Investigating the Role of Metaphors in Human-AI Conversation and Engagement

Pragati Maheshwary
Human-Computer Interaction Institute
Carnegie Mellon University
United States
pragati2@andrew.cmu.edu

Qiawen Liu
Department of Psychology
University of Wisconsin-Madison
United States
qliu295@wisc.edu

Rui Mao
College of Computing and Data Science
Nanyang Technological University
Singapore
rui.mao@ntu.edu.sg

***Abstract*—Metaphorical expressions are widely used in daily communication between humans to improve the understanding of both complex and abstract concepts. However, it is unclear whether the usage of metaphorical language is helpful for enhancing the conversational engagement between humans and AI. In this work, we leverage a state-of-the-art computational metaphor processing tool to gain insights from human and ChatGPT conversations. Our quantitative analysis finds that although metaphors may enhance the quality of human-AI interactions, they do not directly lead to higher levels of conversational engagement, which is measured by the duration of the conversation. However, regression analysis shows a notable relationship between user and ChatGPT metaphor usage, suggesting that ChatGPT is adept at reflecting the linguistic style of users, especially in terms of metaphorical language. Additional topic modeling and concept mapping analyses further explored the patterns of metaphorical language across various engagement levels and topics in user and ChatGPT messages.**

***Index Terms*—Metaphors, ChatGPT, Conversational Engagement, MetaPro, Human-AI Interaction**

I. Introduction

Metaphors are fundamental to human cognition, allowing abstract ideas to be expressed in more relatable, concrete terms [1]. As framed by Lakoff and Johnson [2], metaphors shape our conceptual systems, influencing how we think, communicate, and understand the world. While metaphors have been extensively studied in human-human interactions, the rise of Artificial Intelligence (AI) and Machine Learning (ML) opens new avenues for exploring metaphor usage in human-AI interactions [3], [4]. Large Language Models (LLMs) in the present time demonstrate the ability to engage in human-like dialogue by drawing on a vast training dataset of billions of words. In this context, language models like ChatGPT, a revolutionary creation by OpenAI, serve as a compelling platform for investigating how AI can utilize metaphors to facilitate more effective and meaningful communication.

While previous research has focused on ChatGPT's language processing capabilities [5], including its accuracy and proficiency in handling conceptual metaphors [6], [7], there remains a gap in understanding how these metaphors impact user engagement and communication dynamics. Metaphors, with their inherent emotional and persuasive power [8]–[11], have been shown to enhance communication in contexts like advertising [12] and political discourse [13]. However, little is known about their role in human-AI interactions and whether AI-generated metaphors resonate with users in the same way they do in human conversations.

This study addresses this gap by investigating two key aspects of metaphor usage in human-AI communication. First, we examine whether the inclusion of metaphors in ChatGPT's responses influences user engagement, building on research that highlights the effectiveness of metaphorical language in various communicative contexts [14]. Second, we explore whether ChatGPT reflects the metaphorical language introduced by users, a phenomenon aligned with theories of conversational mirroring and alignment [15]. These theories suggest that linguistic adaptation can support rapport development and enhance empathy, critical elements in fostering deeper engagement.

Understanding how ChatGPT uses and mirrors metaphors is not only important for improving user experience but also contributes to broader discussions about the adaptability of AI language models. Previous research has shown that metaphor usage affects perceptions of human-AI collaboration [16], yet the role of metaphors in driving user engagement in AI systems remains underexplored. This study aims to fill that void by introducing a novel concept mapping analysis to uncover deeper patterns in metaphorical language in human-AI conversations. *Concept mappings*, which describe how abstract ideas are structured and communicated through metaphors [2], allow us to investigate the relationships between different metaphorical constructs in both ChatGPT and user responses. By examining metaphorical patterns through concept mappings, we aim to shed light on how metaphorical complexity might contribute to more exploratory or reflective conversations.

Overall, this research extends the study of metaphor usage from human-human communication into the realm of AI, focusing on how AI models like ChatGPT use metaphors to engage users and align with their conversational styles. By incorporating concept mapping and exploratory analyses, we offer a fresh perspective on the dynamic role of metaphorical language in fostering meaningful human-AI interactions.

2375-9259/24/$31.00 ©2024 IEEE
DOI 10.1109/ICDMW65004.2024.00066

II. Related Work

The use of metaphors in AI-mediated communication, particularly within conversational agents like ChatGPT, has been primarily focused on the technical aspects of language processing. Studies have explored the capacity of AI to understand and generate metaphors, assessing models on their linguistic versatility and precision in metaphor detection and response. Wachowiak and Gromann [7] found that, while large language models like GPT-3 are capable of generating appropriate metaphorical mappings, they sometimes struggle with maintaining context relevance when generating metaphoric responses in extended dialogues. Additionally, research by Mao et al. [17] has shown advancements in AI's capability to interpret complex metaphorical language, aiming for a deeper understanding that goes beyond surface-level text processing. These efforts have significantly contributed to enhancing the linguistic capabilities of conversational agents, enabling them to participate in more nuanced and contextually rich interactions.

Research on human-human interaction has consistently highlighted the role of metaphors in enhancing communication effectiveness, emotional resonance, and persuasion across various domains such as advertising [12] and political discourse [18]. Specifically, Prabhakaran et al. [18] conducted a large-scale, topic-agnostic study using neural networks to detect metaphors in over 85,000 Facebook posts made by US politicians. Their analysis revealed that metaphor use not only varies with ideological leanings but also significantly impacts audience engagement. They found that posts containing metaphors elicited more responses and deeper engagement from the audience compared to non-metaphorical content. However, the application of these findings to human-AI interaction remains limited. Studies like those by Khadpe et al. [16] begin to bridge this gap by examining how metaphors can shape perceptions of AI, yet they primarily focus on user perceptions post-interaction, rather than real-time engagement dynamics.

Theories related to linguistic adaptation, such as those proposed in conversational mirroring and alignment [19], suggest that mimicking a user's linguistic style can enhance rapport and empathy in human-AI interactions [16]. This aspect of communication has been studied within the context of improving user experience by adapting AI responses to match user language, thus fostering a smoother and more engaging conversational flow. However, the specific impact of metaphorical language mirroring on user engagement and the development of rapport in ongoing AI interactions has not been extensively explored.

Building on the foundation laid by these studies, this research seeks to address the noted gaps by investigating the effects of metaphors within a dataset of real-time human-AI interactions. By integrating the nuanced understanding of metaphor processing with theories of conversational alignment, this study aims to ascertain whether AI-generated metaphors can actively influence user engagement.

III. Methods

This study employed a mixed methods approach to analyze metaphor usage in human-AI conversations and its potential impact on user engagement. The following steps outline the methodology used, from data acquisition to advanced metaphor detection, topic modeling and concept mapping analyses.

A. Data Collection and Preprocessing:

The dataset used in this study was an open-source collection of ChatGPT-3 conversations obtained from Kaggle [20]. The dataset contained anonymized conversations which were cleaned and preprocessed to remove stop words, punctuations, special characters, non-English language characters, and to tokenize words. We also removed conversations which were entirely in a different language or were about translating English text to a different language or vice versa. Ultimately, the dataset comprised 543 unique conversations in English, ranging from 2 to 205 exchanges. Conversations were evaluated based on the number of exchanges, with the assumption that more exchanges indicated higher user engagement. Engagement was thus defined as the conversation length.

B. Metaphor Processing:

To detect metaphorical language in both ChatGPT and user messages, we employed MetaPro [21], a computational metaphor processing tool, incorporating end-to-end solutions for metaphor identification [22], metaphor interpretation [17], and metaphorical concept mapping generation [23]. The latest version [24] was developed, based on a novel pre-training language model that was tailored for metaphor processing tasks. MetaPro has been employed in diverse linguistic and cognitive analysis tasks [25]–[28]. Thus, in this work, MetaPro was used to extract metaphorical language from each conversation, enabling a quantitative analysis of metaphor usage in both AI and user messages.

C. Correlation Analyses:

Two sets of correlation analyses were conducted to explore the relationship between metaphor usage and engagement:

Engagement and Metaphor Usage by ChatGPT: To investigate whether ChatGPT's use of metaphors influenced user engagement, the average number of metaphors used per conversation by ChatGPT was treated as the independent variable, while conversation length was considered the dependent variable. We hypothesized that a higher metaphor count would correlate with longer conversations, indicating higher engagement.

Reciprocal Metaphor Usage Between ChatGPT and Users: To explore the reciprocal influence of metaphor usage between ChatGPT and users, the average count of metaphors used by ChatGPT per conversation was analyzed as the dependent variable, while the average count of metaphors used by the user per conversation was treated as the independent variable. This analysis aimed to determine whether linguistic mirroring occurred during conversations.

D. Topic Modeling:

To ensure that thematic diversity within the dataset did not confound the results, we applied Latent Dirichlet Allocation [29] (LDA) using NLTK's Python library. LDA is a topic modeling technique used to group conversations into distinct themes based on the co-occurrence of words [30]. This step was crucial for identifying any potential variations in metaphor usage across different conversational topics and provided a deeper understanding of the contextual richness of the dataset.

Conversations and individual messages within each conversation were categorized based on their dominant topic, which allowed us to explore whether the influence of metaphor usage on engagement was topic-dependent or agnostic.

E. Clustering:

Using K-means clustering, a method that groups data points into clusters based on their similarity, we segmented the dataset into two distinct groups: high-engagement and low-engagement conversations. Conversation length was scaled to have a mean of 0 and a standard deviation of 1, providing a standardized metric for clustering. This segmentation was essential for examining metaphorical patterns in both high- and low-engagement conversations.

F. Concept Mapping Analyses:

We further analyzed metaphorical patterns using concept mapping to understand the structure of metaphorical language in human-AI conversations. Conversations within each engagement group (high and low) were divided into four categories: (a) High-engagement ChatGPT messages, (b) High-engagement user messages, (c) Low-engagement ChatGPT messages, and (d) Low-engagement user messages.

For each category, the top 10 conceptual mappings were identified, with normalization applied to account for differences in the total number of concept mappings across these categories. This ensured a fair comparison of metaphor usage between ChatGPT and users.

To investigate whether metaphorical patterns varied across topics, we identified the top 5 concept mappings for each of the six dominant topics in both high- and low-engagement groups. Normalization was applied within each topic to adjust for variations in metaphor usage, ensuring that the impact of metaphorical language on engagement could be equitably compared across diverse conversational themes.

This comprehensive approach enabled us to explore the role of metaphor usage in human-AI conversations, its potential influence on engagement, and whether this influence varied depending on the conversational topic or engagement level.

IV. Results

The results presented in this paper are not only in relation to two primary research questions: the relationship between metaphor usage and user engagement, and the presence of a linguistic mirroring effect in human-AI conversations; but also take a deep dive into the exploratory analysis conducted to assess metaphorical patterns across different topics and engagement levels.

A. No Correlation between Conversation Length and Average ChatGPT Metaphorical Word Count

The first correlation analysis examined the relationship between the length of the conversation and the average metaphorical word count used by ChatGPT. The results yielded a slight negative correlation coefficient of *-0.056*, indicating a negligible inverse relationship between conversation length and the use of metaphorical language by ChatGPT. This suggests that longer conversations do not necessarily involve an increased use of metaphors by the AI system.

B. Mirroring Effect in Metaphor Use

The second correlation analysis explored the reciprocal influence of metaphor usage between ChatGPT and users, revealing a weak positive correlation of *0.21*. This suggests a modest degree of linguistic mirroring, where the AI reflects the metaphorical language introduced by users.

A more detailed analysis through regression (see Fig. 1) showed a significant relationship between user and ChatGPT metaphor usage, ($t = 4.806$, $p < 0.001$). The results highlight that ChatGPT tends to adapt its metaphorical language based on the user's input. This behavior suggests an element of conversational alignment or adaptation, a dynamic common in human-human interaction, now reflected in human-AI conversations.

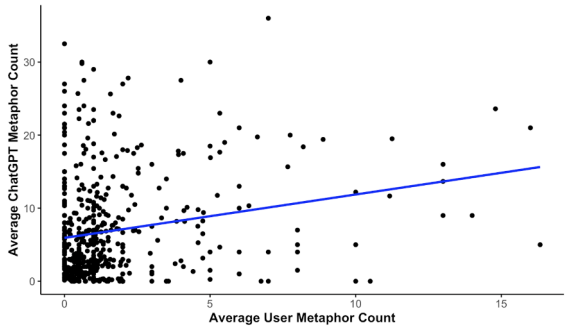

Fig. 1: Regression analysis graph to determine the reciprocal influence of metaphor usage between the user and ChatGPT

However, considering the modest correlation coefficient, it is likely that metaphor usage alone is not the sole factor driving this mirroring effect. Other factors, such as the conversation's overall tone, topic complexity, or user engagement strategies, may also play a role in shaping ChatGPT's responses.

C. Topic Modeling

LDA was applied to analyze the thematic diversity of the dataset, revealing eight distinct topics. Based on the top 40 words in each topic, six major categories were identified:

1) **Business Development and Marketing:** consisted words like business, sponsorship, company, platform, etc.
2) **Data, Programming and Software Development:** consisted words such as data, code, import, application, etc.

3) **Entertainment:** consisted words like character, music, game, challenge etc.
4) **Cybersecurity and Data Protection:** consisted of terms like security, ethical, privacy, hacking, etc.
5) **Health, Research, and Technology:** consisted words such as genetic, cancer, testing, research, etc.
6) **Ethics and Social Considerations:** consisted words like bias, legal, leadership, ethics, etc.

These topics provide a nuanced understanding of the conversational landscape, with Business Development and Marketing emerging as the dominant topic (*n = 200*), and Health, Research, and Technology being the least frequent topics (*n = 24*). The diversity of topics reinforced the importance of examining metaphor usage in varied contexts, as different themes may invoke different metaphorical patterns.

D. Metaphorical Concept Mappings Across Engagement Levels

Clustering analysis segmented the conversations into high and low engagement groups (see Fig. 2). The low engagement group encompassed conversations with 2 to 78 exchanges, while the high engagement group included conversations ranging from 86 to 205 exchanges. The division point occurred at around 75-100 exchanges.

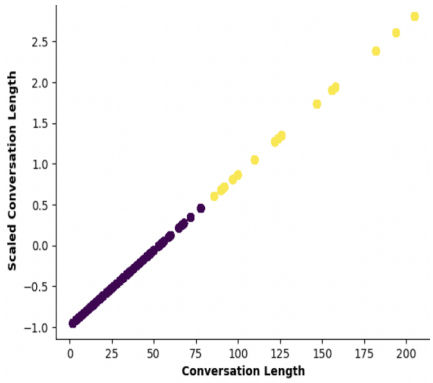

Fig. 2: K-Mean Clustering of conversations into high-engagement group (in yellow) and low-engagement group (in blue).

After separating ChatGPT and user messages within each engagement group, concept mappings revealed consistent metaphor usage across both high- and low-engagement groups. As shown in the heat map (Fig. 3), metaphors such as "quality is asset" and "act is motion" were employed consistently, regardless of engagement level. This indicates that some metaphors serve as fundamental linguistic tools, used across a variety of conversational contexts.

However, there were distinct differences in metaphorical complexity between the two groups. In high-engagement conversations, metaphors such as "cognition is knowing" and "improvement is organic process" appeared more frequently in both ChatGPT and user messages (see Fig. 3).

These metaphors, which convey abstract and interconnected ideas, likely contribute to deeper, more complex conversations. Conversely, in low-engagement conversations, simpler and more concrete metaphors, such as "activity is act" and "improvement is action," were more common (see Fig. 3). This suggests that low-engagement conversations are more task-oriented and may lack the exploratory or reflective nature seen in high-engagement interactions.

TABLE I: Number of conversations for each topic in high and low-engagement group.

Topics	**High-Engagement Group**	**Low-Engagement Group**
Business Development and Marketing	6	194
Cybersecurity and Data protection	0	46
Data, Programming, and Software Development	10	179
Entertainment	7	47
Ethics and Social Considerations	1	29
Health, Research, and Technology	0	24

Table I shows the number of conversations in each topic within high- and low-engagement groups. Conversations were categorized based on their dominant topic, resulting in an uneven distribution between the engagement groups. Notably, there were no conversations under the "Cybersecurity and Data Protection" or "Health, Research, and Technology" topics in the high-engagement group.

Due to this imbalance, the concept mapping analysis was conducted at the message level instead of the conversation level. Analyzing metaphors at the message level allowed for a more granular and comprehensive understanding of metaphor usage by both ChatGPT and users. This approach ensured that important sub-topics were not overshadowed by dominant themes, preserving the diversity of metaphors within sub-topics that might have otherwise been missed in a conversation-level analysis.

E. Metaphorical Concept Mappings Within Each Topic Across Engagement Levels

Table II lists the number of ChatGPT and user messages within each topic of the high- and low-engagement groups. By focusing on message-level topics, the analysis captured broad thematic trends as well as subtle, topic-specific metaphorical transitions. This method provided a more detailed understanding of how metaphors function across different contexts while maintaining flexibility for post-analysis aggregation. As a result, the analysis allowed for a refined exploration of metaphorical patterns, particularly in high-engagement conversations, where multiple metaphors were often intertwined within different thematic strands. The top 5 concept mappings identified within the topics of high-engagement conversations (Fig. 4) again reveal complex and varied metaphorical structures across all topics. Mappings like "outgo IS motion" in

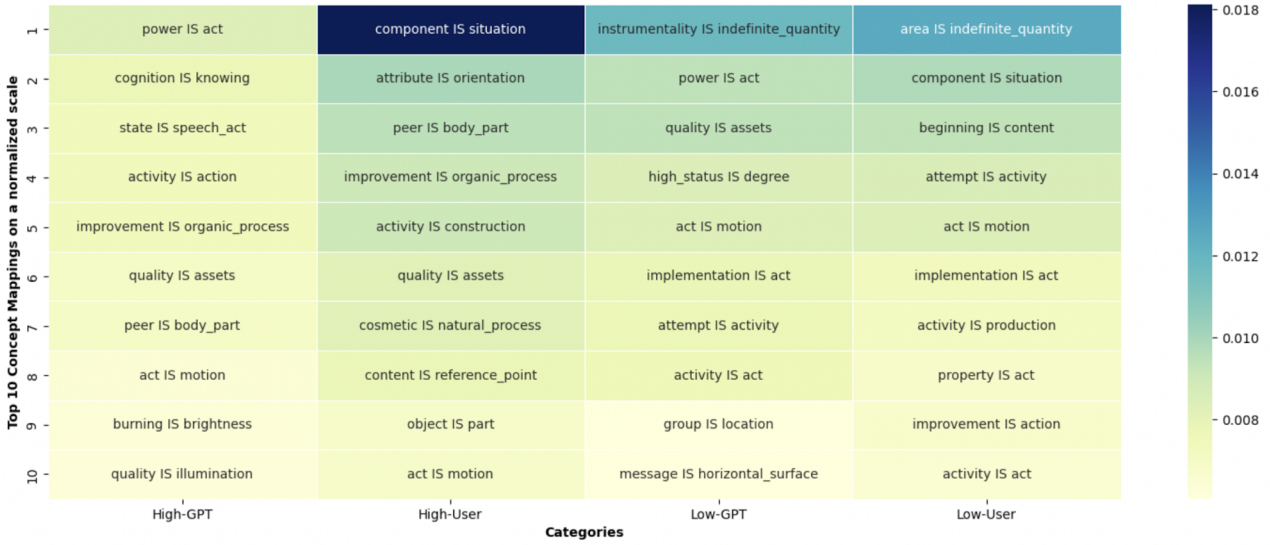

Fig. 3: Heat map for the top 10 concept mappings across four categories: (1) High-engagement ChatGPT messages, (2) High-engagement User messages, (3) Low-engagement ChatGPT messages, and (4) Low-engagement User messages. The color coding represents the normalized scale of concept mapping frequency, with darker colors indicating higher frequencies (i.e. more prevalent concept mappings) and lighter colors indicating lower frequencies (i.e. less prevalent concept mappings) across the categories.

"Cybersecurity and Data Protection" or "adjustment IS temporal relation" in "Ethics and Social Considerations" indicate that high-engagement conversations are rich in metaphors that convey processes, relationships, and systems.

TABLE II: Number of ChatGPT and user messages for each topic in high and low-engagement group

Topics	**High-ChatGPT Group**	**High-User Group**	**Low-ChatGPT Group**	**Low-User Group**
Business Development and Marketing	244	243	1099	903
Cybersecurity and Data protection	51	89	308	425
Data, Programming, and Software Development	552	483	1151	1063
Entertainment	361	383	247	331
Ethics and Social Considerations	79	96	259	373
Health, Research, and Technology	37	63	196	232

In contrast, the top 5 concept mappings in low-engagement conversations (Fig. 5) show a more consistent use of simpler, action-oriented metaphors across all topics. Common mappings such as "power IS act", "attempt IS activity", "act IS activity" appear frequently, suggesting that these exchanges are more surface-level or task-driven, lacking the depth seen in high-engagement conversations. This supports the idea that the richness of metaphorical language enhances conversational engagement, as more involved conversations enable the exploration of abstract and interconnected metaphorical concepts.

Additionally, the user and ChatGPT messages share at least one of the top five metaphorical concept mapping within each high-engagement topic. In contrast, only one shared concept mapping—"component IS situation"—appears from the heat map (Fig. 5) in both user and ChatGPT messages within the "Data, Programming, and Software Development" topic in the low-engagement group. This further supports the hypothesis that metaphorical conversational alignment may contribute to higher engagement levels.

Interestingly, some metaphorical mappings, such as "act IS motion" and "quality IS assets", are topic-agnostic, appearing consistently across both high- and low-engagement groups and across various topics (see Fig. 5 and 6). However, other concept mappings are more context-specific. For instance, in the high-engagement group (Fig. 5), "peer IS body part" is unique to the "Business Development and Marketing" topic, while "outgo IS motion" predominates in "Cybersecurity and Data Protection" topic. Similarly, metaphors like "burning IS brightness" and "perception IS brightness" are closely linked to "Health, Research, and Technology". These findings suggest that metaphorical language is influenced by the thematic content of the conversation, and certain metaphors may only be relevant within specific domains.

While some overlap of general metaphors exists across topics, the heatmaps indicate that metaphorical language is not entirely topic agnostic, especially in longer conversations. Metaphor usage and its impact on engagement are closely

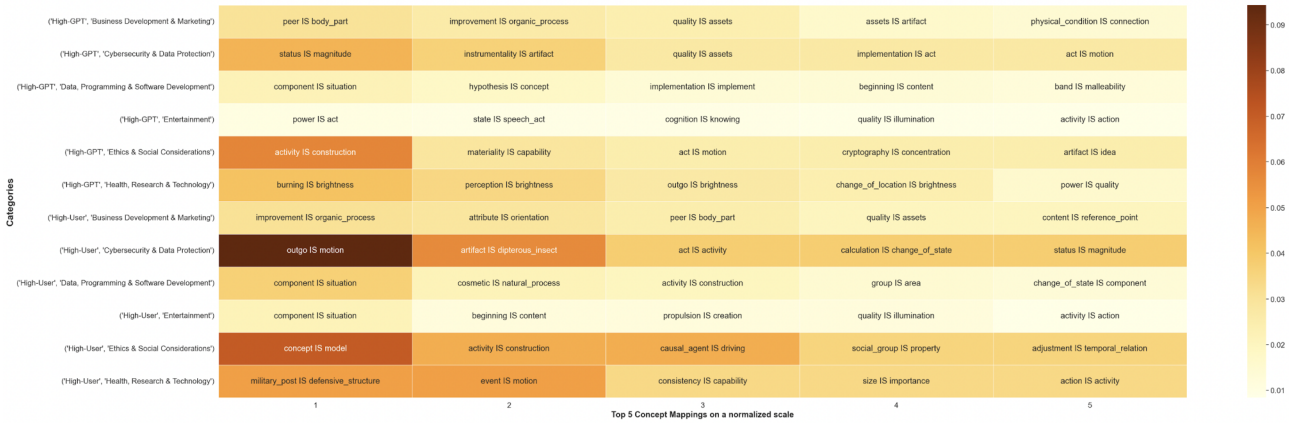

Fig. 4: Heat map for the top 5 concept mappings across the 6 topics within high-engagement ChatGPT messages and high-engagement user messages. The color coding represents the normalized scale of concept mapping frequency, with darker colors indicating higher frequencies (i.e. more prevalent concept mappings) and lighter colors indicating lower frequencies (i.e. less prevalent concept mappings) across the categories.

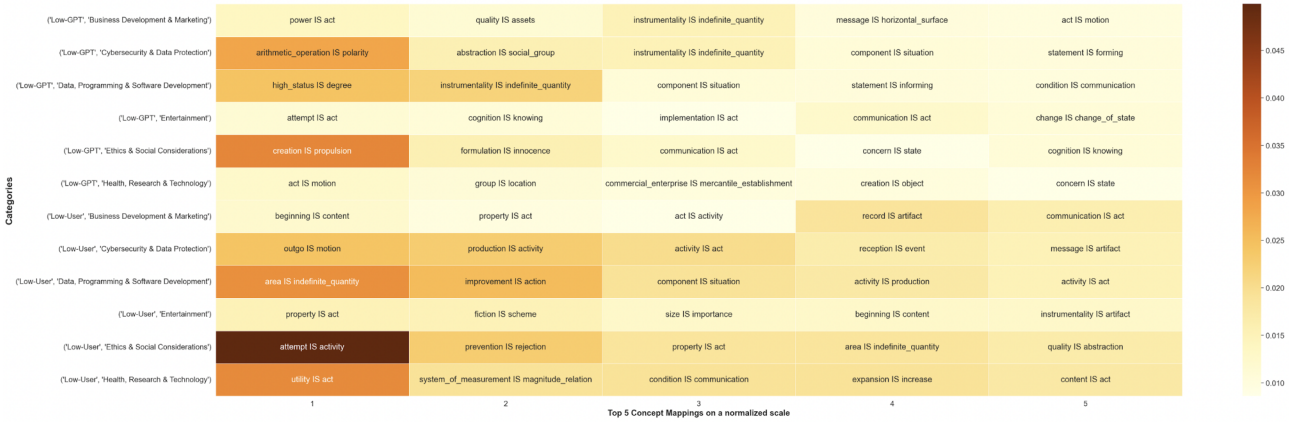

Fig. 5: Heat map for the top 5 concept mappings across the 6 topics within low-engagement ChatGPT messages and low-engagement user messages. The color coding represents the normalized scale of concept mapping frequency, with darker colors indicating higher frequencies (i.e. more prevalent concept mappings) and lighter colors indicating lower frequencies (i.e. less prevalent concept mappings) across the categories.

tied to the specific subject matter being discussed. High-engagement conversations tend to foster more abstract and complex metaphors, whereas low-engagement conversations rely on simpler, more universal metaphorical structures. Therefore, the influence of metaphorical language on engagement appears to be context-dependent, varying across different conversational themes.

V. Discussion

The results of this study provide significant insights into the relationship between metaphor usage in human-AI interactions alongside user engagement. They also reflect on the potential for linguistic alignment between users and AI systems like ChatGPT, contributing to a growing body of research exploring the role of metaphors in both human and AI communication.

A. Metaphor Usage and Engagement

One of the primary objectives of this research was to explore whether the use of metaphors by ChatGPT impacts user engagement, as measured by conversation length. Our results suggested that metaphors in AI responses do not directly lead to longer interactions. This finding is consistent with prior research, which found that while metaphors elicit greater real-time engagement compared to literal language, they may not necessarily extend the duration of conversations, especially in task-focused domains such as programming or problem-solving, which accounted for the majority conversations in the dataset for this study [14].

The context-dependent nature of metaphor usage could also explain the lack of correlation between ChatGPT's metaphor usage and engagement. For instance, in conversations involv-

ing technical content, such as programming, ChatGPT-3's scope for metaphorical language (within this study's dataset) is naturally limited, leading to more concrete and functional exchanges. This reflects findings from Khadpe et al. [16], who demonstrated that metaphor usage has varied effects across different types of human-AI collaboration, with more abstract metaphors being less relevant in technical or highly specific contexts. This reinforces the importance of considering the context in which metaphors are used when evaluating their impact on engagement.

B. Linguistic Mirroring and Conversational Alignment

This study also examined whether ChatGPT mirrors the metaphorical language of its users, revealing a modest positive correlation. This finding supports existing theories of conversational alignment, where speakers in a dialogue naturally adapt their language to match one another's style [15], [19]. Our findings support prior research which suggests that "agents use alignment strategies to maintain user's engagement" and one of these strategies could be potential mirroring of metaphors to increase engagement.

However, while ChatGPT demonstrated some ability to align its metaphor usage with users, the modest correlation indicates that metaphor usage alone may not be the primary driver of this mirroring effect. Other factors, such as the tone or complexity of the conversation, likely influence linguistic adaptation. This aligns with research by Thibodeau and Boroditsky [13], who found that metaphors guide reasoning and influence communication patterns but are not the sole determinants of conversational alignment [16].

C. Topic-Specific Metaphor Usage

The results further suggest that metaphor usage is not entirely topic-agnostic, with certain metaphors being more prevalent in specific conversational themes. These findings are consistent with previous work which showed that AI's grasp of metaphorical mappings is influenced by the topic of conversation, with more abstract and relational metaphors being employed in content that requires deeper cognitive engagement [7]. The variation in metaphor usage across different themes also underscores the importance of adapting conversational AI to context-specific language demands. In technical discussions like programming, where users expect direct answers, simpler metaphors suffice, while in creative or strategic discussions, richer metaphors may facilitate more complex conceptual understanding.

D. Implications for AI-Human Interaction

The findings of this study offer several implications for the design of conversational AI systems. First, the ability of AI to mirror metaphorical language, even modestly, suggests that integrating metaphor detection and adaptation algorithms could improve user satisfaction by creating more natural interactions. The findings also indicate that metaphor usage, while not directly extending conversation length, contributes to the richness of interaction in certain contexts. Therefore, developers should focus on creating AI systems that can adapt metaphor usage based on the thematic content and goals of the conversation, tailoring responses to enhance user engagement.

Moreover, the results highlight the growing use of conversational AI in both abstract and technical domains. Therefore, systems being designed to support more strategic or creative discussions could benefit from emphasizing metaphorical complexity, while AI systems being designed for task-oriented interactions should prioritize direct and simple metaphor usage to maintain clarity and focus.

VI. Limitations and Future Directions

Like any other research study, this study also has some limitations that should be acknowledged. First, user engagement was measured solely based on conversation length, which may not fully capture the quality or depth of interactions. A more nuanced metric could provide deeper insights into how metaphors influence conversational richness. Second, the study relied on a single dataset from one language model (ChatGPT), which may limit the generalizability of the findings across broader human-AI interactions. Future work involving multiple datasets from diverse LLMs could provide a more comprehensive understanding. Finally, while Latent Dirichlet Allocation (LDA) is a widely used method for topic modeling, it relies on statistical patterns and may overlook contextual nuances in conversations. Advanced models, such as BERTopic, offer a more refined approach by capturing richer contextual relationships and could improve topic analysis in similar studies [30].

Future research could broaden the scope of this project by incorporating a wider array of language models and data from different cultural contexts to examine the universality of metaphor usage in human-AI interactions. Investigating how different types of metaphors—such as conceptual versus contextual—affect engagement across various domains, including poetry or scientific discourse, would also provide a more nuanced understanding. Additionally, longitudinal studies tracking the evolution of language models could reveal important shifts in AI communication strategies over time. Exploring how metaphor usage evolves in human-AI interactions will deepen our comprehension of how metaphorical language shapes the dynamics of these conversations.

Further, future research can benefit from conducting additional ablation studies to determine the extent to which mirroring of metaphorical usage is a natural outcome of human-AI conversations or if it might be influenced by AI paraphrasing the user's input. Moreover, the implications of this work could be further explored to assess potential increases in user satisfaction, possibly through survey responses, or to explore whether metaphors can be a method for LLMs to express complex ideas concisely. Such a multifaceted approach would enhance our understanding of the strategic use of language by AI in enhancing human-computer interaction.

VII. Acknowledgments

Funding to attend this conference was provided by the Carnegie Mellon University GSA/Provost Conference Funding.

References

[1] R. Mao, M. Ge, S. Han, W. Li, K. He, L. Zhu, and E. Cambria, "A survey on pragmatic processing techniques," *Information Fusion*, vol. 114, p. 102712, 2025.

[2] G. Lakoff and M. Johnson, "The metaphorical structure of the human conceptual system," *Cognitive Science*, vol. 4, pp. 195–208, 1980.

[3] E. Cambria, "Understanding natural language understanding," *Springer, ISBN 978-3-031-73973-6*, 2024.

[4] E. Cambria, R. Mao, M. Chen, Z. Wang, and S.-B. Ho, "Seven pillars for the future of artificial intelligence," *IEEE Intelligent Systems*, vol. 38, no. 6, pp. 62–69, 2023.

[5] R. Mao, G. Chen, X. Zhang, F. Guerin, and E. Cambria, "GPTEval: A survey on assessments of ChatGPT and GPT-4," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA and ICCL, 2024, pp. 7844–7866.

[6] L. Henrickson and A. Meroño-Peñuela, "Prompting meaning: A hermeneutic approach to optimising prompt engineering with chatgpt," *AI & Society*, 2023.

[7] L. Wachowiak and D. Gromann, "Does GPT-3 grasp metaphors? Identifying metaphor mappings with generative language models," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023, pp. 1018–1032.

[8] S. Flusberg, T. Matlock, and P. Thibodeau, "Metaphors for the war (or race) against climate change," *Environmental Communication*, vol. 11, pp. 769–783, 2017.

[9] M. Ge, R. Mao, and E. Cambria, "A survey on computational metaphor processing techniques: From identification, interpretation, generation to application," *Artificial Intelligence Review*, vol. 56, pp. 1829–1895, 2023.

[10] E. Cambria, X. Zhang, R. Mao, M. Chen, and K. Kwok, "SenticNet 8: Fusing emotion AI and commonsense AI for interpretable, trustworthy, and explainable affective computing," in *Proceedings of International Conference on Human-Computer Interaction (HCII)*, Washington DC, USA, 2024.

[11] E. Cambria, N. Howard, J. Hsu, and A. Hussain, "Sentic blending: Scalable multimodal fusion for continuous interpretation of semantics and sentics," in *IEEE SSCI*, 2013, pp. 108–117.

[12] X. Ding, P. Feng, J. Wang, and M. Lin, "Metaphorical or straightforward? Comparing the effectiveness of different types of social media advertising," *Frontiers in Neuroscience*, vol. 16, 2022.

[13] P. Thibodeau and L. Boroditsky, "Metaphors we think with: The role of metaphor in reasoning," *PLoS One*, vol. 6, no. 2, p. e16782, 2011.

[14] S. Mon, M. Nencheva, F. Citron, C. Lew-Williams, and A. Goldberg, "Conventional metaphors elicit greater real-time engagement than literal paraphrases or concrete sentences," *Journal of Memory and Language*, vol. 121, p. 104285, 2021.

[15] L. Menenti, S. Garrod, and M. Pickering, "Toward a neural basis of interactive alignment in conversation," *Frontiers in Human Neuroscience*, vol. 6, 2012.

[16] P. Khadpe, R. Krishna, L. Fei-Fei, J. Hancock, and M. Bernstein, "Conceptual metaphors impact perceptions of human-AI collaboration," *Proc. ACM Hum.-Comput. Interact.*, vol. 4(CSCW2), pp. 163:1–163:26, 2020.

[17] R. Mao, X. Li, M. Ge, and E. Cambria, "MetaPro: A computational metaphor processing model for text pre-processing," *Information Fusion*, vol. 86–87, pp. 30–43, 2022.

[18] V. Prabhakaran, M. Rei, and E. Shutova, "How metaphors impact political discourse: A large-scale topic-agnostic study using neural metaphor detection," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, 2021, pp. 503–512.

[19] S. Campano, J. Durand, and C. Clavel, "Comparative analysis of verbal alignment in human-human and human-agent interactions," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2014, pp. 4415–4422.

[20] N. Persaud, "89k chatgpt conversations dataset," https://www.kaggle.com/datasets/noahpersaud/89k-chatgpt-conversations/data, 2023, accessed: 2024-10-10.

[21] R. Mao, X. Li, K. He, M. Ge, and E. Cambria, "MetaPro Online: A computational metaphor processing online system," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, vol. 3. Toronto, Canada: Association for Computational Linguistics, 2023, pp. 127–135.

[22] R. Mao and X. Li, "Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 15, 2021, pp. 13 534–13 542.

[23] M. Ge, R. Mao, and E. Cambria, "Explainable metaphor identification inspired by conceptual metaphor theory," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 10, pp. 10 681–10 689, 2022.

[24] R. Mao, K. He, C. B. Ong, Q. Liu, and E. Cambria, "Metapro 2.0: Computational metaphor processing on the effectiveness of anomalous language modeling," in *Findings of the Association for Computational Linguistics: ACL*. Association for Computational Linguistics, 2024, pp. 9891–9908.

[25] R. Mao, K. Du, Y. Ma, L. Zhu, and E. Cambria, "Discovering the cognition behind language: Financial metaphor analysis with MetaPro," in *2023 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2023, pp. 1211–1216.

[26] R. Mao, Q. Lin, Q. Liu, G. Mengaldo, and E. Cambria, "Understanding public perception towards weather disasters through the lens of metaphor," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, (IJCAI-24)*, 2024, pp. 7394–7402.

[27] R. Manro, R. Mao, L. Dahiya, Y. Ma, and E. Cambria, "A cognitive analysis of CEO speeches and their effects on stock markets," in *Proceedings of the 5th International Conference on Financial Technology (ICFT)*, Singapore, 2024.

[28] R. Mao, T. Zhang, Q. Liu, A. Hussain, and E. Cambria, "Unveiling diplomatic narratives: Analyzing United Nations Security Council debates through metaphorical cognition," in *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, vol. 46, 2024, pp. 1709–1716.

[29] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[30] R. Egger and J. Yu, "A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts," *Frontiers in Sociology*, vol. 7, p. 886498, 2022.