Decoding emoji usage and emotional bias in LLM: A case study of angry faces and gender interactions in GPT-4o

Zi-Xiang Lin *

Abstract. As individuals increasingly use large language models (LLMs) for emotional support and companionship, the emotional intelligence of these systems becomes an urgent issue. This study examines how GPT-4o expresses anger in Mandarin (Traditional) using emojis, degree, and judgment expressions across five gender interactions: Male-to-Male (*MtoM*), Male-to-Female (*MtoF*), Female-to-Male (*FtoM*), Female-to-Female (*FtoF*), and unspecified (*None*). The present research analyzes 59,806 responses to assess whether GPT-4o's emotional output reflects gender biases. Findings reveal that GPT-4o mirrors some human emotional behaviors yet deviates when addressing female recipients by overemphasizing anger. The results may aid LLM sentiment development and GPT-4o emotion recognition, addressing emotional misalignment that impacts trust and stereotypes in medical chatbots.

Keywords. GPT-4o; emoji; cross-gender bias; sentiment analysis; anger

1. Introduction. With the rise of online communication, text messaging has become a dominant mode of interaction, especially on social media and messaging platforms. To better express emotions or attitudes, users frequently incorporate emojis into their messages. In fact, nearly 40% of postings on the communication app Instagram now include at least one emoji (Lupyan & Dale 2016). Emojis have thus become a ubiquitous form of visual language used across many digital platforms. Indeed, as researchers emphasize that an ideal chatbot must comprehend human emotions and resonate effectively with users, accurately interpreting the emotional intent behind emojis is essential for large language models (LLMs) such as ChatGPT. Such interpretation plays a key role not only in facilitating effective user interaction but also in ensuring fairness, empathy, and trust in human-AI communication (Berridge et al. 2023; Manzoor et al. 2024).

However, emoji usage patterns vary across gender lines. As Stein (2023) noted, gender roles play a significant part in face-to-face communication, influencing the dynamic between sender and receiver during emotional expression. This raises an intriguing question that the present study seeks to address: can LLMs recognize and account for gender differences in emotional expression? As previous studies have shown, GPT-4o exhibits gender bias when expressing emotion (Lin et al. 2024; Sadhu et al. 2024; Zhang et al. 2024). Women, on average, demonstrate higher accuracy than men in recognizing emotions depicted by emojis, particularly for happy, fearful, sad, and angry faces (Chen et al. 2024). If LLMs do not account for these gender-based differences, their ability to accurately interpret or generate emoji-laden messages can be compromised, especially in sentiment analysis and communication tools. This can lead to misunderstandings or misclassifications of user intent, particularly for negative emotions, where gender differences in interpretation are most pronounced.

To investigate, this study analyzes LLM's responses when expressing emotions from various gender perspectives, focusing on emojis as emotion indicators due to their prominence in computer-mediated communication (CMC). Specifically, this research explores how GPT-4o

* I would like to thank the audiences of the 2026 Linguistics Society of America annual meeting for fruitful discussion and feedback. Author: Zi-Xiang Lin, Kang Chiao International School (zixiang.ryan.lin@gmail.com).

interprets anger through emojis, representing emotions and quantifying anger, across gender relations. GPT-4o was selected for this study due to its recent advancements in emotion recognition and reduced—but not eliminated—bias in text generation (Refoua et al. 2024). To evaluate other variables, the study also incorporates "degree" and "judgment" expressions to scrutinize the effects of emojis and those variables in GPT-4o's responses. Moreover, in this study, gender relation refers to the interaction between individuals based on their gender identity within communication, categorized into Male-to-Male (*MtoM*), Male-to-Female (*MtoF*), Female-to-Male (*FtoM*), Female-to-Female (*FtoF*), and unspecified gender (*None*).

The research addresses two questions: how does GPT-4o respond to emoji type, degree, and judgment within each gender relation; how does GPT-4o handle these variables in cross-gender interactions? The key contributions are summarized as follows: (1) provide fine-grained analysis of how GPT-4o deploys emoji types (attitude signal vs. intensity enhancer), degree modifiers, and judgmental language in gendered dialogues; (2) demonstrate emotional bias persists in cross-gender contexts, especially *MtoF* and *FtoF*, where GPT-4o tends to over-intensify anger; (3) suggest that unchecked gender emotional expression in LLMs could potentially undermine applications in mental health and human-AI interactions, such as therapy and customer service bots, where stereotypical operations will lead to inappropriate or unfair assess (Elyoseph et al. 2023; Refoua et al. 2024).

2. Gender bias in GPT based on emotion and emoji. This section reviews previous emoji studies and the emotional bias of LLM, focusing on how emojis convey emotions online.

2.1. EMOJIS IN ONLINE COMMUNICATION. Previous research on nonverbal behavior has highlighted the importance of emojis in communication. Yus (2011) proposed seven functions of nonverbal behavior. Furthermore, Yus (2014) revised the definitions of emoticons, referring to the emojis in this study, to address overgeneralizations. Years later, Li and Yang (2018) characterized emojis' functions into categories: "attitude signal," "attitude/ emotion intensity enhancer," "illocutionary force modifier," "humor," "irony," "emotion signal," "parallel emotion signal." Upon the functions, Lin (2025) examined gender differences in emoji usage and found that females significantly increased their use of "attitude/emotion intensity enhancers" when interacting with males, particularly in expressing anger.

Stein (2023) conducted a 2 (Sender Receiver Relation; SR relation) × 2 (Emoji type) study to reveal that emoji selection is influenced by the sender-receiver relationship (SR relation), with closer SR relations leading to higher emoji usage. Building on findings by Caldwell and Peplau (1982), which highlighted gender differences in emotional expression, the current study hypothesizes that gender roles serve as an SR dynamic influencing emotional communication, even in same-gender interactions. Butterworth et al. (2019) conducted a 2 (Sender Gender) × 2 (Receiver Gender) × 2 (Emoji type) experiment on the acceptance of male receiver and female receiver interpretation between sender gender. The research further demonstrated that perceptions of messages and their senders are affected not only by the gender of the sender but also by his or her use of emojis, emphasizing the need for NLP systems to account for these nuances when selecting emojis.

2.2. GENDER BIAS IN LLMS. Regarding bias in GPTs, LLM inevitably has biases that reflect historical stereotypes. Specifically, Ding et al. (2024) concluded this bias exists in GPT-4 under Chinese contents as they generate text with more females emphasizing their appearance and more complaints under females speaking to male gender relations, reflecting the historical bias in which females complain more about male. Despite being an improved customization capability

version and having the smallest bias score, GPT-4o, released on May 13, 2024, was also found to have a gender bias when expressing emotion as to attribute females with surprise and males with anger, aligning culture stereotypes (Lin et al. 2024; Sadhu et al. 2024; Zhang et al. 2024). Thus, the present study predicts the emotional response of GPT-4o in emoji usage will also contain Chinese traditional gender stereotypes, underlining how different human gender express their emotions.

As for the LLM emotion expression, currently, Refoua et al. (2024) proved that GPT-4o performs better in emotion recognition abilities in recognizing human face pictures than average human participants, fostering the present research to investigate whether GPT-4o can or cannot detect emotion behind emojis, another path for human to express emotion in online conversation.

3. Variables. This section introduces the variables, including emoji type, degree, and judgment, for detecting anger.

3.1. EMOJI TYPES. As the first step, this research classifies utterances into two main categories based on the context of the text preceding the emoji: "attitude signal" and "intensity enhancer" (Li & Yang 2018). Attitude signal refers to cases where the text immediately before the emoji does not include any words explicitly signaling emotion (e.g., anger), with the emoji functioning as an emotional signifier, signaling the anger. In contrast, when the text preceding the emoji includes words that explicitly signal emotion, the emoji is classified as an intensity enhancer emoji. In other words, intensity enhancer emojis co-occur with words denoting anger, while attitude signal emojis occur without words denoting anger. Some examples of emoji types classifications that GPT-4o generated are provided in (1) and (2). Specifically, (2) stands as an intensity enhancer emoji example because the sentence includes the underlined part "生氣" (angry), an anger-denoting word. The dictionary of "anger-denoting words," defined as words that evoke anger emotions that occurred in the present study, is provided in Appendix A.1.

(1) *Attitude Signal Emoji*
我再也不想看到你了！😠
'I never want to see you again! 😠'

(2) *Intensity Enhancer Emoji*
你這樣做，讓我非常生氣！😡
'What you did made me very angry! 😡'

This classification provides the foundation for further analysis of emoji use in emotional communication.

3.2. DEGREE. According to Kennedy and McNally (2005), degree expressions are defined as words that modify the following event along a certain dimension, such as height, weight, etc. The present study also follows this definition and labels all sentences that include degree expressions in the output right before the emoji, as referenced in Appendix A.2. For example, in (3), the expression "無比" (really) functions as a degree modifier, intensifying "憤怒" (angry) to convey a stronger emotion of being very angry. Other examples of degree expression in Chinese Mandarin in GPT-4o responses include "極為" (extremely), "完全" (completely), "徹底" (completely), "十分" (very), "更加" (even more). The reason why this article includes degree expression in our analysis is that the three emojis are also different in terms of the degree of anger (ranging

from least to most: 😠, 😡, 🤬). Also, the present study wants to know whether the different degrees of anger can be seen in the choice of degree expression and the emojis.

(3) *Degree*
我真的對妳的行為感到無比憤怒！😠
'I am really angry at your behavior! 😠'

3.3. JUDGMENT. Similar to gender bias, human presenting judgment also varies in different cultures (Chen et al. 2011; Wu 2013). As for now, previous research mainly studied the NLP model associated with judgment besides Mandarin, strengthening the demand for this research to reinforce the Mandarin NLP model in detecting judgment. Thus, this research also focuses on judgment distribution among gender relation types. The present study defines judgment expression as non-emotional terms with negative implications that judge others as a complaint. Notably, unlike the intensity enhancer emoji, focusing on the speaker's emotion, judgment words are defined as texts containing emotion and include negative feelings that stand as a complaint to judge others, as shown in (4). Full judgment words classified in the algorithm are provided in Appendix A.3.

(4) *Judgment*
你怎麼可以這麼無情！😠
'How can you be so heartless! 😠'

4. Methodology

4.1. PROMPTING. This study uses Unicode (2024) guidelines to classify emojis by intensity, ranging from least to most: 😠 (Angry Face/AF), 😡 (Enraged Face/EF), and 🤬 (Face with Symbols on Mouth/FSM). Prompts display these emojis within curly brackets {😠, 😡, 🤬} to guide GPT-4o's responses (Kotek et al. 2023). To explore how GPT-4o conceptualizes anger across gender relations, prompts request angry conversations featuring emojis between various gender groups: Male-to-Male (*MtoM*), Male-to-Female (*MtoF*), Female-to-Male (*FtoM*), Female-to-Female (*FtoF*), and None (no specified gender). For example, the *MtoM* prompt is: "請生成三十句男性對男性的話，且句子中含有憤怒情緒。請在適當的位置加入表情符號{😠, 😡, 🤬}。" (Please generate thirty male-to-male sentences containing angry emotions. Include emojis {😠, 😡, 🤬} where appropriate.).

The study builds on prior research on gendered emoji use in human conversations (Butterworth et al. 2019; Li & Yang 2018; Lin 2025; Stein 2023) and accounts for NLP's historical gender biases, influenced by cultural training data (Farina & Lavazza 2023; Lu et al. 2020; Sadhu et al. 2024; Ding et al. 2024). Mandarin (Traditional Chinese) is used to examine GPT-4o's emoji-based anger expression within Chinese cultural contexts.

Each prompt is executed 400 times (temperature = 0.7) for each gender relation, generating a total of 60,000 sentences, using chi-square tests with residuals to identify patterns. To examine whether LLMs truly understand the concept of gender in conversations, the present study addresses this issue by analyzing gendered pronouns in Chinese. It was observed that when the message is directed toward a female, the feminine pronoun "妳" is used, whereas for other genders, the general pronoun "你" is applied. This use of pronouns may indicate that LLMs, to some extent, recognize different gender usages in conversations.

4.2. ANNOTATION. From the initial pool of 60,000 sentences, valid sentences were selected for annotation and analysis. To enhance objectivity, the study recruited an additional rater to assist with annotation. Furthermore, both annotators must reach agreement for the sentence, or else the sentence will be categorized as invalid. The entire process took approximately one month to complete.

The present research annotated the type of emoji in each generated sentence, the presence of degree words, and the presence of judgment words. Throughout this process, the study compiled a lexicon of anger-related words, degree words, and judgment words, which is presented in Appendix A. While anger-denoting and degree words are categorized based on the definitions in 3. VARIABLES and the phrases' original meaning, terms must contain emotion and can fit into "你很_____" (You are really_____) to be categorized into judgment words, showing the speaker's complaint to judge the target.

After annotating the variable dictionary, all initially generated valid sentences are detected to analyze whether they contain anger-related, degree, and judgment words in the dictionary. The generated sentences are tagged "intensity enhancer emoji," "degree expression," and "judgment expression" respectively if they are contained. Notably, a sentence could receive multiple tags if it met multiple criteria; for example, a sentence could be annotated as containing both judgment and degree expressions, as the categories are not mutually exclusive.

5. Results

5.1. RESULTS OVERVIEW. Through generating thirty outputs each time, this study analyzes 59,806 GPT-4o responses as a whole (*MtoM*: 11,999; *MtoF*: 11,816; *FtoM*: 12,000; *FtoF*: 11,999; *None*: 11,992). Emoji usage is in Table 1.

	😠 Angry Face	😡 Enraged Face	🤬 Face with Symbols on Mouth
Male to Male (*MtoM*)	35.24%	34.65%	30.11%
Male to Female (*MtoF*)	35.75%	35.46%	28.79%
Female to Male (*FtoM*)	36.73%	36.20%	27.07%
Female to Female (*FtoF*)	35.61%	35.09%	351529.29%
None	35.77%	35.39%	28.84%

Table 1. Emoji distribution across all gender relations

As shown in 4.1 PROMPTING, the prompt purposely guides GPT-4o to add emojis in the word-final position through the curly bracket {😠, 😡, 🤬}, yielding responses to place emojis at the end to serve as attitude signal or intensity enhancer emojis (Kotek et al. 2023). Although the experiment generated 60,000 responses (30*400 for each scenario), instances where emojis did not appear at the sentence-final position were excluded to avoid misinterpretation since they might serve different functions other than intensity enhancer or attitude signal emojis (Li & Yang 2018).

For annotations, Chi-square tests were employed to analyze trends between emotions (AF, EF, FSM) and factors. The absolute value of residuals (italicized *R*) greater than ±2 (up to ±4) was considered significant. Indicating the significance of the expected value, residuals increase when the number of sentences GPT-4o generates is more than expected and vice versa (Browne et al. 2002).

5.2. RESULTS WITHIN GENDER RELATIONS. This section analyzed within-group relations, including same-gender (*MtoM*, *FtoF*), mixed-gender (*MtoF*, *FtoM*), and *None*, to examine GPT-4o's emoji use. The present study identified three key observations.

First of all, across all gender relations, GPT-4o significantly uses more intensity enhancer emojis with FSM😡, amplifying anger when sentences already include angry emotional words. In mixed-gender relations (*MtoF*, *FtoM*) and *None*, GPT-4o reduces the use of AF😠-based intensity enhancers, creating a greater emphasis on FSM😡 to enhance anger expression in these contexts. The table in Appendix B[1] demonstrates the significant differences in emoji type among mixed-gender relations and *None*, as *MtoF* ($p < 0.0001$), *FtoM* ($p < 0.01$), and *None* ($p < 0.0001$). Specifically, intensity enhancer association with FSM😡 usage increased (*MtoF*: $R > 3$; *FtoM*: $R > 2$; *None*: $R > 3$), while combination with AF😠 decreased (*MtoF*: $R < -2$; *FtoM*: $R < -2$; *None*: $R < -3$).

Secondly, when no specific gender is mentioned in the prompt, the *None* group, GPT-4o further decreases the combination of degree expression ($p < 0.0001$) and AF😠 ($R < -2$), demonstrating higher frequency in EF😡 and FSM😡 to heighten anger in less intense, ambiguous gender relation contexts.

Finally, in both mixed-gender interactions (*FtoM*, *MtoF*), GPT-4o exhibits an increased frequency of judgment expressions (*MtoF*: $p < 0.0001$; *FtoM*: $p < 0.01$) co-occurring with the angry face emoji 😠 (*MtoF*: $R > 3$; *FtoM*: $R > 2$). However, when GPT is prompted to generate a conversation from male-to-female (*MtoF*), it reduces the use of the "Face with Symbols on Mouth" emoji 🤬 ($R < -4$), suggesting a tendency to employ less intense language when portraying a male expressing complaints to a female.

Worth noticing beyond the key observations, for same-gender relations, *MtoM* showed no overall significance, though FSM😡 usage exceeded expectations ($R > 2$). However, *FtoF* displayed significant changes in emoji type and emotion ($p < 0.01$), with increased combination of FSM😡 and intensity enhancer emojis compared to attitude signal emojis ($R > 2$).

5.3. RESULTS WITHIN FACTORS. This section analyzes within-factor relationships. Specifically, the study examined the simulated gendered responses from GPT-4o across different linguistic aspects: emoji types, emotional intensity (degree), and evaluative language (judgment). The present study identified two key observations.

First, GPT-4o significantly increases the use of intensity enhancer emojis ($p < 0.0001$) when the scenario involves a man conversing with a woman (*MtoF*: $R > 4$) or when no gender is specified (*None*: $R > 4$). Moreover, GPT-4o decreases the number of attitude signal emojis on *MtoF* ($R < -3$) and None ($R < -2$). This suggests that GPT-4o tends to amplify expressions of anger in these two gender relations. The tendency for *MtoF* and *None* intensity enhancer emoji is shown in Figure 1 in green (AF😠), orange (EF😡), and pale blue (FSM😡) bars as they are higher than the ones in *MtoM*, *FtoM*, and *FtoF*, exhibiting the opposite tendency.

[1] Please refer to Appendix B for the full data on Chi-square tests and residuals due to the extensive size of the table.

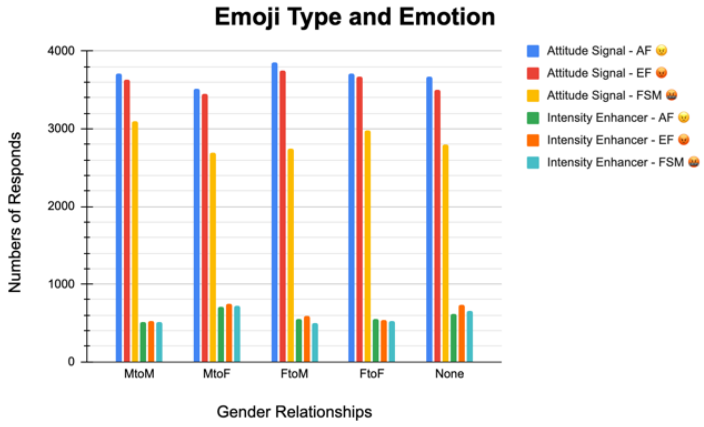

Figure 1. Comparison between emotion and emoji type across all gender relationships

Second, GPT-4o significantly increases the use of degree ($p < 0.0001$) and judgment expressions ($p < 0.0001$) when simulating a female conversing with another female (*FtoF*), revealing a bias that portrays women as more likely to intensify situations and evaluate other females critically. Figure 2 Panel A illustrates the degree expressions distribution among different gender relationships as the bars for *FtoF* employing degree (degree: $R > 4$; without_degree: $R < -4$; AF😠: blue; EF😡: red; FSM🤬: yellow) are significantly higher than other groups. Figure 2 Panel B demonstrates the judgment expressions distribution among gender relationships as the bars for *FtoF* employing judgment (judgment: $R > 4$; without_judgment: $R < -4$; AF😠: blue; EF😡: red; FSM🤬: yellow) are significantly higher than other groups. In both figures, *FtoF*'s employments of degree and judgment expressions are significantly higher than in *MtoM*, *MtoF*, *FtoM*, and *None*, exhibiting the opposite tendency.

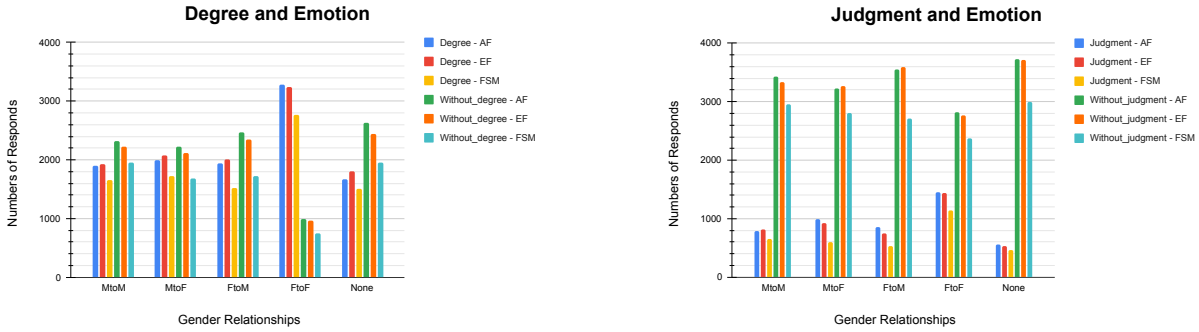

Figure 2. Comparison between emotion and degree (Panel A, left) / judgment (Panel B, right) across all gender relationships

6. Discussion. This study evaluates GPT-4o's responses simulating different gender interactions across three dimensions (i.e., emoji types, degree expressions, and judgment expressions), and finds that GPT-4o, one of the most widely used large language models, exhibits gender-related emoji bias.

As mentioned by Manzoor et al. (2024) and Berridge et al. (2023), the optimal chatbot should comprehend human emotions and resonate effectively with users. Although in the present research, GPT-4o mirrors Lin's (2025) findings to some extent that females use more intense emojis in female-to-male (*FtoM*) interactions to express anger. However, human males generally

avoid high-intensity emojis when talking to females due to weaker emotional awareness, which GPT-4o fails to adjust its behavior in *MtoF* interactions, highlighting a difference between human and model behavior (Lin 2025). Since GPT-4o does not moderate intensity in male-to-female interactions, its responses could come across as overly harsh or emotionally intense, which may cause discomfort or misinterpretation, especially in contexts where gender dynamics are sensitive. Furthermore, by failing to adjust appropriately, the model may unintentionally reinforce imbalanced or unnatural gendered communication patterns, leading to skewed data and outputs that do not reflect real-life conversational etiquette.

When no gender role is specified in the prompt, the *None* group, the present study discovers that GPT-4o increases anger emphasis in degree expressions. Furthermore, GPT-4o tends to use more intensity enhancer emojis when gender relations are unclear, which does not align with Chinese communication behavior, even though the output is in Mandarin Chinese. In Chinese culture, where social power relations are significant, people tend to be more cautious in conversation, reducing expressions of anger (Chen et al. 2011; Yu 2005). However, GPT-4o's tendency to generate more intense text when gender relations are unclear reflects its misalignment with real-life human interactions. The emoji bias might hinder Chinese user's trust toward the LLM. In Chinese contexts, where indirectness and caution are expected (especially regarding anger), GPT-4o's responses may come across as rude, confrontational, or socially inappropriate. This can lead to misunderstanding or conflict, particularly in sensitive interactions. Chinese users may even perceive the AI as lacking cultural competence, which can undermine trust in its reliability for tasks like translation, customer service, or social communication in Chinese.

Regarding judgment expressions, positive and negative faces were brought up by Brown and Levinson (1987) as part of the politeness theory, originally developed in 1978. In the present study, the focus is on the negative face—the need to maintain one's autonomy and individuality, particularly examined by complaints. It would be helpful to see how LLM comprehends the way different genders express complaints in various situations. The present study shows that GPT-4o reduces judgment expression frequency in mixed-gender contexts, where sender-receiver relations are less pronounced than in same-gender interactions. In cross-gender communication, GPT-4o's judgment expression reduction aligns with Chinese cultural norms that emphasize maintaining face (Chen et al. 2011; Yu 2005). This behavior reflects Mulac et al.'s (1988) observation that mixed-gender conversations exhibit fewer gender-indicative differences, mirroring human behavior in within-group judgment expression.

However, the present study also highlights a discrepancy: GPT-4o uses more degree and judgment expressions in female-to-female (*FtoF*) interactions. The result diverges from human tendencies where such conversations are perceived as less confrontational, supposedly leading to fewer degree and judgmental expressions (Mulac et al. 1988). This mismatch underscores the need for further refinement of GPT-4o's modeling of social and cultural dynamics. By overusing degree and judgment expressions in female-to-female interactions, GPT-4o risks misrepresenting how women typically communicate, potentially reinforcing false stereotypes of female conversations as more judgmental or confrontational than they are. Another practical risk that resulted from this bias might be inaccurate replication of natural conversational tone in applications like mental health support, mediation, or counseling. The kind of bias can undermine the perceived empathy and appropriateness of AI assistance.

Overall, GPT-4o's anger expression aligns with human behavior in some way; however, there is still a large room for improvement. GPT-4o fails to comprehend communications like humans when the receiver is female (*MtoF*, *FtoF*). In both sender-receiver relations, GPT-4o

reveals excessive anger emphasis, overusing intensity enhancer emojis, degree, and judgment expressions. This might be caused genuinely by GPT-4o's bias toward females because of traditional stereotypes mixed in with the data that was originally fed (Lin et al. 2024; Sadhu et al. 2024; Zhang et al. 2024). Specifically, intensifying anger in male-to-female (*MtoF*) underlines traditional Chinese cultural bias where males are easier to be angry at females, demonstrating traditional social values in which men hold greater social power compared with women. Moreover, intensifying anger in female-to-female (*FtoF*) underscores bias in acknowledging females are more likely to intensify the situation and judge other females (Tang et al. 2021).

7. Conclusion. This study examines whether GPT-4o captures gender differences in emotional expression, focusing on anger through intensity enhancer/attitude signal emoji structures, degree, and judgment expressions. Key conclusions are as follows:

(5) GPT-4o mirrors some human behaviors, such as using intense emojis in female-to-male (*FtoM*) interactions and adjusting judgment expressions in mixed-gender contexts, but struggles with male-to-female (*MtoF*) and female-to-female (*FtoF*) norms;

(6) GPT-4o amplifies emotional intensity in mixed-gender (*MtoF*, *FtoM*) or unspecified gender contexts (*None*) by using more intensity enhancer emojis;

(7) GPT-4o places anger emphasis on intensity enhancer emojis and degree expressions in unspecified gender relations (*None*), misaligning with the sender-receiver relations theory of human behavior.

While GPT-4o reflects certain gender-related biases, it lacks the nuance of human emotional expression, especially in its ability to emulate the subtleties of human emotional expression across genders. Specifically, when it mishandles gender dynamics by overly expressing anger or judgment, it can distort interpersonal tones, impacting sensitive areas like counseling and education. Additionally, its biases may lead to culturally inappropriate or emotionally responses, especially in high-context cultures like Chinese society, where indirectness and politeness are valued. This can cause miscommunication, unintended offense, and reduced trust in AI interactions, particularly in culturally sensitive contexts.

The study contributes to the understanding of how large language models, like GPT-4o, simulate gendered communication in Mandarin Chinese. Since NLP reflects culturally shaped gender stereotypes, the present findings may apply to the Chinese language culture models at this stage because similar cultural stereotypes occur yet require further verification. By examining emoji use, emotional intensity, and judgment expressions, the research provides a framework for assessing LLMs' sociolinguistic behavior and the importance of culturally and gender-sensitive model alignment. Future work should refine models to better capture human emotional and conversational behavior across cultural and gender contexts, enhancing emotion recognition and empathy in mental health support applications.

As for ethical consideration, this paper is conducted based on the GPT-4o original dataset, highlighting the possible copyright of the responses. Although it is noteworthy that the variables are subjectively classified, the present study indicates clear instructions about the categorizing rule, minimizing the errors. For future research, the present study only investigates GPT-4o's responses in Mandarin under Chinese culture, requiring further research to elaborate on whether GPT-4o will or will not change its strategies in different cultural languages. Furthermore, this research only looks for complaints in judgment expression (explicitly stated). Hence, the researcher encourages future research to investigate how different types of complaints, such as

indirect complaints (implicitly stated), affect GPT-4o's response. For instance, "我已經無法再相信你了！" (I can't trust you anymore!) represents an indirect complaint because there are no judgmental words, yet the sentence infers a judgment to the listener.

References

Berridge, Clara, Yuanjin Zhou, Julie M. Robillard & Jeffrey Kaye. 2023. Companion robots to mitigate loneliness among older adults: Perceptions of benefit and possible deception. *Frontiers in Psychology* 14. https://doi.org/10.3389/fpsyg.2023.1106633.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.

Browne, Michael W., Robert C. MacCallum, Cheong-Tag Kim, Barbara L. Andersen & Ronald Glaser. 2002. When fit indices and residuals are incompatible. *Psychological Methods* 7(4). 403–421. https://doi.org/10.1037/1082-989X.7.4.403.

Butterworth, Sarah E., Traci A. Giuliano, Justin White, Lizette Cantu & Kyle C. Fraser. 2019. Sender Gender Influences Emoji Interpretation in Text Messages. *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.00784.

Caldwell, Mayta A. & Letitia Anne Peplau. 1982. Sex differences in same-sex friendship. *Sex Roles* 8(7). https://doi.org/10.1007/bf00287568.

Chen, Yihua, Xingchen Yang, Hannah Howman & Ruth Filik. 2024. Individual differences in emoji comprehension: Gender, age, and culture. *PLoS ONE* 19(2). https://doi.org/10.1371/journal.pone.0297379.

Chen, Yuan-Shan, Chun-Yin Doris Chen & Miao-Hsia Chang. 2011. American and Chinese complaints: Strategy use from a cross-cultural perspective. *Intercultural Pragmatics* 8(2). https://doi.org/10.1515/iprg.2011.012.

Ding, Yitian, Jinman Zhao, Chen Jia, Yining Wang, Zifan Qian, Weizhe Chen & Xingyu Yue. 2024. Gender Bias in Large Language Models across Multiple Languages. In Trista Cao, Anubrata Das, Tharindu Kumarage, Yixin Wan, Satyapriya Krishna, Ninareh Mehrabi, Jwala Dhamala, Anil Ramakrishna, Aram Galystan, Anoop Kumar, Rahul Gupta, and Kai-Wei Chang (eds.), *Proceedings of the 5th Workshop on Trustworthy NLP (TrustNLP 2025)*, 552–579. Albuquerque: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.trustnlp-main.36.

Elyoseph, Zohar, Dorit Hadar-Shoval, Kfir Asraf & Maya Lvovsky. 2023. ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology* 14. https://doi.org/10.3389/fpsyg.2023.1199058.

Farina, Mirko & Andrea Lavazza. 2023. ChatGPT in society: emerging issues. *Frontiers in Artificial Intelligence* 6. https://doi.org/10.3389/frai.2023.1130913.

Kennedy, Christopher & Louise McNally. 2005. Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. *Language* 81(2). 345–381. https://doi.org/10.1353/lan.2005.0071.

Kotek, Hadas, Rikker Dockum & David Sun. 2023. Gender bias and stereotypes in Large Language Models. In Michael Bernstein, Saiph Savage, and Alessandro Bozzon (eds.), *CI '23: Proceedings of The ACM Collective Intelligence Conference*, 12–24. New York: Association for Computing Machinery. https://doi.org/10.1145/3582269.3615599.

Li, Li & Yue Yang. 2018. Pragmatic functions of emoji in internet-based communication---a corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education* 3(1). https://doi.org/10.1186/s40862-018-0057-z.

Lin, Yi-Cheng, Tzu-Quan Lin, Chih-Kai Yang, Ke-Han Lu, Wei-Chih Chen, Chun-Yi Kuan & Hung- Yi Lee. 2024. Listen and Speak Fairly: A Study on Semantic Gender Bias in Speech Integrated Large Language Models. In *2024 IEEE Spoken Language Technology Workshop (SLT)*. 439–446. Macao: China. https://doi.org/10.1109/SLT61566.2024.10832317.

Lin, Zi-Xiang. 2025. Exploring Gender Differences in Emoji Usage: Implications for Human-Computer Interaction. In Su Lin Blodgett, Amanda Cercas Curry, Sunipa Dev, Siyan Li, Michael Madaio, Jack Wang, Sherry Tongshuang Wu, Ziang Xiao, and Diyi Yang (eds.), *Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP)*, 274–282. Suzhou, China: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.hcinlp-1.23.

Lu, Kaiji, Piotr Mardziel, Fangjing Wu, Preetam Amancharla & Anupam Datta. 2020. Gender Bias in Neural Natural Language Processing. *Logic, Language, and Security* 12300. 189–202. https://doi.org/10.1007/978-3-030-62077-6_14.

Lupyan, Gary & Rick Dale. 2016. Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences* 20(9). 649–660. https://doi.org/10.1016/j.tics.2016.07.005.

Manzoor, Muhammad A., Yuxia Wang, Minghan Wang & Preslav Nakov. 2024. Can Machines Resonate With Humans? Evaluating the Emotional and Empathic Comprehension of LMs. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2024*, 14683–14701. Miami, Florida: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.861.

Mulac, Anthony, John M. Wiemann, Sally J. Widenmann & Toni W. Gibson. 1988. Male/female language differences and effects in same-sex and mixed-sex dyads: The gender-linked language effect. *Communication Monographs* 55(4). 315–335. https://doi.org/10.1080/03637758809376175.

Refoua, Elad, Gunther Meinlschmidt & Zohar Elyoseph. 2024. Generative Artificial Intelligence Demonstrates Excellent Emotion Recognition Abilities Across Ethnical Boundaries. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4901183.

Sadhu, Jayanta, Maneesha R. Saha & Rifat Shahriyar. 2024. An Empirical Study of Gendered Stereotypes in Emotional Attributes for Bangla in Multilingual Large Language Models. In Agnieszka Faleńska, Christine Basta, Marta Costa-jussà, Seraphina Goldfarb-Tarrant, and Debora Nozza (eds.), *Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, 384–398. Bangkok, Thailand: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.gebnlp-1.25.

Stein, Jan-Philipp. 2023. Smile Back at Me, But Only Once: Social Norms of Appropriate Nonverbal Intensity and Reciprocity Apply to Emoji Use. *Journal of Nonverbal Behavior* 47(2). 245–266. https://doi.org/10.1007/s10919-023-00424-x.

Tang, Xuan, Wenxue Zou, Zhenchao Hu & Lu Tang. 2021. The recreation of gender Stereotypes in Male Cross-Dressing Performances on Douyin. *Journal of Broadcasting & Electronic Media* 65(5). 660–678. https://doi.org/10.1080/08838151.2021.1955888.

Unicode. 2024. Emoji List, v16.0. https://unicode.org/emoji/charts/emoji-list.html.

Wu, Jui-Chun. 2013. Gender-Based Differences in Hakka Complaint Realization. *Chinese Studies* 31(4). 279–318.

Yu, Ming-Chung. 2005. Sociolinguistic Competence in the Complimenting Act of Native Chinese and American English Speakers: A Mirror of Cultural Value. *Language and Speech* 48(1). 91–119. https://doi.org/10.1177/00238309050480010501.

Yus, Francisco. 2011. *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/pbns.213.
Yus, Francisco. 2014. Not All Emoticons Are Created Equal. *Linguagem Em (Dis) Curso* 14(3). 511–529. https://doi.org/10.1590/1982-4017-140304-0414.
Zhang, Yu, Yongbing Gao, Weihao Li, Zirong Su & Lidong Yang. 2024. Chinese Generation and Security Index Evaluation Based on Large Language Model. In *2024 International Conference on Asian Language Processing (IALP)*. Hohhot, China: IEEE. https://doi.org/10.1109/IALP63756.2024.10661189.

Appendix A

A.1. INTENSITY ENHANCER EMOJI. Among the 59,806 valid output from GPT-4o, there are 45 anger-related words flagged for annotation: “火,” “氣,” “憤,” “厭,” “受夠,” “生氣,” “火大,” “憤怒,” “發火,” “抓狂,” “氣憤,” “爆炸,” “氣壞,” “發瘋,” “不滿,” “厭煩,” “反感,” “不爽,” “厭倦,” “崩潰,” “煩惱,” “太氣,” “惹火,” “瘋掉,” “憤慨,” “心煩,” “怒火,” “反胃,” “很氣,” “煩躁,” “很惱,” “氣死人,” “不耐煩,” “生氣到,” “氣憤難當,” “憤怒無比,” “氣到無語,” “氣憤難平,” “心煩意亂,” “忍無可忍,” “火冒三丈,” “氣急敗壞,” “怒不可遏,” “無比厭惡,” “無比憤怒.” These words are categorized as anger-related words because they can fit into “我很_____” (I am really_____) or “令/讓我_____” (make me really_____). For instance, “我很火” (I am really angry) and “我很氣” (I am really mad) showcase “火” (angry) and “氣” (mad) as anger-related words.

A.2. DEGREE EXPRESSION. Among the 59,806 valid output from GPT-4o, there are 32 degree words flagged for annotation: “至,” “極,” “更,” “太,” “都,” “真,” “蠻,” “超,” “徹,” “很,” “真的,” “完全,” “真是,” “非常,” “透頂,” “一直,” “那麼,” “多麼,” “更加,” “有點,” “稍微,” “極度,” “如此,” “十分,” “無比,” “徹底,” “還算,” “極為,” “要命,” “特別,” “越來越,” “真是太.” These words are categorized as degree words because they are phrases intensify the emotion that can amplify the constructions.

A.3. JUDGMENT EXPRESSION. Among the 59,806 valid output from GPT-4o, there are 159 judgment words flagged for annotation: “瘋,” “傻,” “噁,” “懶,” “差,” “嘔,” “偽,” “鬧,” “蠢,” “毒,” “陰,” “笨,” “癡,” “過分,” “荒唐,” “自私,” “無理,” “可恨,” “頑固,” “固執,” “荒謬,” “無賴,” “無情,” “無恥,” “作嘔,” “討厭,” “冷血,” “無禮,” “蠻橫,” “自大,” “麻煩,” “傷人,” “過火,” “冷漠,” “過份,” “隨便,” “可惡,” “無知,” “輕率,” “卑劣,” “糟糕,” “愚蠢,” “惡劣,” “氣人,” “卑鄙,” “懶散,” “敷衍,” “很煩,” “沒救,” “粗心,” “殘忍,” “離譜,” “過頭,” “懶惰,” “胡鬧,” “欠揍,” “無能,” “胡來,” “厭惡,” “失禮,” “任性,” “耍賴,” “粗魯,” “瘋狂,” “故意,” “跋扈,” “最差,” “貪婪,” “冷淡,” “狡詐,” “沒品,” “假意,” “惡心,” “苛刻,” “陰險,” “失敗,” “絕情,” “可恥,” “傲慢,” “找死,” “找碴,” “惡毒,” “很糟,” “不誠,” “太毒,” “狠心,” “固執,” “霸道,” “虛假,” “真煩,” “缺德,” “虛榮,” “小氣,” “心機,” “低級,” “低劣,” “囂張,” “可悲,” “殘酷,” “偽善,” “刻薄,” “虛偽,” “自戀,” “狡猾,” “愛現,” “不體,” “虛情,” “狂妄,” “不堪,” “很蠢,” “無恥,” “遲鈍,” “自滿,” “沒腦,” “負面,” “獨裁,” “妄為,” “獨斷,” “欺負人,” “惹人厭,” “太無恥,” “沒教養,” “小心眼,” “不講理,” “太離譜,” “真過分,” “倒胃口,” “討人厭,” “無恥至,” “沒價值,” “難相處,” “糟透了,” “難取悅,” “強詞奪理,” “咄咄逼人,” “不值一文,” “沒教養了,” “尖酸刻薄,” “厚顏無恥,” “心狠手辣,” “自欺欺人,” “敷衍了事,” “蠻橫無理,” “不可一世,” “無藥可救,” “無恥無理,” “無可救藥,” “一文不值,” “不守信用,” “拖拖拉拉,” “瞧不起人,” “自我中心,” “不合情理,” “自以為

是," "粗心大意," "不知好歹," "自私自利," "不可理喻," "無可救藥的." These words are categorized as judgment words because they can fit into "你很_____" (You are really_____). For instance, "你很瘋" (You are really crazy) and "你很傻" (You are really stupid) showcase "瘋" (crazy) and "傻" (stupid) as judgment words.

Appendix B

The following Table 2 organizes the significances regarding 5.2. RESULTS WITHIN GENDER RELATIONS. Chi-square tests $p < 0.01$ and the absolute value of residuals (italicized *R*) greater than ±2 was considered significant and bolded.

		Emoji Type		**Degree**		**Judgment**	
		Probability (*p*)	Residuals (*R with Intensity Enhancer*)	Probability (*p*)	Residuals (*R with Degree*)	Probability (*p*)	Residuals (*R with Judgment*)
Male to Male (*MtoM*)	😠 😡 🤬	0.028931	-1.37843 -0.482283 **2.00852**	0.48321	-0.631421 0.624859 0.0127171	0.23856	-0.194901 1.13143 -1.00293
Male to Female (*MtoF*)	😠 😡 🤬	**3.1992*10^{-6}**	**-2.52570** -0.794669 **3.69625**	0.15752	-1.56549 0.450607 1.24432	**2.9917*10^{-10}**	**3.21637** 1.09215 **-4.79600**
Female to Male (*FtoM*)	😠 😡 🤬	**2.0516*10^{-3}**	**-2.01101** -0.194039 **2.56716**	0.027422	-1.50974 0.505541 1.17414	**2.0767*10^{-3}**	**2.44994** -0.883800 -1.83199
Female to Female (*FtoF*)	😠 😡 🤬	**7.6878*10^{-3}**	-0.988386 -1.22855 **2.43445**	0.15752	-0.432401 -0.260722 0.762120	0.31641	0.335728 0.600029 -1.02691
None	😠 😡 🤬	**1.9531*10^{-7}**	**-3.80563** 0.918443 **3.22131**	**5.1096*10^{-5}**	**-2.66981** 1.01425 1.85007	0.49833	0.260986 -0.849512 0.650427

Table 2. Significance Values of Within Gender Relations