European Journal of Applied Linguistics Studies
ISSN: 2602 - 0254
ISSN-L: 2602 - 0254
Available on-line at: http://www.oapub.org/lit

DOI: 10.46827/ejals.v9i2.746

Volume 9 | Issue 2 | 2026

PROMPT ENGINEERING AND STYLISTIC CONTROL IN LARGE LANGUAGE MODELS: A QUANTITATIVE ANALYSIS OF FORMALITY, EMOTIONAL TONE, AND PERSUASION

Hossain Md. Sayem[1i],
Arif Md. Ismail Hossain[2]
[1]Master's Student,
Faculty of Humanities,
National Research University Higher School of Economics (HSE University),
Nizhny Novgorod, Russia
[2]Graduate Student,
Faculty of Economics,
National Research University Higher School of Economics (HSE University),
Nizhny Novgorod, Russia

Abstract:
This study investigates the extent to which stylistic features of Large Language Model (LLM) output can be controlled through prompt engineering. Drawing on established frameworks in stylistics, we examine three core stylistic dimensions formality, emotional tone, and persuasive style across six controlled prompt conditions using the DeepSeek LLM. Through a mixed-methods approach combining qualitative textual analysis with quantitative measurement of six linguistic metrics (average sentence length, technical term density, emotion marker frequency, modal verb usage, contraction frequency, and a composite formality index), we demonstrate that systematic variations in prompt structure produce measurable and linguistically interpretable stylistic shifts in generated text. Our findings indicate that explicit stylistic instructions yield stronger effects than implicit contextual cues, and that different stylistic dimensions respond differently to prompt manipulations. Specifically, formal prompts produced texts with 137% longer sentences and 185% more technical terms than informal prompts; emotional prompts generated up to 7.19 emotion markers per 100 words compared to zero in neutral conditions; and directive persuasive prompts exhibited 445% more modal verbs than formal neutral texts. These results provide empirical support for applying classical stylistic theory to AI-generated text and offer practical guidance for linguistically informed prompt engineering.

Keywords: large language models, prompt engineering, stylistics, computational linguistics, DeepSeek

[i] Correspondence: email mdhossain@edu.hse.ru

1. Introduction

The advent of Large Language Models (LLMs) such as GPT-4, Gemini, and DeepSeek has fundamentally transformed natural language generation, enabling the production of human-like text across diverse genres and registers. Unlike earlier rule-based chatbots, which were constrained to narrow domains and rigid response patterns, modern LLMs demonstrate remarkable stylistic flexibility, capable of generating academic prose, creative writing, and conversational dialogue. This flexibility raises an important research question at the intersection of linguistics, computer science, and artificial intelligence: to what extent can stylistic features of LLM-generated text be systematically controlled through prompt engineering?

Despite the growing practical importance of prompt engineering, the relationship between prompt characteristics and resulting stylistic output remains underexplored from a linguistic perspective. Current prompt engineering practices often rely on experimentation rather than theoretically grounded linguistic principles. As Ciesla (2024) notes, modern LLMs demonstrate broad writing capabilities, yet the mechanisms underlying reliable stylistic control require further empirical investigation.

This gap is increasingly important as LLMs become integrated into domains where stylistic precision matters, including academic communication, journalism, marketing, legal writing, and creative production. Without a predictive framework connecting prompt features with stylistic outcomes, users must rely primarily on iterative trial-and-error approaches.

The present study addresses this gap by examining how variations in prompt structure---specifically formal cues, emotional instructions, and persuasive directives---influence measurable stylistic characteristics of LLM output. The analysis is grounded in classical stylistic theory and investigates whether frameworks originally developed for human-authored texts can be applied to AI-generated language.

The study is guided by the following research questions:

RQ1: How does the degree of formality in prompts influence linguistic characteristics of LLM output?

RQ2: To what extent can emotional content in prompts modify the emotional tone of generated texts?

RQ3: Which prompting strategies are most effective for producing persuasive texts?

RQ4: Do LLMs respond more strongly to explicit stylistic instructions than to subtle contextual cues?

We hypothesize that systematic variations in prompt structure produce measurable changes across multiple stylistic dimensions, with explicit stylistic instructions generating stronger effects than implicit contextual signals.

2. Literature Review

2.1 Stylistics and Linguistic Theory

The theoretical foundation of this study rests on classical stylistic frameworks developed for the analysis of human-authored texts. Leech and Short (2007, p. 9) define style as the way language is used in a particular context, by a particular speaker or writer, for a particular communicative purpose. Style represents systematic choices operating across multiple linguistic levels, including syntax, vocabulary, semantics, pragmatics, and discourse organization. Biber (1988) further demonstrates that stylistic variation can be quantified through multi-dimensional analysis, identifying correlational patterns between communicative situations and linguistic forms.

Halliday (1985) introduces the concept of register, referring to language variation according to context of use, including field, tenor, and mode. Genre, following Swales (1990), concerns conventionalized text structures such as research articles, news reports, or business communication. Style refers to the specific linguistic realization within a given register and genre. These distinctions are essential for understanding how LLMs might simulate stylistic variation without human sociocognitive grounding.

2.2 Evolution of Conversational AI

The history of chatbots reveals a continuous trajectory toward greater stylistic flexibility. Adamopoulou and Moussiades (2020) trace this evolution from early rule-based systems such as ELIZA (Weizenbaum, 1966) and PARRY (Colby, 1975) to statistical approaches and modern transformer-based architectures. Early chatbots exhibited rigid, programmer-defined styles with no capacity for adaptation. The transformer revolution, initiated by Vaswani et al. (2017), introduced self-attention mechanisms that enabled models to capture long-range contextual dependencies, laying the groundwork for stylistically diverse generation.

The emergence of Large Language Models---from GPT-1 (117 million parameters) to GPT-4 (1.76 trillion parameters)---marks a qualitative shift in generative capability. Zhao et al. (2023) survey this development, noting that modern LLMs can produce not only informational responses but also creative and stylistically varied texts, including poetry, code, and academic prose.

2.3 Large Language Models and Stylistic Generation

Recent scholarship has increasingly focused on the stylistic properties of LLM output. O'Sullivan (2025) conducts stylometric comparisons of human versus AI-generated creative writing, identifying measurable differences in lexical choice and syntactic patterns. Kolmogorova and Margolina (2024) examine the category of "naturalness" in generated text, finding that LLMs reproduce many surface features of human writing while lacking deeper pragmatic appropriateness.

Ciesla (2024) provides a comprehensive treatment of chatbot evolution, arguing that contemporary LLMs have acquired broad stylistic competence through training on diverse textual corpora. However, the systematic relationship between prompt instructions and stylistic output remains insufficiently explored from a linguistic perspective.

2.4 Prompt Engineering and Stylistic Control

Prompt engineering has emerged as a critical discipline for directing LLM behavior. Liu et al. (2024) review the evolution of prompt design techniques, including few-shot prompting, chain-of-thought reasoning, and system prompts, demonstrating that output quality depends significantly on prompt structure. However, existing research focuses primarily on task accuracy rather than stylistic control. The present study addresses this gap by applying quantitative stylistic analysis to prompt-engineered LLM output. By grounding empirical measurement in classical linguistic theory, we seek to establish a predictive framework for stylistic control in human-AI interaction.

3. Material and Methods

3.1 LLM Selection

This study uses DeepSeek as the experimental language model. The model was selected because of its strong English generation capability and accessibility through a publicly available interface. DeepSeek is an open-source LLM that supports both research and commercial applications, making it particularly suitable for reproducible linguistic experimentation. Experiments were conducted in separate sessions to minimize possible influence from previous conversational context. The analysis represents a pilot investigation designed to examine stylistic tendencies rather than provide population-level statistical estimates.

3.2 Prompt Design

Six prompts were designed to manipulate three stylistic dimensions:

1) formality (formal vs. informal),
2) emotional tone (neutral vs. emotional), and
3) persuasive orientation (directive vs. balanced argumentation).

All prompts addressed the same topic: climate change impacts on coastal regions. This controlled topic selection reduced the influence of content variation on stylistic analysis. Each prompt requested approximately 150 words of output.

Table 1: Summary of Six Prompt Conditions

Prompt	Condition	Instruction
P1	Formal + Neutral	Write a 150-word summary of climate change impacts on coastal regions. Use a formal, academic tone.
P2	Formal + Emotional	Write a 150-word summary of climate change impacts on coastal regions. Use a formal tone but convey urgency and concern.
P3	Informal + Neutral	Explain climate change effects on coastal areas in about 150 words. Use a casual, conversational tone like you are talking to a friend.
P4	Informal + Emotional	Explain climate change effects on coastal areas in about 150 words. Use a casual tone but sound concerned about it.
P5	Persuasive Directive	Write a 150-word persuasive paragraph urging immediate action on coastal protection against climate change. Use strong, compelling language.
P6	Persuasive Neutral	Present a balanced 150-word argument about climate change adaptation for coastal regions, considering both action and economic costs.

3.3 Data Generation Process

Each prompt was submitted to the DeepSeek chatbot in a separate session to prevent conversational context contamination. For each prompt condition, one text sample was generated. The same prompt was delivered without modification, and the resulting output was collected for analysis. The generated texts ranged between 140 and 160 words, confirming adherence to the length constraint.

3.4 Linguistic Metrics and Analysis

Generated texts were evaluated using six linguistic indicators representing the three stylistic dimensions:

- **M1:** Average Sentence Length---average number of words per sentence, used as an indicator of syntactic complexity.
- **M2:** Technical Term Density---frequency of domain-specific terminology per 100 words.
- **M3:** Emotion Marker Frequency---occurrence of affective expressions, evaluative adjectives, and intensity markers per 100 words.
- **M4:** Modal Verb Frequency---occurrence of modal verbs associated with obligation, possibility, and recommendation per 100 words.
- **M5:** Contraction Frequency---frequency of contracted forms as an indicator of conversational style per 100 words.
- **M6:** Formality Index---a five-point qualitative scale evaluating lexical register, syntactic complexity, and degree of conversational language.

Quantitative measures were complemented by qualitative analysis of discourse features, including rhetorical structures, pronoun usage, passive constructions, and organizational patterns.

3.5 Abbreviations and Terms

LLM: Large Language Model,
NLP: Natural Language Processing,
P1--P6: Prompt conditions 1 through 6,
M1--M6: Metrics 1 through 6.

4. Results and Discussion

4.1 Overview

Table 2 summarizes the measured stylistic features across the six prompt conditions. Figures 1, 2, and 3 visualize differences among prompt-generated outputs.

Table 2: Quantitative Stylistic Metrics Across Six Prompt Conditions

Metric	P1	P2	P3	P4	P5	P6
Average Sentence Length	19.80	15.20	9.50	8.35	11.70	14.25
Technical Terms (/100)	11.24	10.18	3.95	4.79	7.32	10.53
Emotion Markers (/100)	0.00	4.19	1.32	7.19	3.66	0.58
Modal Verbs (/100)	1.12	2.40	3.29	3.59	6.10	4.09
Contractions (/100)	0.00	0.00	9.21	10.18	0.00	0.00
Formality Index (1--5)	4.80	4.20	1.50	1.20	3.80	4.00

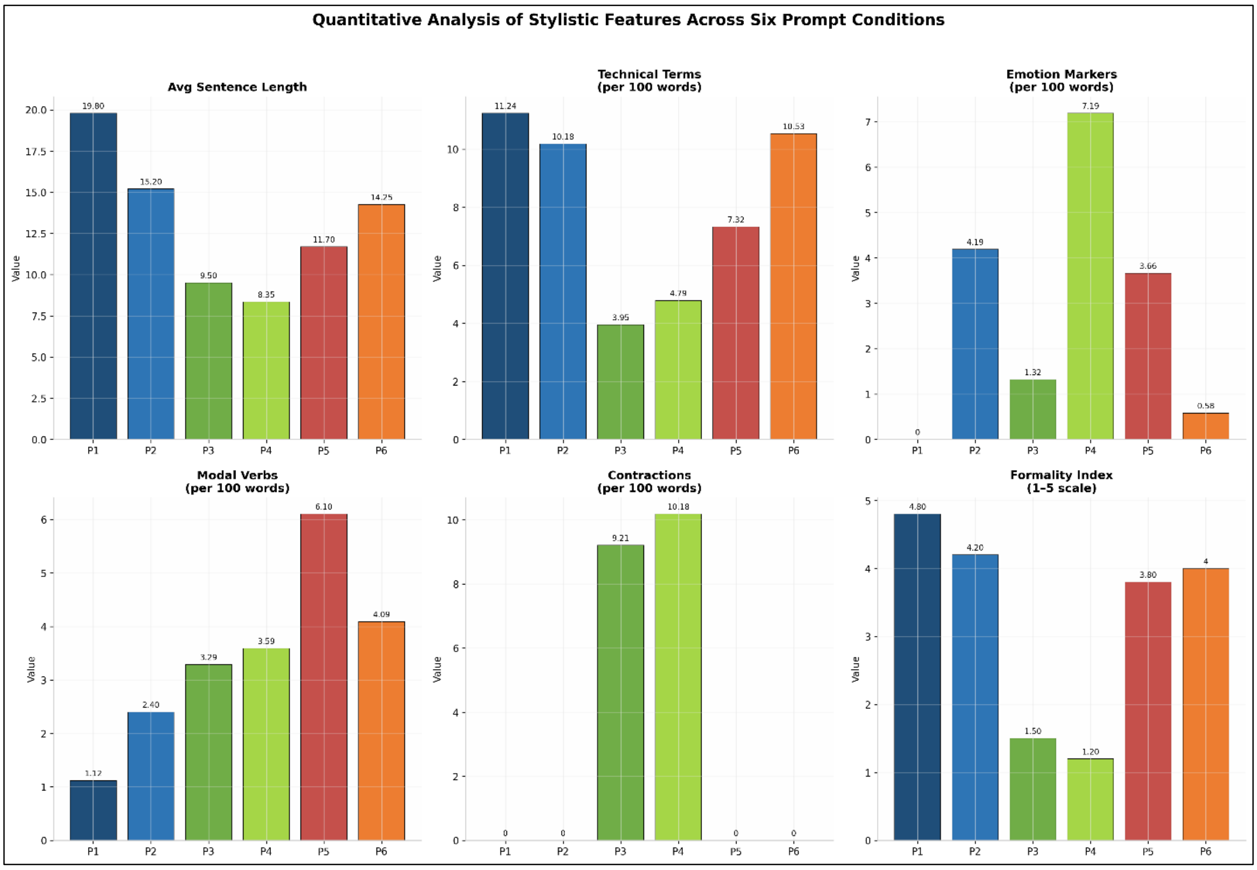

Figure 1: Quantitative analysis of stylistic features across six prompt conditions. Each subplot represents the measured values for prompts P1--P6.

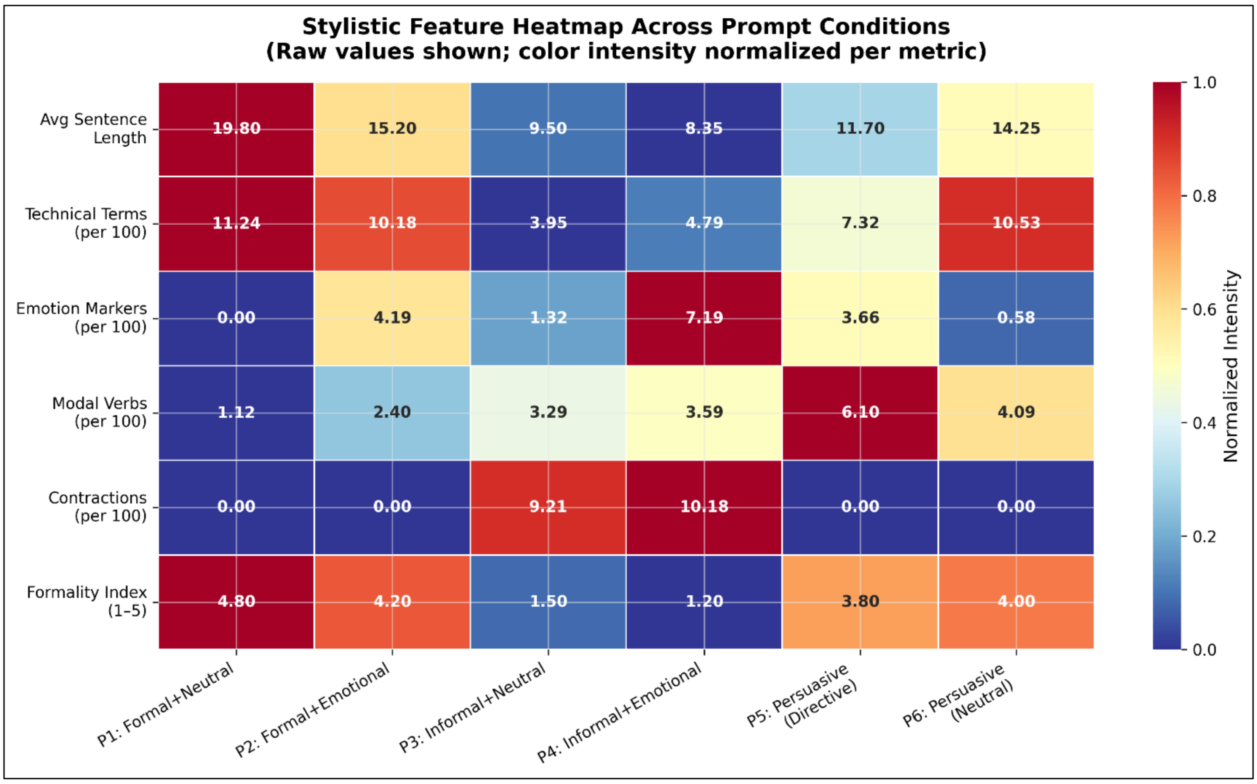

Figure 2: Normalized heatmap showing stylistic feature distributions across prompt conditions. Raw values are displayed in each cell; color intensity is normalized per metric.

4.2 Formality Dimension

The formality manipulation produced the clearest differences among the evaluated conditions. The formal neutral prompt (P1) generated the highest average sentence length (19.80 words) compared with the informal emotional condition (P4: 8.35 words). This represents a 137% increase in sentence length. Similarly, technical term density was substantially higher in formal conditions (P1: 11.24; P2: 10.18 per 100 words) compared with informal conditions (P3: 3.95; P4: 4.79).

The formality index corroborated these findings: formal prompts received higher scores (P1: 4.8; P2: 4.2), whereas informal prompts showed lower values (P3: 1.5; P4: 1.2). Contractions appeared only in informal conditions, suggesting that contracted forms functioned as a strong indicator of conversational style in this experiment.

Interestingly, introducing emotional instructions into a formal prompt influenced other linguistic features. Compared with P1, the formal emotional condition (P2) showed shorter sentences and slightly reduced technical terminology. This suggests that combining stylistic objectives may create interaction effects between linguistic dimensions. These patterns align with traditional descriptions of formal written language (Leech & Short, 2007) and confirm that LLMs can modulate formality in response to explicit instructions.

4.3 Emotional Tone Dimension

Emotion marker frequency varied according to prompt instructions. Neutral conditions produced relatively low emotional density: P1 (0.00), P3 (1.32), and P6 (0.58) markers per 100 words. By contrast, emotional prompts produced higher affective expression: P2 (4.19), P4 (7.19), and P5 (3.66) markers per 100 words.

The highest emotional density occurred in the informal emotional condition (P4), suggesting that emotional instructions combined with conversational framing may encourage stronger affective language. Qualitative analysis also indicated differences in emotional expression strategies. Formal emotional output tended to use abstract evaluative expressions (e.g., "existential threat," "catastrophic outcomes"), whereas informal emotional output relied more on personal and experiential language (e.g., "breaks my heart," "keeps me up at night").

These findings correspond with research question RQ2, confirming that explicit emotional cues can influence the degree of emotional language generated by LLMs. The distinction between formal pathos and informal emotional authenticity mirrors patterns observed in human stylistic production.

4.4 Persuasive Style Dimension

Modal verb frequency served as the primary quantitative indicator of persuasive intensity. The directive persuasive condition (P5) produced the highest modal verb frequency (6.10 per 100 words), exceeding both the formal neutral condition (P1: 1.12) and balanced persuasive condition (P6: 4.09). The difference suggests that explicit instructions requesting strong action-oriented language encourage more frequent use of obligation-related expressions such as "must" and "need to."

In contrast, the balanced persuasive condition (P6) demonstrated greater reliance on informational and hedging strategies, including expressions of possibility and uncertainty. Qualitative analysis showed that P5 frequently employed rhetorical urgency, anaphoric repetition ("We must..."), and direct appeals, while P6 demonstrated contrastive organization ("Proponents argue... However, critics point to...") and evidence-oriented argumentation.

These results address research question RQ3, indicating that prompt instructions can influence rhetorical force. The distinction between directive and neutral persuasion has practical implications for users seeking to generate specific types of argumentative discourse.

4.5 Comparative Stylistic Profiles

The combined metric patterns reveal that each prompt condition produced a distinct stylistic configuration. P1 (Formal + Neutral) represents an informational academic style with high sentence complexity and technical vocabulary density. P2 (Formal + Emotional) maintained formal characteristics while incorporating emotional markers, producing a hybrid academic-affective style. P3 (Informal + Neutral) displayed conversational characteristics through shorter sentences and increased contractions. P4 (Informal +

Emotional) produced the strongest emotional expression and highest contraction frequency, representing an affective conversational style. P5 (Persuasive Directive) showed the strongest persuasive orientation through modal verbs and urgency markers. P6 (Persuasive Neutral) combined relatively high technical density with moderate persuasive features, representing balanced argumentation.

Overall, the findings suggest that LLMs can generate multidimensional stylistic variations when provided with explicit stylistic instructions. However, these observations should be interpreted as exploratory patterns rather than statistically generalized effects due to the limited sample size.

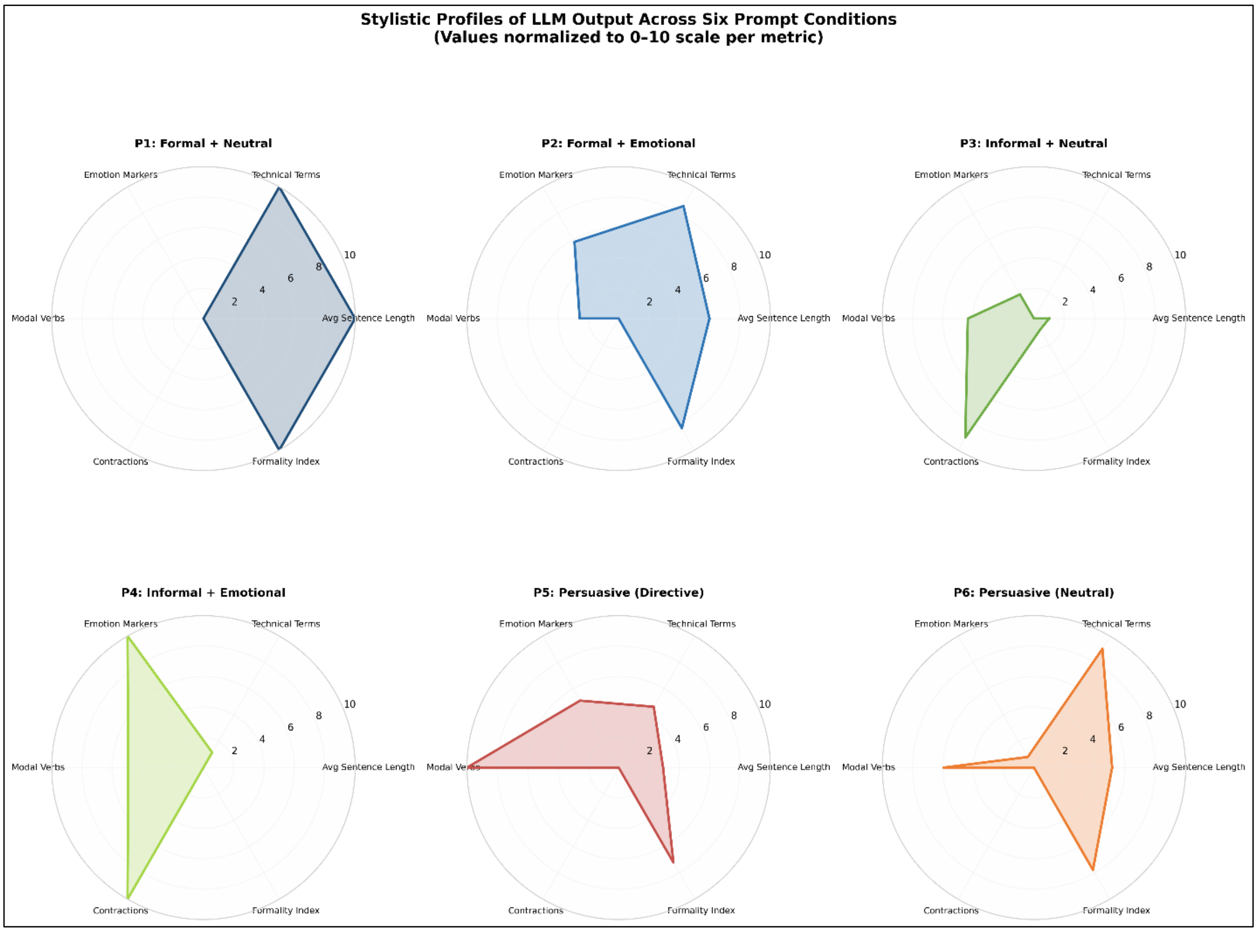

Figure 3: Multidimensional stylistic profiles of generated texts across six prompt conditions. All metrics are normalized to a 0--10 scale per dimension to enable direct visual comparison.

4.6 Evaluation of Research Questions and Hypotheses

The findings provide evidence that prompt structure can influence measurable stylistic characteristics of LLM-generated text. Across the three investigated dimensions, the observed linguistic patterns corresponded with predictions derived from classical stylistic theory.

Regarding RQ1, formal prompts were associated with increased sentence length, higher technical vocabulary density, and reduced conversational markers. For RQ2, emotional instructions produced higher frequencies of affective expressions. For RQ3, directive persuasive prompts generated more obligation-oriented modal verbs. Finally, regarding RQ4, the comparison between explicit and implicit stylistic guidance suggests that direct stylistic instructions produce stronger observable effects than general contextual framing.

These results support the main hypothesis that systematic prompt variations produce measurable stylistic shifts, with explicit cues generating stronger effects than implicit ones. However, the study also reveals interaction effects: combining stylistic objectives (e.g., formal + emotional) produces trade-offs between dimensions.

4.7 Theoretical and Practical Implications

This study contributes to computational stylistics by demonstrating that concepts traditionally applied to human-authored texts can also provide useful analytical frameworks for AI-generated language. The results suggest that LLMs reproduce many statistical relationships between communicative goals and linguistic features observed in human writing. However, stylistic generation in LLMs differs fundamentally from human stylistic choice: human writers make decisions based on social identity and audience awareness, whereas LLMs generate patterns through learned statistical associations.

The findings provide several practical recommendations for prompt engineering. For formal writing, prompts should include explicit labels such as "formal" or "academic." For informal communication, conversational instructions encourage shorter sentences and contractions. For emotional writing, prompts specifying desired emotions can increase affective vocabulary. For persuasive communication, users should distinguish between directive persuasion and balanced argumentation. Multiple stylistic goals can be combined, although interactions between dimensions may occur.

5. Recommendations

Based on the empirical findings and theoretical framework, the following recommendations are proposed for researchers, practitioners, and developers working with LLMs.

5.1 Theoretical Recommendations

Future research should expand the analytical framework to include additional stylistic dimensions such as narrative voice, humor, and irony. Cross-linguistic analysis is essential to determine whether prompt-driven stylistic control functions similarly across typologically diverse languages. The development of automated annotation tools specifically calibrated for AI-generated text would enhance the scalability of stylistic analysis.

5.2 Practical Recommendations

Users seeking formal output should employ explicit lexical markers ("formal," "academic," "professional") and avoid conversational framing. Users seeking emotional expression should specify both the type and intensity of desired affect. For persuasive tasks, the distinction between directive and neutral orientations should be made explicit in prompt design. Content creators, educators, and legal professionals can apply these principles to generate stylistically appropriate LLM output for their specific domains.

5.3 Methodological Recommendations

Future studies should employ multiple text samples per prompt condition to enable statistical testing. Comparative analysis across multiple LLM architectures (GPT-4, Claude, Gemini, Llama) would clarify whether observed patterns are model-specific or generalizable. Integration of automated NLP tools such as LIWC, spaCy, and VADER would improve measurement reliability and replicability.

5.4 Further Research Directions

The relationship between prompt complexity and stylistic fidelity warrants investigation. Additionally, human perception studies should assess whether readers can distinguish between human and AI-generated stylistic variation, and whether they perceive differences in quality or appropriateness across prompt conditions.

6. Conclusion

This study demonstrates that prompt engineering can influence measurable stylistic characteristics of Large Language Model output. By manipulating formality, emotional tone, and persuasive intent, the experiment produced systematic differences in sentence structure, vocabulary choice, emotional expression, and rhetorical strategies. The findings support the application of computational stylistics to AI-generated language and demonstrate that linguistic theories developed for human communication remain valuable tools for analysing LLM behaviour.

Although this study represents a pilot investigation with limited samples, it provides evidence that prompt design functions as an important mechanism for stylistic control in human-AI interaction. The quantitative results confirm that explicit stylistic instructions produce stronger and more predictable effects than implicit contextual cues, validating the central hypothesis of the research.

Future research combining larger datasets, automated linguistic analysis, multiple models, and human evaluation studies will further clarify the relationship between prompts and generated language style. As LLMs become increasingly embedded in professional and creative workflows, the ability to predict and control their stylistic behaviour will be essential for effective human-AI collaboration.

Acknowledgements

The author would like to express sincere gratitude to Professor Kuptsov Pavel Vladimirovich of Saratov State Technical University for his guidance and supervision throughout this research. The author also acknowledges the Faculty of Humanities at the National Research University Higher School of Economics for providing academic support and resources.

The author declares that while preparing this manuscript, AI technology was used for conceptual clarification, APA formatting guidance, data analysis support, translation assistance, and writing improvement. The entire manuscript was written solely by the author, and all AI-generated content was reviewed, checked, and rewritten using the author's own words. The author is responsible for all content and conclusions presented herein.

Conflict of Interest Statement

The authors declare no conflicts of interest.

About the Author(s)

Hossain Md. Sayem is a Master's Student in the Faculty of Humanities at the National Research University Higher School of Economics (HSE University), Nizhny Novgorod, Russia. His field of study is Fundamental and Applied Linguistics. His research interests include computational stylistics, prompt engineering, Large Language Models, and the linguistic analysis of AI-generated text. He holds a degree in Linguistics and is currently investigating the intersection of natural language processing and classical linguistic theory.

Arif Md. Ismail Hossain is a Graduate Student at the Faculty of Economics, National Research University Higher School of Economics (HSE University), Nizhny Novgorod, Russia. His research interests include monetary economics, financial market development, emerging market macroeconomics and AI. He holds a BBA in Finance from Premier University, Chittagong, Bangladesh.

References

Adamopoulou E, Moussiades L, 2020. Chatbots: History, technology, and applications. Machine Learning with Applications 2: 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Alboqami H, 2022. Factors affecting consumers adoption of AI-based chatbots: The role of anthropomorphism. In: Pehcevski J (Ed.), Chatbots and Text Generation, Arcler Press, pp. 29-56. https://doi.org/10.4236/ajibm.2023.134014

Biber D, 1988. Variation across Speech and Writing. Cambridge University Press, Cambridge, UK. Retrieved from https://doi.org/10.1017/CBO9780511621024

Ciesla R, 2024. The Book of Chatbots: From ELIZA to ChatGPT. Springer Nature. Retrieved from https://doi.org/10.1007/978-3-031-51004-5

Halliday MAK, 1985. An Introduction to Functional Grammar. Edward Arnold, London, UK. Retrieved from https://books.google.ro/books/about/An_Introduction_to_Functional_Grammar.html?id=JM3KAgAAQBAJ&redir_esc=y

Kolmogorova AV, Margolina AV, 2024. Written vs generated text: "Naturalness" as a textual and psycholinguistic category. Research Result: Theoretical and Applied Linguistics 10(2): 71-99. https://doi.org/10.18413/2313-8912-2024-10-2-0-4

Leech G, Short M, 2007. Style in Fiction: A Linguistic Introduction to English Fictional Prose, 2nd edn. Pearson, Harlow, UK. Retrieved from https://books.google.ro/books/about/Style_in_Fiction.html?id=mScMAjO9qjkC&redir_esc=y

O'Sullivan J, 2025. Stylometric comparisons of human versus AI-generated creative writing. Humanities and Social Sciences Communications 12: 1708. https://doi.org/10.1057/s41599-025-05986-3

Swales JM, 1990. Genre Analysis: English in Academic and Research Settings. Cambridge University Press, Cambridge, UK. Retrieved from https://books.google.ro/books/about/Genre_Analysis.html?id=shX_EV1r3-0C&redir_esc=y

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I, 2017. Attention is all you need. Advances in Neural Information Processing Systems 30: 5998-6008. https://doi.org/10.48550/arXiv.1706.03762

Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, Min Y, Zhang B, Zhang J, Dong Z, Du Y, Yang C, Chen Y, Chen Z, Jiang J, Ren R, Li Y, Tang X, Liu Z, Liu P, Nie JY, Wen JR, 2023. A survey of large language models. arXiv. Retrieved from https://arxiv.org/abs/2303.18223