Research and Application of Transformation-based Generative AI Technology

Yongtao Yuan

School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China

Yongtao.Yuan22@student.xjtlu.edu.cn

Abstract. In recent times, generative AI has cropped up with many potential areas where its implementation can promise great efficiency. Above all, there is the Transformer model for which a detailed report is required for elaboration on principles, advantages, and disadvantages. To begin with, in this paper, the basic structure and function of the Transformer model will be examined in depth by discussing the effect of the Transformer model on natural language processing and image generation. Case studies showing practical applications are done by the chapter in content creation, marketing, customer support, virtual assistance, and mental health services by showing how this technology will redefine the fields. The paper finally reviews the literature at the end towards pointing to the future for the formulation of actionable recommendations with the view of enhancing both capabilities and applications related to generative AI technologies. This work underlines the profound impact of Transformer-based generative AI in innovation and efficiency across several dimensions.

Keywords: generative AI, Transformer, NLP, GPT.

1. Introduction

In the digital age, technology is changing our way of life and work by Artificial Intelligence (AI) faster than ever. In response to this influx, AI has branched out into many fields including intelligent assistants and even your soon-to-be autonomous car while driving economic growth as well as social advances [1]. It is now making some serious breakthroughs with the likes of deep learning, natural language processing, and computer vision as computational power advancements are paired up with data influx platforms that in turn pave the way for training machines to learn human languages (including high-level reading comprehension), understand images videos covering decision-making aspects. Generative AI technology is one such advancement which is characterized by huge potential and applicability in varied application areas. In particular, the Transformer model and its descendants have been at the forefront in terms of performance on tasks from natural language processing to image generation [2].

In this paper, we endeavor to explore generative applications of the Transformer and provide an in-depth explanation as to how advances can be leveraged in nearly state-of-the-art results. In this part of the blog, we will discuss their strengths and weaknesses for generative tasks by taking an in-depth look at someone's mind. We will then delve into the practical case studies of one which use case that Transformer is employed for Image Generation and Text Generation applications respectively. Second, we will propose the directions for future research and suggest action-oriented recommendations — that could benefit advancing generative AI technology based on a post-hoc review of these materials.

1.1. The development background

The algorithms we work with are getting better and better, but that's because over the last several decades now we've been able to be a lot more sophisticated about programming behavior of sorts and just nowhere near as simple as AI people like to put it in neat tidy little boxes. In the 2010s, deep learning advancements arrived with breakthrough change — CNN/RNN (convolutional/recurrent neural networks) that allowed computers to comprehend vast image-, speech- & text-based information which planted the seed for creating generative models. 2014 [3]: Ian Goodfellow et al. Generative Adversarial Networks(GANs): — This was the idea of GANS, basically a whole new different Idea where NN actually compete with each other (2 NN : real vs fake) A Generator and Discriminator during training on DATA to generate high-quality samples as output by simultaneously being trained iteratively led to this breakthrough in generative AI—threat for medical imaging translation saw quite glamorous applications too like crazy image generation(audio synthesis) [4].

Also, the area of Natural Language Processing (NLP) has seen a major leap [5]. The GPT (Generative Pre-trained Transformer) model, released by OpenAI in 2018 has marked the first step to a new era of text generation through pre-training and transfer learning [6]. Thanks to the improvements in computational power and big datasets, these models are very often employed for training generative AI which opened the door towards their application for artistic creation, game development but also medical imaging purposes.

1.2. Research purpose and significance

This work is organized into six sections, to revisit applications and developments concerning generative artificial intelligence technologies using Transformers in several fields. First, an in-depth study of the structure and principles of the Transformer model will be conducted to discuss the advantages and disadvantages in the performance of tasks such as natural language processing and image processing. Second, it focuses on the optimization of the Transformer model in light of the features of generative AI technology and enhances the performance of generating texts, hence increasing its possibilities of application.

In addition, according to the case studies, the effectiveness and potential of Transformer-based generative AI technologies are verified in an intelligent dialogue system, machine translation, and text summarization. To this end, the present study aims to enlighten researchers and engineers in related fields, arousing further creation and application of generative AI technologies to practical areas.

2. Principle analysis

Generative AI is AI-based technology that generates new data generated from scratch, based mostly on deep learning algorithms. The core ideas involve Generative Adversarial Networks, VAEs, and autoregressive [7].

GAN consists of a generator and a discriminator, two different neural networks. The generator generates synthetic data, while the discriminator decides whether the data is real or artificially generated. Through adversarial training, both networks continue improving their performance in such a way that the generator keeps enhancing the authenticity of the generated samples to generate high-quality images, text, and more in the process [8].

Variational auto-encoders basically work by mapping the input data to a latent space using an encoder, then sampling from that space, and at last, reconstructing the data through a decoder. VAE makes use of variational inference so that the model can maintain diversity in the generated samples while being effective at capturing characteristics from the distribution of data.

These are a set of autoregressive models-GPT included-that generate data by predicting the next element in the sequence, using context for creating coherent texts. This would form language patterns learned from a vast number of training data resulting in high-quality text generation [9].

The Transformer model is the best representative of the generative approach to AI because of the ability to process in parallel incoming data and represent long-range dependencies within this model.

The main parts of the Transformer are a self-attention mechanism and feedforward neural networks; therefore, the model is so good at working with NLP, machine translation, and image processing.

2.1. Self-attention mechanism

While being based on deep learning, Transformer models overwhelmingly rely on a mechanism known as self-attention [10]. The essential intuition of self-attention lies in the very idea of assessing of how much influence is wielded by each element in the input sequence, such as a word, on every other element. This is done through the following key steps:

- **Input Representation**: Each element of the input sequence will be a representation vector. Assuming the input sequence is:

$$X = [x_1, x_2, \ldots, x_n] \tag{1}$$

- **Linear Transformation**: For every input vector, three different vectors are created: Query, Key, and Value. Usually, it is done by multiplying the input with the parameter matrices:

    Query Vector:

$$Q = X \cdot W_Q \tag{2}$$

    Key Vector:

$$K = X \cdot W_K \tag{3}$$

    Value Vector:

$$V = X \cdot W_V \tag{4}$$

    Where W_Q, W_K, and W_V are the learned weight matrices.

- **Calculating Attention Scores**: Computation of Attention Score: The attention scores are calculated by the dot product of the Query vector and the Key vectors. This score would give the amount of attention that the particular element in the input sequence pays to others:

$$Attention\_scores = \frac{QK_T}{\sqrt{d_k}} \tag{5}$$

    Where d_k is the dimension of the Key vectors, used for scaling.

- **Applying Softmax**: Apply the Softmax function to the attention scores to get a distribution of weights over each element:

$$Attention\_weights = Softmax(Attention\_scores) \tag{6}$$

- **Weighted Sum**: Weight and sum the Value vectors with the weights coming from the previous step to produce an output vector:

$$Output = Attention\_weights \cdot V \tag{7}$$

With this, one can gauge various advantages of self-attention mechanisms. Long-range feature dependency captures are one of the prominent strengths, where they play an important role in understanding intricate structures or contextual positions of a language. While RNNs often fail to retain information after processing long sequences, self-attention models successfully capture relationships between the farthest elements of a sequence.

Besides, compared with the recursive structure, self-attention makes use of parallel computation-the whole sequence can be processed all at once- and is much more efficient in computation. In this self-attending mechanism, dynamically adjusting attention weights can emphasize important parts of the input with adaptation to context [11].

In essence, self-attention gives the Transformer its mighty feature extraction and representation capabilities. The flexibility and efficiency have achieved great success on many tasks such as machine translation, text generation, and even image processing.

2.2. The structure and working principle

2.2.1. Transformer model
Two parts of the Transformer model- The Encoder and the Decoder with identical layers stacked on top [12]. In this particular case, the structure is:

2.2.1.1. Encoder

- **Input Embedding**: This layer generates a vector representation of input text.
- **Positional Encoding**: Include the position (reverence of elements) in each input vector.
- **Self-Attention Layer**: Calculates the relationship between each word of input sequences with other words and generates context-aware representations.
- **Feed-Forward Neural Network**: Normally, a feed-forward neural network consists of two linear transformations and an activation function after the output of self-attention block undergoing nonlinear transformation.
- **Residual Connection and Layer Normalization**: The output of each sub-layer is added to its input (residual connection), and then the layer normalizes to reduce gradient explosions which helps the model be more robust.

2.2.1.2. Decoder

- Input Embedding: The same encoder converts a target sequence to vectors.
- Masked Self-Attention Layer: This layer ensures the model is not able to look in the future when predicting the next word.
- Self-Attention Layer (connected to the encoder): Receives output from the encoder and incorporates context information from the target sequence.
- Feed-Forward Neural Network: like the feed-forward network in the Encoder.
- Output Layer: Converts the decoder output into a probability distribution using a linear layer followed by a softmax function, which is used to generate the next word.

2.2.2. Working principle
The Transformer works in a four-step process: input processing, processing by the encoder, processing by the decoder, and generation of the output [13].
  A. The input processing would involve tokenization of the input text, changing it into embedding vectors, and adding positional encodings to make up the input sequence.
  B. The encoder would process this input sequence through multiple layers:

  - **Self-Attention Computation**: The representation of each word is decided based on the impact of all other words in the sequence.
  - **Feed-Forward Network Processing**: After going through the self-attention layer, the output results pass through a nonlinear transformation to get richer features.

  C. While generating words in the target sequence one by one, the decoder does the following to the words:

  - **Masked Self-Attention**: Here, the model is supposed to "keep its eyes" only on the words already generated.
  - **Encoder-Decoder Attention**: The output from the encoder has been used to capture information from the input sequence.

- **Feed-Forward Network Processing**: Similar to that performed in the encoder.

D. Finally, this decoder output is projected into the probability distribution over the next word using a linear layer and then a softmax function and the complete output sequence is generated step by step until the final output is produced.

2.3. Application of Transformer in NLP field

First introduced in 2017, the Transformer model has since rapidly emerged as one of the core architectures in NLP. Novel self-attention combined with parallel processing capability made the Transformer perform uniquely well in several tasks within NLP [14].

Applications of the Transformer in NLP include machine translation, text generation, summarization of text, question-answering systems, sentiment analysis, dialogue systems, chatbots, language modeling, semantic search, relation extraction, information extraction, and multimodal tasks [15].

The wide adoption of the Transformer model in NLP has driven not only technological advancement but also massive enhancement in performance regarding several various tasks. It is expected that as research goes deep and technology further develops, Transformers and their variants will play an important role in the future of Natural Language Processing.

3. Model Analysis of Transformer-based generative AI

3.1. Overview of generative AI technology

Generative AI can be understood to be a class of artificial intelligence technologies that are creative in generating new content. The technology could potentially generate different types of content, including text, images, audio, and video [16]. Fundamentally, generative AI reflects on the ability of the system to understand the distribution of data to learn from it, and generate new data that may be similar but different from the original data. The main concepts, technical principles, major models, and application scenarios of Generative AI are introduced below.

- **Generative Models**: These models constitute the bedrock for generative AI in capturing the true distribution of data and generating new samples that would resemble the training data. Common generative models include Generative Adversarial Networks, Variational Autoencoders, and Transformer-based models [17].
- **Autoregressive Models**: The model works its way by generating one thing at a time, conditioned on the output generated so far. Examples of autoregressive generative models include GPT, standing for Generative Pre-trained Transformer.
- **Conditional Generation**: This is based on the generation of content on certain conditions. It could be a generation of images based on some description in a textual form or, a generation of responses based on the given input, among others. This method is often employed in an attempt to try and achieve more targeted generative tasks.

3.2. Comparative analysis of Transformer-based generative AI models

3.2.1. Generative Pre-trained Transformer (GPT)

The GPT model is an autoregressive generative model based on the transformer architecture for text generation tasks [18]. Its philosophy is to have the model comprehend and generate natural language through pre-training and fine-tuning. In the pre-training process, GPT is trained on huge volumes of text, and unsupervised learning of the structure and semantics of the language is done by predicting the next word. In contrast, the fine-tuning stage optimizes the model for the end tasks with a small quantity of labeled data.

The most robust capability of GPT is in the rich context within which it understands and generates text that is both coherent and logically consistent. Excellent implementation in generating dialogues, content writing, and even programming tasks is being found for this. The autoregressive nature of GPT

means each step of the generation process depends on previous outputs, keeping coherence in the generated content. The architecture of GPT makes it exceptionally performant for natural language generation, with the generated text turning out both coherent and creative. Large-scale pre-training, as well as the generative ability of GPT-3 [19] and GPT-4 [20], finds huge applications in dialog systems, code generation, and content creation, among others.

3.2.2. Bidirectional Encoder Representations from Transformers (BERT)

BERT is a deep bidirectional encoder model inspired by the pre-training and fine-tuning strategy [21]. It gets pre-trained on two major tasks: Masked Language Modeling and Next Sentence Prediction. In Masked Language Modeling, some percentage of input words gets masked randomly, and a model should predict these masked words. This approach allows BERT to capture much richer contextual information, while Sentence Prediction allows the model to learn the relationships between sentences.

BERT is a deep model that performs most of the natural language processing tasks, such as text classification, sentiment analysis, and question answering, with excellence. The bidirectional nature of BERT gives it high superiority about comprehension of complex sentence structures and semantics. However, BERT is not suited for generative tasks because it is an encoder model basically intended for comprehension rather than generation. Although BERT itself does not generate any text, its variants such as BART and T5 further incorporate encoder-decoder structures for effective text generation. For instance, BART generates autoregressively on the input text, thereby making it suitable for summarization and translation tasks.

3.2.3. Text-to-Text Transfer Transformer (T5)

In the T5 model, this is done through pre-training via a well-known "fill-in-the-blank" task where the model is being trained to predict the masked word in an input text. It is not only this method that captures the structure and semantics of the language, but it is also the reason for the enhancement of multi-task performance in the model. Unlike other models, the T5 prepares a text-to-text format for the fine-tuning stage. It does so whereby both the input prompts and the target outputs shall be in text form. This design makes it suitable for a range of tasks: translation, summarization, question answering, and text classification [22].

Essentially, T5 provides a unified formulation for different NLP tasks as a text-to-text format and is thus much more flexible. The main philosophy of this is treating every language processing task as a text generation problem so that model design would be simpler, with better generalizability [23].

3.3. Application case analysis of Transformation-based generative AI technology

3.3.1. Use cases for text generation tasks

The Transformer-based generative AI technology has extensive applications for the text generation task in quite a few fields [24]. Key applications include:

- **Content creation**: Media organizations and bloggers make use of AI models, including GPT-3, when writing articles, blogs, and even social media posts. The model can create high-value content in very little time and help a writer get out of a creative block and ensure a continuous flow of material.
- **Marketing and Advertising**: AI-generated text helps companies create engaging product descriptions, promotional emails, and ad copy. Besides saving time, AI helps marketers scale their marketing with personal messages that linger in the minds of target audiences.
- **Recommendation Systems**: E-commerce websites make use of AI-powered technology in making personalized product recommendations and reviews based on their usage and browsing history, therefore enhancing customer interaction and conversion rates.
- **Educational Tools**: AI models in education are used to bring personalized learning tools, quizzes, and complex topic summaries to match diverse learning needs for students in general, improving educational experiences.

3.3.2. Use cases in the dialog generation task

Another domain in which the Transformer-based generative AI has shown its potential is in the case of dialogue generation. Key applications are listed below.

- **Customer Support Chatbots**: Businesses are using AI-powered chatbots to instantly reply to customers. Those bots utilize NLP to understand the context and generate an appropriate, helpful response, thus enhancing customer satisfaction and reducing response times [25].
- **Virtual Assistants**: Virtual assistants embed AI models into their functionality, like Siri, Alexa, and Google Assistant, which would allow them to converse-like talking and, therefore, answer and execute requests as a user needs. This sets the path for better user interaction.
- **Mental Health Support**: Mental Health Support-AI-powered applications are coming up to facilitate mental health services. These applications come with conversational agents that offer supportive dialogue and resources. They permit users to share their thoughts and, as an added advantage compared to traditional therapy, suggest ways of coping with the problems.
- **Storytelling Interactive**: Within the gaming and entertainment industries, this is used in dynamic narration, where players get to interact with characters that respond according to their choices, thereby enriching the experience.

3.3.3. Application cases in other fields

The applications of Transformer-based Generative AI are varied, although the two most common ones remain Text Generation and Dialogue Generation [26]. Other important applications are listed below:

- **Translation Services**: AI models working on real-time translation facilitate cross-linguistic communication. Technology such as Google Translate makes it easier to navigate around in a globalized environment.
- **Summarization**: AI is being used to summarize long documents, reports, and articles in law, research, and journalism. It contributes to the professionals' capturing all information in a nutshell without having to go through a plethora of texts.
- **Code Generation**: Enabling tools like GitHub Copilot to use Transformer models in generating code snippets with comments or function names given rapidly speeds up the development by avoiding repetitive programming tasks.
- **Creative Arts**: The creative domain applies AI in the construction of poetry, song lyrics, and even to the point of describing visual arts. This can work for artists, but also it enables human-machine collaboration on new artistic frontiers.

4. Challenges and Limitations

4.1. Ethics of Generative AI

Generative AI has many significant ethical considerations, though most of them are centered on the propensity to misuse the technology [27]. Creating misleading information, deepfakes, and unauthorized content can enable misinformation and an erosion of trust. There are also legal issues, including responsibility for the output produced through AI systems. Developers and organizations should take ethical considerations in the consequences of their models on privacy, autonomy, and societal norms. Some of such risks could be set off by drafting regulations and standards for responsible use of AI [28].

4.2. Data Biases and Fairness Issues

One big problem with generative AI, however, is biases in data. Since the models are trained on skewed data, they quite often show stereotypical behavior and discrimination. Fairness, therefore, by necessity requires the thorough examination of training data, whereby removal strategies offer the potential to make the outcomes equitable in AI-generated content [29,30].

4.3. Balancing Creativity and Consistency in Generated Content
While this was exciting in the prospect that generative AI might produce genuinely innovative outputs, in reality, the trade-offs that would have to be made between creativity and consistency remain unresolved. The model can either be too creative, in which case the results are incoherent, or it can be too conservative, in which case originality is not introduced. It is a subtle balance, and applications of creative industrial productions require meeting both artistic and functional standards [31,32].

5. Future Directions and Potential Developments

5.1. New Architectures to Be Explored for Generative AI
Generative artificial intelligence in the future needs to be developed through the investigation of new architecture so that the performance and innovation of models can be improved. This might integrate the advantages of transformers and GANs in the development of better and more creative models. Therefore, new architectures will contribute to improving the quality and individuality of generated content [33].

5.2. Improvement of Generative AI Models with Multi-modal Input
Future directions in developing generative AI will also involve a move toward multimodal inputs. Many other creative outputs could be envisioned based on present models in text, images, and audio. For example, video creation from written descriptions might further enhance user experiences and unleash creativity in education and entertainment, among many other industries [34,35].

5.3. Human-AI Collaboration in Generative AI
Generative AI in the near future will be more about human-machine collaboration: not using AI as a tool, but more like a collaborator in creation. With instinctive interface construction, users will better tap into the power of AI while retaining control over their creations. Such collaboration will bring revolutionary changes to the creative industry, amplifying human creativity with increased efficiency [36,37].

6. Conclusion

In the process, this paper has tried to closely consider the broad and developing uses of Transformer-based generative AI. It started with a close look into the Transformer structure and its working principles, outlining its huge advantages in capturing long-range relations, pivotal in natural language processing and image generation tasks. We then examined a series of case studies, which would, quantitatively and qualitatively, manner, demonstrate the effectiveness of Transformers across domains such as content creation, marketing, customer service, and virtual assistants. This analysis showed unmistakably that these AI models are now functioning to fundamentally change these professions by bringing productivity and creativity to entirely new dimensions.

Finally, we developed directions for future research and actionable recommendations for further optimization and capability expansion of generative AI. The wide-ranging investigation into generative AI underlines the prospective transformational power of such technologies and their promise to turn out truly innovative drivers of various sectors.

References
[1]   Feuerriegel S, et al., 2024 Generative AI. Business & Information Systems Engineering, 66(1):111-126
[2]   Mao X, Li Q, Mao X, et al., 2021 Generative adversarial networks (GANs)[J]. Generative Adversarial Networks for Image Generation: 1-7
[3]   Goodfellow I, Pouget-Abadie J, Mirza M, et al., 2014 Generative adversarial nets[J]. Advances in neural information processing systems, 27

[4]   Sauvola J, et al., 2024 Future of software development with generative AI. Automated Software Engineering, 31(1): 26

[5]   Feng S Y, et al., 2021 A survey of data augmentation approaches for NLP. arXiv preprint arXiv: 2105.03075

[6]   Topal M O, Bas A and van Heerden I 2021 Exploring transformers in natural language generation: Gpt, bert, and xlnet[J]. arXiv preprint arXiv:2102.08036

[7]   Brynjolfsson E, Li D and Raymond L R 2023 Generative AI at work. No. w31161. National Bureau of Economic Research

[8]   Goodfellow I, et al., 2020 Generative adversarial networks[J]. Communications of the ACM, 63(11): 139-144

[9]   Wang B, et al., 2023 DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models. *NeurIPS*

[10]   Zhang H, et al., 2019 Self-attention generative adversarial networks. International conference on machine learning. PMLR, 7354-7363

[11]   Pan X, Ge C, Lu R, et al., 2022 On the integration of self-attention and convolution. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 815-825

[12]   Ji Y, et al., 2021 CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. Information Sciences, 546: 835-857

[13]   Badrinarayanan V, Kendall A and Cipolla R 2017 Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12): 2481-2495

[14]   Nath S, et al., 2022 New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology. British Journal of Ophthalmology, 106(7): 889-892

[15]   Kang Y, et al., 2020 Natural language processing (NLP) in management research: A literature review. Journal of Management Analytics, 7(2): 139-172

[16]   Theis L, Oord A, and Bethge M 2015 A note on the evaluation of generative models. arXiv preprint arXiv: 1511.01844

[17]   Salakhutdinov R 2015 Learning deep generative models. Annual Review of Statistics and Its Application, 2(1): 361-385

[18]   Liu X, et al., 2023 GPT understands, too. AI Open

[19]   Floridi, L and Chiriatti M 2020 GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30: 681-694

[20]   Achiam J, et al. 2023 Gpt-4 technical report. arXiv preprint arXiv: 2303.08774

[21]   Koroteev M V 2021 BERT: a review of applications in natural language processing and understanding. arXiv preprint arXiv: 2103.11943

[22]   Raffel C, et al., 2020 Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140): 1-67

[23]   Du Z, et al., 2021 All nlp tasks are generation tasks: A general pretraining framework. arXiv preprint arXiv: 2103.10360,18

[24]   Gatt A and Krahmer E 2018 Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. Journal of Artificial Intelligence Research, 61: 65-170

[25]   He Z, Wang J, and Chen J 2020 Task-Oriented Dialog Generation with Enhanced Entity Representation. INTERSPEECH, 3905-3909

[26]   Dubey S R and Singh S K 2024 Transformer-based generative adversarial networks in computer vision: A comprehensive survey[J]. IEEE Transactions on Artificial Intelligence

[27]   Farina M, Yu X, and Lavazza A 2024 Ethical considerations and policy interventions concerning the impact of generative AI tools in the economy and in society. AI and Ethics, 1-9

[28]   Rana N P, et al., 2024 Assessing the nexus of Generative AI adoption, ethical considerations and organizational performance. Technovation, 135: 103064

[29]   Chen P, Wu L, and Wang L 2024 AI fairness in data management and analytics: A review on challenges, methodologies and applications. Applied Sciences, 13(18): 10258

[30]   Anzum F, Asha A Z and Gavrilova M L 2022 Biases, fairness, and implications of using AI in social media data mining. International Conference on Cyberworlds (CW). IEEE, 251-254

[31]   González-Sendino R, et al., 2023 A review of bias and fairness in artificial intelligence

[32]   Mazilu L, et al., 2020 Fairness in data wrangling. IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI). IEEE, 341-348

[33]   Chen H, et al., 2024 Multi-Modal Generative AI: Multi-modal LLM, Diffusion and Beyond. arXiv preprint arXiv: 2409.14993

[34]   Long X, et al., 2024 Generative multi-modal knowledge retrieval with large language models. Proceedings of the AAAI Conference on Artificial Intelligence, 38(17): 18733-19841

[35]   Ramisa, Arnau, et al., 2024 Multi-modal Generative Models in Recommendation System. arXiv preprint arXiv: 2409.10993

[36]   Fui-Hoon N F, et al., 2023 Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3): 277-304

[37]   Grabe I, et al., 2022 Towards a framework for human-AI interaction patterns in co-creative GAN applications. Joint Proceedings of the ACM IUI Workshops