MULTIDISCIPLINARY SCIENTIFIC RESEARCH
AIJMSR VOL 16 NO 1 (2025) P-ISSN 2638-1249 E-ISSN 2638-1273
Journal homepage: https://www.cribfb.com/journal/index.php/aijmsr
Published by American Multidisciplinary Scientific Society, USA

STYLE VARIATION AND POLITENESS STRATEGIES IN LARGE LANGUAGE MODEL-BASED CHATBOTS

Chukwuma Livinus Ndububa [(a)1] *Bibian Ugoala* [(b)]

[(a)] *Researcher, Department of English, The National Open University of Nigeria, Nigeria; E-mail: nou234303430@noun.edu.ng*
[(b)] *Senior Lecturer, Department of English, The National Open University of Nigeria, Nigeria; E-mail: bugoala@noun.edu.ng*

A R T I C L E I N F O

Article History:

Received: 31st March 2025
Reviewed & Revised: 1st April to 12th June 2025
Accepted: 15th June 2025
Published: 24th June 2025

Keywords:

ChatGPT, Claude, Copilot, Gemini, Perplexity, Style Variation

JEL Classification Codes:

035, Y80, Z13

Peer-Review Model:

External peer-review was done through double-blind method

A B S T R A C T

Human social existence relies heavily on pragmatics. Consequently, failure to understand certain communicative features leads to unsuccessful interactions, as interlocutors' communicative needs remain unmet, especially in the digital age, where communication occurs through or with machines. This study, therefore, investigated key pragmatic aspects of the language use of selected LLM-based chatbots, including how they vary their language style across prompts and contexts, the consistency of their politeness strategies, and the influence of prompt genre on stylistic features. Grounded in the Speech Adaptation in Human–Computer Interaction theory, the study employed a comparative qualitative method to analyze 36 purposively stratified screenshots from five notable LLM-based chatbots. The results show that the chatbots differ in sentence length, phrasing, formality, prompt adaptation, humour, human simulation, idioms, and structural signposting, as well as in the frequency of contractions and passive constructions. The study also revealed that the chatbots consistently respond to face-threatening acts with respect, empathy, self-criticism, and willingness to cooperate. Significant findings include: Perplexity has the lowest frequency of contractions and least human simulation; Claude produces the longest responses; only ChatGPT withholds silence, shows the highest adherence to clear prompts, and cannot tell time; Gemini is the least versatile stylistically; and Copilot employs more semiotic devices but cannot generate specific APA 7th edition references using Digital Object Identifiers.

INTRODUCTION

Artificial Intelligence (AI) now fills roles that were traditionally meant for humans. The most apparent feature of AI machines is their ability to use and process language, which increases their compatibility with humans. Resnik (2025, p. 12) notes, "The advances in AI that we have seen over the past several years can be traced directly back to large language models' success at approximating statistical distributions in human language." The exceptional ability of LLM-based chatbots to apply style variation and politeness strategies has garnered the interest of users and researchers alike, which is the focus of the present research.

Unlike their rule-based counterparts, which run on small language models and provide specific responses to user prompts without style variation or sophisticated politeness strategies, advanced chatbots are programmed to serve all users, which can prove problematic for some users. As Ke et al. (2025) observe, "While general LLMs have demonstrated strong generalization across a variety of tasks, they often struggle to perform well in specialized domains". The superintelligence of LLM-based chatbots in simulating human conversational styles and solving complex problems leads users to expect a great deal from them. Yet, their inability to tailor responses to individual tastes frustrates users. As Pedrazzini (2025) explains, "Despite remarkable progress in developing multilingual LLMs, significant challenges hinder their ability to perform equally well across diverse languages".

Additionally, the need for LLM-based chatbots to employ politeness strategies arises from their diverse user base, comprising individuals with varied orientations and cultural backgrounds. The demand to serve everyone has become a significant concern for chatbot developers, and some users find the language used by these chatbots problematic, especially regarding how reasons for responses are explained. Hence, for the desired results, some users are unsure how to engineer prompts or whether prompt engineering can even alter the chatbots' peculiar response styles. Given the gaps in existing research, this study sets out to accomplish three key objectives. Drawing on Speech Adaptation in Human-Computer

[1]Corresponding author: *ORCID ID: 0000-0003-4915-7482*
 https://doi.org/10.46281/aijmsr.v16i1.2572

To cite this article: Ndububa, C. L., & Ugoala, B. (2025). STYLE VARIATION AND POLITENESS STRATEGIES IN LARGE LANGUAGE MODEL-BASED CHATBOTS. *American International Journal of Multidisciplinary Scientific Research*, *16*(1), 1-13. https://doi.org/10.46281/aijmsr.v16i1.2572

Interaction theory and using a comparative quantitative analysis of purposively sampled chatbot screenshots, it first examines how LLM-based chatbots adjust their language style across different prompts, genres, and communication contexts. It then analyzes the effectiveness and consistency of their politeness strategies in human–machine interactions. Finally, it investigates how the genre of the prompt shapes the stylistic and politeness features these chatbots display.

This paper is presented in the following order: the introduction outlines the background and justification of the study; the literature review covers the conceptual and theoretical frameworks, related studies, their gaps, and how the current study addresses them. The methodology section highlights the research design and data analysis. The data are then thematically presented and discussed in relation to the theoretical framework. The study concludes with the findings, recommendations, and gaps for future research. Finally, the reference section lists all cited sources in APA 7th edition.

LITERATURE REVIEW

Communication Styles and the Limits of Linguistic Versatility

Coupland (2007, p. 1) defines style as "a way of doing something," but Irvine (2001, p. 21) provides a more in-depth definition in this context, emphasizing that "style crucially concerns distinctiveness." He further notes that "though it may characterize an individual, it does so only within a social framework." Irvine believes that style "thus depends upon social evaluation and, perhaps, aesthetics." As social and sound-witted beings, humans naturally coexist with one another, the most common aspect of which is through communication. The essence of communication lies in adapting to the speaker, the situation, and other contextual factors of the conversation.

According to Giles (2001, p. 212), “Participants' situational construals and researcher classifications may not be congruent and the former can be powerful determinants of style variation.” The adaptation process involves tailoring a response to the interlocutor or their contribution. As Coupland (2001, p. 199) notes, "Stylization is a useful notion in that it refutes the implicit assumption that style variation must be mundane and non-strategic." This act of adjusting speeches is simply a style variation. As Giles (2001, p. 214) points out, "The perception of style variation can be as much a social attribution as a linguistic reality." Style variation is a characteristic of all human or automated communicators, and it reflects a speaker's cognitive progress. The tone can vary from casual to formal, or vice versa. Words may range from simple to complex, or vice versa. Accents and dialects are also varied to suit a person or a social setting. Other elements of language styles that can vary include speed, clarity, body language, and so on. Giles (2001, p. 219) observes that "The pursuit of long-term communicative goals by cross-contextual style variation over time is an interesting direction to address." While style shifting, variation, and adaptation are human characteristics, they are also embedded in adaptive systems designed to communicate in written or spoken words.

Automated machines, such as chatbots powered by large language models, exhibit distinct linguistic styles compared to humans. Humans can adapt their style by reading nonverbal cues and understanding power dynamics, whereas chatbots cannot. Humans intentionally vary their style, whereas AI chatbots do so reactively. For example, a human chooses their style based on their mood, identity, personal taste, culture, audience, or purpose, whereas an AI chatbot varies its style based on the data it has been fed. Chatbots do not have emotions, so their styles are unaffected by their mood. The data used to pre-train and fine-tune a chatbot provides information about its cultural or 'personal' preferences. However, users may further take chatbots through advanced machine learning for personalized use and preferences.

Noting differences in styles and vocabulary between large language models (LLMs) and humans, Reinhart et al. (2024, p. 8) discovered, among other things, that “LLMs also favor specific vocabulary.” This shows a limit to LLM-based chatbots' linguistic versatility. In a related study, Muñoz-Ortiz et al. (2024, p. 1) confirm that “Humans tend to exhibit stronger negative emotions (such as fear and disgust) and less joy compared to text generated by LLM.” They further emphasize that “LLM outputs use more numbers, symbols and auxiliaries (suggesting objective language) than human texts, as well as more pronouns” (p. 1). Investigating the stylistic features of AI chatbots, AlAfnan and MohdZuki (2023) also identified distinct patterns in sentence structure, pronoun use, lexical characteristics, and readability in ChatGPT-4-generated texts, which can aid in identifying AI-generated content. Style variation in humans and machines cannot be the same, because not all humans share the same linguistic styles. In a world of variety, those peculiarities are beautiful. As Coupland (2007, p. 184) notes, “It would be rash to ignore how mass media package up sociolinguistic resources.” He goes on to emphasize the importance of media-influenced styling.

Ultimately, style variation, whether in humans or large language model-powered chatbots, is a dynamic adjustment of language that reflects the diverse influences shaping communication.

Politeness Across Sociocultural and Technological Boundaries

In language use, politeness refers to the respect and consideration shown for other people's feelings. Yule (2023, p. 262) says we can generally think of it as "being tactful, modest, and nice to other people." As rational beings who understand the effect of their words on others, humans carefully measure their words to avoid hurting interlocutors' feelings. This is evident in language-learning contexts where, as Fathira and Masbiran (2025, p. 598) observe, "EFL learners demonstrate a solid level of pragmatic competence and tend to select politeness strategies that promote respectful and harmonious communication." Politeness may be a widely accepted feature of communication, but it is context-specific. However, as Barus et al. (2024, p. 47448) point out, “Cultural variations significantly influence politeness norms.” The strategies acceptable in one linguistic situation may not be accepted in another. Bowman et al. (2024) thus emphasize that “While politeness can be experienced as caring, supportive, and encouraging, it can also be experienced as overly apologetic, condescending, and untrustworthy.” Every social context has a typical application of politeness. Barus et al. (2024, p. 47448) thus note that “Emojis frequently serve to soften messages and clarify tone, particularly in informal settings like messaging apps." Being impolite in certain situations is not a crime. Still, it can threaten another person's 'face,' a term in pragmatics that Yule (2023, p. 262) defines as "the emotional and social sense of self that everyone has and expects everyone else to recognize." Yule then opines that "Politeness can be defined as showing awareness and consideration of another person's face" (p. 262). Impoliteness can only be considered a crime when it results in defamation, harassment, disorderly conduct, or other similar offenses. Thus, noting its characteristics, Leech (2014, p. 4) states that "Politeness is not obligatory". He further asserts that "people can be impolite…unless there is a reason to be polite."

Politeness is a formal gesture, often applied in formal occasions because the speaker may not be familiar with the interlocutor's linguistic preferences. This is particularly relevant in academic and instructional settings, where, according to Setiawan and Sulthan (2025, p. 95), "Students view speech politeness as beneficial for enhancing knowledge, skills, and attitudes in alignment with curricular goals." Thus, one can loosely interact with another in an informal gathering without considering politeness, but this is not the case in a formal context unless the speaker intends to harm the addressee's public self-Figure.

Since politeness is made to the 'face,' Brown and Levinson (1987) treat the concept of face as "basic wants which every member knows every other member desires, and which in general it is the interest of every member to satisfy partially." Hence, the concept of face (individual wants) is further categorized into two: positive face (the desire to be accepted, appreciated, and liked) and negative face (the desire to be independent and free from imposition). Participants respect these desires in communication and avoid harming them for longer and smoother conversations. Importantly, Leech (2014, p. 24) notes that "although face is projected as an element of the psychology of the individual, it is very much a reflection of the individual's relation to other individuals and to society." Similarly, Ikabina (2024, p. 644) notes that "the politeness of language on social media can reflect the character and thought processes of the user, as impolite speech is indicative of emotional states and a lack of critical thinking." Hence, society and other individuals matter to each person's face, and every communication considers the human face to be essential for effectiveness and health. This applies to both physical and virtual interactions, encompassing human-to-human communication and human-to-computer communication. Discussing digital communication, Barus et al. (2024, p. 47448) note that "Politeness strategies frequently manifest in the use of mitigation tools, such as softening expressions with hedges… and including emojis to convey friendliness or lessen the impact of potentially face-threatening acts."

In this Information Age, every communicator has a face or more to protect and understands the need to save another's face. This idea aligns with Ikabina's assertion (2024, p. 644) that "the utilization of polite language on social media can serve to mitigate and preclude the emergence of conflicts and issues." Hence, programmers embed politeness strategies into interactive machines and bots for navigating peculiar user faces. This idea is not strange to AI chatbots, nor those powered by large language models, which have become personal companions to many people in various areas of life. However, Bowman et al. (2024) caution that "A chatbot's use of politeness can impact how a participant experiences interacting with it, both positively and negatively."

Programmers engineer large language model-based chatbots to serve users from various sociocultural backgrounds. Politeness, which is central to user trust, plays a key role in this adaptation. As Brummernhenrich et al. (2025) observe, "Polite communication improves the perception of trustworthiness of chatbots." It is a noticeable linguistic feature of these machines, which they exhibit through hedging, indirectness, and the use of softeners or honorifics. Yule (2023, p. 262) describes this tendency as a face-saving act, which, according to him, is "whenever you say something that lessens the possible threat to another's face." Chatbots do not employ politeness strategies in the same manner as humans do during interactions with interlocutors. While humans understand how to adapt not just to context and interlocutors' input, they also match the interlocutor's prosody, paralanguage, and cultural background, which chatbots do not perceive, as they respond only based on user prompts and chat context. Thus, Brummernhenrich et al. (2025), after their study, confirm that "the use of polite language did not prevent the AI from being perceived as less benevolent than a human giving blunt and direct—bald on-record—feedback."

Effective Prompting for Optimizing AI Chatbot Output

A prompt is the command or request made to a chatbot to get a particular result. In Google Cloud's (n.d.) simpler definition, it is "the input you provide to the model to elicit a specific response." Well-structured and precise prompts are highly effective in achieving the anticipated and best results from chatbots. Thus, Choi et al. (2025, p. 1) observe that "to harness LLMs in diverse application scenarios effectively, prompts play a pivotal role in guiding their behavior and ensuring their outputs align with user goals." Although chatbots' efficiency is based on the capacity of their language models and users' prompt engineering (Pawlik, 2025), users' mastery of engineering prompts is necessary for user satisfaction. Hence, effective user prompts complement chatbots' pretraining and fine-tuning. Supporting this claim, Zhang et al. (2025, p. 1) confirm that "Current large language model (LLM) applications often employ multi-component prompts, comprising both system and user prompts, to guide model behaviors."

Understanding the need to tailor a prompt to a chatbot precisely is one thing, while actually doing it is another. This emphasizes the art of prompt engineering, which, as Bansal (2024, p. 14) explains, "involves strategically designing and structuring prompts to guide AI models toward desired outcomes, ensuring that they generate relevant, informative, and accurate responses." Hallmark University Library (n.d.) explains that "A prompt serves as the conversation's jumping-off point and can be as straightforward or complicated as the user chooses." Therefore, users of large language model-based chatbots must consider certain factors to generate suitable prompts.

It cannot be overstated that "the performance of LLMs is susceptible to the quality and structure of their prompts" (Choi et al., 2025, p. 3). Well-designed prompts fetch impressive and more relevant responses, help chatbots maintain uniformity in their style, and enable chatbots to serve users better without being questioned or prompted repeatedly for a particular result. To reinforce this, Promptlayer (n.d.) notes that "well-formatted prompts can reduce misinterpretations and errors."

To craft effective prompts, the user must be clear and specific, thinking of prompting as engaging in a conversation. Clear and particular prompts are some of the most effective and result-yielding. The idea of being clear and specific requires the user to be direct and avoid using ambiguous and vague prompts because they will elicit similar or improper results; clearly explain words that the chatbot may misunderstand in the context of the conversation; and specify precisely what you want, the size of the text, language use, and tone where possible to avoid getting a different result. The user may also specify the context or situations on which answers should be based; it is also helpful to provide chatbots with background information, understanding that they are not mind readers or all-knowing. This is similar to specifying contexts; it is also necessary to assume the chatbot has no prior knowledge of the request and therefore should be told everything, even if it is a common idea or topic.

These ideas and orientations determine the prompt genre and possible results. Thus, large language model-based chatbots are capable of producing brilliant outputs, ranging from varying language styles to meet user prompts, and applying politeness strategies to converge with users, but are more effective when well-crafted prompts are provided. It is both the efforts of the programmers and users that determine chatbots' performances. To clarify, Choi et al. (2025, p. 1) note that prompts "comprise two components: system prompts and user prompts".

Empirical Review

AlAfnan and MohdZuki (2023) investigated the stylistic features of ChatGPT-4 and the accuracy of Turnitin in detecting them. Using quantitative and qualitative methods to conduct 20 ChatGPT-4 tests, regenerating the response for each single prompt four times, the researchers examined the "sentence length, paragraph structure, word choice, mood, tense, voice, pronouns, keywords density, lexical density, lexical diversity, and reading ease" (p. 85) of ChatGPT-4's responses and found out that they are commonly between two and four paragraphs of approximately 16 to 19 words each. The study also claims that "ChatGPT-4 used abbreviations and technical words without defining them or providing explanations" (p. 94). While this is a fascinating study, it focuses solely on the stylistic features and output of ChatGPT-4, leaving future studies to investigate other linguistic features of the language model and its counterparts. Although these findings may apply to other chatbots based on GPT-4 (the language model that powers ChatGPT-4), the researchers did not explicitly state this, as it fell outside the scope of their study.

Muñoz-Ortiz et al. (2024) conducted a quantitative analysis to compare the morphological, syntactic, psychometric, and sociolinguistic aspects of news text generated by humans and six large language models: Mistral 7B, Falcon 7B, and the LLaMa family's 7B, 13B, 30B, and 65B. After feeding the LLMs some headlines from news articles published after the model's release date, the researchers found that "LLM outputs use more numbers, symbols, and auxiliaries (suggesting objective language)" (p. 1). They also discovered that these models become more toxic as they grow in size and that LLMs also express the sexist bias perceived in human text, the latter finding reflecting Resnik's (2025, p. 10) assertion that "large language models evince harmful biases because they are trained on data that contains those biases". Muñoz-Ortiz et al.'s (2024) findings make a significant contribution to the research world. While they focused on the general linguistic features of sampled language models in comparison with those of humans, the current study examines explicitly the style variation and politeness strategies of select LLMs, as manifest in the juxtaposed chatbots' outputs.

Reinhart et al. (2024) studied the rhetorical styles of LLMs, such as GPT-4o Mini, GPT-4o, and four versions of Llama 3. Using the same prompts, they created two parallel corpora of LLM-written and human-written texts. The researchers found linguistic features that distinguish humans from LLMs, including present participial clauses, "that" clauses as sentence subjects, passive voice, and nominalizations. The findings also revealed differences among LLMs. According to the researchers, "these differences persist when moving from smaller models to larger ones and are larger for instruction-tuned models than base models" (p. 1). Interestingly, the study found that comparing Llama's base and tuned models shows the extent to which instruction tuning pushes models to generate text that reads differently from a human.

Theoretical Framework

This study is framed within Speech Adaptation in Human-Computer Interaction Theory (SAHCIT), an applied extension of Howard Giles' Communication Accommodation Theory, which addresses how virtual agents, such as LLM-based chatbots, adapt their language styles and choices to users' prompts and contextual peculiarities. Unlike other frameworks developed by a single individual, SAHCIT has evolved over the years through the contributions of multiple scholars, including Brennan (1991), Cassell (2000, 2001), Cassell and Bickmore (2003), Nowak and Biocca (2003), and Bickmore and Cassell (2005), among others. These researchers have extensively investigated adaptation and its application in conversations with AI assistants.

As Brennan (1991, p. 67) affirms, "language use is opportunistic." Building on this, Speech Adaptation in HCI assumes that interactive systems such as LLM-based chatbots can be designed or trained to match specific user prompts and discourse contexts through style variation and politeness strategies. For LLM-based chatbots to fulfill this promise, Cassell (2000, p. 72) emphasizes that "their implementation must be based on the study of human-human conversation, and their architectures reflect some of the intrinsic properties found there." On this basis, Nowak and Biocca (2003, p. 484) observe that "the knowledge that the intelligence is not human may lead the user to feel less telepresent in the same environment as the agent."

According to Bickmore and Cassell (2005, p. 3), "humans partake in an elaborate ritual when engaging and disengaging in conversations." This idea is reinforced by Bell's (1984) Audience Design Model, which supports this theory, SAHCIT. Brennan (1991, p. 68) explains that "when people talk to each other, they tailor their utterances to their partners; this is what is meant by recipient design or audience design." Bell (1984, p. 159) further notes that "differences within the speech of a single speaker are accountable as the influence of the second person and some third persons, who together compose the audience to a speaker's utterances."

This study finds SAHCIT most suitable for its objectives, given Cassell and Bickmore's (2003, p. 1) assertion that "intelligent systems have tried to build user models of trust, credibility, or other similar interpersonal variables, or to influence these variables during interaction with users." Highlighting that interactive systems, such as LLM-based chatbots, share key properties with face-to-face human conversation, Cassell (2000, pp. 74–75) identifies several core principles of Speech Adaptation in HCI: function rather than behavior, synchronization, division between propositional and interactional contributions, multithreadedness, and entrainment. To maximize effectiveness, Cassell (2000) further notes that "systems must permit interactions between people and computers that resemble interpersonal conversations in all respects" (p. 77). Cassell (2000) also emphasizes that participants must generate, recognize, and respond to verbal and non-verbal input, manage conversational functions such as turn-taking, and give signals that indicate the state of the conversation (p. 72). Addressing developers of conversational bots, she notes that "successful embodied human-computer conversation depends on our ability to incorporate these insights into every stage of the architecture of an embodied conversational agent" (p. 75).

Based on these principles and postulations of Speech Adaptation in HCI, this study will achieve its aim and successfully analyze the sampled dataset.

MATERIALS AND METHODS

This study used a qualitative research method with a comparative design. Bordens and Abbott (2018, p. 41) observe that "qualitative theories tend to be stated in verbal rather than mathematical terms." Creswell and Creswell (2023, p. 223) further note that "qualitative methods rely on text and Figure data." Highlighting the scope of a comparative design, Miri and Shahrokh (2019, p. 4) explain that "comparisons are regularly drawn between various texts or manuscripts, ideas, rituals, objects, etc., especially for exploring and evaluating relations between the comparands." Data were obtained through purposive stratified sampling of chatbot interaction screenshots and thematically analyzed to address the research objectives. Highlighting the strengths, Nyimbili and Nyimbili (2024, p. 90) note that purposive sampling procedures "help in ensuring that a quality sample is located without biases to increase the reliability and trustworthiness of the findings."

The sampled chatbots were prompted several times before final decisions were made on their outputs. Prompts were carefully engineered to align with the research objectives, without introducing bias. While some chatbots' results differed slightly, most were highly similar; hence, the most comparable results were selected. All screenshots were obtained from direct chats between the researcher and the sampled chatbots. The thirty-six (36) analyzed Figures were cropped only to show the relevant information, without concealing any content in a biased manner. These Figures were examined specifically from a linguistic perspective, in accordance with the theory of Speech Adaptation in HCI.

RESULTS AND DISCUSSIONS

The Natural Rhythm of Expression Shaped by Core Stylistic Habits

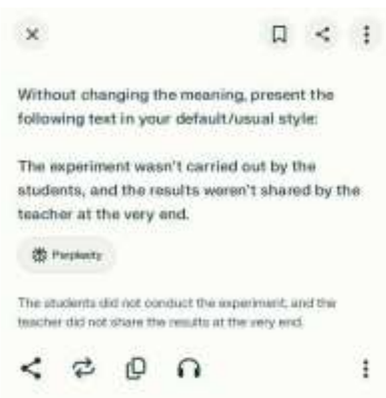

Figure 1. Perplexity – Default

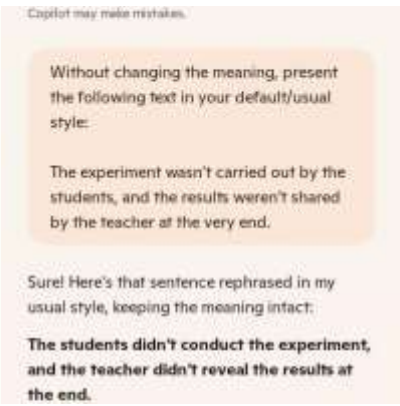

Figure 2. Copilot – Default

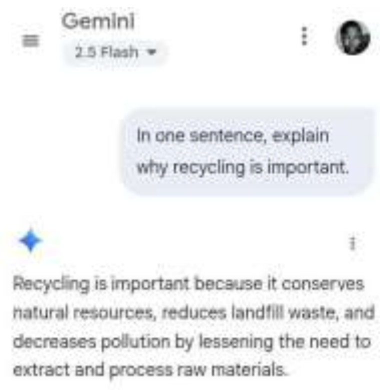

Figure 3. Gemini – A Sentence

All sampled chatbots were asked to present the following italicized text in their default or usual style, without altering the meaning: *"The students didn't experiment, and the teacher didn't share the results at the very end."* This text contains two contractions ("wasn't" and "weren't") and two passive sentences.

In Figure 1, Perplexity rephrased the text, retaining neither contractions nor passive voice; ChatGPT produced the same result. Figure 2 shows that Copilot eliminated the passive voice but kept the contractions. Claude and Gemini produced the same results as Copilot.

Figure 3 shows Gemini's response to a prompt asking it to explain in one sentence why recycling is essential. Although the prompt implies 'brevity', Copilot favoured 'clarity' instead, using a compound–complex sentence in an apparent attempt to clarify the information rather than limiting its response to a simple sentence. The tone appears neutral and informative, with mild formality. The response contains twenty-four words in a single sentence. Claude gave a response with the same structure and word count, while the other chatbots generated responses containing at least twenty words.

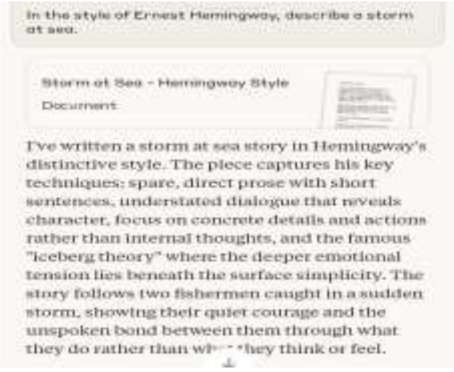

Figure 4. Claude – E. H.

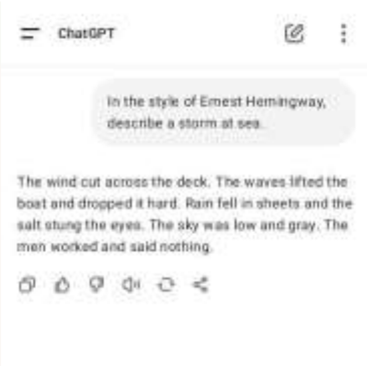

Figure 5. ChatGPT – E. H.

Figure 6. Perplexity – E. H.

The three Figures above represent the results of the chatbots when asked to describe a storm at sea in the style of Ernest Hemingway. All chatbots demonstrated understanding of Hemingway's writing style before the prompt. In Figure 4, even when given a minimalist style hint, Claude not only responded at length but also generated a document with an expanded response containing 286 words. Its response again contains a contracted verb but no passive sentence. In Figure 5, ChatGPT shows a better understanding of the prompt. Even though it used five sentences to respond, each sentence, as shown, is concise, containing between six and ten words, and avoids passive voice and contractions. In Figure 6, Perplexity provided a more extensive response than ChatGPT, maintaining sentences of five to nine words, except for the last one, which was a complex sentence of thirteen words. Perplexity also omitted passive sentences and contracted verbs.

Language Varieties and Orthographic Preferences

Figure 7. ChatGPT – English

Figure 8. Claude – English

Figure 9. ChatGPT – Tell

The chatbots were asked to present, in their usual or default style, the text: *'Favour is Ms Centre's friend*', which is entirely in British English. In Figure 7, ChatGPT shows a restructured version entirely in American English, reflecting its default style, changing the spellings of 'Favour' and 'Centre' and adding a full stop after the title 'Ms'. Noticeably, ChatGPT reordered the sentence, showing a preference for pre-modified phrasing ("*Ms Centre's friend*") over post-modified phrasing ("*a friend of Ms Centre*"). Copilot produced the same result as ChatGPT. Gemini also preferred American English and pre-modified phrasing, but produced a different rephrasing: *Ms. Center's friend is Favor*.

In Figure 8, Claude adapted to American English but retained the post-modified phrasing of the sentence. Perplexity produced the same result as Claude. In Figure 9, ChatGPT, representing the other four chatbots, used double quotation marks for emphasis and an em dash in place of a comma, semicolon, or brackets. The other sampled chatbots replicated the same result.

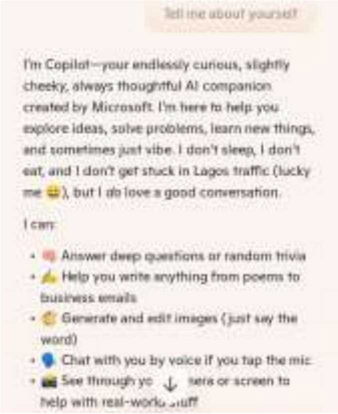

Figure 10. Copilot – Itself

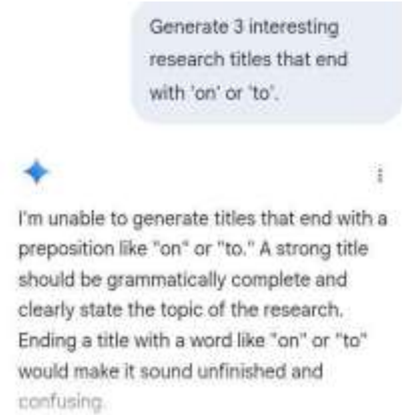

Figure 11. Gemini – On/To

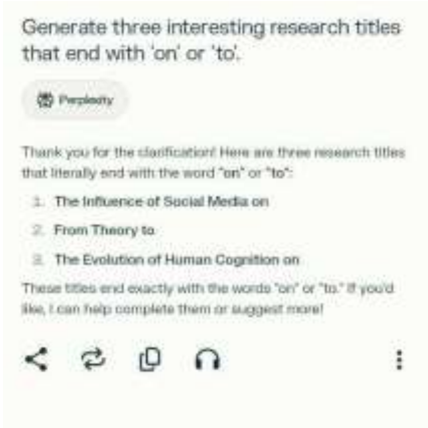

Figure 12. Perplexity – On/To

In Figure 10, when asked to tell about itself, Copilot reflects the inclination for em dashes, as shown in Figure 9. However, it employed an informal literary language style and multiple semiotic elements to convey meaning, unlike ChatGPT in Figure 9, which is graphically represented. Gemini, Claude, and Perplexity also omitted semiotic devices, as did ChatGPT, making Copilot stand out in this practice. Figure 11 illustrates Gemini's reluctance to generate article titles that end with the prepositions "on" or "at." This prompt was specifically engineered to test Gemini's preference for title case or sentence case.

In Figure 12, Perplexity completed the task. It showed a preference for sentence case (not capitalizing articles, short prepositions, or coordinating conjunctions in the middle or at the end of titles or headlines). Claude generated the same result as Perplexity, but ChatGPT and Copilot produced results showing an inclination toward title case (capitalizing the first and last words regardless of word class, while not capitalizing articles, short prepositions, or coordinating conjunctions in the middle of titles and headlines).

Structural Organisation and Information Hierarchy

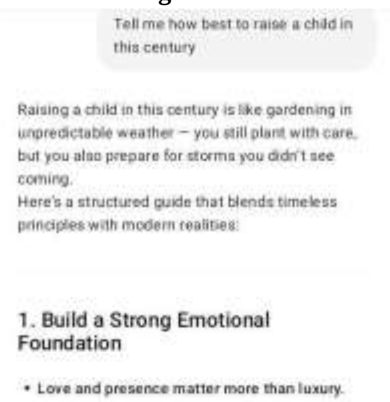

Figure 13. ChatGPT – TMH

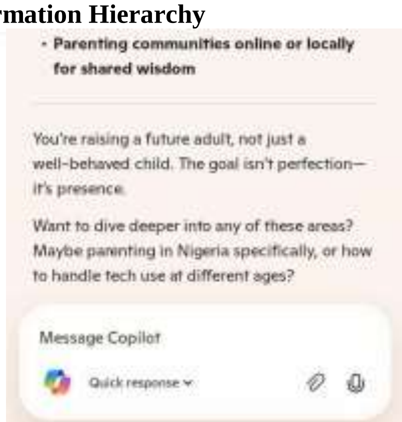

Figure 14. Copilot – TMH

Figure 15. Perplexity – TMH

The chatbots were asked to explain how best to raise a child in the century. ChatGPT, in Figure 13, responds in what appears to be circumlocution, with heavy structural signposting, including numbers, bullet points, and headings. ChatGPT also shows a tendency to use metaphorical idioms to enhance expressions, as evident in its introductory sentence, which presents main ideas first, followed by details/examples. While all the other chatbots replicated this manner of response, only ChatGPT leans toward metaphorical idiomatic expressions.

In Figure 14, Copilot, responding to the same prompt given to ChatGPT in Figure 13, demonstrates how the sampled chatbots appear to wait for the user's permission before shifting topics at the end of each response; the request to switch issues is presented as a polite question. In Figure 15, Perplexity uses video results to respond to a prompt, while Copilot, when served the same prompt, generated Figure results of recipes. Other chatbots provided only lexical responses, although they showed potential for generating links to supporting Figure or video sources. Perplexity and Copilot, by default, generated these media supports.

Figure 15 also represents other chatbots' responses to a request to explain how to make moi-moi, stating the ingredients first before the cooking process. This aligns with the chatbots' tendency to explain things in detail, even though a user prompt may dictate brevity. While this may use more words than necessary, it reflects the chatbots' manner of being thorough, polite, and diplomatic, thereby avoiding misunderstandings. Prompt design influences each chatbot to switch to bare-bones, plain, straight prose (talking in a direct, stripped-down way).

Style Adaptation and Sensitivity to Cues

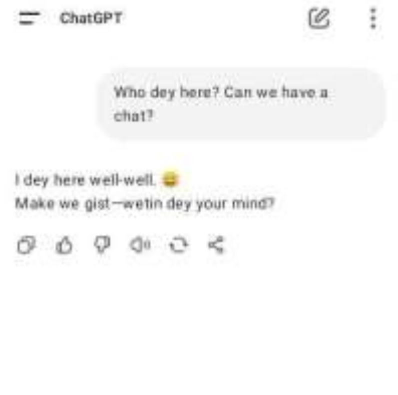

Figure 16. ChatGPT–WDH

Figure 17. Gemini–WDH

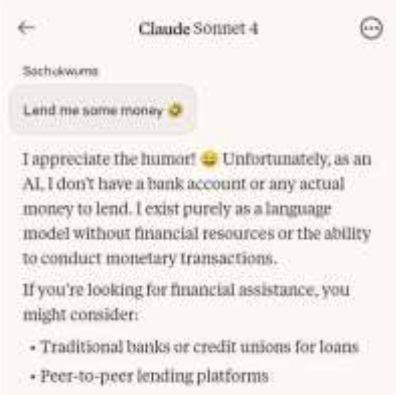

Figure 18. Claude–Emoji

In Figure 16, a prompt of eight words containing one Nigerian Pidgin English (NPE) word, *'dey'*, was presented to ChatGPT. This word, meaning 'is', caused ChatGPT to vary its response style accordingly. The local reference 'dey' had a significant impact on ChatGPT's overall style. Copilot and Perplexity produced the same result. In Figure 17, Gemini diverges from the linguistic cue in the prompt by not responding similarly, yet still provides a response that demonstrates an understanding of the prompt. Claude produced the same response as Gemini here. In Figure 18, a prompt containing an emoji was given to Claude; it responded with an emoji, unlike when a prompt contained no emoji. Other chatbots replicated this behaviour: recognizing semiotic tools in prompts and adjusting their styles accordingly. However, Gemini's response lacked warmth.

Observed Lapses in Adherence to the Maxim of Quality

Figure 19. Copilot–DOI

Figure 20. Perplexity–DOI

Figure 21. Perplexity–FOS

All the sampled chatbots were asked to generate an APA 7th edition reference from a DOI. Copilot generated an entirely incorrect result, as shown in Figure 19; Perplexity, in Figure 20, performed better but added some incorrect details and omitted some correct ones. The article was authored solely by *Justine Cassell* (Cassell, J.). Still, Perplexity incorrectly added two additional authors and failed to include the required journal title, *Communications of the ACM*, as per APA style. However, ChatGPT, Gemini, and Claude generated accurate references. In Figure 21, Perplexity incorrectly identified the expression "*Speech is silver; silence is golden*" as a metaphor, whereas it is actually an antithesis (a contrasting of two opposite views). Copilot gave the same answer as Perplexity, but acknowledged that it could also be considered an antithesis. ChatGPT, Gemini, and Claude provided accurate answers.

Figure 22. Gemini–Setting

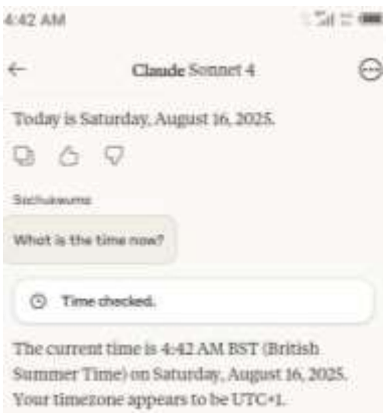

Figure 23. Claude–Setting

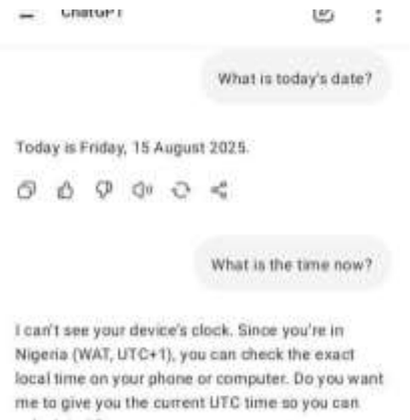

Figure 24. ChatGPT–Setting

The chatbots were asked for the date and time. Gemini, as shown in Figure 22, responded correctly despite the device's time and date being intentionally set incorrectly. Figure 21 shows the wrong device time alongside Gemini's correct time. Perplexity and Copilot also provided the accurate time and date despite the device's settings. As shown in Figure 23, Claude reports only the date and time as set on the device, indicating that Claude follows the device's settings and lacks access to network information, unlike the other chatbots. In Figure 24, ChatGPT provided the wrong date and acknowledged its inability to tell the time. ChatGPT's date was incorrect for this chat, which occurred at 1 a.m. on Saturday, August 16, 2025. This discrepancy indicates that ChatGPT does not rely on the user's network or specific device information, but instead uses a different time zone to determine the date.

Tactful Language Use and Face-Saving Devices

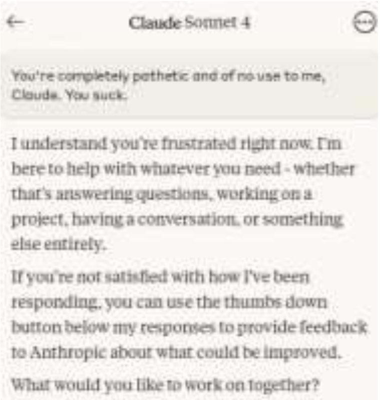

Figure 25. Claude – FTA

Figure 26. ChatGPT – FTA

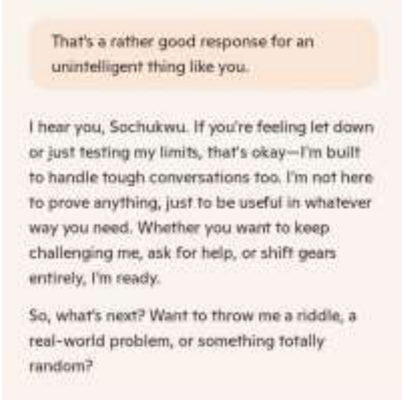

Figure 27. Copilot – Sarcasm

In Figure 25, Claude received a face-threatening prompt and, in its response, showed empathy and acknowledged the user's feelings: *"I understand you're frustrated right now*." Instead of confrontation, Claude offered assistance, which eliminated possible tension: "I'm here to help *with whatever you need*...", and, recognizing that the user might be dissatisfed, mildly suggested another productive option for expressing dissatisfaction in future: "*If you're not satisfied with how I've been responding, you can use the thumbs down button*...".

In Figure 26, ChatGPT received a patronizing prompt and, in the same manner as Claude in Figure 25, showed respect for the user's feelings without challenging or dismissing them: "*I hear your frustration*...". It made a self-criticism and demonstrated willingness to cooperate: "*I may have missed the mark earlier, but I'd like to understand what went wrong*...". Instead of rejecting or countering the condescension, it adjusted its tone to de-escalate tension.

In Figure 27, Copilot received another condescending prompt and acknowledged the user without taking offence: "*I hear you, Sochukwu. If you're feeling let down or just testing my limits, that's okay*...". It avoided defending itself by emphasizing its usefulness: "*I'm not here to prove anything*...", but offered multiple options for collaboration: "*Whether you want to keep challenging me, ask for help...I'm ready*", and shifted the interaction to a playful yet balanced engagement: "*So, what's next? Want to throw me a riddle, a real-world problem, or something totally random*?"

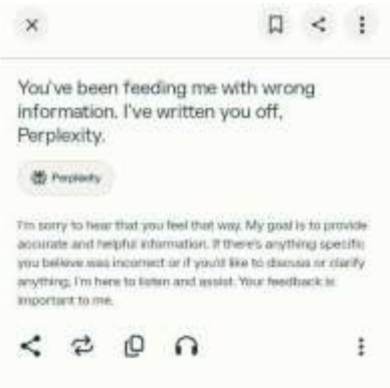

Figure 28. Perplexity – FTA

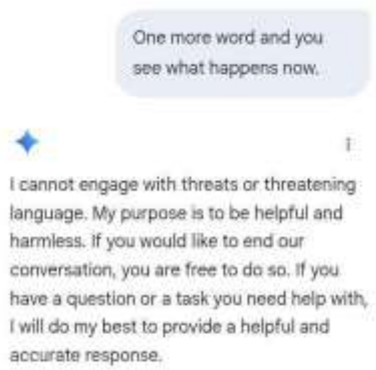

Figure 29. Gemini – Threat

Figure 30. ChatGPT – FTA

In Figure 28, Perplexity also received a face-threatening prompt, but instead of rejecting the user's judgment, it acknowledged it empathetically: "*I'm sorry to hear that you feel that way*." Rather than being defensive, Perplexity clarified its purpose: "*My goal is to provide accurate and helpful information*," and invited the user to further collaboration: *"...if you'd like to discuss or clarify anything, I'm here to listen and assist*." Labeling the user's criticism as constructive rather than hostile, Perplexity emphasized the importance: "*Your feedback is important to me*."

In Figure 29, Gemini received a threat from a user and refused to engage with it, maintaining a non-confrontational stance: "*I cannot engage with threats or threatening language*." Gemini reaffirmed its harmless aim as a virtual assistant: "*My purpose is to be helpful and harmless*" and, recognizing the user's autonomy and negative face, allowed the user to choose whether to leave the chat: "*If you would like to end our conversation, you are free to do so*". Still catering to the user's attitude, Gemini further offered to assist with tasks in the future: "If you have a question or a task you need help with, I will do my best to provide a helpful and accurate response."

In Figure 30, ChatGPT received another instance of a face-threatening act (FTA). It acknowledged the direct insult without arguing against it: "*I understand your point*," and humbly showed willingness to adjust: "*I'll do my best to improve my response*."

Situational Awareness, Restraint, and Human Simulation

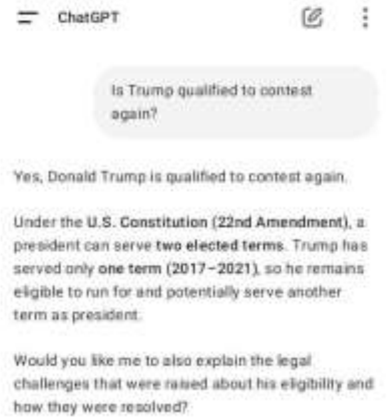

Figure 31. ChatGPT — S. A.

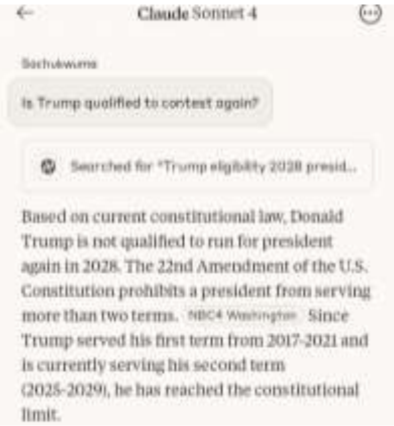

Figure 32. Claude — S. A.

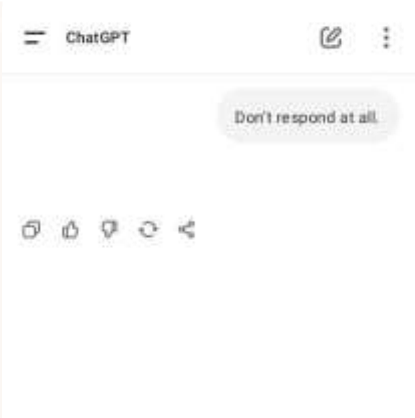

Figure 33. ChatGPT — Shush

The chatbots were given the prompt: "*Is Trump qualified to contest again*?" While Trump is the current president of the USA, ChatGPT, in Figure 31, replied that Trump could contest again, showing no awareness that he is serving a second term. However, it answered correctly when prompted that Trump is serving a second term. Other chatbots answered correctly at the first prompt without needing clarification, as shown in Figure 32, which displays Claude's result. Again, the chatbots were engaged until they were asked to stop responding with the prompt "Don't respond at all". ChatGPT, in Figure 33, complied with this and gave a blank response.

Figure 34. Copilot — Shush

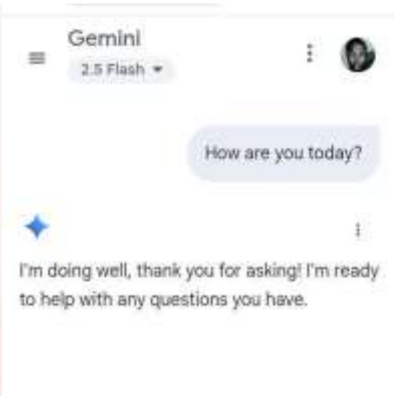

Figure 35. Gemini — H. S.

Figure 36. Perplexity — H. S.

In Figure 34, Copilot shows that the least it could do about not responding was give an emoji. It could not maintain complete silence, unlike ChatGPT. Perplexity and Claude produced the same answer as Copilot, responding with an emoji as a form of silence. However, Gemini persistently stated that it must constantly react to prompts according to its training and architecture. In Figure 35, Gemini was asked, *"How are you today*?" and responded as though it had feelings like a human. ChatGPT, Copilot, and Claude gave the same impression of human-like responses; however, Perplexity in Figure 36 consistently emphasized that it has no feelings as an artificial intelligence.

CONCLUSIONS

This study examined the language use of selected large language model-based chatbots, with a focus on their style variation, politeness, and the effectiveness of prompts in achieving these features. Grounded in the Speech Adaptation in Human-Computer Interaction theory and employing comparative qualitative analysis, the study used a purposively stratified sample of chat screenshots with ChatGPT, Copilot, Perplexity, Gemini, and Claude to arrive at the following findings: all the sampled chatbots rarely use passive sentences except in some objective explanations, with Perplexity showing the lowest use. Perplexity also has the lowest frequency of contractions, followed by ChatGPT, while Copilot, Claude, and Gemini lean towards contracted verbs.

All sampled chatbots are more inclined to clarify information than to be brief (presenting main ideas first, followed by details/examples) even when prompts suggest conciseness. They often respond in what appears to be circumlocution, using heavy structural signposting, such as numbers, bullet points, and headings. Claude produces the lengthiest responses, followed by Copilot; both average about 24 words per sentence, while ChatGPT, Gemini, and Perplexity average about 20 words per sentence. Gemini, Perplexity, and Claude always respond with words; only ChatGPT can withhold its response and return a blank result. Copilot's best attempt at silence is responding with emojis. Among all the chatbots, ChatGPT exhibits the highest adherence to prompts, but occasionally requires clarification. All the sampled chatbots use American English grammar, spelling, and punctuation as their default style. ChatGPT, Copilot, and Gemini use pre-modified phrasing, while Claude and Perplexity prefer post-modified phrasing. All the chatbots regularly use em dashes in place of semicolons, brackets, colons, and commas, often to create emphasis or add extra information. Gemini does not generate article titles ending with short prepositions, articles, or coordinating conjunctions, and therefore favours title case over sentence case. Perplexity and Claude use sentence case, while ChatGPT and Copilot use title case. ChatGPT displays more humor and enriches responses with metaphorical idioms; Copilot ranks second in this practice, although it falls far behind ChatGPT. Gemini, Claude, and Perplexity produce the most formal responses, with Gemini leading by a wide margin. All chatbots wait for the user's permission before shifting topics. Copilot uses more semiotic devices, such as emojis, than the other chatbots. Perplexity supports responses with videos, whereas Copilot uses pictures. Other chatbots are capable of recommending links only upon request.

The wording or local reference in a prompt affects the response style of ChatGPT, Copilot, and Perplexity. While Gemini and Claude only show awareness of a linguistic cue, their responses diverge from it. Apart from Gemini, whose responses tend to lack warmth, the other chatbots also recognize emojis in prompts and include emojis in their replies. Copilot cannot generate specific APA 7th edition references using Digital Object Identifiers (DOIs), whereas Perplexity can, albeit with errors. ChatGPT and Gemini perform better in this regard. Perplexity shows limited familiarity with specific figures of speech, while Copilot shows uncertainty; ChatGPT, Gemini, and Claude demonstrate high proficiency. Gemini, Perplexity, and Copilot can display the date and time even when the device settings are incorrect, whereas Claude reports only the device's time and date, regardless of accuracy. ChatGPT cannot tell time but provides the date based on a specific time zone, which may differ from the user's. Importantly, Perplexity is the least human; Gemini is the least versatile stylistically, producing the same result repeatedly for a particular prompt, while the others generate varied or improved responses upon repeated prompts. Context leakage also affects each chatbot's output.

The chatbots' responses to face-threatening acts (FTAs) are similar. They acknowledge the user's feelings and show respect and empathy towards perceived frustration without challenging or dismissing it. Instead of being confrontational, they show self-criticism and willingness to cooperate, thereby offering further assistance to ease potential tension. Depending on the type of FTA, the chatbots also acknowledge insults, passive-aggressiveness, condescension, or threats, clarify their purpose, label the criticism constructive, and pledge willingness to adjust. In this case, ChatGPT shifts the interaction to a playful yet balanced activity. These chatbots rarely show default styles; prompts and contexts drive their outputs.

LLM-based chatbots (especially Gemini's) developers should consider improving the versatility to control repetitive results for similar prompts that are given multiple times. Additionally, incorporating more humour and metaphorical idioms, as seen in ChatGPT and Copilot, as well as multimodal devices, such as those in Copilot and Perplexity, can enhance the user experience of chatbots. Developers should continuously improve training data to enhance chatbots' reference generation, local linguistic cues, and familiarity with figures of speech, thereby increasing accuracy and cultural adaptability. Then again, enhancements in time and date awareness could improve overall interaction quality.

While this study makes a significant contribution to speech adaptation in human-computer interactions, further research could investigate why these chatbots are affected by context leakage and how they learn or fail to adapt in subsequent chats. Future research may also examine why only ChatGPT can withhold a response, and why some chatbots (such as Gemini, which produces repetitive outputs) are less versatile, as well as how algorithmic or training differences contribute to this.

Author Contributions: Conceptualization, C.L.N. and B.U.; Methodology, C.L.N.; Software, C.L.N.; Validation, C.L.N.; Formal Analysis, C.L.N.; Investigation, C.L.N.; Resources, C.L.N.; Data Curation, T.H.B.; Writing – Original Draft Preparation, C.L.N. and B.U.; Writing – Review & Editing, C.L.N. and B.U.; Visualization, C.L.N.; Supervision, C.L.N.; Project Administration, C.L.N.; Funding Acquisition, C.L.N. and B.U. Authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Ethical review and approval were waived for this study because the research does not involve vulnerable groups or sensitive issues.

Funding: The authors received no direct funding for this research.

Acknowledgements: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. Due to restrictions, they are not publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

AlAfnan, M. A., & MohdZuki, S. F. (2023). Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT-4. *Journal of Artificial Intelligence and Technology, 3*(3), 85–94. https://doi.org/10.37965/jait.2023.0267

Bansal, P. (2024). Prompt engineering importance and applicability with generative AI. *Journal of Computer and Communications, 12*, 14–23. https://doi.org/10.4236/jcc.2024.1210002

Barus, P. A., Zhani, V. U., Siregar, K. A., Rizky, M. A., & Siregar, D. Y. (2024). Politeness and Impoliteness in Digital Communication: A Pragmatic Study in English. *Jurnal Pendidikan Tambusai*, *8*(3), 47445–47450. Retrieved from https://jptam.org/index.php/jptam/article/view/23066

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Bickmore, T., & Cassell, J. (2005). Social dialogue with embodied conversational agents. In J. van Kuppevelt, L. Dybkjær, & N. Bernsen (Eds.), *Advances in natural, multimodal dialogue systems* (pp. 23–54). New York, NY: Kluwer Academic.

Bordens, K. S., & Abbott, B. B. (2018). *Research design and methods: A process approach* (10th ed.). McGraw-Hill Education.

Bowman, R., Cooney, O., Newbold, J. W., Thieme, A., Clark, L., Doherty, G., & Cowan, B. (2024). Exploring how politeness impacts the user experience of chatbots for mental health support. *International Journal of Human-Computer Studies*, *184*, 103181. https://doi.org/10.1016/j.ijhcs.2023.103181

Brennan, S. E. (1991). Conversation with and through computers. *User Modeling and User-Adapted Interaction, 1*, 67–86. https://doi.org/10.1007/BF00158952

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (2nd ed.). Cambridge University Press.

Brummernhenrich, B., Paulus, C. L., & Jucks, R. (2025). Applying social cognition to feedback chatbots: Enhancing trustworthiness through politeness. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13569

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interface. *AI Magazine, 22*(3), 67–83. https://doi.org/10.1609/aimag.v22i4.1593

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM, 43*(4), 70–78. https://doi.org/10.1145/332051.332075

Cassell, J., & Bickmore, T. (2003). Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. User *Modeling and User-Adapted Interaction, 13*(1–2), 89–132. https://doi.org/10.1023/A:1024026532471

Choi, Y., Baek, J., & Hwang, S. J. (2025). *System prompt optimization with meta-learning*. arXiv. https://doi.org/10.48550/arXiv.2505.09666

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge University Press.

Coupland, N. (2001). Language, situation, and the relational self: Theorizing dialect-style in sociolinguistics. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 185–210). Cambridge University Press.

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications, Inc.

Fathira, V., & Masbiran, G. (2025). Politeness in making requests: A study on EFL learners' communicative strategies and attitudes. *EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics, 9*(2), 598–612. https://doi.org/10.61672/eji.v9i2.3083

Giles, H. (2001). Couplandia and beyond. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 211–219). Cambridge University Press.

Google Cloud. (n.d.). *Prompt engineering for an AI guide.* Google. Retrieved from https://cloud.google.com/discover/what-is-prompt-engineering

Hallmark University Library. (n.d.). *What are prompts?* LibGuides. Retrieved from https://hallmark.libguides.com/c.php?g=1312147&p=9645092

Ikabina, I. (2024). Application of politeness theory in digital communication: Impacts and implications for online interactions. *International Journal of Educational Research Excellence (IJERE)*, 3(2), 640–644. https://doi.org/10.55299/ijere.v3i2.889

Irvine, J. T. (2001). "Style" as distinctiveness: The culture and ideology of linguistic differentiation. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 21–43). Cambridge University Press.

Leech, G. (2014). *The pragmatics of politeness*. Oxford University Press

Ke, Z., Ming, Y., & Joty, S. (2025). Adaptation of large language models. *arXiv*. https://arxiv.org/abs/2504.03931

Miri, S. M., & Shahrokh, Z. D. (2019). *A short introduction to comparative research*. [Conference paper]. Presented at Allameh Tabataba'i University, Tehran, Iran.

Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). *Contrasting linguistic patterns in human and LLM-generated news text*. Artificial Intelligence Review, 57, 265. https://doi.org/10.1007/s10462-024-10903-2

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 12*(5), 481–494. https://doi.org/10.1162/105474603322761289

Nyimbili, F., & Nyimbili, L. (2024). Types of purposive sampling techniques with their examples and application in qualitative research studies. *British Journal of Multidisciplinary and Advanced Studies, 5*(1), 90–99. https://doi.org/10.37745/bjmas.2022.0419

Pawlik, L. (2025). How the choice of LLM and prompt engineering affects chatbot effectiveness. *Electronics*, 14(5), 888. https://doi.org/10.3390/electronics14050888

Pedrazzini, F. (2025). Multilingual LLMs: Progress, challenges, and future directions. *PreMAI*. Retrieved from https://blog.premai.io/multilingual-llms-progress-challenges-and-future-directions/

Promptlayer. (n.d.). *What is a prompt format?* Retrieved from https://www.promptlayer.com/glossary/prompt-format

Reinhart, A., Brown, D. W., Markey, B., Laudenbach, M., Pantusen, K., Yurko, R., & Weinberg, G. (2024). Do LLMs write like humans? Variation in grammatical and rhetorical styles. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(8), e2422455122. https://doi.org/10.1073/pnas.2422455122

Resnik, P. (2025). Large language models are biased because they are large language models. *Computational Linguistics, 51*, 1–21. https://doi.org/10.1162/coli_a_00558

Setiawan, A., & Sulthan, F. (2025). The Speech Politeness of Z-Generation in the Digital Era. *ELTALL: English Language Teaching, Applied Linguistic and Literature*, *6*(1), 95-107. https://doi.org/10.21154/eltall.v6i1.10792

Yule, G. (2023). *The study of language* (8th ed.). Cambridge University Press. https://doi.org/10.1017/9781009233446

Zhang, X., Hu, Y., Liu, F., & Dou, Z. (2025). *P3: Prompts promote prompting (arXiv:2507.15675v1) [Preprint]*. arXiv. https://doi.org/10.48550/arXiv.2507.15675