Prompting AI in Co-Creation: The Role of Syntax and Sentiment in Shaping AI-Generated Content

Giancarlo Colloca, Jeba Rezwana
Computer and Information Science Department
Towson University
8000 York Road, Towson, MD 21252 USA
gcolloc1@students.towson.edu; jrezwana@towson.edu

Abstract

As Generative AI has advanced at an incredible pace and has been increasingly integrated into creative domains, understanding how users' prompting styles shape AI-generated creative content is critical. Despite a growing interest in human-AI co-creation, there is limited research exploring how linguistic and emotional cues in user prompts affect AI-generated responses in creative domains. This study addresses this gap by analyzing the syntactic and sentiment patterns of user prompts and their impact on ChatGPT's responses in a co-creative storytelling study with 100 participants. We found different stylistic patterns in both prompts and responses, while ChatGPT often amplified users' sentiment and positivity. The findings offer insights to enhance our understanding of how AI can be optimized for more appropriate interactions in co-creation.

Introduction

Generative AI (GenAI) is transforming the landscape of creative expression, offering new possibilities for collaboration between humans and machines. Large language models (LLMs) are becoming increasingly integrated into creative workflows, democratizing the generation of creative artifacts. In co-creative domains, users and AI systems engage in dynamic interactions where each party contributes to the evolving narrative (Davis 2013), which makes understanding how human input shapes AI output critical. Much attention has been given to the capabilities and limitations of these models, but comparatively little research has examined the linguistic and emotional characteristics of user prompts and their influence on the nature and quality of AI-generated content. The structure and emotional tone of user prompts play a pivotal role in shaping the responses of LLMs in creative contexts like storytelling (Ul Huda et al. 2024). Prior research has shown that minor variations in prompt phrasing can lead to substantial differences in output style, content, and coherence (Brown et al. 2020; Reynolds and McDonell 2021). Linguistic cues often carry implicit emotional and contextual information that can guide the model's interpretation of user intent (Chung et al. 2024). In co-creative settings, such cues are especially important, as they can determine whether the AI responds passively, takes initiative, or mirrors the user's creative direction.

Recent research has highlighted the critical role of prompt engineering in guiding LLMs toward more accurate, coherent, and contextually appropriate outputs (Brown et al. 2020; Wei et al. 2022). These efforts have focused on measures like reasoning ability, factual accuracy, and task success, but the more nuanced, co-creative dimensions of human-AI collaboration remain underexplored. Work on prompt tuning and dynamic adaptation has advanced internal model optimization (Xiao et al. 2025; Zhang et al. 2024), yet it typically overlooks how users engineer their prompts to interact with models in creative contexts. Concerns about the diversity and authenticity of model outputs have also been tied to prompt phrasing (Holtzman et al. 2019), and prompt robustness has been studied in adversarial contexts (Liu et al. 2021). However, there has been limited exploration into how prompts influence model behavior in terms of the sentiment of user prompts, as well as the structural and stylistic aspects of user prompts. Studies have examined prompt performativity (how prompts elicit desired responses) (Shin et al. 2020) and generation control (guiding model output) (Dathathri et al. 2020), however these analyses often focus on token-level or dataset-driven approaches, overlooking the informal nature of human-AI co-creation, where users engage with the model through dynamic prompts. Recent studies have begun to map the prompt engineering landscape (Chen et al. 2023; Sahoo et al. 2024), and exploratory work on prompt literacy (Hwang, Lee, and Shin 2023) indicated the growing need for more user-centric approaches. The interpretive space where human intention meets AI output remains largely unmapped, in particular, how linguistic and emotional cues in user prompts influence AI responses in co-creative storytelling.

Motivated by this gap, this study investigates the linguistic structure and sentiments in user prompts and how these components shape model outputs during a co-creative writing task. We conducted a user study with 100 participants who engaged with ChatGPT in a creative writing task. Focusing on prompt-response pairs, we examined how syntax, style, and sentiment in prompts influence the variability, creativity, and alignment of AI-generated content. We present findings that could be utilized in deepening our understanding of the dynamics and structures of prompts and responses in co-creativity and supporting future work in prompt literacy and personalized interaction design.

Related Work

Human-AI co-creation is a rapidly growing research area where humans and AI collaborate on creative tasks to produce creative artifacts or ideas (Davis 2013). Generative AI has been heavily used in human-AI co-creation lately, which serves as an effective co-creation technology (Bozkurt and Sharma 2023) by enhancing our creative abilities and transforming our experience of how we interact with AI systems. The efficacy of generative AI is influenced by its underlying algorithms and training data, but also by the way users interact with it (Rezwana and Maher 2022), for example, the quality of the prompts AI systems receive (Liu et al. 2023). Generative AI offers immense potential, but its effectiveness depends on well-crafted prompts (Cheung 2023). Well-designed prompts can elicit meaningful responses, while a poorly constructed prompt may lead to irrelevant outputs. Yet, many users struggle with prompt engineering due to the novelty of these technologies (Bozkurt and Sharma 2023). This highlights the need for research into effective prompting strategies.

Previous studies emphasize that prompt structure plays a critical role in AI performance. For instance, Brown et al. (2020) introduced few-shot prompting as a means of guiding language models, while Wei et al. (2022) demonstrated that structured chains of thought can elicit logical reasoning. These approaches highlight how task framing can significantly impact model behavior and effectiveness. In creative contexts, creative generation using prompt engineering goes beyond correctness, requiring expressiveness, engagement, and emotional alignment (Reynolds and McDonell 2021). In the domain of co-creativity, Jakesch et al. (Jakesch et al. 2019) explored human perception of AI-generated content, emphasizing factors like trust and social resonance. Serban et al. (Serban et al. 2015) examined how user expectations shape dialogue interactions—primarily in task-oriented settings—but their findings on conversational alignment offer insights relevant to co-creative communication. Recent research (Bhandarkar et al. 2024) highlights the growing need for adaptive models that align with not just content but user sentiment and style—factors increasingly relevant in co-creative systems. Other work (Smith et al. 2025) has begun to examine multimodal and cross-domain prompt structures, especially in artistic and educational settings, though few studies link linguistic traits directly to co-creative engagement.

The literature shows that the way users interact with AI significantly influences the creativity that emerges from co-creation (Karimi et al. 2020). We can identify prompt attributes that lead to specific creative outcomes, which can inform effective prompt engineering across various creative domains. Additionally, the field of human-centered NLP (Bender and Koller 2020) emphasizes the importance of considering the social and affective aspects of AI responses. However, there is a gap in computational creativity research regarding the structure and sentiment of prompts and their impact on AI responses. Our research contributes to this area by exploring how the style and sentiment of user prompts influence AI outputs in creative writing/storytelling.

Study Methodology

Procedure

This study explores users' prompt writing styles and how they influence AI-generated responses in creative contexts. Participants engaged in a co-creative storytelling task using ChatGPT-3.5. Participants were provided with the following narrative prompt: "A little girl named Alice, who is 7, wants to go on an adventure with her dog Lola. One day when both her parents went out for a quick errand, Alice and Lola. . ." Participants had 7 minutes to continue a story and prompt ChatGPT to finish or improve it, providing screenshots of the prompt and final story as proof of completion. The study was approved by the Institutional Review Board.

Using Prolific, an online participant recruitment platform, we recruited 100 participants (56 men, 40 women, and 4 non-binary) who were at least 18 years old, with an average age of 27.0 (Med = 27.5, SD = 6.88). As our aim was to gather a diverse set of data across various demographics, we kept the inclusion criteria to a minimum, requiring only that participants be at least 18 years old.

Data Collection and Analysis

We collected users' prompts and ChatGPT responses through the screenshots participants submitted.

We conducted a series of statistical analyses. Prompts and AI responses' syntactic structures were analyzed using Parts-of-speech (POS), categories to which each word is assigned in accordance with its syntactic functions (Jurafsky and Martin 2024). We focused on POS features as they provide an interpretable framework for analyzing language patterns across diverse users. POS tags capture the grammatical roles of words, allowing us to quantify the expressive, descriptive, and structural tendencies present in both human and machine-generated texts. We utilized POS Tagger of the Stanford CoreNLP package version 4.5.8 (Manning et al. 2014) to categorize words into nine features: nouns (NOUN), verbs (VERB), adjectives (ADJ), adverbs (ADV), pronouns (PRON), conjunctions (CONJ), adpositions (ADP), numbers (NUM) and determiners (DET).

To investigate stylistic patterns in the users' prompts and AI responses, we conducted k-means clustering with the nine POS tags as the features. To visualize and interpret clusters, we applied Principal Component Analysis (PCA) to reduce the dimensionality of POS tag distributions to two key dimensions.

Additionally, we used paired t-tests to assess differences in syntactical structures between clusters of user prompts and AI responses, and one-way ANOVA to examine the most influential cluster features. Additionally, we used Chi-square tests to identify the association between several variables, such as clusters of users' prompts and AI responses.

Sentiment analysis was conducted using Text2Data (Tex 2025), an online text analysis tool, which has been utilized in the literature for sentiment analysis (Gandy et al. 2025; Ermakova, Henke, and Fabian 2021). Text2Data generates sentiment polarity scores measuring a text's emotional tone (from -1 indicating negative to 1 indicating positive).

Results

In this section, we present the findings from our analysis of user prompts and AI-generated responses in the co-creative storytelling task. We begin by examining the syntactic patterns using POS distributions, followed by clustering analysis that identifies distinct stylistic patterns across prompts and responses as well as examines how these clusters relate to one another. Finally, we report on the sentiment analysis, comparing the emotional tone between user inputs and ChatGPT's outputs.

POS Tags	User Prompts			ChatGPT Responses			Comparisons	
	COUNTS	Freq. %	SD	COUNTS	Freq. %	SD	Freq. Diff. Δ%	t-test p-value
ADJ	380	6.68	2.50	1833	9.00	2.87	2.32	0.0055
ADV	269	4.73	2.12	853	4.19	1.98	-0.54	0.4685
CONJ	294	5.17	2.23	934	4.59	2.08	-0.58	0.4262
DET	660	11.61	3.21	2278	11.19	3.15	-0.42	0.6391
NOUN	1619	28.47	4.52	5811	28.54	4.52	0.07	0.9194
NUM	267	4.69	2.11	60	0.29	0.54	-4.40	<0.0001
ADP	622	10.94	3.11	2795	13.73	3.47	2.79	0.0006
PRON	493	8.67	2.80	1920	9.43	2.92	0.76	0.3501
VERB	1083	19.04	3.93	3876	19.04	3.93	0.00	1

Table 1: Comparison of part-of-speech (POS) tag distributions between user prompts and AI responses, with results of paired t-tests.

Comparison between Structural Patterns of User Prompts and AI Responses

To understand how linguistic structure varies between users and ChatGPT during co-creative storytelling, we analyzed the POS tag distributions in user prompts and corresponding AI-generated responses. Table 1 presents the mean percentage and standard deviation for each POS category across all samples, along with the results of paired t-tests comparing prompts and responses. Table 1 reveals clear syntactic differences between participant prompts and ChatGPT responses. ChatGPT uses significantly more adjectives (ADJ) ($\Delta = 2.32\%, p = 0.0055$) and adpositions (ADP) ($\Delta = 2.79\%, p = 0.0006$), suggesting a tendency to enhance descriptions and relational clarity. In contrast, it drastically reduces NUM ($\Delta = -4.40\%, p < 0.0001$) that were part of the task for the user prompts, which may strip away some contextual or quantitative detail. These shifts highlight how the model transforms factual, structured input into more elaborative and expressive language.

Cluster Analysis of Prompts and Responses

To identify stylistic patterns in both user prompts and ChatGPT responses, we applied k-means clustering using distributions across nine POS categories. As visualized, we identified two distinct clusters (See Figure 1) in both participant Prompts and ChatGPT Responses. Principal Component Analysis (PCA) was used to reduce dimensionality and capture the dominant syntactic variance, with the first two dimensions explaining the majority of variance in both prompts and responses.

For user prompts, PCA Dimension 1 is driven by VERB, DET, and NOUN, reflecting content density related to ac-

tions and subjects. PCA Dimension 2 is shaped by CONJ, NUM, and ADP, capturing structural complexity. For Chat-GPT responses, Dimension 1 similarly reflects content richness, dominated by NOUN, VERB, and DET, while Dimension 2, influenced by ADJ, ADV, and CONJ, highlights descriptive and connective elaboration.

In prompt clusters, Cluster 0 (n = 23) includes more expressive prompts, using more ADJ, ADV, and PRON, suggesting a personal tone. Cluster 1 (n = 77) prompts relied more on NOUN, DET, and CONJ, indicating a more descriptive, scene-setting approach. A similar pattern emerged in ChatGPT response clusters: Cluster 0 involves 41 instances that used more ADJ, ADP, and PRON, aligning with an expressive style, while Cluster 1 involves 59 instances that used more NOUN, reflecting a simpler, clearer structure.
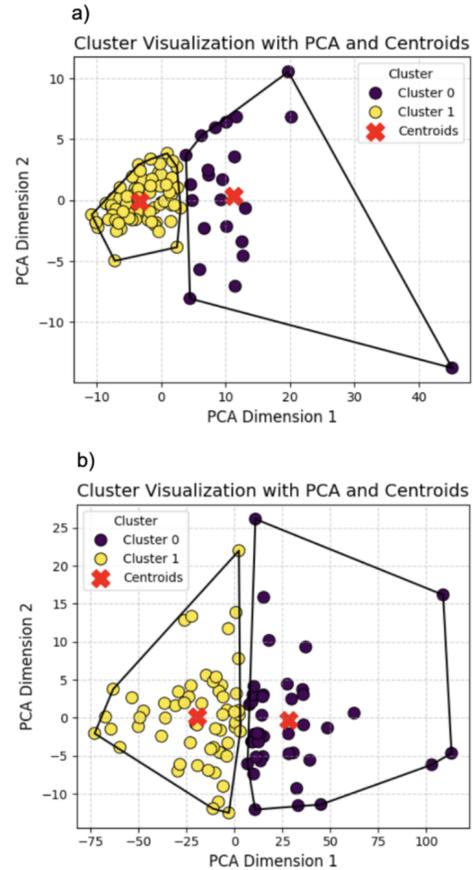

Figure 1: Clustering of user prompts (a) and ChatGPT responses (b) based on part-of-speech (POS) tag distributions using k-means. Purple dots (Cluster 0) represent more expressive language patterns and yellow dots (Cluster 1) reflect a more descriptive, scene-setting style.

We conducted one-way ANOVA tests to identify which POS features significantly differed between the clusters; all were statistically significant ($p < 0.001$) except for NUM, which did not show a meaningful difference. Among the most distinguishing features were NOUN, DET, and VERB,

with higher frequencies observed in Cluster 0 of the prompt data—a pattern that also appeared in the response clusters. PRON (personal pronouns) showed particularly strong differences: in prompts, Cluster 0 had a mean value of 7.57 compared to 4.14 in Cluster 1; in ChatGPT responses, the difference was even larger (25.80 vs. 14.61). ADP (adpositions) exhibited the greatest divergence in the response clusters (38.00 vs. 20.97). Similarly, ADV (adverbs) showed a notable difference in prompts (5.70 vs. 1.79). ADJ (adjectives) and CONJ (conjunctions) followed similar trends across both datasets, with higher averages in Cluster 0, indicating a more expressive and syntactically rich style.

Despite these clear stylistic patterns within prompts and responses, no significant associations were found between the clusters of prompts and responses in terms of syntactic styles. This finding is somewhat unexpected, as one might assume the structure of user input would influence the syntactic style of the AI output. Additionally, we also found no significant associations between the prompt/response clusters and participant demographics (age, gender, race, country of origin, formative background, disability, education level, field of work/study, occupation, income, political affiliation, or AI expertise).
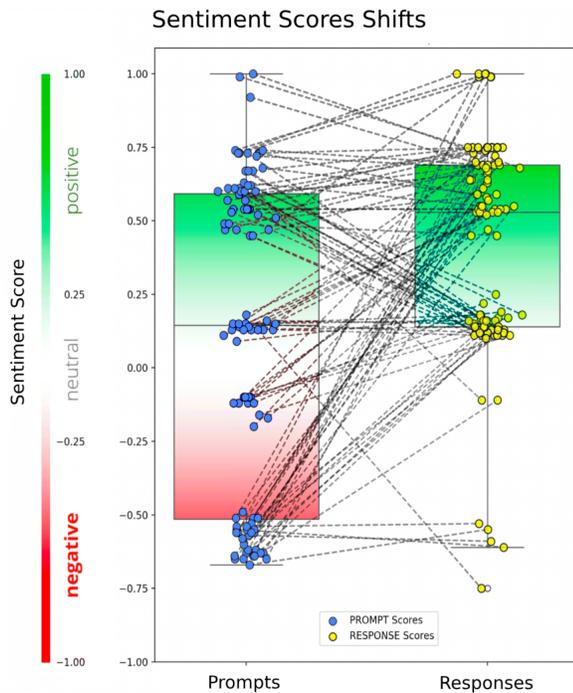
Sentiment Analysis

Figure 2: Comparison of sentiment scores between user prompts and ChatGPT responses. Each line connects a user prompt (blue dot) to its corresponding AI response (yellow dot), showing that ChatGPT responses generally exhibit increased positivity.

We observed a pattern of positive sentiment shift in ChatGPT's responses compared to user prompts (Figure 2). Sentiment analysis revealed that user prompts encompassed a wide range of emotional tone, with 44 categorized as positive, 27 as neutral, and 29 as negative. In contrast, ChatGPT responses exhibited a clear shift toward positivity, comprising 58 positive, 37 neutral, and only 5 negative instances.

A paired t-test confirmed a statistically significant increase in sentiment scores from prompts to responses ($p < 0.001$), indicating that ChatGPT tends to respond with more positive sentiment regardless of the prompt's tone. Figure 2 visualizes these changes. Each dot represents an individual sentiment score, with blue dots indicating sentiment scores for user prompts and yellow dots for ChatGPT responses. Dashed lines connecting the pairs illustrate the direction of sentiment shifts, and boxplots provide an overview of the distribution in each category.

In nearly half the cases (47%), ChatGPT responses showed a sentiment increase of over 0.25, indicating a tendency to elevate emotional tone. In 34% of cases, sentiment remained largely stable (≤ 0.25 change). In 19% of cases, responses were more negative, reflecting moments when ChatGPT adopted a more critical or somber tone.

Overall, prompts displayed broader emotional variability, whereas ChatGPT's responses were more tightly clustered around the positive end of the sentiment scale. This trend reaffirms that while users bring a wide range of emotional tones to the interaction, ChatGPT tends to focus on more positive emotional expressions.

Discussion and Conclusions

As the use of GenAI rapidly grows in creative domains, understanding the stylistic, structural, and sentiment patterns in user prompts—and how these patterns influence AI-generated responses—is crucial to unlocking the full potential of GenAI. This study explored stylistic and emotional patterns in prompts and AI-generated responses and how user prompts influence ChatGPT's linguistic and emotional responses in a collaborative storytelling context with 100 participants. We present the findings and discuss the implications in this section.

The syntactic analysis revealed that ChatGPT responses contained significantly more adjectives (ADJ) and adpositions (ADP) than user prompts, indicating that the AI enriches its language with descriptive and relational elements. This suggests an attempt to create more vivid and engaging responses, potentially aligning with users' implicit requests for expressiveness. In contrast, numerical expressions (NUM) were much less frequent in ChatGPT responses, implying that the AI prioritizes narrative clarity and emotional tone over factual precision in storytelling, which aligns with the existing research (Breithaupt et al. 2024).

The clustering analysis reveals two distinct linguistic styles in both user prompts and AI-generated responses. Cluster 0 in user prompts, involves more expressive prompts, uses a higher frequency of adjectives, adverbs, and pronouns, suggesting a personal, emotional tone. In contrast, Cluster 1 relies more on nouns, determiners, and conjunctions, indicating a descriptive, factual approach. AI responses followed a similar pattern, with Cluster 0 exhibiting more adjectives, adpositions, and pronouns, aligning with

the expressive tone of the prompts. Cluster 1's responses mirrored the user prompts' simpler, more structured style. These contrasting styles highlight different rhetorical strategies, representing the AI's ability to adjust its narrative structure, suggesting it can enhance engagement in creative or emotionally sensitive contexts while maintaining clarity in more factual or instructional interactions.

Human input is more stylistically diverse compared to AI responses, as seen in the greater spread of points for the user prompts across PCA dimensions. AI responses, while reflecting the general structure of the prompts, remain more standardized and homogeneous, aligning with the AI's tendency to prioritize grammatical correctness and coherence over stylistic variation. This distinction underscores the syntactic regularity and homogeneity in AI responses, potentially limiting its adaptability, which aligns with existing research echoing concerns (Liu et al. 2024) about "AI homogenization" in creative fields. Identifying mechanisms to increase response diversity without compromising coherence could improve co-creation balance.

Interestingly, no significant association was found between the clusters of prompts and responses, which is surprising given the expected direct influence of user input on AI-generated content. This finding suggests that ChatGPT may not always perfectly align with the user's intended tone or style. The lack of correlation could be due to the model's tendency to generate responses based on a range of factors beyond just syntactic alignment, such as the model's training data, its programming to optimize for clarity, and its inherent biases in generating emotionally positive outputs.

Sentiment analysis revealed a significant shift toward more positive sentiment in ChatGPT responses compared to user prompts. While user prompts spanned a wide range of emotional tones, from highly negative to positive, ChatGPT's responses tended to be more positive, with a focus on neutral to mildly positive sentiment, highlighting ChatGPT's potential to modulate emotional tone in a way that might help to maintain positive engagement. Such tendencies are likely influenced by reinforcement learning from human feedback (RLHF), which rewards outputs perceived as helpful, engaging, or agreeable (Ouyang et al. 2022). However, this shift also raises questions about the implications of AI modifying users' emotional expressions. For instance, in settings where a user may express frustration or dissatisfaction, ChatGPT's tendency to amplify positive sentiment could be seen as dismissive or unresponsive to the user's emotional state. This aspect of AI behavior warrants further investigation, particularly in creative contexts such as storytelling where emotional nuance is important.

It is also important to recognize that the interpretability and behavior of language models are continually evolving. Different versions of LLMs, or the inclusion of multimodal capabilities, may generate markedly different responses even when given identical inputs. This suggests there may not be a singular or universal mapping between user intention and AI output. Human intention itself is deeply complex and inherently subjective, posing further challenges in aligning AI responses with nuanced human goals or emotional states. Future studies should investigate how varying model architectures and user contexts shape this interpretive space, and how models might better account for the richness of human intent.

Limitations and Future Work

This study has a few limitations. One is the use of a fixed story prompt and a constrained interaction time, which may not fully capture the dynamic nature of user-AI interactions. Future work should explore open-ended interactions with evolving prompts to better understand how style and sentiment develop in more spontaneous and extended contexts. The study focused on a single creative task with a specific system that might limit the generalizability. Future research should examine whether similar stylistic and emotional patterns emerge in a broader range of creative tasks.

Another limitation is the participant sample, which may not reflect the diversity of users who engage with AI in creative settings. Future studies should include more diverse demographics to explore how different user groups engineer prompts. Further research could also examine the range and types of sentiments involved in human-AI interactions to better understand how emotional tone is expressed, transformed, or constrained in these interactions.

Finally, future research should investigate whether AI's tendency to shift responses toward positive sentiment might inadvertently limit the emotional depth and authenticity of co-created content. In creative collaborations, human expression often encompasses a broad spectrum of emotions, including negative and ambiguous tones. If AI systems consistently steer outputs toward optimism or neutrality, there is a risk of diminishing the nuanced emotional landscapes that are essential to authentic creative expression (Ashkinaze et al. 2024). Understanding how AI's sentiment modulation shapes the co-creative process is critical for designing tools that support—not flatten—human creativity.

Acknowledgments

References

Ashkinaze, J.; Mendelsohn, J.; Qiwei, L.; Budak, C.; and Gilbert, E. 2024. How ai ideas affect the creativity, diversity, and evolution of human ideas: evidence from a large, dynamic experiment. *arXiv preprint arXiv:2401.13481*.

Bender, E. M., and Koller, A. 2020. Climbing towards nlu: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, 5185–5198.

Bhandarkar, A.; Wilson, R.; Swarup, A.; and Woodard, D. 2024. Emulating author style: a feasibility study of prompt-enabled text stylization with off-the-shelf llms. In *1st Workshop on Personalization of Generative AI Systems*, 76–82.

Bozkurt, A., and Sharma, R. C. 2023. Generative ai and prompt engineering: The art of whispering to let the genie out of the algorithmic world. *Asian Journal of Distance Education* 18(2):i–vii.

Breithaupt, F.; Otenen, E.; Wright, D. R.; Kruschke, J. K.; Li, Y.; and Tan, Y. 2024. Humans create more novelty than chatgpt when asked to retell a story. *Scientific Reports* 14(1):875.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33.

Chen, B.; Zhang, Z.; Langrené, N.; and Zhu, S. 2023. Unleashing the potential of prompt engineering in large language models: a comprehensive review. *arXiv preprint arXiv:2310.14735*.

Cheung, K. S. 2023. Real estate insights unleashing the potential of chatgpt in property valuation reports: the "red book" compliance chain-of-thought (cot) prompt engineering. *Journal of Property Investment & Finance*.

Chung, H. W.; Hou, L.; Longpre, S.; Zoph, B.; Tay, Y.; Fedus, W.; Li, Y.; Wang, X.; Dehghani, M.; Brahma, S.; et al. 2024. Scaling instruction-finetuned language models. *Journal of Machine Learning Research* 25(70):1–53.

Dathathri, S.; Madotto, A.; Lan, J.; Hung, J.; Frank, E.; Molino, P.; Yosinski, J.; and Liu, R. 2020. Plug and play language models: A simple approach to controlled text generation.

Davis, N. M. 2013. Human-computer co-creativity: Blending human and computational creativity. In *Artificial Intelligence and Interactive Digital Entertainment Conference*.

Ermakova, T.; Henke, M.; and Fabian, B. 2021. Commercial sentiment analysis solutions: a comparative study.

Gandy, L. M.; Ivanitskaya, L. V.; Bacon, L. L.; and Bizri-Baryak, R. 2025. Public health discussions on social media: Evaluating automated sentiment analysis methods. *JMIR Formative Research* 9(1):e57395.

Holtzman, A.; Buys, J.; Du, L.; Forbes, M.; and Choi, Y. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Hwang, Y.; Lee, J. H.; and Shin, D. 2023. What is prompt literacy? an exploratory study of language learners' development of new literacy skill using generative ai.

Jakesch, M.; French, M.; Ma, X.; Hancock, J. T.; and Naaman, M. 2019. Ai-mediated communication: How the perception that profile text was written by ai affects trustworthiness. In *Proceedings of the 2019 CHI*, 1–13.

Jurafsky, D., and Martin, J. 2024. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech cognition. draft (2023).

Karimi, P.; Rezwana, J.; Siddiqui, S.; Maher, M. L.; and Dehbozorgi, N. 2020. Creative sketching partner: an analysis of human-ai co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*.

Liu, J.; Takanobu, R.; Wen, J.; Wan, D.; Li, H.; Nie, W.; Li, C.; Peng, W.; and Huang, M. 2021. Robustness testing of language understanding in task-oriented dialog. In Zong, C.; Xia, F.; Li, W.; and Navigli, R., eds., *59th Annual Meeting of the Association for Computational Linguistics*.

Liu, Y.; Deng, G.; Xu, Z.; Li, Y.; Zheng, Y.; Zhang, Y.; Zhao, L.; Zhang, T.; Wang, K.; and Liu, Y. 2023. Jailbreaking chatgpt via prompt engineering: An empirical study. *arXiv preprint arXiv:2305.13860*.

Liu, Q.; Zhou, Y.; Huang, J.; and Li, G. 2024. When chatgpt is gone: Creativity reverts and homogeneity persists. *arXiv preprint arXiv:2401.06816*.

Manning, C. D.; Surdeanu, M.; Bauer, J.; Finkel, J. R.; Bethard, S.; and McClosky, D. 2014. The stanford corenlp natural language processing toolkit. In *52nd annual meeting of the association for computational linguistics: system demonstrations*.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35:27730–27744.

Reynolds, L., and McDonell, K. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended abstracts of the 2021 CHI*, 1–7.

Rezwana, J., and Maher, M. L. 2022. Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Transactions on Computer-Human Interaction*.

Sahoo, P.; Singh, A. K.; Saha, S.; Jain, V.; Mondal, S.; and Chadha, A. 2024. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.

Serban, I. V.; Lowe, R.; Henderson, P.; Charlin, L.; and Pineau, J. 2015. A survey of available corpora for building data-driven dialogue systems. *arXiv preprint arXiv:1512.05742*.

Shin, T.; Razeghi, Y.; IV, R. L. L.; Wallace, E.; and Singh, S. 2020. Autoprompt: Eliciting knowledge from language models with automatically generated prompts.

Smith, B. E.; Shimizu, A. Y.; Burriss, S. K.; Hundley, M.; and Pendergrass, E. 2025. Multimodal composing with generative ai: Examining preservice teachers' processes and perspectives. *Computers and Composition* 75:102896.

2025. Text2data. https://text2data.com/ [Accessed: (2025)].

Ul Huda, N.; Sahito, S. F.; Gilal, A. R.; Abro, A.; Alshanqiti, A.; Alsughayyir, A.; and Palli, A. S. 2024. Impact of contradicting subtle emotion cues on large language models with various prompting techniques. *International Journal of Advanced Computer Science & Applications* 15(4).

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35.

Xiao, Z.; Yan, S.; Hong, J.; Cai, J.; Jiang, X.; Hu, Y.; Shen, J.; Wang, Q.; and Snoek, C. G. 2025. Dynaprompt: Dynamic test-time prompt tuning. *arXiv preprint arXiv:2501.16404*.

Zhang, J.; Wu, S.; Gao, L.; Shen, H. T.; and Song, J. 2024. Dept: Decoupled prompt tuning. In *Conference on Computer Vision and Pattern Recognition*, 12924–12933.