Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 16, No. 2, 2025

Research on Optimizing Virtual Reality User Experience Based on Large Language Models

Xingyu Liu

Social Science Research Institute, Duke University, Durham, NC 27708, United States

Abstract: With the rapid development of virtual reality (VR) technology, how to further improve the user's experience in this field has become a research hotspot. Based on Large Language Model (LLM), this paper discusses its application and optimization path in VR field. Firstly, the basic principle and core technology of LLM are expounded, and its working mechanism is analyzed emphatically. Then, the application of LLM in VR field is discussed, including virtual assistant, intelligent recommendation, natural language interaction and multi-modal collaboration. Finally, a path for optimizing virtual reality user experience based on LLM is proposed, aiming to improve the accuracy of voice interaction, realize personalized content recommendation, optimize the interaction quality of dialogue system and strengthen multi-modal data fusion, so as to enhance the immersion and interactivity of virtual reality.

Keywords: Large language model; Virtual reality; User experience optimization.

1. Introduction

In recent years, the rapid development of virtual reality (VR) technology has made immersive experience more extensive, in which the optimization of user experience has become the core factor to enhance its application value. With the rise of large language models (LLMs), artificial intelligence technology has introduced entirely new modes of interaction to VR environments. This model can understand and generate natural language, and then provide strong technical support for intelligent assistants, dialogue systems and content recommendation in VR space. Large language models show unique development potential in optimizing virtual reality user experience, and their applications in voice interaction, personalized content generation, and environment adaptation create new possibilities for the future progress of VR technology.

2. Fundamentals of large language models

2.1. The working mechanism of large language models

Through deep learning and massive data training, large language models (LLMs) demonstrate the ability to understand and generate natural language text. Its working mechanism relies on neural networks, particularly "Transformer" architecture, which enables efficient parallel processing and long-range dependency modeling (Figure 1). In the training stage of the model, it adopts the self-supervised learning method to learn large-scale text data, and then masters the language syntax, semantics and context association information. For example, the training process of GPT is a typical self-supervised learning process. It is trained on the "predict the next word" task, where given a piece of text, the model needs to predict the most likely words to follow. Each time the difference between the generated prediction words and the real words is adjusted through backpropagation, gradually improving the accuracy of the prediction until the model can better understand and generate syntactic and semantic text. Specifically, LLM represents the word embeddings of the input sequence and processes the input data through multiple layers of encoder and decoder modules. The model can determine the relationship between different words through a layer-by-layer attention mechanism, and continuously optimize the model parameters through backpropagation to improve the accuracy and fluency of the generated language text. Its core idea is that the model does not treat each input word with equal attention to all other words, but rather gives different weights to each word according to its correlation with each other. This approach allows the model to better understand long-distance dependencies and the importance of certain words in the current context.

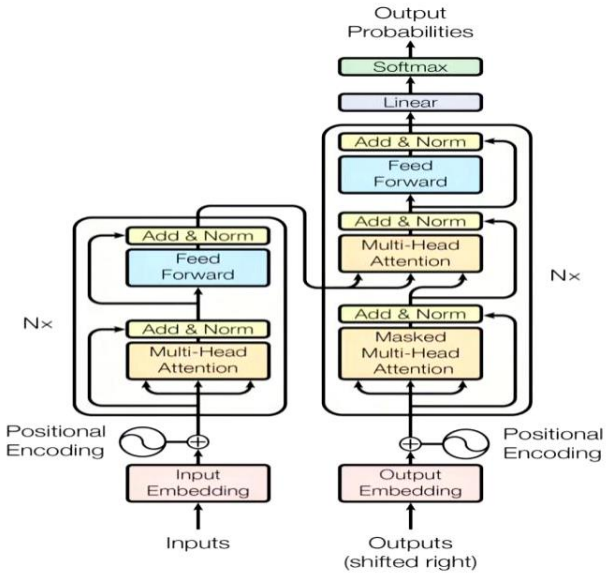


Figure 1. Technical principles of large language models

2.2. Core technologies of large-scale language models

Through multi-level neural network structure, deep learning can automatically extract features and patterns from large-scale data, so as to realize language understanding and generation. During the training process, the model not only

learns the surface syntax structure of the text, but also can grasp the complex semantic and contextual information. This enables large language models to show higher accuracy and flexibility than traditional methods in handling various language tasks. As one of the core technologies of large-scale language models, natural language processing (NLP) is responsible for processing and analyzing various complex structures in language. NLP covers many aspects such as text analysis, emotion recognition, semantic reasoning, etc. The key tasks include syntax analysis, named entity recognition, text generation, etc. The power of NLP enables machines to understand the polysemous and complex nature of human language, enabling natural interaction with users.

Among them, GPT is one of the most representative large-scale language model architectures. GPT is based on the Transformer architecture and uses a self-attention mechanism to capture long distance dependencies in the input language. This mechanism enables GPT to take contextual information into account when generating language, resulting in more coherent and context-appropriate text. Through pre-training and fine-tuning, GPT models can learn a wide range of language patterns from massive corpora and adjust them according to different application scenarios. This powerful language generation capability makes GPT one of the key technologies for optimizing user experience in virtual reality.

Combined with virtual reality technology, the benefits of deep learning and NLP can significantly enhance the interactive experience in VR environments. For example, GPT can analyze the user's voice input in real time through natural language processing technology, generate accurate system feedback, reduce the complexity of user operations, and improve the response speed of the system. At the same time, GPT and NLP can give virtual characters in virtual reality a more natural and personalized dialogue ability, making the interaction of virtual characters more real and increasing the user's immersion. Through the combination of these technologies, VR user experience has been significantly optimized, and users can interact with the virtual world more intuitively and flexibly, promoting the development and popularization of virtual reality technology.

3. Application of large-scale language model in virtual reality

3.1. Virtual Assistant and Dialog System

In the field of VR, virtual assistants and dialogue systems are increasingly becoming one of the key technologies to enhance user experience. As an advanced natural language processing tool, Large Language Models (LLMs) provide a more natural and intelligent experience in VR environments. In the VR environment, users usually have to interact with virtual characters, scenes or objects, and the traditional interaction methods mostly rely on fixed buttons or menus, making it difficult to meet the personalized needs of users. LLMs are able to parse the user's voice input, understand the semantics and intent behind it, and generate context-appropriate responses. This way of interaction greatly reduces the operational barriers between users and the system. Large language models also enable virtual assistants to provide more personalized and dynamic feedback in VR environments through their powerful language generation capabilities. For example, when a user asks a question in a virtual environment, the LLMs can not only understand the user's question, but also generate detailed and relevant answers based on contextual information, and even adjust the tone and intonation to match the user's mood or needs. In addition, LLMs enable the virtual assistant to remember the user's preferences and historical interaction records, thereby providing more accurate services and improving the coherence and immersion of the interaction. Thus, a more immersive and personalized virtual environment is created, which enhances the user's sense of participation and experience.

3.2. Intelligent recommendation and personalized content generation

Table 1. Intelligent recommendation application of LLMs in virtual reality

Application scenario	Function description	User experience enhancement
Education and training	Recommend personalized learning tasks and materials based on students' learning progress and interests	Improve learning efficiency and increase user immersion and participation
Entertainment and games	Recommend personalized game missions to interact with characters based on the player's historical behavior	Increase the fun and engagement of the game
Virtual shopping	Recommend customized product displays based on user preferences and purchase history	Improve the shopping experience, increase purchase intention and user satisfaction
Fitness and sports training	Recommend a personalized training plan based on the user's physical fitness and exercise records	Improve the training effect and enhance the user's sense of accomplishment and motivation to participate

In virtual reality (VR), the quality of the user experience is closely related to the degree of personalization of the content. Through intelligent recommendations and personalized content generation, LLMs can adjust and optimize content in virtual environments in real time based on user preferences, behaviors, and contexts to enhance immersion and interactivity. Based on the analysis of user data, the intelligent recommendation system provides users with personalized virtual content. For example, LLMs can infer the user's points of interest and needs by analyzing the user's historical behavior, language input, or interaction data in the virtual environment, so as to recommend relevant virtual scenes, tasks, or people. Such personalized recommendations not only increase user engagement, but also enable users to feel a more natural and personal experience in the virtual environment. In addition, another important application of LLMs in virtual reality is content generation. In virtual environments, users often expect to encounter virtual characters and plots with autonomy and personalization. LLMs can generate highly customized content in real time based on user input, such as generating dialogue, task clues, or plot developments. This content generation can not only

enrich the diversity of the virtual world, but also provide a unique experience path according to the needs of the user, thus avoiding the limitations of a single content.

Through deep learning of data in virtual reality, large language models not only optimize the user experience, but also make the virtual environment highly adaptable and interactive. This way of personalized content generation has gradually become an important way to improve the quality of virtual reality experience.

3.3. Adaptive adjustment of natural language interaction and virtual environment

Natural language interaction is an important function of large language models in virtual reality (VR) applications, which allows users to communicate seamlessly with the virtual environment through voice or text. This interactive approach not only enhances the immersion of the user experience, but also promotes the dynamic adaptation of the virtual environment. By understanding the user's natural language input, the VR system can respond in real time and adjust the elements of the virtual environment to meet the user's individual needs and behavior patterns. For example, in response to the user's specific instructions, the system can adjust the lighting, sound effects, object positions or character interactions of the scene to enhance personalization and interactivity. Table 2 below shows the main application areas of natural language interaction in virtual environment and the specific performance of adaptation.

Table 2. Application of natural language interaction and virtual environment adaptation

Application field	Adjustment mode	Actual effect
Scene setting	Adjust the lighting, temperature, and background sound of the virtual environment	Enhance the user's sense of immersion and increase the sense of reality of the virtual environment
Role interaction	Change role behavior based on user instructions	Make characters more intelligent and increase the variety and authenticity of interactions
Session management	Adjust the conversation based on voice input	Provide personalized conversations to increase user engagement
Task accomplished	Adjust task difficulty based on user feedback	Optimize the task flow, improve the sense of task completion and challenge

This adaptive adjustment based on natural language interaction can break the static setting in traditional VR and improve the intelligence level of the system. The user's interaction with the virtual world becomes more intuitive and flexible, greatly enhancing the immersive experience of virtual reality.

3.4. Multi-modal interaction and collaboration

Table 3. Multimodal interaction and collaboration applications and LLM role

Application scenario	Primary interaction mode	LLM role	User experience enhancement
Voice controlled virtual environment	Speech, gesture, vision	Provides voice recognition and command generation	Improve the accuracy and naturalness of speech interactions
Multi-user collaborative task	Speech, gesture, haptic feedback	Real-time voice communication, context understanding and conversation generation	Enhance teamwork and communication
Avatar dialogue	Voice, facial expression, movement	Generate the avatar's natural dialogue and behavioral responses	Enhance immersion and emotional connection

In the application scenario of virtual reality technology, multi-modal interaction and collaboration greatly enhance the immersion and interactivity of user experience. Multimodal interaction refers to the reception and feedback of information through the combination of multiple perceptual channels, such as speech, vision, touch, etc. This approach simulates complex interactions in the real world, allowing users to interact more naturally and intuitively in virtual environments. The application of LLMs in multimodal interaction, especially in speech recognition and generation, is of great significance. Multimodal interaction refers to the integration of a variety of input and output methods, such as voice, image, gesture, etc., to achieve a more natural and intuitive communication between people and the virtual environment. By integrating LLM, virtual reality systems can more accurately recognize voice commands and provide immediate feedback, streamline interactions and improve real-time responsiveness. In the multi-modal collaboration in the VR field, users can collaborate to complete various tasks through a variety of interactive ways such as voice, gestures, and eye movements. Table 3 below shows several key applications of multimodal interaction and collaboration in virtual reality.

By incorporating large language models into virtual reality, users can enjoy a more smooth and natural interactive experience, while the realization of multi-modal collaboration can also promote the development of virtual reality technology to a higher level of intelligence and humanization.

4. Virtual Reality user experience optimization path based on large-scale language model

4.1. Improve the accuracy and response speed of voice interaction

As an important way of human-computer interaction, the accuracy and response speed of voice interaction directly affect the user's immersion and experience. In order to

improve the accuracy of speech interaction, it is necessary to strengthen the ability of speech recognition system to adapt to accent, noise and context changes. With the deep learning ability of large language models, the speech input of different languages and dialects can be better understood. The neural network architecture adopted by LLMs, especially the model based on Transformer, has strong sequence processing ability, which can carry out deeper understanding and reasoning of the speech input. This allows the model to not only recognize the surface features of the language (such as phonemes and vocabulary), but also infer implicit semantic information from the context, thus improving the accuracy of speech recognition. In addition, responsiveness is also a core element of optimizing the user experience. In virtual reality, users have high requirements for real-time feedback, and any form of delay can destroy immersion. The use of distributed computing and edge computing technology can effectively reduce the response delay, but also to ensure the rapid response after voice input. A voice interaction system is a technology that allows users to interact with a computer or virtual environment through voice input. In traditional speech recognition systems, the user issues voice commands through a microphone, and the system converts those commands into text and responds according to preset rules or algorithms. Table 4 below shows the performance improvement after the introduction of a large language model to optimize the speech interaction system.

Table 4. Performance comparison before and after speech recognition optimization

Before optimization	post-optimization	Lifting range
Speech recognition accuracy: 80%	Speech recognition accuracy: 96%	+ 16%
Voice response time: 300ms	Voice response time: 210ms	- 30%

This application scenario refers to the user's interaction with objects, roles, or systems in a virtual environment through voice. This scenario can include many aspects, such as character dialogue in the game, task execution by virtual assistants, real-time question-and-answer in VR teaching, and real-time voice control in complex situations. These results show that using the advantages of large-scale language models can not only significantly improve the accuracy of speech recognition, but also effectively shorten the interaction response time, and greatly improve the user experience in virtual reality.

4.2. Implement personalized content recommendation and context adaptation

As one of the core ways for users to interact with the system, the accuracy and response speed of voice communication directly affect the quality of user experience. In order to improve the efficiency of speech interaction, LLMs can be used for technical optimization. Through large-scale corpus training, the model can obtain stronger natural language understanding (NLU) ability. Unlike traditional speech recognition technology, LLM can effectively reduce the error rate through context analysis and deep semantic understanding. For example, using a pre-trained model such as GPT or BERT can improve the accuracy of the model's recognition of common commands in VR by adding domain-specific data training. Its optimization formula can be expressed as.

$$\hat{y} = \arg\max\left(P\left(y|x;\theta\right)\right) \quad (1)$$

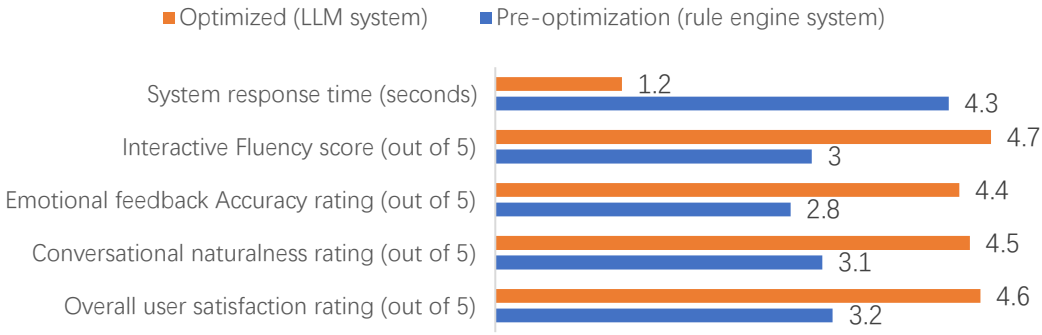
Among them,x Represents the incoming speech signal,θ Is the model parameter,$P(y|x;\theta)$ Is a probability distribution based on context, $\hat{y}$ For predictive voice commands. By introducing context information, the recognition accuracy of voice commands can be improved under noise interference. In addition, the increase in response speed depends on efficient inference engine optimization for LLM. The traditional inference process often takes a long time to calculate due to the large number of model parameters, which affects the fluency of user interaction. By means of simplified model, quantitative technique and model pruning, the inference speed can be greatly improved to meet the requirement of real-time response. The simplified model mainly reduces the complexity of the model to improve the operational efficiency and real-time response ability. In the application of LLMs, simplification is usually expressed by reducing the number of layers, the number of neurons, or the size of the parameters of the model. Quantization reduces memory usage and computing resource consumption by representing the model's parameters as low-precision numbers, such as converting floating points to integers. Model pruning is a technique that reduces the size of a model by removing unimportant neurons or connections. The model is constructed mainly by weight pruning, structured pruning and dynamic pruning techniques. Optimized response time$T_{response}$ It can be expressed by the following formula.

$$T_{response} = \frac{T_{processing}}{Acceleration\ Factor} \quad (2)$$

Among them, $T_{processing}$ For processing time, $Acceleration\ Factor$ It is an improved acceleration ratio after optimization.$Acceleration\ Factor$In the process of model optimization, the acceleration factor can improve the overall efficiency by reducing computational complexity and memory usage. Pruning is the reduction of computational effort and memory requirements by removing unimportant neurons or connections in the model. After the pruning operation is performed, the size and computational complexity of the model are reduced, so the amount of data and computation steps to be processed are also reduced, thus increasing the processing speed. Because pruning removes redundancy and reduces computational complexity, it often results in systems being able to complete tasks with fewer computational steps, resulting in faster response times. Which means, Processing time $T_{processing}$ will shorten significantly, this in turn raises $Acceleration\ Factor$ in the formula. The result is a faster, smoother response to voice interactions. The optimization of speech interaction in large language models can not only improve the accuracy of speech recognition, but also improve the response speed through inference optimization, which brings significant improvement to the user experience in virtual reality.

4.3. Optimize the quality of interaction between virtual assistants and intelligent dialogue systems

The performance of virtual assistant and intelligent dialogue system is the core factor that determines the quality of user experience. Optimizing the interaction quality of these systems can greatly enhance the user's immersion and operation efficiency. Take, for example, the virtual assistant launched by VRWorld, a virtual reality gaming platform that originally used a rules-based dialogue system to guide players through in-game tasks. However, users frequently report that the virtual assistant performs poorly in complex situations, especially when dealing with unstructured conversations. In order to improve the interactive quality of virtual assistants, VRWorld decided to upgrade its virtual assistant system to a large-scale language model based on GPT-4. The upgraded system is able to better understand the user's intentions, analyze the context of the conversation, and adjust the tone based on the user's emotional state. To measure the effectiveness of the optimization, VRWorld conducted a before-and-after survey of user satisfaction and collected relevant user feedback data (see Figure 2).



	Overall user satisfaction rating (out of 5)	Conversational naturalness rating (out of 5)	Emotional feedback Accuracy rating (out of 5)	Interactive Fluency score (out of 5)	System response time (seconds)
Optimized (LLM system)	4.6	4.5	4.4	4.7	1.2
Pre-optimization (rule engine system)	3.2	3.1	2.8	3	4.3

Figure 2. Comparison before and after virtual assistant optimization

In order to further improve the quality of dialogue, the key is to enhance the context awareness of the system, so that the virtual assistant can accurately grasp the actual needs of users. For example, using deep learning-based natural language processing (NLP) models that use contextual information to make accurate guesses about user intent can significantly improve interaction efficiency. Suppose the input to the dialog system is I_t (current user input), the historical dialog is H_{t-1}, Output as O_t ,then the optimization objective can be expressed as.

$$O_t = f(I_t, H_{t-1}; \theta) \tag{3}$$

Among them, f Represents the language model's handling of user input and historical conversations, θ Is the model parameter. Through the support of these real case data and optimization formulas, the user's immersion and experience quality are greatly enhanced. With the continuous progress of technology, large-scale language models will play an increasingly important role in the field of virtual reality in the future, promoting the further improvement of VR user experience.

4.4. Strengthen multi-modal data fusion and immersion

In order to achieve deeper immersion, multi-modal data needs to be efficiently fused, which includes multiple perceptual data including sight, sound and touch. Large-scale language model (LLM) is used to process and analyze massive cross-modal data, which can respond to the changes of user behavior and emotion in real time, and then enhance the immersive experience of virtual environment. The key of multimodal data fusion lies in the collaborative processing of multiple dimensions of data such as language, image, sound and tactile information, so that the virtual world can interact with users in a more natural and coherent way. Language models can not only handle natural language interactions, but also generate context-related visual and auditory feedback when users input information, thereby enhancing the realism and interactivity of virtual environments. Through intelligent recognition and understanding of user voice, movement and emotional changes, combined with the feedback mechanism of VR technology, the intensity of immersion can be more accurately adjusted to achieve a more personalized user experience.

5. Conclusion

In summary, LLMs is able to better understand the effects of various accents, dialects, and environmental noise by training on large amounts of diverse speech data. Through deep learning, LLMs can recognize nuances in speech, such as syllable pronunciation, intonation changes in speech, and context between words, resulting in a 16% increase in speech recognition accuracy from 80% to 96%. The experiment shows that in terms of speech response speed, the introduction of LLMs in the experiment significantly improves the system response time, reducing the original 300 ms to 210 ms, with an increase of 30%. This is because the LLMs architecture (such as Transformer) itself has strong parallel computing

capabilities, which means that it is able to process multiple data levels and dimensions simultaneously when processing voice input, rather than relying on serial processing as traditional models do. Traditional speech recognition and response systems typically need to analyze input data layer by layer, and the processing process is slow. The parallelization capability of LLMs greatly accelerates this process and reduces the latency of system response. Secondly, the application of LLMs in speech recognition is not simply the conversion of audio signals into text, but also involves the context understanding and semantic reasoning of speech input. In traditional speech recognition systems, input is usually processed through preset rules or fixed dictionaries, while LLMs uses deep learning and NLP technology to parse the user's speech input and quickly understand and respond. This context-understanding approach allows LLMs to complete more computing tasks in a shorter time, which improves response speed.

References

[1] Douglas M R .Large Language Models [J].Communications of the ACM, 2023, 66:7 - 7.

[2] Giachos I , Batzaki E , Papakitsos E C ,et al.A Natural Language Generation Algorithm for Greek by Using Hole Semantics and a Systemic Grammatical Formalism[J].Journal of Computer Science Research, 2023, 5(4):27-37.

[3] Feng B .Research on the Application Effects of Artificial Intelligence in Personalized Marketing[J].Journal of Computer and Communications, 2024, 12(11):10.

[4] Rong Z .Application of Natural Language Processing in Virtual Experience AI Interaction Design[J].Journal of Intelligent Learning Systems and Applications, 2024, 16(4):15.

[5] Meena Y K , Arya K V .Multimodal interaction and IoT applications[J].Multimedia Tools and Applications, 2023, 82(4):4781-4785.