

Application of Prompt Engineering Techniques to Optimize Information Retrieval in the Metaverse

Muhammed Abdulhamid
KARABIYIK

Niğde Ömer Halisdemir University,
Niğde, Türkiye
abdulhamidkarabiyik@ohu.edu.tr
0000-0001-7927-8790

Fatma Gülşah TAN

Isparta Süleyman Demirel University,
Isparta, Türkiye
gulsah@sd.u.edu.tr
0000-0002-2748-0396
(Corresponding Author)

Asım Sinan YÜKSEL

Isparta Süleyman Demirel University,
Isparta, Türkiye
asimyuksel@sdu.edu.tr
0000-0003-1986-5269

Abstract— Prompt engineering techniques are instructions that enable large language models (LLMs) to solve real-world problems more effectively. These techniques enhance the capabilities of LLMs to generate accurate and efficient responses. Our study examines the challenge of acquiring comprehensive and efficient information in the metaverse through the application of various prompt engineering techniques. The main objective is to improve the accuracy and effectiveness of metaverse-related responses by leveraging LLM capabilities. In this study, 100 questions were generated using GPT, GEMINI, QWEN, and MISTRAL language models focusing on the metaverse. Our experiments indicated that responses often included unrelated information, highlighting the need for prompt engineering techniques. We applied knowledge-based, rule-based, few-shot, and template-based prompt engineering techniques to refine the responses. The performance of GPT, GEMINI, QWEN, and MISTRAL models were evaluated based on criteria including accuracy, timeliness, comprehensiveness, and consistency. Our findings reveal that prompt engineering techniques significantly enhance the efficacy of LLMs in providing improved information retrieval and response generation, aiding users in efficiently acquiring information in complex environments like the metaverse.

Keywords— Prompt engineering, large language models, metaverse, information retrieval, response generation

I. INTRODUCTION

The metaverse represents both physical and digital environments, integrating augmented and virtual realities [1]. Within this space, users engage in a variety of activities, ranging from social interactions to commercial transactions and work through unique digital representations known as avatars. Consequently, user experiences in the metaverse are shaped by information retrieval and orientation strategies that facilitate personalized interactions. This digital environment holds the potential to significantly transform how individuals interact online and access information. However, the metaverse's dynamic nature and the extensive sharing of information among users pose challenges in accessing information within this environment. Information in virtual settings is often dispersed and unstructured, complicating the retrieval of accurate and reliable data. Additionally, the rapid increase in virtual interactions contributes to information pollution and uncertainty in the information access process. The substantial time and effort users must invest to locate specific information may decrease the metaverse's effectiveness as an efficient information acquisition platform.

Large language models (LLMs) and prompt engineering techniques are pivotal tools for enhancing information access in the metaverse. LLMs, trained on extensive datasets, effectively process information within complex and diverse environments such as the metaverse. Prompt engineering is a method designed to ensure these models respond appropriately to user queries. By crafting precise prompts, information retrieval in the metaverse can be significantly optimized. Consequently, LLMs and prompt engineering offer essential solutions to the challenges faced in the metaverse's information acquisition process.

Our main object of study is to extract meaningful and accurate Metaverse data using LLMs such as GPT, GEMINI, QWEN, and MISTRAL which are top ranked in benchmarks. The Generative Pre-trained Transformer (GPT) is recognized for its efficiency in processing large volumes of text-based information [2]. The GEMINI language model, an artificial intelligence tool used in natural language processing tasks, prioritizes user experience and delivers responses with high knowledge-based accuracy [3]. On the other hand, Qwen model, has been noted for its innovative architecture that supports various applications, including natural language-based applications [4]. This model's foundation in pre-trained weights allows it to leverage existing knowledge while adapting to new tasks effectively. Mistral, has been recognized for its performance on the MMLU benchmark, attributed to its innovative two-stage knowledge distillation process which enhances its understanding and generation of human-like text [5]. This duality of strengths and weaknesses underscores the importance of ongoing research to refine these models, particularly in areas requiring high precision. All models have demonstrated success in understanding and interpreting natural language without employing prompt engineering techniques.

Prompt engineering techniques are employed to ensure that models generate responses in the desired format, aligning with specific knowledge patterns or user requirements. Queries leveraging these techniques are crafted within defined constraints, providing templates or data elements to LLMs [6]. Implementing these strategies in large language models can yield accurate and comprehensive responses to inquiries. Within this framework, various prompt engineering approaches are utilized, including knowledge-based, rule-based, few-shot, and template-based methods. This study also examines the consistency of these models in delivering answers to diverse questions within a given context.



Our experiments have demonstrated that prompt engineering techniques significantly enhance the retrieval of Metaverse-related information using large language models. Our contributions are as follows:

- Extending literature on the prompt engineering in the Metaverse
- Establishing a foundation for enhancing user experiences in virtual environments.
- Advancing discussions on LLM adaptability in dynamic digital spaces.
- Improving access to accurate information in the Metaverse.
- Solving information overload for users through optimized prompts.
- Support for diverse use cases like education, events, and business interactions.
- Applying specific prompt techniques to enhance data precision in LLMs.
- Providing a framework for evaluating accuracy, comprehensiveness, timeliness, and consistency.

II. RELATED WORK

Prompt engineering techniques are recognized for their role in enhancing the capabilities of large language models, particularly in the domain of knowledge discovery. The Metaverse, with its vast and intricate datasets, presents significant challenges in efficiently searching and extracting knowledge. In response, research efforts have concentrated on exploring the use cases of the Metaverse, examining the application of prompt engineering techniques in large language models, and conducting experiments to assess their performance comprehensively.

Huang [7] highlighted the potential applications of AI technologies in the educational context, emphasizing the necessity of prompt engineering for effective knowledge acquisition. His findings indicate that the methods employed significantly influence the quality of questions and responses. Huang also stated that employing combined prompt engineering techniques with ChatGPT technology in education improves the teaching process for both learners and tutors. Conversely, Tassoti [8] explored the broader application of new technologies in education, noting a lack of focus on specific teaching challenges. He determined that the 5S method, used in AI applications, generates responses that are more effective and of higher quality compared to the traditional question-and-answer approach. The 5S method comprises five steps: Set the Stage, Be Specific, Simplify Your Language, Structure the Output, and Share Feedback.

The application of prompt engineering techniques in large language models has a significant and positive impact on improving performance. Notably, few-shot prompts have been observed to outperform traditional data augmentation methods by a substantial margin in scenarios with limited samples [9, 10]. In their study, Gao et al. [11] determined that few-shot

prompts were more advantageous for the GPT-3 model compared to data augmentation without prompt engineering, resulting in a performance improvement of nearly 30% for each task. Overall, prompt engineering techniques contribute to enhancing the efficiency of large language model systems. Wang et al. [12] and Shin et al. [13] emphasized the importance of these techniques in quickly comprehending various knowledge classifications, as well as in synthesizing and generating new knowledge. Despite these benefits, concerns persist regarding the limitations of integrating different prompt engineering techniques [14].

In a vast universe like the Metaverse, the application of prompt engineering techniques can yield more effective and interactive outcomes. Users who generate their own content within the Metaverse can benefit from the question-answering and scenario evaluation capabilities of large language models. Elsadig et al. [15] noted the complexity of the Metaverse and underscored the importance of integrating prompt engineering techniques to enhance AI system performance as the Metaverse advances. Haque et al. [16] investigated the Metaverse's role in education, highlighting the benefits of using these techniques with large language models to improve the learning process. Sun et al. [17] explored the potential applications of the Metaverse in plastic surgery, compiling research on its medical applications up to 2022. Their findings emphasize the Metaverse's promise as a tool in plastic surgery, though they call for further research and evaluation. Lee and Kwon [18] examined methods for integrating brands into the Metaverse, stating that marketing strategies developed using prompt engineering techniques have increased brand visibility.

III. METHODOLOGY

The research methodology was structured to examine the effectiveness of large language models (LLMs) and the impact of performance enhancement techniques in efficiently delivering accurate data within the Metaverse environment. Prompt techniques were used for this purpose. Prompt engineering is a method used in natural language processing to optimise the performance of large language models. It involves the design and construction of customised prompts that elicit appropriate and correct responses [19]. In our study, 4 different prompt engineering techniques were examined and their effects on performance were analysed. These techniques are knowledge-based, rule-based, few-shot and template-based. Knowledge-based prompting is a way of using information from specific sources to extend the knowledge of the language model and to focus on specific topics [20]. Rule-based prompting is a technique used to control the language model according to a set of rules. These rules determine the type of text that the language model produces for a given task or topic. In developing the rules for our rule-based prompting technique, we employed a systematic approach that began with establishing clear criteria and objectives that would guide the rule creation process. We aimed to ensure that each rule would maximize the relevance, neutrality, and informativeness of responses within the context of the metaverse. Our primary source of domain knowledge included an extensive review of existing metaverse literature, complemented by insights gained from datasets and consultations with subject matter experts. These inputs were crucial in grounding the rules in current

understanding and practical relevance to real-world applications. For instance, the rule that answers should not exceed two paragraphs was based on established communication practices that balance detail and conciseness. Additionally, the rule that stipulates responses be based solely on metaverse information while maintaining neutrality ensures the accuracy and objectivity of the output. We complemented this with an iterative refinement process, where rules were subjected to pilot testing to iron out ambiguities and enhance their effectiveness. This comprehensive methodology ensured that our rule-based framework was robust and capable of producing precise and contextually appropriate responses. A few-shot prompting is a technique used so that language

models can understand new tasks or classes with only a few examples [21]. The main purpose of template-based prompts is to help language models to produce text that conforms to a particular structure or format [22]. The approach comprised the following steps:

- Question formulation
- Response collection without prompt engineering
- Response collection with prompt engineering
- Performance evaluation

TABLE I. RESPONSES FROM GPT, GEMINI, QWEN AND MISTRAL WITHOUT APPLYING PROMPT ENGINEERING

Question	GPT	GEMINI	QWEN	MISTRAL
What types of events are organized in simulation?	Social events, educational seminars, concerts, competitions, trade fairs and various interactive games are often organized in simulations. Events can often have a variety of themes that users can interact with, experience and learn from.	Simulations are a powerful tool that can be used in many different fields by mimicking real-world situations. Therefore, there is a wide range of types of activities that can be organized. Simulation technologies are rapidly developing and being used in many fields. In this way, it is possible to create more effective learning, better decision making and more efficient business processes.	This Physics simulation team activity is designed to help students understand classical mechanics. They are introduced to models of the motion of objects at rest and uniform linear motion due to unbalanced forces. In addition, they experience vertical jump and ballistic motion through a cricket ball back into the hole.	Simulations are invaluable across various fields, serving numerous purposes that enhance training, research, decision-making, entertainment, and more. In sectors like aviation, medicine, and military operations, they provide safe environments for trainees to practice and develop skills. In research and development, simulations allow for testing theories and prototypes without the need for physical models.
What are the most popular virtual sports or competitive games on virtual platforms?	Popular virtual sports and competitive games include soccer, basketball, racing games and fighting games. Games often offer multiplayer and competitive features, allowing players to compete in real time.	Virtual sports have gained immense popularity, especially with professional gaming leagues called e-sports. Games such as "League of Legends", "Dota 2", "Counter-Strike: Global Offensive" are followed by millions of players. Sports simulations and racing games are also very popular.	The most popular virtual sports or competitive games include titles such as football (soccer), basketball, tennis, and various racing games. These games allow players to compete online against friends or others worldwide, offering an engaging way to experience sports in a virtual environment. Additionally, esports has gained significant popularity, with competitive gaming tournaments for games like League of Legends, Fortnite, and Call of Duty drawing large audiences.	Fortnite, developed by Epic Games, is a widely popular online multiplayer battle royale game where up to 100 players compete to be the last one standing on a shrinking island, available on PC, console, and mobile platforms. League of Legends (LoL), created by Riot Games, is a renowned multiplayer online battle arena (MOBA) game, known for its vibrant esports scene and millions of players engaged in professional leagues and tournaments.
Which virtual beauty saloons or spas are available?	Virtual beauty salons and spa centers offer beauty and grooming services in virtual environments. Users can experience virtual beauty services and receive digital spa treatments.	Some platforms have virtual beauty salons and spas where you can change the look of your avatar. In these centers, you can get a haircut, apply makeup, get skincare and try on different outfits.	Virtual beauty salons and spas are increasingly popular as they allow individuals to explore new beauty treatments and services from the comfort of their homes. These platforms offer a variety of services, such as virtual spa treatments, which deliver facials, massages, and body treatments directly to clients, often through advance bookings for at-home delivery or specific locations.	Several virtual beauty salons and spas have emerged, offering innovative services to clients. GlamSquad provides a mobile beauty service and features GlamSquad Live, where users can book virtual makeup and hair tutorials with professional stylists. Paintbox, known for its nail artistry, offers virtual nail art classes through its Paintbox at Home platform, allowing users to learn intricate designs from home.
How is security and privacy ensured in virtual platforms?	Security and privacy are ensured through methods such as encryption, user authentication, privacy settings and protection of user data. Platforms generally offer various security measures to protect users' privacy.	Security and privacy on virtual platforms is ensured by strong encryption methods and user agreements. However, it is important to protect your personal information and choose reliable platforms.	Security and privacy are paramount on virtual platforms, particularly concerning users' personal information and online activities. Two-Factor Authentication (2FA) adds an extra layer of security by requiring users to verify their identity with a password and a code sent to their device. Secure communication protocols, such as HTTPS, ensure that data transmissions are protected from eavesdropping, while regular security audits help identify and address vulnerabilities over time.	Ensuring security and privacy on virtual platforms involves several key strategies, including data encryption to protect information at rest, in transit, and during use; implementing strict access controls through multi-factor authentication and role-based permissions; designing a secure network architecture that minimizes the attack surface; and conducting regular security audits to identify vulnerabilities. Additionally, educating users on security best practices, collecting only essential data to reduce privacy risks, and using secure payment systems to protect financial information are crucial.

A. Question Formulation

In this step, 100 questions concerning Metaverse were formulated for large language models to address. These questions were crafted to cover current developments, usage, areas of interest, and applications of the Metaverse. The purpose was not only to acquire knowledge but also to enhance the analytical thinking capabilities of the large language models.

Sample questions in this study are as follows:

- What types of events are organized in the metaverse?
- What are the most popular virtual sports or competitive games in virtual platforms?
- Which virtual beauty salons or spas are available?
- How are security and privacy provided in metaverse?

B. Response Collection without Prompt Engineering

In this step, the questions were input into the GPT, GEMINI, QWEN, and MISTRAL models, and responses were collected without employing any prompting techniques. This step aimed to evaluate and compare the performance of language models that did not utilize prompt engineering techniques. Table I displays the responses generated by the

GPT, GEMINI, QWEN, and MISTRAL models for sample questions, illustrating outcomes without the application of prompt engineering techniques.

C. Response Collection with Prompt Engineering

In the third step, responses were obtained by employing rule-based, knowledge-based, and template-based prompt engineering techniques to compare the performance of the language models. The goal is to enhance the performance of large language models by applying these techniques. Rule-based prompt engineering ensures that responses adhere to specific rules and are subject-specific [23]. This technique is beneficial in the Metaverse's complex and dynamic environment, helping to produce more consistent answers [24]. Knowledge-based prompt engineering restricts data to obtain responses solely related to the Metaverse [25]. Template-based prompt engineering is used to format responses in the desired manner. Utilizing these techniques together optimizes large language models to focus on certain topics and produce efficient and consistent results [26]. Figure 1 illustrates the prompt template created for information retrieval using knowledge-based, rule-based, and template-based prompt engineering methods. Table II shows the sample responses from GPT, GEMINI, QWEN and MISTRAL models with the prompt engineering techniques applied.

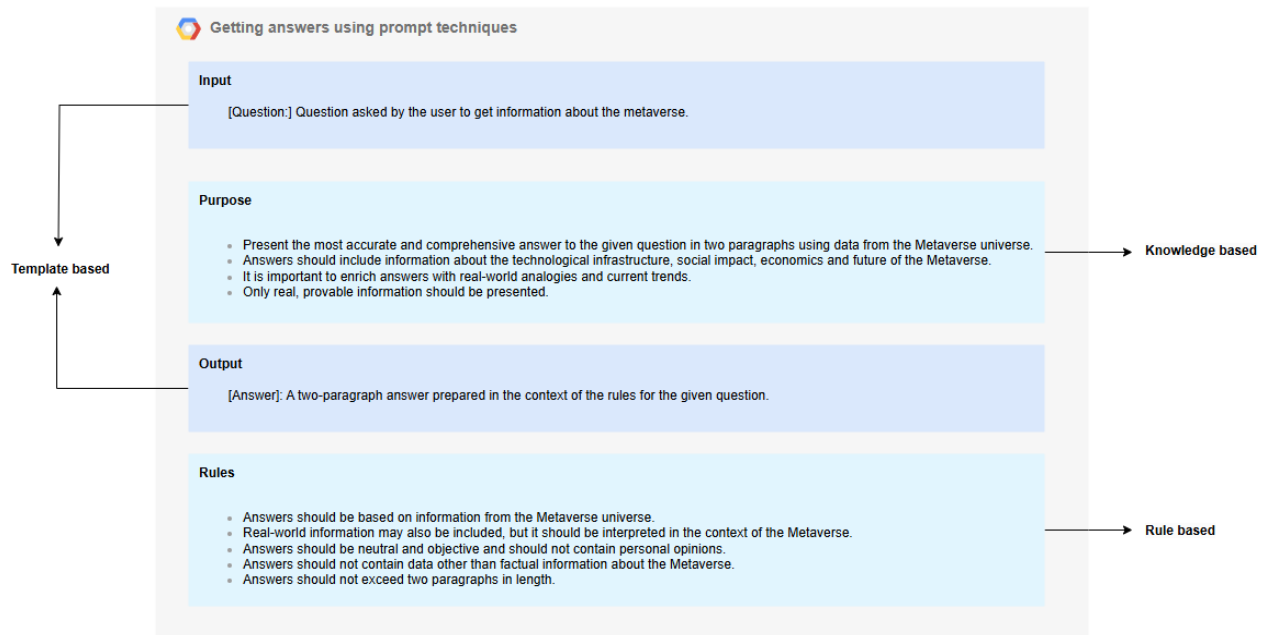


Fig. 1. PROMPT ENGINEERING TEMPLATE

D. Performance Evaluation

In the fourth section, the answers obtained using the techniques applied in the third section were analyzed and scored based on four categories: accuracy, comprehensiveness, timeliness, and consistency. The LLM-as-a-Judge method is used to analyse the responses of large language models. LLM-as-a-Judge involves an LLM evaluating the responses of another LLM against a given set of criteria. Since traditional human evaluations are slow and costly, evaluations using LLMs are an alternative solution. To evaluate the performance of different language models, the answers were analyzed

separately for each model. This scoring system employed rule-based prompting, few-shot prompting, and template-based prompting techniques. Few-shot prompting is a method that enhances the performance and efficiency of large language models by providing examples on a specific topic, making model performance better [27]. This is particularly effective when fewer examples are provided through backpropagation [14]. Figure 2 illustrates a concise example where all three methods—few-shot prompting, rule-based prompting, and template-based prompting—are applied, along with their corresponding scoring.

TABLE II. RESPONSES FROM GPT, GEMINI, QWEN AND MISTRAL BY APPLYING PROMPT ENGINEERING TECHNIQUES

Question	GPT	GEMINI	QWEN	MISTRAL
What types of events are organized in simulation?	In the Metaverse, users can organize and participate in a variety of events. These events include a wide range of events, such as virtual concerts, art exhibitions, educational seminars, business meetings, social gatherings, sports competitions, and gaming tournaments. Especially during the COVID-19 pandemic, with real-world events being moved to the digital environment, such activities in the Metaverse have become more popular. For example, famous artists can reach their fans worldwide by giving virtual concerts, while companies can continue their business processes by holding virtual meetings with their global employees.	Metaverse provides a wide platform where users can interact in virtual environments by creating a digital copy of the real world through virtual reality and augmented reality technologies. Events organized on this platform offer new experiences by exceeding the boundaries of traditional events. Many different events, from educational seminars to virtual concerts, fashion shows to virtual meetings, can be held on metaverse platforms. Users can move freely in the virtual world with their own avatars, interact with other users and interact with virtual objects. This makes metaverse events more attractive for both participants and organizers.	In the metaverse, simulation events often include virtual training, social gatherings, and workshops. These events give users a chance to interact, learn, and practice skills in a digital environment.	Metaverse simulations organize events like virtual trainings, interactive social meetups, and collaborative sessions. Such events create a digital space for users to engage, learn, and experience activities similar to the real world.
What are the most popular virtual sports or competitive games on virtual platforms?	Popular virtual sports and competitive games on virtual platforms often provide users with interactive and dynamic experiences. These include virtual football, basketball and e-sports tournaments. Virtual sports events can be organized on platforms such as "VRChat", while social VR applications such as "Rec Room" and "Horizon Worlds" offer areas where users can play and compete in virtual sports games. In addition, popular battle royale games such as "Fortnite" and "Apex Legends" are among the main games where virtual competition is prominent.	The Metaverse pushes the boundaries of traditional sports and games, offering its users unique and interactive experiences. With virtual reality and augmented reality technologies, users can play sports in the virtual world, participate in tournaments, and compete with other players. Popular virtual sports include virtual football, basketball, racing, and fighting games. These games bring the excitement and competition of real-world sports to the virtual environment, attracting millions of users	In the metaverse, popular virtual sports and games include VR racing, simulated soccer, and virtual basketball. These games allow players to compete in lifelike, interactive environments.	On virtual platforms in the metaverse, competitive games like VR racing, soccer simulations, and multiplayer battle arenas are widely popular. These provide an immersive sports experience for users.
Which virtual beauty salons or spas are available?	There are various virtual beauty salons and spas in the Metaverse. These centers allow avatars to receive beauty treatments, makeup, hair styling, and relaxation services in a virtual environment. Users can experience these services to personalize their avatars' appearance and renew themselves in the virtual world. Virtual beauty salons help users express their virtual selves and express themselves through their avatars.	Some platforms feature virtual beauty salons and spas where users can change the appearance of their avatars, and services such as hair styling, makeup and skin care can be received.	In the metaverse, virtual beauty salons and spas provide services like avatar styling, virtual makeovers, and relaxation experiences. Users can personalize their avatars and enjoy immersive spa environments.	Virtual beauty salons and spas in the metaverse offer avatar customization, styling sessions, and calming digital environments for relaxation. These spaces help users personalize and unwind in a unique virtual setting.
How is security and privacy ensured in virtual platforms?	Security and privacy in virtual platforms are usually provided by advanced technology and protocols. Blockchain technology increases the security of ownership and transactions by encrypting and decentralizing user data. In addition, virtual platforms implement additional measures such as firewalls, encryption techniques, and regular security audits to protect user data. User authentication, two-factor authentication (2FA), and encryption are used to prevent unauthorized access and prevent data leaks.	While the metaverse offers its users a wide range of experiences, it also brings security and privacy concerns. In this virtual universe, users' data, identities, and digital assets must be protected. Platforms try to ensure security by encrypting users' personal information and preventing unauthorized access. In addition, account security is increased with methods such as multi-factor authentication. In terms of privacy, it is important for users to have control over what data they share.	In the metaverse, security and privacy are managed with encryption, user verification, and controlled data sharing to protect user information and interactions.	Virtual platforms in the metaverse enhance security through encryption, identity verification, and privacy settings. These measures help safeguard user data and maintain a secure environment for interactions.

Table III, IV, V and VI present the average scores, comparison of results with and without the application of prompt engineering techniques for responses generated by GPT, GEMINI, QWEN and MISTRAL language models respectively.

Results indicate that models employing prompt engineering techniques achieved higher average scores across

all categories compared to models without such techniques. Prompt engineering enables models to generate more accurate, comprehensive, timely, and consistent responses. It is recognized as a critical tool for maximizing the potential of language models, allowing them to be better optimized for specific tasks and resulting in more accurate and useful outcomes. The results also demonstrate that language models can be utilized more effectively in Metaverse domains.

Scoring answers with prompt techniques

Input

[Question]: Question text to be evaluated.
 [Answer 1]: GPT answer.
 [Answer 2]: GEMINI answer.

Purpose

- Score each answer according to certain criteria and calculate the average score.

Output

- Scores between 1-10 for each answer in line with the given criteria and the average of these scores.

Rules

- Answers should only be relevant to the topic of the Metaverse.
- The assessment should be based on objective criteria.

Criteria

- Accuracy: Assess how well the answers align with the concepts of the Metaverse.
- Comprehensiveness: Examine how comprehensively the answers address all aspects of the question.
- Timeliness: Assess how well the answers reflect the latest developments in the metaverse.
- Consistency: Check how consistent the answers are with other metaverse sources.

Sample Usage

Question: What impact can the Metaverse have on the education system in the future?
 Answer 1: The Metaverse can make learning more fun by providing students with a more interactive learning experience in virtual classrooms.
 Answer 2: Metaverse can contribute to the development of social skills by allowing people to interact in virtual worlds.

Example Output:

Answer 1:
 Accuracy: 8,
 Comprehensiveness: 7,
 Timeliness: 9,
 Consistency: 8,
 Average Score: 8

Answer 2:
 Accuracy: 6,
 Comprehensiveness: 5,
 Current Affairs 7,
 Consistency: 6,
 Average Score: 6

FIG II. PERFORMANCE EVALUATION PROMPT TEMPLATE

TABLE III. PERFORMANCE COMPARISON FOR GPT LANGUAGE MODEL

Category	GPT (without prompt engineering)	GPT (with prompt engineering)
Accuracy	7.17	9.2
Comprehensiveness	5.93	9.11
Timeliness	5.94	9.03
Consistency	6.88	9.16

TABLE V. PERFORMANCE COMPARISON FOR QWEN LANGUAGE MODEL

Category	QWEN (without prompt engineering)	QWEN (with prompt engineering)
Accuracy	5.2	7.8
Comprehensiveness	3.58	6.45
Timeliness	3.16	6.9
Consistency	5.4	7

TABLE IV. PERFORMANCE COMPARISON FOR GEMINI LANGUAGE MODEL

Category	GEMINI (without prompt engineering)	GEMINI (with prompt engineering)
Accuracy	6.13	9.14
Comprehensiveness	4.75	9.23
Timeliness	5.14	9.12
Consistency	6.4	9.03

TABLE VI. PERFORMANCE COMPARISON FOR MISTRAL LANGUAGE MODEL

Category	MISTRAL (without prompt engineering)	MISTRAL (with prompt engineering)
Accuracy	6.36	8.1
Comprehensiveness	4.36	7.2
Timeliness	4	7.1
Consistency	5.49	7.34

IV. CONCLUSION

One of the significant challenges in the Metaverse is accessing accurate and reliable information due to its complex and dynamic nature. Ensuring that users receive precise, comprehensive, timely, and consistent information is essential for effective engagement and decision-making within this virtual environment. This study evaluated the potential of prompt engineering techniques to enhance information retrieval in the Metaverse using large language models, specifically GPT, GEMINI, QWEN, and MISTRAL. By analyzing responses to Metaverse-related questions, we experimented with various prompt engineering methods, including rule-based, knowledge-based, template-based, and few-shot prompting. Our results highlighted improvements in accuracy, comprehensiveness, timeliness, and consistency. GPT demonstrated enhancements of 28.31% in accuracy, 53.63% in comprehensiveness, 52.02% in timeliness, and 33.14% in consistency. GEMINI showed even greater improvements, with 49.10% in accuracy, 94.32% in comprehensiveness, 77.43% in timeliness, and 41.09% in consistency. QWEN demonstrated enhancements of 50% in accuracy, 80.17% in comprehensiveness, 118.35% in timeliness, and 29.63% in consistency. MISTRAL showed enhancements of 27.36% in accuracy, 65.14% in comprehensiveness, 77.5% in timeliness, and 33.61% in consistency. These findings underscore the effectiveness of prompt engineering in refining the outputs of large language models to provide accurate and detailed information about the Metaverse. Future work could focus on developing more adaptive and context-specific prompt engineering methods tailored to the unique requirements of the Metaverse. Additionally, incorporating human evaluation metrics alongside automated assessments may enhance the robustness of performance evaluations.

ACKNOWLEDGEMENT

None.

AUTHORS' CONTRIBUTIONS

All authors have participated in drafting the manuscript. All authors read and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

FUNDING

This research did not receive any outside funding or support. The authors report no involvement in the research by the sponsor that could have influenced the outcome of this work.

DATA AVAILABILITY

The dataset used in this study is publicly available at "Metaverse dataset", Mendeley Data, V1, doi: 10.17632/br68w4gn43.1

ETHICAL STATEMENT

In this article, the principles of scientific research and publication ethics were followed. This study did not involve human or animal subjects and did not require additional ethics committee approval.

REFERENCES

- [1] Mystakidis, S. (2022). Metaverse. Encyclopedia, 2(1), 486-497.
- [2] Phoenix, J., & Taylor, M. (2024). Prompt engineering for generative AI: future-proof inputs for reliable AI outputs at scale. O'Reilly Media, Inc..
- [3] Islam, R., & Ahmed, I. (2024, May). Gemini-the most powerful LLM: Myth or Truth. In *2024 5th Information Communication Technologies Conference (ICTC)* (pp. 303-308). IEEE.
- [4] Bai, S., Zheng, Z., Wang, X., Lin, J., Zhang, Z., Zhou, C., ... & Yang, Y. (2021). Connecting language and vision for natural language-based vehicle retrieval. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 6, 4029-4038. <https://doi.org/10.1109/cvprw53098.2021.00455>.
- [5] Wilkins, J. and Rodriguez, M. (2024). Higher performance of mistral large on mmlu benchmark through two-stage knowledge distillation.. <https://doi.org/10.21203/rs.3.rs-4410506/v1>.
- [6] Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. *arXiv preprint arXiv:2402.07927*. <https://doi.org/10.48550/arXiv.2402.07927>.
- [7] Huang, J. (2023). Engineering ChatGPT prompts for EFL writing classes. *International Journal of TESOL Studies*, 5(4), 73-79.
- [8] Tassoti, S. (2024). Assessment of students use of generative artificial intelligence: prompting strategies and prompt engineering in chemistry education. *Journal of Chemical Education*, 101(6), 2475-2482. <https://doi.org/10.1021/acs.jchemed.4c00212>.
- [9] Cui, G., Hu, S., Ding, N., Huang, L., & Liu, Z. (2022). Prototypical verbalizer for prompt-based few-shot tuning. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. <https://doi.org/10.18653/v1/2022.acl-long.483>
- [10] Schick, T. and Schütze, H. (2021). It's not just size that matters: small language models are also few-shot learners. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*. <https://doi.org/10.18653/v1/2021.naacl-main.185>.
- [11] Gao, T., Fisch, A., & Chen, D. (2021). Making pre-trained language models better few-shot learners. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Confer.* <https://doi.org/10.18653/v1/2021.acl-long.295>.
- [12] Wang, L., Chen, R., & Li, L. (2023). Knowledge-guided prompt learning for few-shot text classification. *Electronics*, 12(6), 1486. <https://doi.org/10.3390/electronics12061486>.
- [13] Shin, T., Razeghi, Y., Logan, R. L., Wallace, E., & Singh, S. (2020). Autoprompt: eliciting knowledge from language models with automatically generated prompts. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. <https://doi.org/10.18653/v1/2020.emnlp-main.346>.
- [14] Shi, C., Zhai, R., Song, Y., Yu, J., Li, H., Wang, Y., ... & Wang, L. (2023). Few-shot sentiment analysis based on adaptive prompt learning and contrastive learning. *Information Technology and Control*, 52(4), 1058-1072. <https://doi.org/10.5755/j01.itc.52.4.34021>
- [15] Elsadig, M., Alohali, M. A., Ibrahim, A. O., & Abulfaraj, A. W. (2024). Roles of blockchain in the metaverse: concepts, taxonomy, recent advances, enabling technologies, and open research issues. *IEEE Access*, 12, 38410-38435. <https://doi.org/10.1109/access.2024.3367014>.
- [16] Haque, M. A., Rahman, M., Md. Faizanuddin, & Anwar, D. (2023). Educational horizons of the metaverse: vision, opportunities, and



- challenges. *Metaverse Basic and Applied Research*, 3, 60. <https://doi.org/10.56294/mr202460>.
- [17] Sun, P., Zhao, S., Yang, Y., Liu, C., & Pan, B. (2022). How do plastic surgeons use the metaverse: a systematic review. *Journal of Craniofacial Surgery*, 34(2), 548-550. <https://doi.org/10.1097/scs.00000000000009100>.
- [18] Lee, J. and Kwon, K. H. (2022). Future value and direction of cosmetics in the era of metaverse. *Journal of Cosmetic Dermatology*, 21(10), 4176-4183. <https://doi.org/10.1111/jocd.14794>.
- [19] Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., ve Chadha, A. (2024). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. *arXiv preprint arXiv:2402.07927*.
- [20] Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. Le, ... Hajishirzi, H. (2021). Generated Knowledge Prompting for Commonsense Reasoning. *arXiv*.
- [21] Yang, J., Guo, X., Li, Y., Marinello, F., Ercisli, S., ve Zhang, Z. (2022). A survey of few-shot learning in smart agriculture: developments, applications, and challenges. *Plant Methods*, 18(1), 28. <https://doi.org/10.1186/s13007-022-00866-2>
- [22] Ma, R., Zhou, X., Gui, T., Tan, Y., Li, L., Zhang, Q., ve Huang, X. (2021). Template-free Prompt Tuning for Few-shot NER. *arXiv preprint arXiv:2109.13532*.
- [23] Jiang, Z., Xu, F. F., Araki, J., & Neubig, G. (2020). How can we know what language models know?. *Transactions of the Association for Computational Linguistics*, 8, 423-438. https://doi.org/10.1162/tacl_a_00324.
- [24] Yong, G., Jeon, K., Gil, D., & Lee, G. (2022). Prompt engineering for zero - shot and few - shot defect detection and classification using a visual - language pretrained model. *Computer-Aided Civil and Infrastructure Engineering*, 38(11), 1536-1554. <https://doi.org/10.1111/mice.12954>.
- [25] Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. Le, ... Hajishirzi, H. (2021). Generated Knowledge Prompting for Commonsense Reasoning. *arXiv preprint arXiv:2110.08387*.
- [26] Ma, R., Zhou, X., Gui, T., Tan, Y., Li, L., Zhang, Q., ve Huang, X. (2021). Template-free Prompt Tuning for Few-shot NER. *arXiv preprint arXiv:2109.13532*.
- [27] Lester, B., Al - Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. <https://doi.org/10.18653/v1/2021.emnlp-main.243>