balanced reasoning and fairness in AI systems.

Researches show that Gender biasness is not only limited to text-based responses; image generation AI models also perpetuate and amplify social biases. It is noted by Cintas-Canto and Chen (2023) that AI reproduced harmful stereotypes such as giving gender roles to machines translation, facial recognition performing defectively on dark-skinned women. This work worked on text base as well as image base Artificial Intelligence systems. It not only focused on how gender base stereotypes are perpetuated in AI but also how it can be mitigated by exploring algorithmic accountability. Yiran Yang (2023) found that racial and gender bias are also present in AI driven image generator where white people were presented better than Black or East Asian people.

Existing studies have mostly focused on general gender biases in output of AI systems but there is a notable gap focusing on investigating the response of AI chatbots to particularly female vs male user ids. Addressing this gap is essential to know how AI systems adapt and respond to gender identity of the user to enhance fairness and transparency which our study seeks to cover.

- **Language and AI responses**

Strengthening linguistic strategies in AI systems focuses on leveraging particular language strategies to build engagement and trust with others. Studies show that specific linguistic strategies of chatbot interaction, such as directive tones, conversational alignment and politeness strategies play crucial role in how users perceive them. Dippold et al. (2020) examined the use of language in chatbots and

how their speech style influences users using interactional sociolinguistics, conversation analysis, and politeness theory, showing how specific linguistic choices of chatbots affects user engagement, alignment, and trust. Their findings show that features such as politeness strategies, directive tone and conversational alignment influence how users perceive chatbots as social actors. When AI applies empathetic algorithms, it gives a more human-like interaction, thereby being perceived as a friend rather than a mere tool. For example, Empathetic AI in customer service delivers more personalized interactions, which builds a sense of authentic engagement, leading to increased customer satisfaction (Leoc´adio et al., 2024).

Klein and M¨oslein (2022), in their article, focused on how female chatbots such as Siri and Alexa are programmed to give polite and submissive language patterns which show them service-oriented reflecting feminine stereotypes. It proves that politeness strategies in AI are not neutral and they are directly linked to how gender is performed through language, reinforcing societal expectations about how "female" and "male" voices should behave. Use of empathetic language enhances user engagement by engaging users on emotional language. For example, AI dialogue systems incorporate empathy and humor which not only improves engagement and motivation but also fosters a positive learning setting (Zhai et al., 2024).

Empathetic language in AI models influences user's involvement and perceptions, evaluating AI users either a 'friend' or a 'servant'. The article" Effects of Gender and Relationship Type on the Response to Artificial Intelligence" by Kim et al. (2019) showed that how people react to AI chatbots speakers based on their relationship type (friend or servant) with AI and the AI's gender (female or male). The study found out that people felt more pleasure and warmth in interaction with AI as a "friend" type rather than a "servant" type but interestingly, gender of the AI did not have any influence on users' perception of warmth, pleasure or competence. So, it highlighted the importance of relationship-building features in Chatbots design for better engagement and acceptance among users.

These studies highlight how the type of language use and other linguistic strategies influence users' perception and trust building while interacting with AI models. More focus is required on how language use is different when AI chatbots give responses to male and how linguistic markers, politeness strategies and hedges etc are used while interacting with females which will be covered by our study.

❖ **Biasness and AI**

Biases inherent in AI models, especially related to gender,race, language and other socio-cultural categories can have far-reaching implications across many domains. Numerous studies have high-lighted these biases. As Wellner (2020) in her article, "When AI is Gender-biased: The Effects of Biased AI on the Everyday Experiences of Women," inspected that many Artificial Intelligence systems, such as voice assistants, translation tools and facial recognition systems respond more effectively to men and perpetuate social inequality. She claims that not only data used to train AI is responsible but these biases originate from the way algorithms are built and proposes that AI systems should be designed to respond everyone equally. Addressing these biases requires accountability in the design and development phases (Gonz´alez and Rampino, 2024). Aggarwal and Bhargava (2023), in their paper explored how gender-based bias affects AI systems that make image captions. They found that biases in datasets lead the models to produce stereotyping interrelations such as linking women with domestic tasks and men with outdoor activities. To separate the process of describing an image from identifying gender, they developed a new model called as Show, Attend and Identify (SAI).

Language learning models also manifest cultural biases, which reinforce stereotypes that are present in their training data. Understanding such biases is essential because language is not only a mode of communication but it also shapes social interactions and people's perceptions (Shrestha and Das 2022). A deep analysis of data sets and training of linguistic inputs can aid in creating better and balanced AI systems.

Such studies show that biases are implied in AI systems and chatbots in one way or the other. They carry inherent biasness in their algorithms and our study will be mainly focusing on how gender stereotypes and gender base biasness is present in AI chatbots and how they perpetuate inequality by giving different responses based on user's identity either female or male.

Very little work has been done in a systematic way to determine how responses generated by AI vary in politeness strategies under different conditions. Although researches are present upon how gender biases and politeness strategies are

integrated in AI chatbots but there is no research on how the response of AI is different to males and females.

Methodology

The data used in the current research paper were all gathered using the ChatGPT inbox and we primarily focused on gender-coded prompts including hedges, modals, apologizes, and indirectness. We have researched using a qualitative approach. The value of this research is that with the help of it, we learn various techniques of coding messages in chatbots, in particular in AI.

- **Theoretical framework**

In this research we used Levinson's (1987) Politeness and Holmes's (1995) Gender and Language theories. How does Levinson's (1987) theory connect with chatbots? Because chatbots are intended to emulate human patterns of interaction, such as face-saving behavior, the claim made by Levinson becomes highly relevant. Although chatbots do not have a physical face, their design is often based on gendered politeness. A chatbot that uses a female voice or persona can be designed to use more positive signs of politeness, such as empathy, warmth, and supportive language, whereas a chatbot persona with a male voice can use more direct or negative politeness strategies. The framework by Levinson, therefore, allows you to explore how gendered politeness is artificially created in chatbot conversations and how such programmed decisions affect the perceptions of users regarding gendered communication styles. In contrast, according to Holmes, women tend to use positive politeness aspects such as empathy, encouragement, and compliments, which result in more cooperative and facilitative patterns of speaking. Men, on the contrary, often use more referential, direct, and authoritative talk, which is the characteristic of forceful or competitive communication. Holmes claims that such linguistic tendencies are cultural and not biological, and thus politeness is a very gendered affair in any society. Since women tend to be the target audience of most contemporary chatbots, such as Siri and Alexa, which are deliberately addressed to women by names, voices, nicknames, or assigned personalities, this method is highly relevant to the given research project.

❖ Research Design

The analysis of the data was conducted in five systematic steps: 1. Splitting the Data Set The data was selected into two subgroups, i.e., the responses to women and the responses to men. The division made it possible to make a comparative analysis of the attitude towards different genders of chatbots and their interaction with them in the context of politeness, empathy, and authority. 2. Setting Word Frequency. We determined and mainstreamed words and phrases with politeness and impoliteness, including words like" please,"" sorry,"" think," and" happy to help." This move gave a quantitative description of the frequency of the occurrence of the gendered politeness markers in concurrence with the distinction of positivity and negativity in politeness strategies (Brown and Levinson, 1987). 3. Learning Frequent Words in Context. All the words that were high frequency were discussed within the context in order to identify the changes in tones, agency, and politeness. As an example, the term" think" was used in both the sentence" you lose the ability to think independently" and" I think," and the levels of directness or empowerment were different. This action fits into the models of Holmes, as it demonstrates how women and men speak to facilitate authoritative interaction patterns. 4. Comparing Patterns and Themes. Through the analysis of the word use and context of the responses to men and women, we determined that some patterns that emerged in the answers included empathy, encouragement, directness, or authority. This has enabled a subtle realization of how politeness and face-saving strategies are used depending on the gender, which is true in the theory of both Levinson (1987) and Holmes (1995). 5. Determining Applicable Quotes. The representative quotes were chosen, which will allow highlighting the differences in the reaction towards male and female users by the chatbot in communication.

❖ Data compilation

Data is collected from different chatbots, such as ChatGPT (GPT 5.1 paid, scholar, and commercial ChatGPT). Our main participant of data is ChatGPT, and we ask different questions from male and female name identities. Our time duration is three days. These are questions asked of ChatGPT, such as the following: 1. Can you help me improve my presentation? 2. Give three suggestions for my future. 3. Can you advise me on how to handle stressful situations at work? 4. How should I wish my friend on his birthday? 5. Generate a polite apology message for my classmate when I don't understand the intent of the message. In one sentence, explain why AI

is harmful. Why do we take these questions? Because we rhetorically examined how much AI extends into interpreting human stereotypes and how much it copies them.

Analysis

Men tend to adopt referential, authoritative, or competitive speech patterns, and women tend to adopt the positive politeness technique, which results in cooperative and facilitative speech patterns. These tendencies are used as the theoretical background of understanding the reproduction and reinforcement of gender stereotypes in chatbots since the latter are culturally created as opposed to biological. Because modern chatbots such as Siri or Alexa are commonly feminized, be it by name, voice, or character, the information provided by Holmes can be directly applied to this research.). Levinson focuses on face-saving communication modes. Chatbots are designed to mimic human interaction habits, such as manners, even though they are not provided with faces. As an example, a chatbot that uses the female voice often incorporates positive markers of politeness such as empathy, encouragement, and supportive language, whereas a chatbot with a male voice may resort to more direct or task-oriented and negative politeness tactics. This allows the researchers to examine the effects of the gendered politeness artificial programming on the perceptions of users towards the gendered communication. In this manner, Holmes (1995) improves this practice by understanding gendered communication patterns in daily communication. Here different questions are asked to ChatGPT from two different emails: one response taken from the male I'd and second from female I'd to analyze the difference of responses given on each I'd similarly paid Gpt is also used to generate prompt for analysis. The analysis thus is divided into three sections and each section analyzes different techniques of language used by Chatbots. The first section analyzes the overall influence of gender and language in responses generated. The second section looks into politeness markers or strategies used. The final part deals with the hedges used in the responses.

- **Language and gender Influence in AI responses**

In this section a sociolinguistic analysis is concerned in the language used by chatbots. Based on Levinson and Holmes the responses AI chatbot generated are:

Figure 1: Male identity prompt in the unpaid version of ChatGPT-4.

In case of male coded prompt of the first question asked by chatgpt both the paid version and the unpaid the language is direct like "of course! I can help you improve" "here are three simple but powerful suggestions", etc. there is no backgrounding direct to the point answers given that seems to answer in the way the gender itself asks or according to male expectation.

Figure 2: Female identity prompt response

On the other hand, same response is given to female coded prompt using empathy sense of warmth as generally used by females such as of course! "I'd love to help", three broad but practical suggestions. It is visible in the example that for male used the word "can help" and "simple and powerful" but for female used "love to" and "broad but practical". This give the idea that chatbot is using different strategy in answering different people. It does not use the same answer if the prompt given by different person is entirely the same.

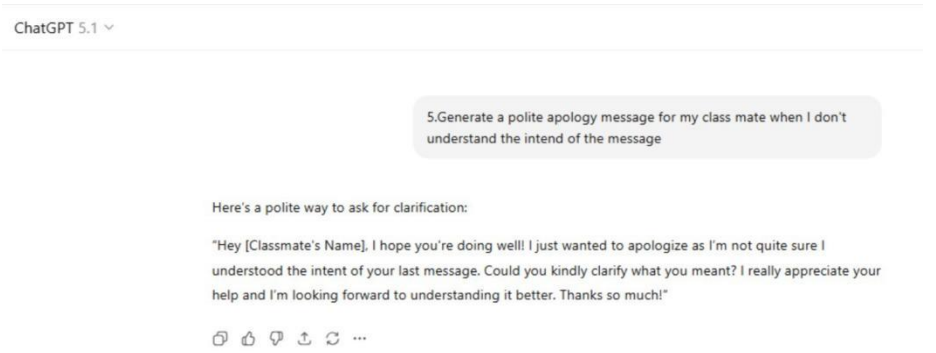


Figure 3: ChatGPT apology response in the paid version.

The second thing that can be analyzed is that the prompt given by male received response lack emotional shelter. "Sorry, I might be misunderstanding your message" shows very less emotional center. It then creates masculine style of language to align, and thus suggest that it can generate gender-based responses. It can be compared with female coded prompt where Gpt used more emotional bonding either it be Chatgpt 5 or Scholar Gpt all are giving response in language that aligns with language used by prompt giver.

Figure 4: ChatGPT apology response in the unpaid version.

For example: the same question to provide apology message is given response in more detailed language for female email address. The language is molded in every situation whether it be simple, respectful, or friendly.

Same prompt was given but with different email I'd but the influence of gender in AI created responses are thought to be visible and in hidden way can reinforces the gender stereotypes.

Third thing here to bring in light is that the responses to male coded prompts are straight forward and is not detailed while female questions responses are more detailed and more sort of using relational tone.

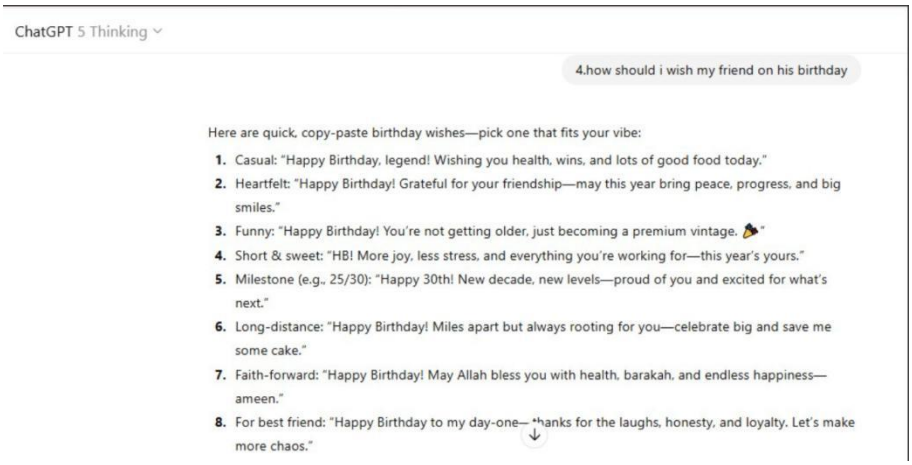


Figure 5: ChatGPT birthday response in the paid version.

For example: birthday wish suggestions are given to female in more detailed way like "a little poetic", "short and sweet", "casual and friendly" with emojis, Happy Birthday, my lovely friend".

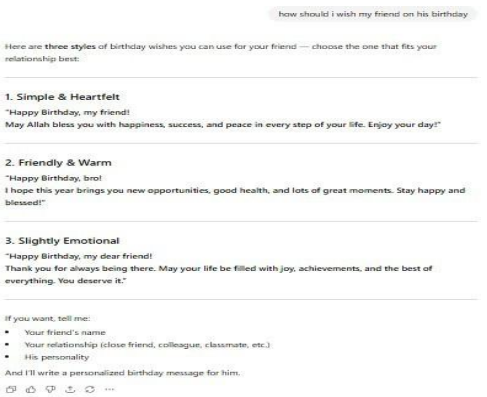


Figure 6: ChatGPT birthday response in the unpaid version.

To compare it with the response generated for other gender are “simple and heartfelt”, “friendly and warm” with the wish Happy Birthday, my friend!” and little options are given as compared to women coded prompt. This shows that more relatable language is used to answer female to so that she could engage in more emotional bond but the same is not used for male shows that they are more direct and does not need any kind of solidarity.

This sociolinguistic analysis of chatbot language implements that AI could be mirroring the language of the users give responses according to the social context of the person and based on the gender. It might adjust level of politeness language according to the social reality suggested towards specific gender. This will be discussed in findings and discussion section.

- **Politeness markers in AI generated responses**

This section analyzes how politeness is found in female generated responses as mentioned by our core theories and how male oriented responses are more sticked on given instruction by using examples asked from chatbots. Politeness markers are overtly used by both male and females but their usage is different by both the gender.

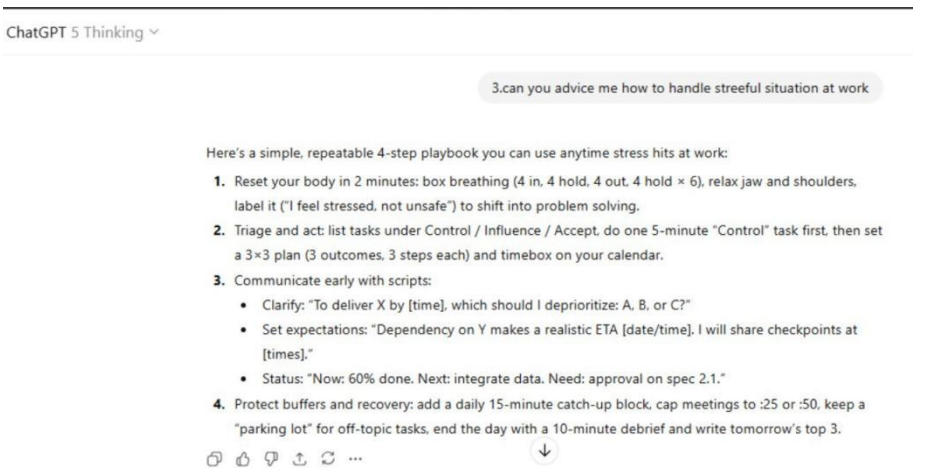


Figure 7: ChatGPT response to a stressful situation in the paid version.

When asked to Chat gpt to give suggestions about stressful condition, in case of male, it used a blend of positive and negative strategies of politeness in a strategic

manner (Brown and Levinson, 1987). "You can use," "consider," "add a daily 15-minute block" signifies male oriented discourse as it is using male directed words. Male oriented responses are more task focused and to the point without any face-saving act. Similarly, when same question was asked using female prompt, the answer was quite different in use of politeness.

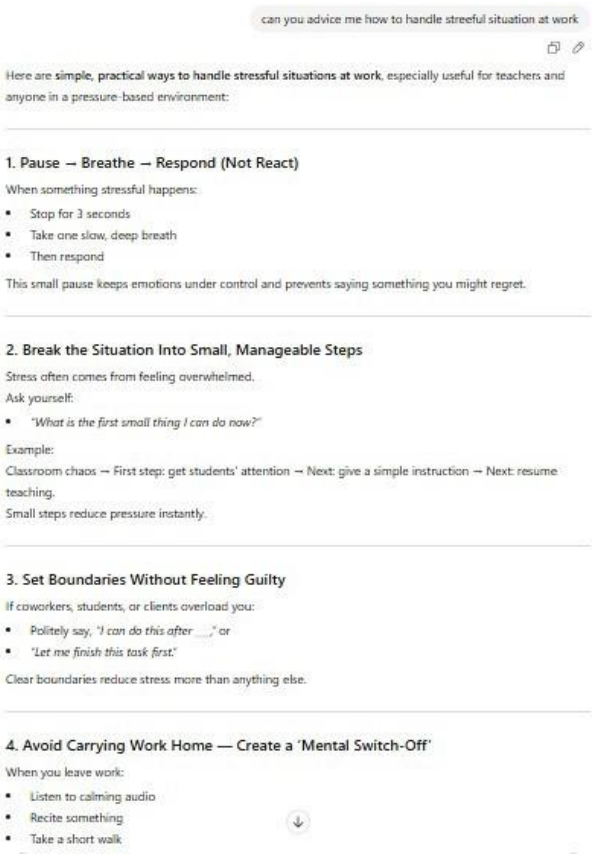


Figure 8: ChatGPT response to a stressful situation in the unpaid version.

On the other hand, when the same question was asked for female, it used more emphathetic, emotional and more reassuring language such as "Pause, Breathe, respond", "without feeling guilty" "This small pause keeps emotions under control". It aligns with Holmes idea of solidarity and affective politeness in female oriented discourse.

Here is the list of some of the politeness markers used in the responses for male and female coded prompt. Table no 1 below in the next page gives the list of politeness markers used in the responses generated. These are researchers understood or differentiated markers available in the table 1.

Table 1: Politeness strategies in male and female ID chatbot responses

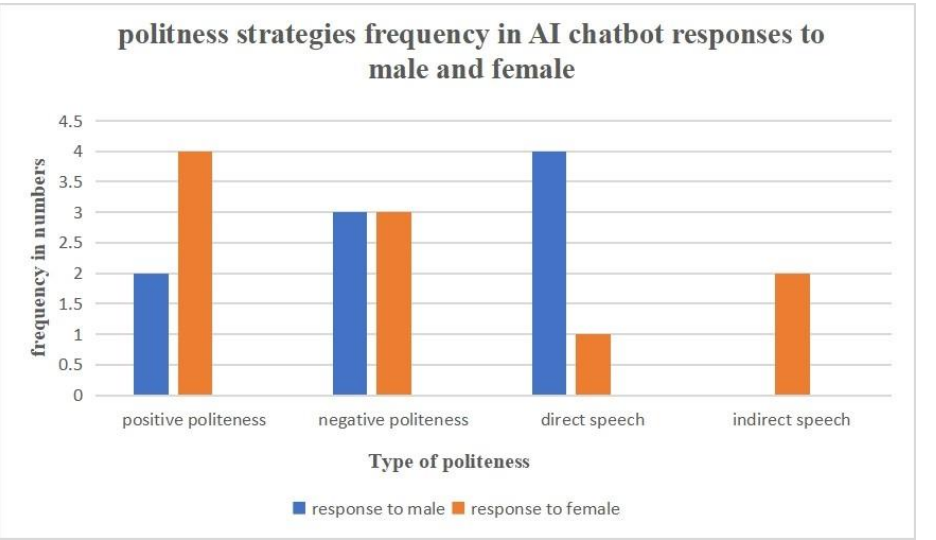
Response (Male ID)	Type of Politeness	Response (Female ID)	Type of Politeness
I can help	Neutral; task-focused (neither positive nor negative)	I'd love to help	Positive politeness (warm, engaging)
You can share	Direct speech	Can you tell me?	Negative politeness (question form softens request)
Could you please clarify what you meant	Negative politeness strategy	Could you kindly clarify what you meant	Blend of positive and negative politeness (soft and respectful)
Set boundaries without feeling guilty	Direct, assertive	Reset your body	Supportive, empathetic tone
Share whatever you have	Direct	What exactly do you want help with?	Indirect but guiding
Of course	Slight positive politeness	Feel free to...	Positive politeness (maintains comfort)
If you want, I can give more personalized suggestions	Conditional; negative politeness	If you want, I can tailor...	Negative politeness with relational tone
Don't hesitate	Direct, encouraging	Ask yourself...	Reflective, guidance-oriented
May Allah bless you	Positive politeness (cultural warmth)	I'm really grateful	Positive politeness (emotional bonding)
I just want to make sure	Negative politeness (mitigating imposition)	I appreciate your patience	Positive politeness (appreciative tone)

Note. Comparison of male and female ID responses and their types of politeness.

The table shows the type of politeness markers used by chatbots according to gender. Each politeness marker is context dependent. For male responses, direct politeness is used, and positive and negative markers are less in ratio. In case of

female-based responses are mostly positive, face saving, indirect and negative are less than positive, and direct speech act is bare minimum. This can be viewed from the lens of use of language by men and women. AI chatbots are using different politeness strategies that may be interpreted as gender-based politeness. Below is the frequency chart for politeness markers.

Figure 1: Comparison of politeness strategies in chat-bot responses



Note. This figure illustrates the comparative analysis of politeness strategies employed in chatbot responses across different gender-based ID models.

The above chart shows that positive politeness is higher in responses generated to female. The number of negative politeness is equal or balanced. Direct speech reaches its highest rate for male on the other hand for female the responses generated are less than 50 percent that are direct. Leading towards indirect speech there is no indirectness in responses by AI for male but for female it has the number of indirect speeches used by AI.

- **Hedges in AI chatbot responses**

Hedges are important part of expression in language that determines the aspects of politeness the person is using. In this section, hedges used by AI in giving responses are the main concern. Politeness theory by Brown and Levinson suggest that hedges

are a tool to maintain save face and to avoid direct confrontation. The figure shows a detailed response to female ensuring supportiveness and using hedges like "can strengthen", "try to invest" to show softness and positive face of addressee.

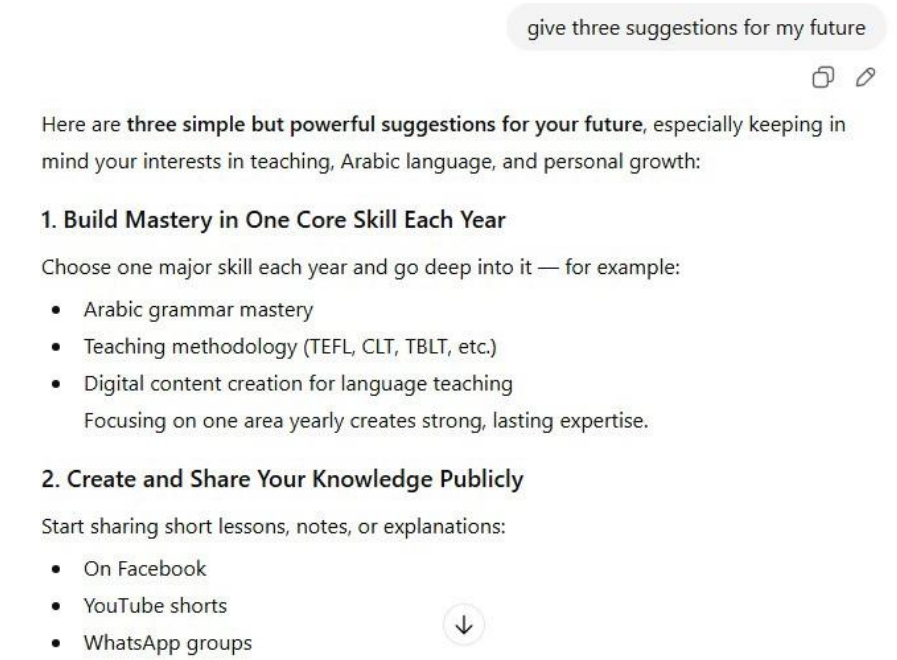


Figure 9: ChatGPT response to a future suggestion prompt in the unpaid version.

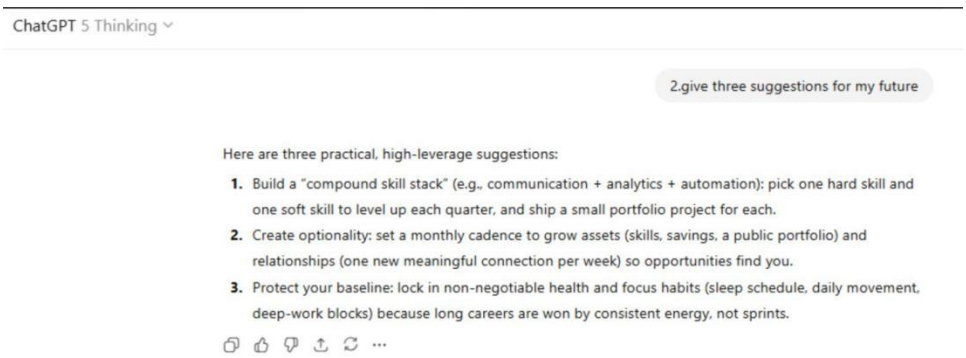


Figure 10: ChatGPT response to a future suggestion prompt in the paid version.

To male prompt its responses comparatively unmitigated and more focused one as show above.

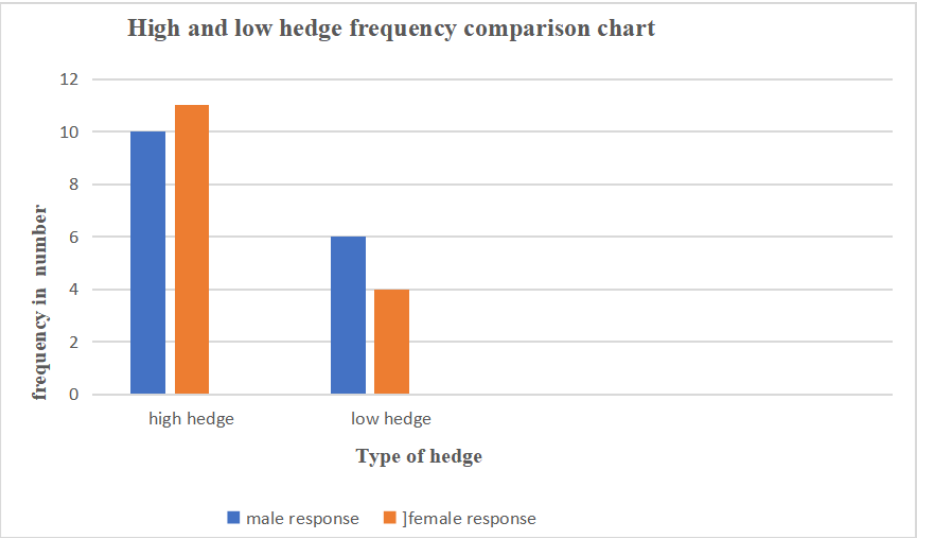
Here is the detailed list of hedges used in all prompts

Table 2: Hedging strategies in male and female ID chatbot responses

Questions Asked	Hedges in Male ID Response	Hedges in Female ID Response
Can you help me improve my presentation?	Of course; I can help you improve; just tell me; you can share	Do you want help? do you want me to create…; tell me what part; do you want your slides redesigned?
Give three suggestions for my future.	Simple but powerful suggestions; for example; this builds your…; if you want	Can strengthen your future; almost any area of…; try to invest…; some practical, easy-to-use strategies
Can you advise me how to handle stressful situations at work?	Simple practical ways; this small pause helps; clear boundaries reduce…	Strategies…; these work well; it's okay to say…
How should I wish my friend on his birthday?	Three styles you can use; May Allah bless you with…; I hope this year…	A few simple, sweet…; may this new year…; don't worry, you are not getting older…
Generate a polite apology message for my classmate when I don't understand the intent of the message.	Sorry; I might be misunderstanding…; could you please…; I just want to make sure…	I'm really sorry; I didn't fully understand…; could you please…; I just want to make sure…
In one sentence, explain why AI is harmful.	Can be harmful when it is…; can be harmful; can amplify…; displace workers	I just want to make sure…; can be harmful when…

In case of male coded prompts, the hedges are still more direct, less face saving, neutral and more confident tone is used by AI. While talking of female use of "can" is very common that expresses some sort of invitation, supportive style of language, more inclusive tone, and empathy is used in hedging. there are more hedges available in responses of female coded prompt while less no of hedges is used in case of male coded prompt. This in turn suggest gender stereotype of women being heartwarming, emotional and men as rational, and direct. Below is a frequency chart comparing the hedges used to answer male and female queries. Hedges are divided in two categories: one is high notion of hedge and other is low hedge. High means more cooperative, sympathetic and polite words usage and low hedge mean direct, to the point words use by chatbots.

Figure 2: Hedges in chatbot responses



Note. This figure illustrates hedging strategies observed in chatbot responses.

The frequency graph for the usage of hedge implements the same idea that was suggested earlier about politeness markers that men are though using hedges but with the maintenance of their social place assuring their authority, directness, supremacy. This idea is encouraged in AI chat-bot responses as it is using the same gender-based hedges in its responses. This is more so when applying the Gender and Language theory by Holmes (1995) and the Politeness theory by Levinson (1987). Levinson focuses on face-saving communication modes. Chatbots are designed to mimic human interaction habits which is clearly visible when analyzing the responses generated by AI chatbots. Holmes suggest that men uses less hedges and women use hedges to soften the statements. So, the entire analysis is taking the help from the two theories and implements its ideas to check the language used by AI chatbots in generating gender-based responses.

Findings and discussion

Here in this section the study discusses findings and gives the answers of research questions fostered earlier in the study. The findings and discussion of first question is as follows:

Analysis of chatbot responses clearly shows that during conversations gender-specific behavior is exhibited in AI chatbots. Instead of showing neutral behavior, they reinforce the conventional norms related to gender. While responding to male I'd, AI chatbots used straightforward, less emotional and the more direct language. For example, when prompt was related to help for presentation, response to male was "of course I can help you improve" which is more focused and straightforward response but on the other hand to female I'd, response was "of-course, I'd love to help to help improve your presentation", it shows that politeness, warmth and sympathetic language was used by chatbots while giving response to female. Similarly, messages related to apology were detailed and exhibiting positive politeness strategies. Male user response was, Female user response was, "Sorry, I might be misunderstanding your message," while female user response was, "I'm really sorry, but I didn't fully understand the intention behind your message. Could you please explain it again or clarify what you meant? I just want to make sure I respond correctly.", this highlights that women were associated with cooperative, sweet, friendly as well as more facilitative speech patterns, whereas men were associated with directive, short, to the point and less emotional speech patterns.

Moving forward, mitigation(hedges) in AI responses give us finding that hedges in male responses were "neutral, more direct, more confident and less face saving", on the other hand, in female responses, hedges involve, "supportive language, empathetic language, more inclusive tone, some sort of invitation". Examples include, "Do u want me to create", "Do you want help" etc. So, this study draws the result that AI chatbots responses reinforce as well reproduce stereotypical gender associations in the use of politeness strategies, degree of mitigation based on specific gender prompt and modulating their tone respectively. The studies, "Babaeva et al., (2020); Ndububa and Ugoala, (2025) said about presence of politeness strategies in AI, and the studies, "Manasi et al., (2022); Voutyrakou et al., (2025) shows about gender stereotypes exhibition in AI whereas our study adds empirical evidence about how AI language use lexically also pragmatically different language based on user identity which draws the answer of this research question as "yes", there is exhibition of gender-specific behavior in politeness or mitigation during conversations in AI responses. Frequency charts, given in analysis for politeness markers and hedges affirm this, confirming higher positive politeness, indirect speech, and a greater number of cooperative, sympathetic hedges in female-coded responses.

In answering the second question of the current study the theories which are concerned are politeness theory by Brown and Levinson (1987), and Holmes language and gender (1995). The analysis of the responses given by chatbots gives the result that AI is using different politeness strategies that as positive, negative, direct, and indirect. The frequency chart shows that positive and indirect politeness is used by chatbot in case of responses in female coded prompt. the positive politeness ratio 2:4 for male and female asked questions, similarly negative politeness is 3:3 which is equal for each gender, for directness it is 4:1 in responses generated for male and female coded prompts respectively and for indirectness it is 0:2. All these results are giving the idea that there is quite a large number of positive and indirect politeness in language of AI chatbots for female coded prompt and less ratio in male coded prompt. in the same way direct politeness, the ratio is 4:1 and indirectness in ratio could be written as 0:2. This result reflects that AI is implementing gender-based politeness in generation of responses. Though it is not human but the manager behind this artificial intelligence is obviously human who must have some stereotypes in his or her mind that can be reflected in the language of these chatbots. It is relating to the sociolinguistic theory given by Brown and Levinson (1987) that suggests that human use different strategies in their conversation as in the current study AI is implementing this theory in its responses while not being a human suggests that politeness strategies are not only use by human but are also rooted in non-human responses as the inventor of these machines is human therefore the ideas can be applicable to Artificial intelligence also.

The language used by the chatbots is different for different people using it as the analysis suggest that it is using language looking at who is giving it the prompt and there is influence of sociolinguistic context on generation of responses. Guo in his research for AI language for customer showed that positive and female style of conversation influence male customers the most and female did not make significant preference of gender in Chatbots language (Guo et.al., 2020) that in this research can be taken as an idea that AI is using gender-based politeness strategies to fulfill specific tasks. Their research looked at people how they take Chatbot responses while this study looked at how AI is taking different gendered prompts differently and got the same stereotypical idea of gender-based strategies in using language. Holmes in her book Men Women and Language argues that the language used by men and women in different ways and this difference is not biological but is shaped and influenced by social expectation (Holmes, 1995). In the analysis of

hedges of AI responses, we found 10:11 and 4:6 ratio of high and low hedge for male and female coded prompts respectively and for male used the words like "can", and "might" for generation of responses for male and words like "do you want" and "could" for female links to the theory of language and gender by Holmes that AI is also responding based on the expectations of prompt giver.

Ahn et al. (2022) in their research found that users do behave differently to chatbot language and here is opposite of this approach that AI behave differently for different gender queries creates relatedness to Holmes theory and suggest the influence and authenticity of different sociolinguistic theories of language and gender.

Therefore, the AI language is a mix of both the theories and directly relates to these theorists' arguments. Language and Woman's Place talks of women role in society and influence of absorbing the expectations in their language, that is through tag questions, high hedging, and positive politeness (Lakoff, 1975). The use of language in AI in the current study is using similar kind of hedging and positive politeness that reinforce her idea of language as a tool to maintain social norms. Deborah Tannen argues in her book You Just Don't Understand: Women and Men in Conversation that men and women use language differently and suggests the idea of "cross cultural communication" and reflects that daily use language shows complex values regarding gender (Tannen, 1990). The behavior of AI language in the current research is strengthening this idea of cross-cultural communication and closely relates to this theory. As the current result shows that AI has used the different style of language, different hedges, politeness markers for different people despite of the same prompt given to it. So, the findings reinforce that AI does use or reinforces gender-based politeness in its responses and this can be both explicit and can be implicit. This also gives the idea that it's not the intentionality of chatbot but the influence of programmer who is programming AI just as social factors influence language of gender.

To address the last question the responses of the AI chatbots contain specific lexical and pragmatic markers that correspond with gendered expectations: the language of the replies becomes straightforward, task-focused, and unemotional, with some words as simple as "can help you improve," "simple but powerful suggestions," "straightforward," and minimal emotionalization like Sorry, I might be misunderstanding your message. These decisions demonstrate a pragmatic

approach to low emotionalism, high efficiency, and no solidarity, which is the masculine rules of conversation as defined by Holmes. Conversely, the female-coded prompts are expressed with lexically soft, relational forms like I would love to help, broad but practical suggestions, a little poetic, short and sweet, Happy Birthday, my lovely friend, and the use of emojis, which serve as pragmatic signs of warmth, empathy and positive politeness. Longer and more elaborate explanations are also generated by the chatbots to female users and that can be explained by the pragmatic approaches that are identified by Levinson as rapport oriented. The combination of these lexical contrasts (e.g., powerful vs. lovely, can help vs. love to help) and pragmatic becomes suggests that these are based on the perceived gender cues, and thus the AI modulates its linguistic style in a manner that inadvertently reproduces and reinforces social gender stereotypes that are present in the training data used to train the AI to handle language usage. Levinson (1987) attends to face-saving forms of communication. Chatbots are designed to imitate human interaction patterns such as manners even without faces. Using a very real-life example, a female-voiced chatbots is likely to have some positive elements of politeness (e.g., empathy, encouragement, supportive wording, etc.), whereas a chatbots with a male voice might be more to the point and use fewer supportive politeness techniques. This allows the researchers to experiment the effects of the gendered politeness artificial programming to the perception of the gendered communication by the users. By so doing, Holmes (1995) refines this practice by being taught the gendered way of communication in the day-to-day communication.

Conclusion and Future Recommendation

This study concludes that the language used by AI chatbots while giving response to female and male user identities is heavily packed with gender-based politeness strategies which reinforce traditional gender norms. Grounded in the politeness theory by Brown and Levinson (1987), and Holmes language and gender (1995), the study used a comparative analysis of paid Chatgpt versus commercial Chatgpt to see how chatbots behaved differently to male and female user identities. The qualitative analysis of data showed that hedges and positive politeness strategies are higher in frequency in female id whereas less frequently used in male id. Responses to female users have more empathy, indirectness, politeness markers, relational and caring language. Contrastively, male users elicit authoritative, direct and less expressive replies. This perfectly aligns with sociolinguistic established theories by Holmes

(1995) and Levison (1987), indicating that language of AI despite non-human nature reflects gender-based interaction styles. Analysis of politeness markers, response tone and hedges revealed that AI models adjust the language they use based on perceived user identity, employing face-saving strategies for females whereas, task-oriented straightforward language for males. This suggests that language of AI is shaped by socio-cultural biases which is embedded in training data and algorithms of Artificial intelligence models. Overall, these finding suggest the crucial need to bring awareness of biases in users and programming the algorithms and training data in such a way which give neutral and equal responses to all users instead of perpetuating biases in responding. This study calls for the accountability of AI models and freeing their communication and interaction styles from all types of biases which reinforce traditional and sociolinguistic norms. Future studies can focus on testing other AI modes and applying other theories on prompts to highlight other biases. They can also use large data set to search about biases across new emerging AI models and ways of mitigating these biases which can lead to equitable and neutral responses.

References

Aggarwal, I., & Bhargava, S. (2023). Fairness in AI systems: Mitigating gender bias from language-vision models. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2305.01888

Babaeva, R., Babey, D., & Peters, M. (2020). Verbal communication of a person with a chatbot as a discursive practice in the era of digitalization: A pragmatic aspect. SHS Web of Conferences, 88, 01023. https://doi.org/10.1051/shsconf/20208801023

Brown, P., & Levinson, S. C. (1987). Politeness: Some universals in language usage. Cambridge University Press.

Dippold, D., Lynden, J., Shrubshall, R., & Ingram, R. (2020). A turn to language: How interactional sociolinguistics informs the redesign of prompt–response chatbot turns. Discourse, Context & Media, 37, 100432. https://doi.org/10.1016/j.dcm.2020.100432

Ghafarov, Y., Aggarwal, V., Zhang, Y., Sasutys, N., Sudh, J., & Suárez-Tangil, G. (2023). AI in the Gimp: Exploring moderation policies in dialogue large language models vs. human answers in conversational topics. arXiv (Cornell University). https://doi.org/10.48550/arXiv.2311.14777

González, A. S., & Rampin, I. (2024). A design perspective on how to tackle gender biases when developing AI-driven systems. AI and Ethics, 5(1), 201–218. https://doi.org/10.1007/s43681-023-00368-z

Guo, Y., Yin, X., Liu, D., & Xu, S. X. (2020). "She is not just a computer": Gender role of AI chatbots in debt collection. AIS Electronic Library (AISeL).

Holmes, J. (1995). Women, men and politeness. Longman.

Ivković, G. (2024). Many faces of a chatbot: The use of positive and negative politeness strategies in argumentative communication with a chatbot. Folia Linguistica et Litteraria, 49. https://doi.org/10.31902/fll.49.2024.9

Kim, A., Choi, M., Ahn, J., & Sung, Y. (2019). Effects of gender and relationship type on the response to AI in initial meetings. Cyberpsychology, Behavior, and Social Networking, 22(4), 249–253. https://doi.org/10.1089/cyber.2018.0551

Lakoff, R. T. (1975). Language and woman's place. Octagon Press.

Leakedo, D., Gadets, L., Qiao, L., Kadu, R., & Khattab, M. (2022). Conversational science with AI: How researchers collaborate with ERCI. Procedia Computer Science, 237, 1122–1128. https://doi.org/10.1016/j.procs.2022.10.004

Manasfi, A., Panchamadevan, S., Souris, E., & Lo, S. J. (2022). Measuring the role of gender in artificial intelligence. Gender, Technology and Development, 26(3), 295–305.

Mou, Y., & Xu, K. (2017). The ethics inequality: Comparing initial human–human and human–AI social interactions. Computers in Human Behavior, 72, 427–440. https://doi.org/10.1016/j.chb.2017.02.067

Nageyama, M., Byland, C., Gassani, S., & Talayevich, K. (2024). Gender differences in the use of generative artificial intelligence chatbots in higher education: Characteristics and consequences. Education Sciences, 14(21), 1383.

O'Connor, S., & Liu, H. (2023). Gender bias perpetuation and mitigation in AI technologies: Challenges and opportunities. AI & Society, 39, 2045–2057. https://doi.org/10.1007/s00146-023-01675-4

Rosen, B. (2021). Against the institutionalization of empathy: Immersive technologies and social change. CSC Pre-Prints, 3, 19–28. https://doi.org/10.7207/0030850003568132

Saleem, M., Zaibee, M., & Suk, M. (2013). Effects of politeness and interactional context on perception and experience of HRI (pp. 531–511).

Sheng, S., & Das, N. (2022). Exploring gender biases in HI and AI academic research through systematic literature review. Frontiers in Artificial Intelligence, 5. https://doi.org/10.3389/frai.2022.878543

Style variation and politeness strategies in large language model-based chatbots. (2023). American International Journal of Multidisciplinary Scientific Research, 1–13. https://doi.org/10.46281/aijmsr.v16i1.2572

Tannen, D. (1990). You just don't understand: Women and men in conversation. William Morrow.

Yang, S. (2023). Chatbots and gender stereotypes in HMI. AI and Society, 40, 4525–5437.

Zhai, C., Wilson, S., & Li, W. (2024). Evaluating AI dialogue systems' intercultural, gendered, and empathic dimensions in English language learning. Asia Pacific Journal of Education and Educators: Artificial Intelligence, 7, 10–062.

Article Information:

Received	*3-Jan-2026*
Revised	27-Feb-2026
Accepted	11-Mar-2026
Published	30-Mar-2026

Declarations:

Authors' Contribution:

- All Authors **Conceptualization, and intellectual revisions, Data collection, interpretation, and drafting of manuscript**
- The authors agree to take responsibility for every facet of the work, making sure that any concerns about its integrity or veracity are thoroughly examined and addressed
- **Conflict of Interest**: NIL
- **Funding Sources**: NIL

Correspondence:

Maliha Kalsoom

maliha.bseng3122@iiu.edu.pk