ISSN 2959-6157

The foundation, current situation and future prospects of pre-training large language models

Haoran Han[1], Siyao Wu[2,*], Jinyao Yang[3] and Yizhuo Zhao[4]

[1]University of Southampton , Southampton, England
[2]School of Environmental Science and Engineering, Tongji University,Shanghai, China
*Corresponding author: 2153661@tongji.edu.cn
[3]School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang, China
[4]Beijing-Dublin international college, Beijing University of Technology, Beijing, China

Abstract:

The field of artificial intelligence has developed rapidly recently, and large language model technology, as a representative technology of it, can provide general knowledge and make many downstream tasks easier and more convenient. However, although many people use large language models to do some work, they still lack a systematically summarized literature. Therefore, in this article, we made a systematic summary. We first wrote about the early large language models, then we presented the development of GPT and how to use the GPT model, then we introduced the advanced GPT models, and finally we mentioned the risks and challenges faced by the GPT model. Our work can help users better use large language models.

Keywords: Large language models; GPT model; advanced GPT models.

1. Introduction

In recent years, the domain of Artificial Intelligence (AI) has experienced a paradigm shift, propelled by significant enhancements in computational power and algorithmic sophistication. Large-scale language models (LLMs), epitomized by the Generative Pretrained Transformer (GPT) series, have been pivotal in this transformation, advancing machine comprehension of human language. These models not only bolster general AI capabilities but also streamline complex tasks across varied sectors, from automated customer support to intricate content generation. This paper delineates the early development of LLMs, tracing their roots from foundational theories and models in natural language processing to their evolution as forerunners of more sophisticated systems. This historical context is crucial for understanding the incremental advancements that have sculpted the present functionalities of LLMs. The paper delves into the genesis and evolution of the GPT series, examining the intricate progressions from GPT-1 through GPT-4, and how each iteration builds upon its predecessor in terms of technological innovation and application diversity. Furthermore, this study offers insights into optimal utilization strategies for these models, emphasizing just-in-time engineering and customization for specific applications. Additionally, this paper addresses the emerging challenges and risks linked with the deployment of advanced GPT models, such as model bias, ethical concerns, and potential misuse, providing a balanced perspective on the advantages and drawbacks of employing this potent AI technology. Ultimately, this paper not only synthesizes existing knowledge concerning LLMs but also proposes directions for future research and development to aid users in navigating the complexities of these models while maximizing their potential benefits.

2. Early large language models

2.1 Large language model and natural language processing

Natural language processing is a branch of artificial intelligence and linguistics. What it studies is how to use artificial intelligence to process human language, so that machines can know what people are saying and discussing. Natural language processing includes many aspects and steps, basically including cognition, understanding, generation, and other parts.

The basic idea of natural language processing is, on the one hand, to let the computer turn the input language into a series of symbols and relationships, and then process it according to the purpose; on the other hand, it is to convert the data output by the computer into natural language that humans can understand and understand.

There are many ways to implement natural language processing, and large language models are one of them. From an artificial intelligence perspective, a large language model is not so much a model as it is a prediction task used to train the model. Simply put, it requires being able to predict what the next word will be based on a given string of text, or completing such a string of text through trained intelligence in a string of missing text. The model itself improves the accuracy of predictions through continuous iteration.

The task of large language models is to develop models that represent language capabilities and language applications, establish a computing framework to implement and improve language models, design various practical systems based on language models, and explore evaluation technologies for these systems [1].

2.2 The development history of early large language models

The development of early large language models has gone through roughly three stages, from artificially defined rules, to relying on mathematical statistical probability, to based on neural networks.

A Recurrent Neural Network (RNNQ for short) is a neural network structure that processes sequence data. It has memory capabilities and can capture temporal information in the sequence. RNN is widely used in natural language processing, time series prediction, speech recognition, and other fields.

The purpose of RNN is to process sequence data. In the traditional neural network model, from the input layer to the hidden layer and then to the output layer, the layers are fully connected, and the nodes between each layer are unconnected.

The reason why RNN is called a recurrent neural network is that the current output of a sequence is also related to the previous output. The specific manifestation is that the network will remember the previous information and apply it to the calculation of the current output, that is, the nodes between the hidden layers are no longer unconnected but connected, and the input of the hidden layer not only includes the output of the input layer Also includes the output of the hidden layer at the previous moment. In theory, RNN can process sequence data of any length. However, in practice, in order to reduce complexity, it is often assumed that the current state is only related to several previous states. The recurrent neural network is expanded in natural language processing like this: the size of the entire vocabulary is V, and for the convenience of unification, three words „BOF“, „EOF“ and „UNK“ are added, which represent the starting word of the sentence and the sentence respectively. Ending word, an unknown word.

The Seq2seq (Sequence to sequence) model is a neural network machine learning model that maps sequence to sequence. This model was originally designed to improve machine translation technology, allowing machines to discover and learn to map sentences (sequences of words) in one language to corresponding sentences in another language [2]. In addition, Seq2Seq can also be widely used in various technologies, such as chat robots, Inbox by Gmail, etc., but it requires a paired text set to train the corresponding model [3]. Application areas include machine translation, image description, conversation models, and text summarization.

The model generally consists of two parts: the first part is the Encoder part, which is used to characterize the input N-length sequence; the second part is the Decoder part, which is used to map the representation extracted by the Encoder to the output M-length sequence.

The decoder has three structures. The first structure directly inputs the context vector obtained by the Encoder as the initial hidden state of the RNN into the RNN structure. The second method uses the context vector obtained by the Encoder as the input of each neural unit, no longer just as the initial hidden state of the first unit. The third type is similar to the second type in the processing of c, but the difference is that the output of the previous neural unit is used as the output of the current neural unit.

3. The Development History of the GPT Series Models

3.1 Generative Pre-trained Transformer 1

GPT-1, is the first iteration of the GPT series developed by OpenAI, introduced in a paper titled „Improving Language Understanding by Generative Pre-Training“ [4]. GPT-1 is a Transformer-based model pre-trained on text data to predict the next word, learning language patterns and structures. It utilizes self-attention for efficiency, contains 12 Transformer layers, and undergoes fine-tuning for NLP tasks to optimize performance by adjusting parameters for specific tasks. GPT-1 shows diverse applications in NLP tasks like text generation, translation, question-answering, summarization, and sentiment analysis. It can generate coherent content, translate languages, provide answers, summarize text, and analyze sentiment, showcasing its broad utility. Despite its initial design limitations, subsequent GPT iterations have significantly advanced in performance and applications. GPT-1 excels in NLP tasks due to its transformer architecture capturing context effectively and unsupervised pre-training on extensive text data, enabling generalization. Despite its strengths in coherent text generation and versatility, limitations like contextual understanding, model size, fine-grained control, factual accuracy, and biases in responses are present.

3.2 Generative Pre-trained Transformer 2

GPT-2 is an enhanced version of the GPT model by OpenAI, with larger size and improved capabilities, targeting enhanced natural language understanding and generation by capturing human language intricacies for contextually relevant text production [5]. GPT-2, following the Transformer architecture, significantly scales up in size and dataset to enhance language understanding and generation. It undergoes fine-tuning on diverse internet text data and utilizes multi-scale training to improve coherence and relevance in text generation. Through masked language modeling and fine-tuning stages, GPT-2 adapts to specific tasks efficiently, achieving better performance across various applications. GPT-2, renowned for its language generation prowess, is applied in diverse fields like text completion, summarization, translation, chatbot development, creative writing, and data analysis. Its versatility extends to social media analysis, sentiment assessment, and code snippet generation, making it a valuable tool across various industries. GPT-2 surpassed GPT-1 by utilizing the Transformer architecture to enhance contextual understanding, increasing model size for better language pattern recognition, introducing prompts for control, addressing biases, and improving overall performance in natural language tasks. GPT-2 marked a substantial improvement over GPT-1, yet faced limitations such as model size, control precision, bias handling, source tracing, and restricted user control in specific contexts, indicating the need for further enhancements and refinements.

3.3 Generative Pre-trained Transformer 3

OpenAI's GPT-3 revolutionizes AI and NLP with its advanced text understanding and generation abilities, including innovative features like zero-shot learning and multimodal capabilities, attracting widespread attention and adoption in academia and industry. GPT-3, with its immense Transformer-based architecture of 175 billion parameters, stands out for its exceptional language modeling and learning capabilities [6]. Its extensive pre-training, zero-shot learning capability, and multimodal features enable it to excel in various tasks, making it a highly versatile and powerful language model. GPT-3 showcases its adaptability and efficacy in diverse domains, excelling in language translation, chatbot development, and creative writing tasks. Its applications span from facilitating cross-cultural communication to enhancing virtual assistant interactions and content creation, signaling its potential to transform industries and elevate user experiences. GPT-3 significantly increasing its model size and capacity, introducing few-shot and zero-shot learning, enhancing contextual understanding, refining fine-grained control through prompt engineering, adding multimodal capabilities, and expanding access and deployment options via an API. While GPT-3 overcame many limitations of its predecessors, by offering enhanced language understanding and generation capabilities, it nonetheless exhibited specific limitations, including: context limitation, cost and accessibility, interpretability and explainability.

3.4 Generative Pre-trained Transformer 3.5

GPT-3.5 is a transitional NLP model by OpenAI, bridging GPT-3 and GPT-4 with optimizations and enhancements to design and performance. It includes subtle adjustments in architecture, training, and adaptability to boost performance and generalization compared to GPT-3. GPT-3.5 advances the capabilities of the Generative Pre-trained Transformers, focusing on refining GPT-3 while setting the stage for enhancements in GPT-4. By optimizing the Transformer architecture and self-attention mechanism, GPT-3.5 boosts model accuracy, speeds up inference, and improves resource efficiency. Through dual-phase training and a larger pre-training dataset, GPT-3.5 enhances language comprehension, generalization, and efficiency, demonstrating a commitment to technological innovation for superior model performance [7]. GPT-3.5's deployment in various sectors, including customer service, creative content generation, education, healthcare, legal, finance, and software development, has revolutionized artificial intelligence applications. The model's capabilities in natural language processing have enabled advancements in chatbots, adaptive learning tools, clinical decision-making, text automation, and software development, showcasing its transformative impact on efficiency and productivity across diverse industries. GPT-3.5 improves upon GPT-3 by enhancing context handling, mitigating biases, increasing computational efficiency, and offering refined fine-tuning capabilities for developers to customize applications more precisely. GPT-3.5, despite its improvements over GPT-3, particularly in handling context and biases, retained limitations that restricted its application and effectiveness, including: context length, knowledge updates, processing efficiency and energy consumption.

3.5 Generative Pre-trained Transformer 4

GPT-4, the latest in the series of generative language models from OpenAI, marks a significant advancement in natural language processing (NLP) by incorporating multimodal capabilities to interpret and generate both text and image-based content. With a robust Transformer-based architecture and billions of parameters, GPT-4 demonstrates improved performance and efficiency, making it a versatile tool for applications in various fields like healthcare and content creation [8]. GPT-4 is a groundbreaking AI language model that excels in processing text and visual content with a focus on scalability and versatility. Its architecture combines advanced technologies and training methods to achieve exceptional contextual understanding and content generation capabilities, making it a benchmark in AI innovation across industries. GPT-4's versatility is showcased across sectors, from healthcare to education and customer service, revolutionizing industries with its advanced natural language processing capabilities [9]. The model supports medical research, personalized education, enhanced customer service through chatbots, creative content creation, and aids developers in coding tasks, demonstrating its profound impact on multiple fields.

GPT-4's enhancements over GPT-3.5 address limitations, making it a more potent model. These include a wider context window for coherence, dynamic knowledge updates, enhanced computational efficiency, advanced reasoning abilities, and refined bias mitigation, ensuring ethical and innovative content generation. GPT-4 surpasses its predecessors in capability and performance metrics but still faces challenges typical of generative pre-trained transformers, including context limitations, computational demands, ethical considerations, and emotional intelligence nuances. While it shows improved content understanding and generation, it struggles with emotional subtleties like sarcasm and empathy, illustrating the gap between AI and human communication complexities.

4. Use GPT models

Although the GPT model has achieved significant results on a wide range of natural language processing tasks, it still has some shortcomings when dealing with specific problems. These problems are mainly in the lack of understanding of complex contexts, the problem of logical coherence of generated content, and the ability to capture long-term dependencies. First, the model often performs poorly in understanding texts containing complex contexts or implicit meanings, partly due to the fact that its training data lacks sufficient contextual depth. Second, although GPT is able to generate grammatically structurally correct text, sometimes the logic and coherence of the content could be improved, especially in long text generation. In addition, despite the adoption of the Transformer architecture, the model‘s capture of long-distance dependencies in long sequences is still unsatisfactory. To address these issues, researchers can enhance the model‘s ability to understand context by introducing more sophisticated attention mechanisms, improve the coherence of content generation by utilizing advanced text summarization techniques, and better handle long-term dependency information by improving the model‘s memory mechanisms. These improvements will help drive the performance of GPT models in a wider range of application domains to better meet the needs of real-world applications.

4.1 Prompt engineering

The concept of prompt engineering emerges as a pivotal technique for optimizing the interaction with and output of generative language models like ChatGPT. This methodology underscores the importance of meticulously constructing queries that guide these AI models toward generating outputs that are not only relevant but are also intricately aligned with specific user-defined objectives. Through the strategic refinement of prompts, practitioners can significantly enhance the precision, relevance, and utility of the outputs generated by models such as ChatGPT, thereby elevating these tools from mere conversational agents to sophisticated instruments capable of performing complex tasks across diverse domains. Louie Giray [10], Prompt Engineering with ChatGPT: A Guide for Academic Writers, points out the art of prompt engineering is grounded in the manipulation of four cardinal components: Instruction, Context, Input Data, and Output Indicator, each serving a distinct purpose in the crafting of an effective prompt.

Instruction: Delineates the explicit task or action desired from the AI model, serving as a direct command that shapes the model‘s focus and output. In the lexicon of computer science, this can be likened to the function call in programming that specifies what operation is to be performed, thereby providing a clear directive for the AI‘s processing.

Context: supplies the AI model with a framework of additional information or background knowledge, akin to the parameters or environment settings in a software program, which enhances the model‘s understanding and the relevance of its response. This component is crucial for embedding the prompt within a specific domain or scenario, ensuring that the AI‘s output is not only accurate but also contextually appropriate.

Input Data: represents the core information or question that the model is to address, acting as the input variable in a function. This is the essence of the prompt, upon which the model bases its entire generative process, analogous to the data fed into an algorithm for processing. The clarity and specificity of this data are paramount for achieving a precise and meaningful output.

Output Indicator: specifies the desired format, structure, or style of the AI‘s response, effectively setting the output constraints of the model‘s operation. This can be compared to defining the return type in a function, guiding the model in structuring its response according to the user's specifications, whether that be in the form of a concise summary, a detailed exposition, or a structured document.

The synthesis of these components into a coherent and strategic prompt is critical for leveraging AI models like ChatGPT effectively, especially within academic and research contexts. An illustrative example can be found in the academic discourse on the utilization of nanotechnology in targeted drug-delivery systems for cancer therapy. Here, the construction of a prompt incorporating clear instructions, relevant context, specific input data, and defined output indicators serves as a paradigmatic example of how prompt engineering can facilitate the generation of sophisticated academic content, tailored to the stringent requirements of scholarly research. The application of prompt engineering transcends academic writing, extending into practical computer science applications such as software documentation, code generation, and algorithm explanation. The detailed case studies presented in Dr. Sabit Ekin's work [11], „ Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices," illustrate the breadth of prompt engineering's applicability, showcasing its potential to enhance customer support, content creation for digital platforms, and the development of interactive narratives in gaming.

In conclusion, prompt engineering represents a crucial intersection between human cognitive skills and AI capabilities, facilitating a more nuanced and effective utilization of generative language models. For students and practitioners within the computer science discipline, mastering the art of prompt engineering not only augments the functionality and applicability of AI tools but also contributes to the advancement of AI research and its applications. This discussion aims to provide a more academically rigorous and professionally relevant perspective on the subject, reflecting the intricate complexities and substantial potential of prompt engineering within the computer science domain.

4.2 Memory

In the design and application of GPT models, „Memory" plays a crucial role. Unlike simple text memory functions, the memory mechanism in GPT models is more complex. It not only saves information from previous inputs but also helps the model abstract and understand long-term dependencies, i.e., textual information that spans a large number of words or sentences, as well as complex relationships between contexts. This capability significantly enhances the model‘s efficiency in handling long text sequences, enabling it to generate text that is more coherent and relevant. Dai et al [12], in their 2019 paper „Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context," demonstrated how to effectively handle longer text sequences through an extended attention mechanism. This study emphasized the role of memory in managing long-term contextual information, proving that the memory mechanism can significantly improve the coherence and relevance of generated text. This finding is particularly important for applications like ChatGPT because it allows the model to maintain topic continuity and context consistency across multiple rounds of dialogue. In practical applications, the memory mechanism enables ChatGPT to review and refer to previous dialogue content, thereby more accurately understanding user needs and providing more relevant responses. Moreover, ChatGPT can integrate knowledge from different sources through its memory mechanism, whether it is information learned from training data or provided by users during dialogue, to produce more accurate and information-rich responses. For example, in extended dialogues, ChatGPT can recall parts of earlier conversations, even after several rounds of dialogue, to make related responses based on previous information. This feature is particularly useful for providing customized suggestions or information, especially when users mention specific interests or needs during a dialogue. Through the memory mechanism, ChatGPT can also gradually learn and adapt to users‘ preferences and styles, offering a more personalized experience.

4.3 Planning

In advanced research and application of GPT models, the introduction of a planning mechanism has become a new trend to enhance model performance. Unlike the memory mechanism that mainly focuses on saving and referencing past input information, the planning mechanism enables GPT models not only to look back at the past but also to predict the future, achieving proactivity and goal orientation in text generation. This capability is especially useful for generating coherent long texts, solving complex problems, and conducting effective dialogues. The planning mechanism allows GPT models to construct a series of actions or paths before generating text or making decisions, with the expectation that these actions or paths will achieve a certain goal or solve a problem. This indicates that GPT models, when processing input information, rely not just on learning from historical data but also on generating the best output strategies based on the current context and set goals. Vaswani et al.‘s „Attention is All You Need" (2017) [13] provided the foundation for understanding how GPT models process and generate text through the self-attention mechanism, which is crucial for integrating complex planning strategies.

By pre-setting the structure and content of the text, the planning mechanism makes the text generated by GPT models more coherent and logically consistent, especially in the creation of long articles or stories. Moreover, for specific tasks such as generating answers and summarizing articles, the planning mechanism enables the model to more precisely identify and achieve task goals. In solving complex problems requiring multi-step reasoning, the planning mechanism helps the model predict and choose the most effective resolution path, enhancing the efficiency and accuracy of problem-solving. Brown et al. [7] demonstrated how the GPT-3 model can effectively learn with very few examples, proving the model‘s capability to formulate plans based on minimal context information. Key steps in implementing the planning mechanism include adding a specialized prediction layer to the GPT model for generating preliminary planning for tasks or text generation, and then the model generates the final output based on this plan. Additionally, adjusting the training objectives of the model to not only learn language patterns but also how to formulate plans based on given goals and context information is crucial. Integrating external knowledge bases into GPT models is also an important direction for implementing the planning mechanism, using external information to assist the planning process, especially in tasks requiring extensive background knowledge.

In summary, introducing a planning mechanism has brought unprecedented capabilities to GPT models, making them more efficient and goal-oriented in generating text, executing specific tasks, and solving complex problems. By comprehensively utilizing the techniques and concepts from the aforementioned literature, the planning mechanism of GPT models can be further perfected, opening new possibilities for future research in natural language processing and artificial intelligence.

4.4 Action

The introduction of an action mechanism marks a significant leap in the capabilities of models, extending their range of applications from purely understanding and generating text to executing specific actions within given contexts. This advancement is particularly important for applications in interactive cloud-based dialogue systems, automated task processing, and interactive gaming environments. It enables GPT models not only to generate text responses to user needs but also to take actions in appropriate situations, such as querying databases, calling APIs, or executing other specific commands. The introduction of this capability not only significantly enhances the interactivity and proactivity of the models but also opens new possibilities for their practical application. At the core of the action mechanism is its ability to allow GPT models to do more than just generate response text when analyzing input information; they can also plan and execute a series of actions based on the current context and predefined goals to achieve specific objectives or complete certain tasks. Implementing this mechanism involves defining the action space available to the model, training an action prediction model, and continuously optimizing the model‘s strategy through feedback from interactions with the environment. Vaswani et al.‘s „Attention is All You Need“ (2017) [13], with the introduction of the Transformer architecture, provides a strong foundation for processing and generating text in GPT models and lays the theoretical groundwork for incorporating more complex action mechanisms. Furthermore, Sutton and Barto‘s work [14] offers a detailed introduction to reinforcement learning, providing key techniques for training GPT models to select and execute optimal actions. Through such technological advancements, the application scope of GPT models has been significantly expanded. They can now take appropriate actions in dialogue systems based on dialogue history and current context to meet user needs, understand task requirements and plan action sequences to achieve goals in automated task processing, and make strategic decisions in games or simulation environments based on game state and rules. The introduction of the action mechanism, therefore, moves GPT models beyond mere text generation to the execution of actual tasks, paving a new path for future research and applications in artificial intelligence.

5. Advanced GPT model

5.1 GPT Medical Big Model

Initially, the GPT model was primarily deployed in natural language processing endeavors such as text generation and comprehension. The medical domain has progressively directed attention towards this potent natural language processing technology, recognizing its potential for integration into healthcare practices. With the established efficacy of GPT, scholarly discourse has commenced to probe the utilization of GPT models within the medical realm. Scholars have embarked on the training of GPT models utilizing medical text datasets for tasks encompassing medical diagnosis, processing of medical records, and dispensation of medical advice.

Nori et al. (2023) deliberated on the proficiency of GPT-4 in tackling medical challenges [15]. The researchers Nori et al. presented the utilization and performance appraisal of the GPT-4 model in the medical sector. They may have scrutinized the efficacy of GPT-4 in addressing medical hurdles, encompassing medical diagnosis, case studies, and recommended treatment strategies. The article possibly centered on the capacity of GPT-4 in resolving medical issues and offering clinical assistance, as well as exploring the prospective applications and obstacles of this model in healthcare. This study might offer crucial insights into evaluating the usefulness and potential of GPT-4 in medicine and provide valuable references for future research and implementation of medical artificial intelligence. The study revealed that GPT-4, even without tailored prompts, surpassed previous models and exceeded the passing threshold on the USMLE by over 20 points. Additionally, GPT-4 exhibited enhanced probability calibration and the capability to elucidate medical reasoning and personalize explanations for learners. Nevertheless, the large medical model also encounters several challenges. Chen et al. (2023) conducted a thorough examination of the evolutionary trajectory of the ChatGPT model [16]. The researchers Chen et al. may have investigated the dialog generation and comprehension capacities of ChatGPT at different stages, as well as its varied behavioral manifestations when confronted with diverse contexts and subjects. They may have utilized extensive conversation datasets for analysis, evaluating the performance fluctuations and behavioral patterns of the ChatGPT model across distinct time frames. This study could report on the adaptability of the ChatGPT model to linguistic usage and dialog context over time, in addition to its performance adjustments in handling novel problems and topics. By delving into the behavioral changes throughout the evolution of the ChatGPT model, the article could furnish critical insights into the developmental trends and avenues for enhancing dialogue systems based on large-scale language models. This study might provide valuable guidance and inspiration for further refining the performance of the ChatGPT model and enhancing the quality of dialog generation and comprehension. Tang et al. (2023) evaluated Large Language Models (LLMs), including GPT-3.5, in medical evidence summarization [17]. The study unveiled that LLMs could produce factually inconsistent summaries and encounter difficulties in identifying salient information, thereby posing potential harm through misinformation. Human evaluations delineated a spectrum of error typologies for medical evidence summarization, underscoring the limitations of LLMs in this sector. In conclusion, the merits of the GPT medical large model, such as GPT-4, are underscored by its performance in medical competency exams, enhanced probability calibration, capability to explicate medical reasoning, and potential utilities in medical education, assessment, and clinical practice.

5.2 Multimodal large model

Multi-modal Large Models (LMM) have garnered significant attention in the field of artificial intelligence. These models are designed to amalgamate diverse modes such as text, images, and speech to fabricate artificial intelligence systems that are more user-friendly and efficient. The advancement of LMM is perceived as a strategic step towards the creation of artificial intelligence systems capable of perceiving and responding to the world akin to humans. The use of multimodal technology has seen significant growth in various applications in recent years. Zhang et al. predominantly deliberated on a pioneering framework named CompositeMap tailored for gauging music similarity [18]. The inception of this framework epitomizes an inventive stratagem for gauging the resemblance amidst musical compositions. Harnessing the CompositeMap framework, the study endeavors to address the quandary of similarity assessment within the realm of music information retrieval, offering an avant-garde approach to quantifying music similitude. Zhang et al. delineated the foundational design principles, operational workflow, and illustrative applications of the CompositeMap framework in the domain of music similarity evaluation. They conceivably conducted meticulous comparative scrutiny of prevailing methodologies for music similarity assessment, showcasing the merits and trailblazing attributes of the CompositeMap framework vis-à-vis conventional modalities. Through this scholarly inquiry, readers are invited to delve deeper into the realm of avant-garde music similarity evaluation methodologies and acquaint themselves with the deployment of the CompositeMap framework to refine the efficacy and precision of music retrieval systems. This technological innovation boasts substantive import within the precincts of music information retrieval, furnishing a cornucopia of insightful notions and methodologies for scholarly investigation and practical implementation within this specialized domain.

6. Conclusion

6.1 Overview

GPT models have garnered significant attention due to their proficiency in generating text resembling human language; nonetheless, they encounter various challenges. Trained on extensive amounts of internet text data, GPT models are susceptible to biases and prejudices, potentially resulting in skewed outputs, encompassing racial, gender, and geographical biases. Despite their adeptness in text generation, GPT models encounter difficulties in navigating complex and abstract subjects and drawing logical inferences, leading to erroneous or nonsensical outcomes. The escalating apprehension revolves around the misuse of GPT models, capable of disseminating misinformation, perpetrating harmful actions, or jeopardizing individuals. The absence of transparency and oversight over the generated outputs by GPT models raises profound ethical dilemmas. Moreover, the computational demands and memory requirements of GPT models pose challenges in their application on small-scale or low-power systems. Furthermore, the cost associated with accessing and employing GPT models can be prohibitive, limiting their availability to numerous organizations. The integration of GPT models into existing systems presents a hurdle, necessitating specialized expertise and skills, encompassing a blend of technical and ethical proficiencies to manage the challenges affiliated with GPT models. While GPT models offer a plethora of functionalities, the scope for customization and scalability remains restricted, hampering their utilization for expansive applications or tailored requirements. Addressing these challenges is imperative to ensure the responsible and efficient utilization of GPT models.

6.2 Key Challenges of GPT

Large language models, such as ChatGPT, have garnered substantial interest across diverse domains, notably in natural language processing due to their remarkable ability to generate human-like responses and comprehend language in various conversational contexts. However, the opaque nature of these LLMs presents a challenge as their performance in practical applications remains largely undisclosed.

While these models offer advantages, they are not devoid of limitations and challenges. We have identified factors such as a decline in robustness, along with particular hurdles like real-time responsiveness and digital acuity. Chen et al. highlighted that GPT-3.5 exhibits decreased performance and inefficiency when tackling specific language understanding tasks, displaying deficiencies when compared to its predecessors. The study delved into the inadequate resilience of GPT-3.5 when confronted with particular tasks or inputs, showcasing its vulnerability to interference and producing unstable results. It is indicated that GPT-3.5 exhibits subpar performance or encounters challenges in certain specialized language comprehension tasks, potentially necessitating further refinement and enhancement.

6.3 Potential Solutions and Future Directions

GPT has significantly advanced the field of natural language processing, demonstrating exceptional proficiency in comprehending and generating human-like language. Leveraging the transformer architecture, a sophisticated neural network tailored for natural language processing tasks, GPT has garnered extensive acclaim within the research and industrial sectors, establishing itself as a prevalent and efficacious model in the domain. An innovative methodology introduced is the „GPT-in-the-loop“ approach, integrating the reasoning prowess of Large Language Models (LLMs) such as GPT with multi-agent systems to facilitate adaptive decision-making. The study delves into harnessing GPT as a decision-support tool to enable adaptive decision-making in multi-agent systems, where GPT is embedded to furnish decision support and guidance to the agents within the system. It examines the adaptive decision-making mechanism founded on GPT, leveraging its generative and interpretative capabilities to aid agents in multi-agent systems in making adaptive decisions to navigate evolving environments and task exigencies. Exploring GPT‘s role in the collaborative decision-making process of intelligent agents aims to foster cooperation and coordination among agent entities within the system, thus fostering heightened system intelligence and flexibility. The discourse encompasses the methodologies and ramifications of GPT application in the decision-making loop, elucidating how GPT can complement or supersede traditional decision-making approaches to enhance the efficiency and precision of decision-making within multi-agent systems. In essence, this article scrutinizes the utilization of GPT in multi-agent systems to harness its generative and interpretative capabilities for supporting adaptive decision-making processes, fostering the enhancement of intelligence and operational efficiency within multi-agent systems.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1]Kammersgaard J. Four different perspectives on human–computer interaction. International Journal of Man-Machine Studies, 1988, 28(4): 343-362.

[2]Hutchins J. The history of machine translation in a nutshell. Retrieved December, 2005, 20(2009): 1-1.

[3]Feng Z. Formal Models of Neural Network and Deep Learning//Formal Analysis for Natural Language Processing: A Handbook. Singapore: Springer Nature Singapore, 2023: 653-752.

[4]Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training. 2018.

[5]Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners. OpenAI blog, 2019, 1(8): 9.

[6]Kaplan J, McCandlish S, Henighan T, et al. Scaling laws for neural language models. arXiv preprint arXiv:2001.08361, 2020.

[7]Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. Advances in neural information processing systems, 2020, 33: 1877-1901.

[8]Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report. arXiv preprint arXiv:2303.08774, 2023.

[9]Rahaman M S, Ahsan M M T, Anjum N, et al. From

ChatGPT-3 to GPT-4: a significant advancement in ai-driven NLP tools. Journal of Engineering and Emerging Technologies, 2023, 2(1): 1-11.
[10]Giray L. Prompt engineering with ChatGPT: a guide for academic writers. Annals of biomedical engineering Prompt engineering for ChatGPT: a quick guide to techniques, tips, 2023, 51(12): 2629-2633.
[11]Ekin S. Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices[J]. Authorea Preprints, 2023.
[12]Dai Z, Yang Z, Yang Y, et al. Transformer-xl: Attentive language models beyond a fixed-length context[J]. arXiv preprint arXiv:1901.02860, 2019.
[13]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in neural information processing systems, 2017.
[14]Sutton R S, Barto A G. Reinforcement learning: An introduction. MIT press, 2018.
[15]Cristani M, Tomazzoli C. A multimodal approach to relevance and pertinence of documents//Trends in Applied Knowledge-Based Systems and Data Science: 29th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2016, Morioka, Japan, August 2-4, 2016, Proceedings 29. Springer International Publishing, 2016: 157-168.
[16]Hirasawa T, Kaneko M, Imankulova A, et al. Pre-trained word embedding and language model improve multimodal machine translation: A case study in Multi30K. IEEE Access, 2022, 10: 67653-67668.
[17]Oppenlaender J, Hämäläinen J. Mapping the challenges of HCI: An application and evaluation of ChatGPT and GPT-4 for cost-efficient question answering. arXiv preprint arXiv:2306.05036, 2023.
[18]Chen X, Ye J, Zu C, et al. How robust is gpt-3.5 to predecessors? a comprehensive study on language understanding tasks. arXiv preprint arXiv:2303.00293, 2023.