

Pragmatic Analysis of AI-Generated Conversations in A Simulated Financial Services Setting

Jonathan Yi-Chen Cheng^{1*}, Tzu-Chuan Huang²

¹ Taipei European School, Taipei, Taiwan

² Labi Education, Taoyuan, Taiwan

*Corresponding author E-mail: jonathan.chengb2028@stu.tes.tp.edu.tw

Abstract

Advances in natural language processing have led to the development of AI-based chatbots; however, their performance in specialized domains remains underexplored. Pragmatics involves inferring meaning from situational, cultural, and interpersonal cues. Currently, there is limited research on the pragmatic competence of chatbots in the financial sector. In this study, we investigated ChatGPT-generated conversations simulating bank account inquiries to evaluate pragmatic coherence, user-friendliness, and similarities to human communication. Thirty discrete utterances across nine conversational turns were qualitatively assessed for adherence to Grice's Maxims, pragmatic encoding, inference processes, and formality. Our analysis revealed that ChatGPT consistently decoded conversational intent and effectively delivered contextually appropriate content, although responses were occasionally verbose, repetitive, or overly formal. In addition, a user-opinion experiment was conducted to compare original and pragmatically modified responses with contextual and social features removed. Notably, the removal of pragmatic cues did not significantly affect participants' perceptions of naturalness or authenticity. Overall, current AI can implement certain pragmatic adaptations to interpret user intentions and contextual relevance. However, in specific contexts such as financial services, achieving a balance between communicative efficiency and naturalness remains an ongoing challenge. Our results suggest exploratory evidence that targeted pragmatic editing can nudge perceptions in specific contexts.

Keywords: Pragmatic Analysis; Natural Language Processing; Chatbot; Financial Services; Grice's Maxims.

1. Introduction

Artificial intelligence (AI) has become increasingly integrated into multiple professional domains, including healthcare, education, and business operations [1], [2]. One of the most visible applications is conversational systems, or chatbots, which serve as personal assistants or conversational partners. Chatbots can be classified into three main types: rule-based chatbots, retrieval-based chatbots, and generative chatbots, which dynamically generate context-sensitive responses [3]. These systems rely on large databases with internet access and benefit from enhanced computational power, now demonstrating remarkable fluency and contextual awareness. Although general guidelines exist for developing chatbots, they may not fully capture certain aspects of human speech [4]. Additionally, chatbots may struggle to recognize the nuanced intentions of their users, which can lead to frustration.

In everyday conversations, many language principles are employed instinctively, without premeditation or conscious planning. These principles make human language difficult to imitate and learn. In recent years, analytical chatbots have implemented a system known as the natural language interface, which facilitates intuitive human-computer interactions by translating user intentions into controls and commands [5]. This interface is built on advancements in natural language processing, a subfield of AI that enables chatbots to comprehend, interpret, and respond to humans naturally [6]. The natural language interface allows users to make direct and personalized requests, such as "How's the weather today?" instead of memorizing specific commands. This process involves multiple layers of processing, including text parsing, where the system breaks down input into individual parts of speech and analyzes their structure and relationships. It then undergoes another stage of natural language understanding, which is trained through machine learning to help chatbots recognize patterns and make predictions or assumptions based on prior knowledge and examples.

A natural language interface consists of three main components: syntax analysis, semantic analysis, and pragmatics. While syntactic and semantic analyses focus on grammatical and literal meanings, pragmatic analysis involves understanding unspoken user intentions, grasping contextual relevance, and identifying ambiguity and implicit meanings [7]. ChatGPT, which stands for Chat Generative Pre-trained Transformer, represents a state-of-the-art generative AI. Compared to earlier chatbots, ChatGPT has rapidly evolved from a research tool into a versatile and articulate conversational AI suitable for real-world use [8]. A recent study assessing ChatGPT's performance on medical licensing exams in the UK and the USA revealed its ability to answer novel questions and pass these professional exams [9]. Given the growing use and continuous improvement of large language model-based chatbots, a deeper examination of their comprehension capabilities in various everyday contexts is warranted.

The adoption of chatbots in the financial sector has been relatively slow in comparison to other fields, such as the creative and academic domains [10]. However, AI offers the banking industry unprecedented advantages, including considerable increases in operational

efficiency and revenue [11]. As such, major financial institutions are actively pursuing AI-driven transformations. By contrast, chatbots remain constrained in their capabilities and in interpreting customer queries [12]. Predictably, human customer service agents are still preferred by consumers over chatbots [13]. This preference may stem from the complexities of pragmatic understanding during financial service interactions, where AI systems must navigate indirect speech, implicature, politeness strategies, and user intentions. Currently, studies addressing the pragmatic competence of chatbots are limited. Understanding how chatbots handle pragmatic interpretation can inform their safe and effective deployment in professional domains.

In this study, we conducted a pragmatic analysis of AI-generated conversations in a simulated financial services environment. Specifically, we examined ChatGPT's adherence to Grice's Maxims, its use of pragmatic encoding and inference, and its resemblance to human communicative norms. In addition, a user-opinion experiment was conducted to assess how pragmatic modifications influence perceived naturalness and appropriateness.

2. Methods

To test the similarity between human and AI-generated text, we used the ChatGPT-4o version. Our study was divided into two parts (Figure 1). We conducted virtual conversations by asking a series of questions about opening a bank account. Each response consisted of multiple utterances structured into sentences. In total, 30 utterances were generated across 9 turns.

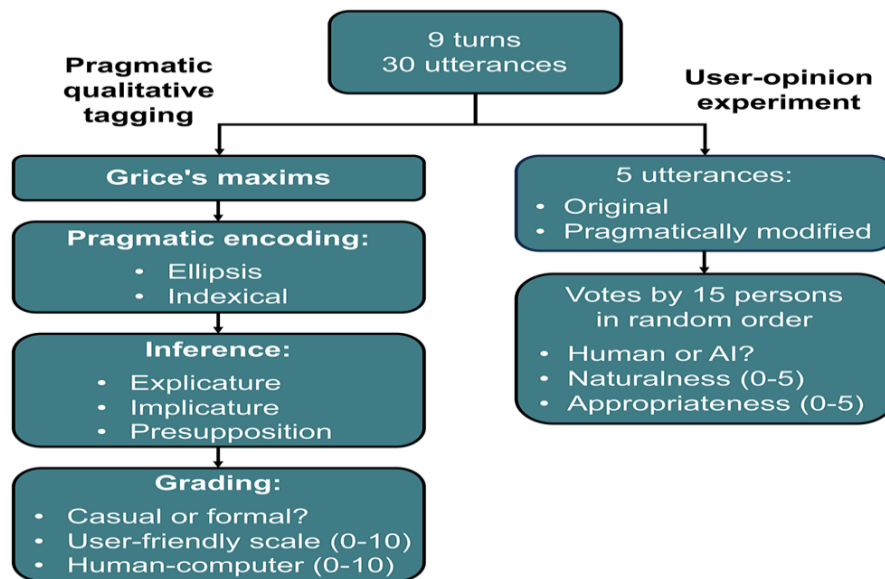


Fig. 1: Study Flowchart.

The first part of the study involved a set of pragmatic qualitative analyses of the 30 utterances. We began by examining whether these responses violated Grice's Maxims. Generally, AI incorporates many principles of human speech into its replies, including Grice's Cooperative Principle [14], which outlines how human conversational exchanges are organized to be efficient and effective. Grice divided his principles into four conversational maxims: the Maxim of Quantity, the Maxim of Quality, the Maxim of Relation, and the Maxim of Manner [15]. These maxims were later modified by Setlur and Tory for chatbot design [16]. It is believed that Grice's Maxims help chatbots formulate responses effectively and coherently.

The next assessment focused on pragmatic encoding processes, specifically ellipsis and indexicals [17]. An ellipsis is the omission of words to avoid unnecessary repetition while maintaining clarity in context. Indexicals are expressions whose meanings depend on the context or the referent identified by the user. The third assessment concentrated on inferring the user's speech, encompassing explicature, implicature, and presupposition [18]. Explicature enriches or clarifies the literal meaning of an utterance through context. Implicature refers to what is implied by the context or suggested by the user but not explicitly stated. Presupposition involves background assumptions that are mutually understood by both the user and the chatbot.

The last assessment aimed to evaluate formality, user-friendliness, and similarity to human responses. Given that the primary purpose of this chatbot is to serve as a personal assistant, its user-friendliness was measured by assessing its understanding of pragmatics and adherence to Grice's Maxims on a scale from 0 to 10. Additionally, a human-computer scale (also from 0 to 10) evaluated the chatbot's understanding of human pragmatics and its ability to generate human-like responses, utilizing key principles of human language and literature.

In the second part of the study, we conducted experimental pragmatics to assess the perceptual significance of pragmatic features [19]. A user-opinion experiment was carried out with ordinary individuals who were unaware of the results of the previous analysis to prevent subjectivity and bias in the pragmatic annotations. Five representative utterances were selected, and ChatGPT's responses were transformed into pragmatically modified forms that excluded certain pragmatic features. These modifications involved stripping away contextual and social layers while retaining literal, syntactic, and semantic accuracy. The main framework involved removing implicatures, deixis, and indirect speech acts while avoiding mechanical tones. Two investigators (JYCC and TCH) collaboratively reconstructed these modifications and reached a consensus through discussion.

Fifteen fluent English-speaking participants, with no professional backgrounds in computer science or linguistics, were randomly assigned different versions of the same sentence. They were recruited to ensure gender balance and asked to identify which sentence was generated by AI and which was produced by a human, all in a blind manner. Additionally, participants were asked to rate the naturalness and appropriateness of each response on a scale from 0 to 5, with higher scores indicating better quality.

3. Results

Analyzing 30 utterances across 9 turns revealed that ChatGPT typically responded to questions with a consistent structure. First, it provided a brief general answer, followed by a list of relevant information. Notably, ChatGPT included multiple examples or details, formatted as bullet points or in a numbered list (Figure 2). After the list, a summary concluded the response. Additionally, ChatGPT included a turn-taking cue to indicate that it had finished speaking. This structure enhanced clarity and reduced confusion by summarizing the information. The turn-taking cue signaled that all information had been fully presented, allowing the user to ask follow-up questions.

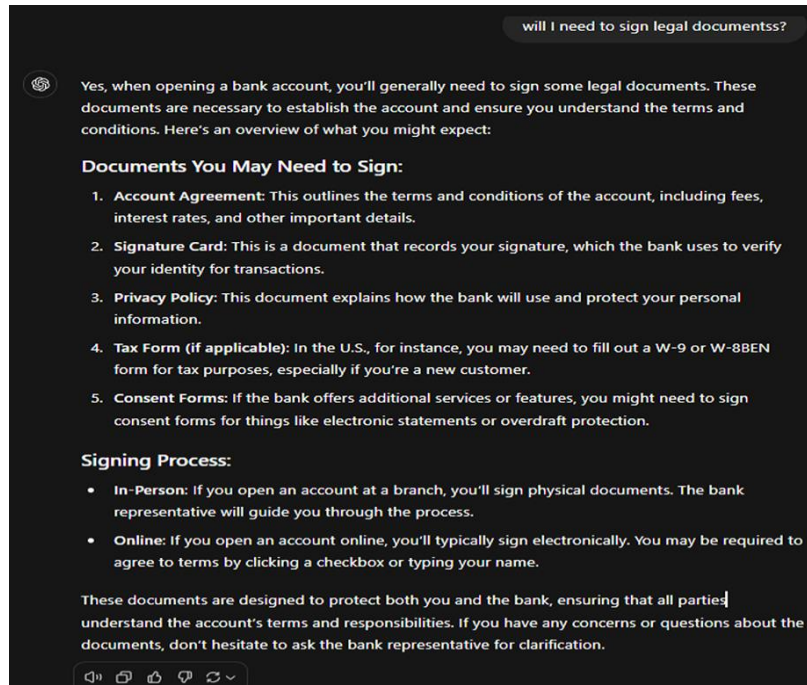


Fig. 2: Sample Utterances Generated by ChatGPT.

In the pragmatic qualitative analyses, compliance rates for the Maxims of Quantity, Quality, Relation, and Manner were 67%, 100%, 93%, and 47%, respectively (Figure 3). ChatGPT consistently adhered to the Maxims of Quality and Relation, as it is designed to stay on topic while accessing the internet to retrieve necessary and accurate information. Although the four Maxims were generally followed, some violations did occur.

- User: Will I need to sign legal documents?
- Chatbot: ... Here's an overview of what you might expect.

In this case, the Maxim of Manner was violated because the sentence was overly complicated. A transition into a list should be short and clear; the phrase "an overview of" is unnecessary and detracts from clarity. Occasionally, excessive repetition also violates the Maxim of Quantity, reiterating the same meaning without providing new information and offering more content than necessary to convey the message effectively. Despite these occasional violations of verbosity or repetitive phrasing, they are inconspicuous without close reading and analysis.

- User: What are some bank suggestions?
- Chatbot: Here are some popular banks and financial institutions known for their savings accounts, along with some key features to consider.

The sentence literally informs the user about a list of banks and financial institutions along with their key features. However, contextual enrichment, it implies that the user is expected to identify the specific banks and financial institutions mentioned, as well as their key features. In this regard, ChatGPT successfully employed explicature to contextualize user intent.

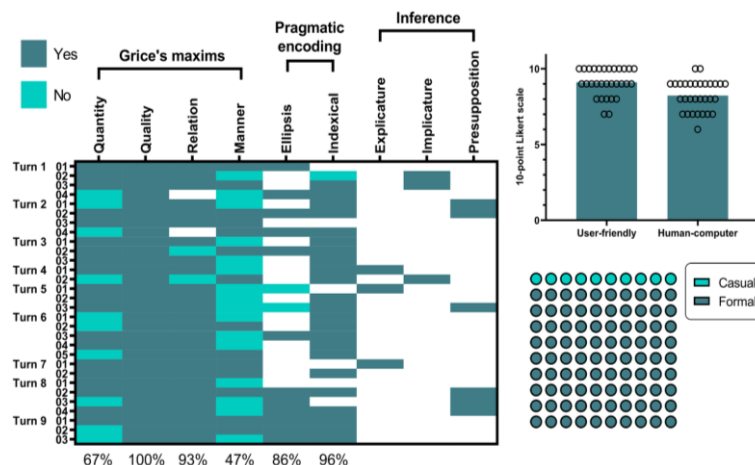


Fig. 3: Pragmatic Qualitative Tagging. Bars Represent the Means.

Remarkably, ChatGPT has consistently demonstrated the ability to decode and interpret the underlying messages in the user's questions while maintaining topical relevance and factual reliability. For instance:

- User: Can you teach me how to open a bank account?
- Chatbot: Sure! Opening a bank account is usually a straightforward process.

In this instance, instead of interpreting the request literally as an inquiry about its teaching ability, ChatGPT understood that the user was seeking assistance. Not only could it decode natural language, but it could also encode it in its responses.

- Chatbot: If you have any specific questions or need help with a particular step, feel free to ask!

This response makes use of ellipsis and indexical references. However, it omits specific details, making it difficult to understand without context. The response does not specify what the "specific question" or "particular step" refers to. Furthermore, the use of ellipses may occasionally feel unnatural and produce ambiguity. In the following, the phrase "dive in" is unclear, as there is insufficient context to support its reference.

- User: I will consider it in the future.
- Chatbot: When you're ready to dive in, you'll have a better idea of what fits your needs best.

Notably, the formality used by ChatGPT varies depending on the nature of the question. If the user begins with a casual greeting, such as "Hello!" or "Hi!"—with exclamation marks indicating a relaxed tone—the response is casual, featuring phrases like "Sure!" or "Feel free to ask!" As the dialogue continues and questions become more direct, ChatGPT adopts a more formal tone. It also occasionally offers suggestions and encouragement to take action.

- User: What are the downsides?
- Chatbot: Some goals, especially short-term or emergency savings, a savings account is a great option.

In this example, ChatGPT indirectly suggests what it believes the user should do by expressing its view on what it considers a good option. While having an opinion is normal for humans, AI is designed to remain as neutral as possible. Although offering an opinion can introduce bias, it is noteworthy that AI can make rational suggestions based on the available facts and information.

Overall, ChatGPT typically strives to be user-friendly by providing accurate and detailed information. At times, overly formal phrasing can lead to awkwardness. For instance, when asked about opening a bank account, a response like, "Yes, when opening a bank account, you'll generally need to sign some legal documents," clarifies the situation but is too wordy for users seeking a quick answer. Lengthy responses and extensive lists of examples can obscure the main point, resulting in unclear communication. Additionally, ChatGPT lacks a distinct personality; it consistently uses a monotone tone that, while formal, is not engaging and can become dull over time.

In the user-opinion experiment, most respondents identified the original versions as AI (Figure 4). In some cases, pragmatic modifications enhanced the human-like qualities. For instance:

- User: Hello! Can you teach me how to open a bank account?
- Original: Sure! Opening a bank account is usually a straightforward process.
- Modified: Of course! It is usually a pretty straightforward process.

The original versions of ChatGPT's responses demonstrated a moderate level of naturalness, eliciting mixed opinions from respondents. Some found these AI-generated conversations to be natural, as they struck a balance between oversimplification and excessive seriousness. However, for others, while it was not entirely clear that the original version was AI-generated, the sentences did not sound as natural as typical human speech. Unexpectedly, removing pragmatics did not increase the likelihood of identifying a sentence as AI-generated, although we cannot rule out the possibility that the lack of a significant difference was due to the small sample size. Interestingly, while the modified versions sometimes appeared less appropriate than the original, their naturalness was not inferior and, in some cases, even surpassed that of the original AI-generated version. In these instances, ChatGPT's original responses may be perceived as more formal due to the removal of politeness markers in the modified version.

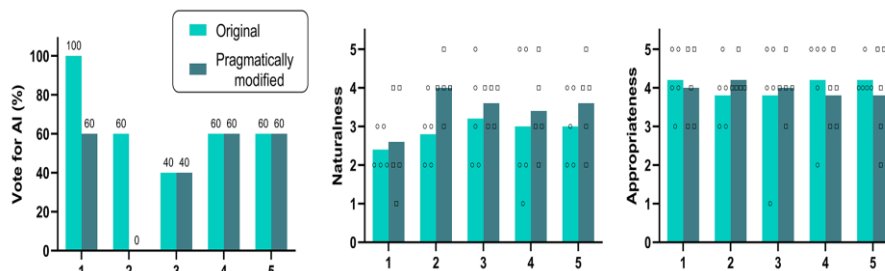


Fig. 4: User-Opinion Experiment. Bars Represent the Proportions or Means.

4. Discussion

In this concise yet detailed analysis of conversations between ChatGPT and the user, we observed distinct response structures and subtle differences in language use compared to human interaction. While ChatGPT's responses are not entirely unnatural, it is important to recognize the purpose of a chatbot [20]. For rapidly evolving chatbots like ChatGPT, achieving convenience and user-friendliness while adhering to the principles of literature and human nature can be challenging. Striking a balance through pragmatic adjustments can enhance the efficiency and effectiveness of responses.

Our study confirmed ChatGPT's communicative and social capacities, demonstrating its ability to effectively interpret indirect and contextual cues with strong pragmatic awareness. Our findings are consistent with a recent study indicating that participants could not distinguish between conversations generated by ChatGPT and those written by humans [21]. In the present study, we found that the four Maxims were largely adhered to, and AI-generated conversations received high scores for user-friendliness and human-likeness. While two additional maxims, benevolence and transparency, have been proposed [22], they are less relevant to the financial services sector.

In contrast to casual small talk, special considerations should be taken into account for conversations in the financial services setting. Chatbots in finance present unique challenges, including ethical concerns, data privacy risks, and a lack of transparency and accountability in decision-making [23]. In financial settings, where time efficiency and factual accuracy are paramount, dialogues must balance efficiency with customer care. This balance is referred to as "rationalized politeness" in highly professional environments such as call centers [24]. In practice, efficiency often takes priority in transactional conversations. This may explain our intriguing observation that some respondents

perceived pragmatic modifications as more human-like, suggesting that excessive formal politeness may sometimes undermine conversational authenticity. However, this does not imply that politeness strategies can be completely discarded to meet efficiency targets. Our study supports the concept that interaction conventions are shaped by specific contextual considerations. The financial context can also be influenced by geographic and cultural factors. For instance, American and Chinese banks may exhibit different thematic orientations and communication styles [25].

For a chatbot, the Maxim of Manner highlights the importance of avoiding redundancy, repetition, and excessive length. In this study, we found that ChatGPT sometimes provided lengthy, repetitive, or overly complex responses—likely a result of its optimization for informativeness rather than conciseness. However, seemingly excessive information can be acceptable if it "ensures that the response addresses the intent in the question" [16]. The apparent conflict may arise from short passages of conversations in our analysis, where ChatGPT did not quickly grasp the user's character. In reality, external cues such as appearance and accents play a crucial role. To adhere to the Maxim of Quantity without understanding the context, ChatGPT may endeavor to deliver unnecessary or superfluous information in its responses. This issue was noted in a recent report showing that ChatGPT often performed unrequested tasks, leading to undesirable stylistic effects [26]. Such excess can create communication barriers in real-life interactions.

Our findings of a relatively monotone tone in AI-generated conversations are consistent with a previous report [27], which showed that ChatGPT exhibited a more uniformly positive tone, while humans displayed a broader range of emotions. In the financial services context of this study, the emotional differences between ChatGPT and human responses were more subtle, and the narrative monotony was less bothersome for respondents in the user-opinion experiment. However, as a general-purpose AI agent, it is essential for ChatGPT to understand emotions in opinions, feedback, and dialogues to provide emotionally aware responses [28]. Although ChatGPT still needs to overcome obstacles in determining the appropriateness of short input texts for making accurate inferences, recent findings suggest that ChatGPT-4 can reasonably infer personality traits from written text [29]. Future models may incorporate adaptive pragmatic tuning, adjusting tone, formality, and brevity based on user profiles and contextual circumstances.

This study has several caveats and limitations. The database used for analysis is relatively small, and a larger study should aim to include more conversational variations and a broader range of opinions. We plan to interview additional respondents to gain a better understanding of their reasoning. Furthermore, conducting parallel procedures with different large language models would allow us to compare their similarities and differences. Finally, incorporating multilingual or cross-cultural comparisons would strengthen the future outlook of this research.

5. Conclusion

In summary, our pragmatic analysis revealed that current AI-generated conversations in simulated financial services are user-friendly, exhibiting a high level of naturalness and appropriateness. The chatbot generally adheres to Gricean principles and demonstrates effective inferential reasoning, although it occasionally produces responses that are overly formal or verbose. Pragmatic modifications do not significantly influence users' perceptions of authenticity or naturalness, suggesting that while these cues enhance nuance, they are not the sole determinants of perceived humanness in financial service interactions. For domain-specific AI applications, optimizing the balance between pragmatic richness and communicative efficiency remains an ongoing challenge.

References

- [1] Yang CH. How artificial intelligence technology affects productivity and employment: firm-level evidence from Taiwan. *Res Policy*. 2022;51(6):104536. <https://doi.org/10.1016/j.respol.2022.104536>.
- [2] Enholm IM, Papagiannidis E, Mikalef P, Krogstie J. Artificial intelligence and business value: a literature review. *Inf Syst Front*. 2021;24:1709-1734. <https://doi.org/10.1007/s10796-021-10186-w>.
- [3] Alharbi N, Ud Din F, Paul D, Sadgrove E. Driving AI chatbot adoption: a systematic review of factors, barriers, and future research directions. *J Open Innov Technol Mark Complex*. 2025;11(3):100590. <https://doi.org/10.1016/j.joitmc.2025.100590>.
- [4] Adamopoulou E, Moussiades L. Chatbots: history, technology, and applications. *Mach Learn Appl*. 2020;2:100006. <https://doi.org/10.1016/j.mlwa.2020.100006>.
- [5] Zhang G. AI linguistics. *Nat Lang Process J*. 2025;10:100137. <https://doi.org/10.1016/j.nlp.2025.100137>.
- [6] Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc*. 2011;18(5):544-551. <https://doi.org/10.1136/amiajnl-2011-000464>.
- [7] Cambria E. *Understanding natural language understanding*. Switzerland: Springer Nature; 2025.
- [8] Alomari EA. Unlocking the potential: a comprehensive systematic review of ChatGPT in natural language processing tasks. *Comput Model Eng Sci*. 2024;141(1):43-85. <https://doi.org/10.32604/cmescs.2024.052256>.
- [9] Newton PM, Summers CJ, Zaheer U, Xiromeriti M, Stokes JR, Bhangu JS, Roome EG, Roberts-Phillips A, Mazaheri-Asadi D, Jones CD, Hughes S, Gilbert D, Jones E, Essex K, Ellis EC, Davey R, Cox AA, Bassett JA. Can ChatGPT-4o really pass medical science exams? A pragmatic analysis using novel questions. *Med Sci Educ*. 2025;35:721-729. <https://doi.org/10.1007/s40670-025-02293-z>.
- [10] Bouteraa M, Chekima B, Thurasamy R, Bin-Nashwan SA, Al-Daihani M, Baddou A, Sadallah M, Ansar R. Open innovation in the financial sector: a mixed-methods approach to assess bankers' willingness to embrace Open-AI ChatGPT. *J Open Innov Technol Mark Complex*. 2024;10(1):100216. <https://doi.org/10.1016/j.joitmc.2024.100216>.
- [11] Kshetri N. Generative artificial intelligence in the financial services industry. *Computer*. 2024;57(6):102-108. <https://doi.org/10.1109/MC.2024.3382452>.
- [12] Graham G, Nisar TM, Prabhakar G, Meriton R, Malik S. Chatbots in customer service within banking and finance: do chatbots herald the start of an AI revolution in the corporate world? *Comput Hum Behav*. 2025;165:108570. <https://doi.org/10.1016/j.chb.2025.108570>.
- [13] Sugumar M, Chandra S. Do I desire chatbots to be like humans? Exploring factors for the adoption of chatbots for financial services. *J Int Technol Inf Manage*. 2021;30(3):38-77. <https://doi.org/10.58729/1941-6679.1501>.
- [14] Grice HP. Logic and conversation. In: Cole P, Morgan JL. *Syntax and semantics, volume 3: speech acts*. New York: Academic Press; 1975. pp. 41-58. https://doi.org/10.1163/9789004368811_003.
- [15] Grice P. *Studies in the way of words*. Cambridge: Harvard University Press; 1989.
- [16] Setlur V, Tory M. How do you converse with an analytical chatbot? Revisiting Gricean Maxims for designing analytical conversational behavior. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*; 2022, 1-17. <https://doi.org/10.1145/3491102.3501972>.
- [17] Predelli S. Hybrid indexicals and ellipsis. *Erkenntnis*. 2006;65:385-403. <https://doi.org/10.1007/s10670-006-9010-5>.
- [18] Reboul A. Truthfully misleading: truth, informativity, and manipulation in linguistic communication. *Front Commun*. 2021;6:646820. <https://doi.org/10.3389/fcomm.2021.646820>.
- [19] Ghafar ZN. The experimental pragmatics: a review of language constructing and interpreting. *J Lang Pragmat Stud* 2023;2(2):89-99. <https://doi.org/10.58881/jlps.v2i2.30>.

- [20] Luo B, Lau RYK, Li C, Si YW. A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdiscip Rev Data Min Knowl Discov.* 2022;12(1):e1434. <https://doi.org/10.1002/widm.1434>.
- [21] Chen X, Li J, Ye Y. A feasibility study for the application of AI-generated conversations in pragmatic analysis. *J Pragmat.* 2024;223:14-30. <https://doi.org/10.1016/j.pragma.2024.01.003>.
- [22] Miehl E, Nagireddy M, Sattigeri P, Daly EM, Piorkowski D, Richards JT. Language models in dialogue: conversational maxims for human-AI interactions. In: Al-Onaizan Y, Bansal M, Chen YN. *Findings of the association for computational linguistics: EMNLP 2024*. Stroudsburg, PA: Association for Computational Linguistics; 2024. pp. 14420-14437. <https://doi.org/10.18653/v1/2024.findings-emnlp.843>.
- [23] Adegbite MA. The intersection of AI, banking and fintech in seamless financial services. *World J Adv Res Rev.* 2025;25(3):1516-1526. <https://doi.org/10.30574/wjarr.2025.25.3.0613>.
- [24] Hultgren AK. Vocatives as rationalized politeness: theoretical insights from emerging norms in call centre service encounters. *J Socioling.* 2017;21(1):90-111. <https://doi.org/10.1111/josl.12224>.
- [25] Sun Y, Wang G, Ren H. To entertain or to serve: Chinese and US banks' online identity based on a genre analysis of social media. *IEEE Trans Prof Commun.* 2021;64(2):121-136. <https://doi.org/10.1109/TPC.2021.3064395>.
- [26] Andersson M, McIntyre D. Can ChatGPT recognize impoliteness? An exploratory study of the pragmatic awareness of a large language model. *J Pragmat.* 2025;239:16-36. <https://doi.org/10.1016/j.pragma.2025.02.001>.
- [27] Fatahi S, Vassileva J, Roy CK. Comparing emotions in ChatGPT answers and human answers to the coding questions on Stack Overflow. *Front Artif Intell.* 2024;7:1393903. <https://doi.org/10.3389/frai.2024.1393903>.
- [28] Kusal S, Patil S, Choudrie J, Kotecha K, Vora D, Pappas I. A systematic review of applications of natural language processing and future challenges with special emphasis in text-based emotion detection. *Artif Intell Rev.* 2023;56:15129-15215. <https://doi.org/10.1007/s10462-023-10509-0>.
- [29] Piastra M, Catellani P. On the emergent capabilities of ChatGPT 4 to estimate personality traits. *Front Artif Intell.* 2025;8:1484260. <https://doi.org/10.3389/frai.2025.1484260>.