УДК 004.9 **doi: 10.26906/SUNZ.2024.4.075**

A. Kapiton[1], D. Tyshchenko[2], A. Desiatko[2], V. Lazorenko[2]

[1] National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine
[2] State University of Trade and Economics, Kiev, Ukraine

EVOLUTION AND DISTRIBUTION ANALYSIS OF MULTIMODAL ARTIFICIAL INTELLIGENCE SYSTEMS

Abstract. The article considers the main aspects of evolution and performs a thorough analysis of the stages of formation of multimodal artificial intelligence systems (AIS). It was determined that in modern realities, artificial intelligence has undergone a transformational shift towards embracing multimodality in large language models. Modern approaches and ways of improving large language models by means of processing and generating a large amount of data are analyzed. The stages of transformation of artificial intelligence in the direction of multimodality of innovative development in large language models have been studied. The issue of verification and interaction of information systems with the surrounding world is considered. It was determined that they are inherently multimodal, multicomponent. Ways of improving large language models with the help of the ability to process and generate different data modalities are analyzed. It has been investigated that modern multimodal artificial intelligence systems are effectively used in various fields of science, education, and economics and require further development and improvement. It was determined that due to the rapid development of information technologies and systems in various spectrums of life, AI is experiencing a rapid modification, where generative models, which are becoming more and more perfect, deserve special attention. An overview of the architecture of the AnyGPT model is performed, where modalities are tokenized into discrete tokens, on the basis of which LLM performs multimodal perception and generation in autoregression. The methodology underlying AnyGPT was found to be multi-component, with the model demonstrating capabilities on par with specialized models in all assessment modalities tested. It has been established that tools designed to detect objects generated by artificial intelligence are in a state of development and are constantly being modified.

Keywords: artificial intelligence, bioengineering, generative models, multimodality.

Introduction

Artificial intelligence has undergone a transformational shift towards embracing multimodality in large language models (LLMs), which has ushered in a new way of looking at how machines perceive and interact with the world around them. This evolution stems from the recognition that human experience is inherently multimodal, encompassing not only text but also speech, images, and music. Hence, enhancing large language models with the ability to process and generate different data modalities holds great promise for increasing their utility and applicability in real-world scenarios. Information technologies and systems play an increasingly important role in today's world. Their influence is felt in all spheres of life, from economy and education to science and transport. Information technologies and systems not only make our work more efficient, but also open up new opportunities for development and innovation. The purpose of this study is to analyze the impact of information technologies and systems on the economy, education, science and transport, identification of key problems and challenges related to the development of these technologies and systems.

Analysis of recent research and publications. The problem of analyzing the evolution of the development of artificial intelligence in the direction of multimodality and transformational development in large language models (llm) has always been in the scientific focus of leading foreign and domestic scientists. It was the study of verification and interaction of information systems with the surrounding world that caused, according to scientists, this evolution of views regarding the perception of certain results that are inherently multimodal and multicomponent. Analysis of the improvement of llm with the help of the ability to process and generate different data modalities in the field of view of a number of foreign scientists. C. Wang, S. Chen, Y. Wu, Z. Zhang, L. Zhou, S. Liu, Z. Chen, Y. Liu, H. Wang, J. Li, L.He, S. Zhao, F. Wei, Z.Tang, Z. Yang, M. Khademi, Y. Liu, C. Zhu, and M. Bansal consider the issue of language models of neural codecs [1, 2].

Y. Wang, Y. Kordi, S. Mishra, A. Liu, N. Smith, D. Khashabi, and H. Hajishirzi investigate the problems of self-learning from the point of view of matching the LLMs with self-created instructions [3].

Sh. Wu, H. Fei, L. Qu, W. Ji, and Tat-Seng Chua They study the main advantages and disadvantages of multimodal Next-gpt [4].

T. Zhang, Y. Wu, T. Berg-Kirkpatrick, K. Chen, Y. Hui and S. Dubnov consider the features of large-scale contrastive pre-learning of speech and audio with feature fusion and keyword addition to captions [5].

N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi explore the features of Soundstream through the lens of an end-to-end neural audio codec [6].

D. Tyshshenko, T. Franchuk, R. Zakharov, V. Moskalenko in their works explore the design of key information management protocols using multimodal AI [7].

According to domestic scientists, one of the main tasks in this dynamic field is the design and development of models capable of seamlessly integrating and processing various types of data. It is the analysis of the creation of dual-modal models that combine different forms of data that is devoted to the works of O. Sukhorebrogo, D. Nenych, A. Kurilekh [8, 9]. Practical applications of the integration of artificial intelligence are investigated by S. Gladkyi, M. Prorok [10].

Main part

AI became the fastest service to reach one hundred million monthly active users. According to PwC's 2023 Emerging Technology Survey, more than fifty four percent of surveyed companies integrated generative AI into their business processes during the year. They was practiced by Blackrock Neurotech, Precision Neuroscience and many others. but Neuralink's main difference is its focus on expanding human capabilities, not just restoring lost ones. The analysis of scientists' research on the main characteristics of multimodal AI technologies made it possible to highlight the key ones are presented on Fig. 1 [11–20].

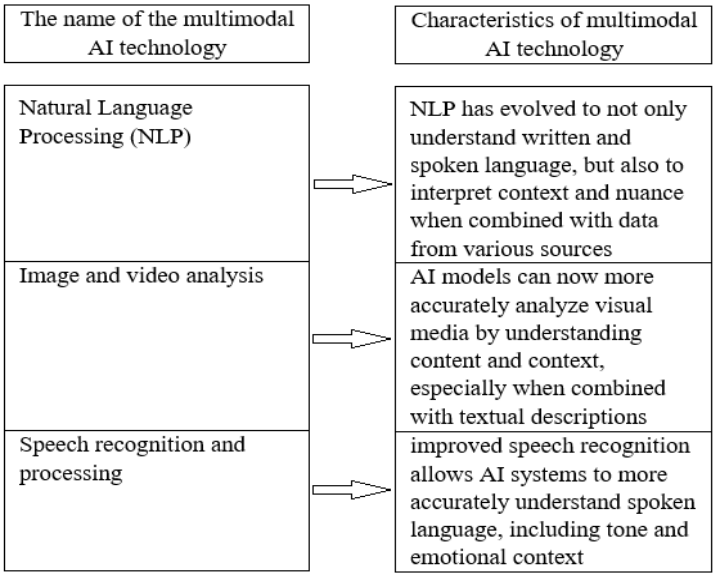

Fig. 1. Key technologies of multimodal AI

Information technologies and systems are rapidly developing, having a significant impact on all spheres of life. AI is experiencing explosive growth, with generative models becoming increasingly sophisticated. Mobile communication is evolving to provide access to communication anywhere in the world. Bioengineering is taking significant steps forward in exploring the possibilities of the brain-computer interface. In our opinion, the best tools for recognizing artificial intelligence AI are presented on Fig. 2 [17–20].

Tools designed to detect AI-generated images are mostly still in a state of development, evolving along with technological advances. The rapid pace of change in the field of AI is impressive, as significant advances are made almost every day. As more AI tools become available, the more they increase our efficiency, and this momentum of progress is expected to accelerate even further.

AnyGPT's performance underscores the efficiency of its design. Thanks to the rapid development of generative AI, creating convincing fake content has become much easier and more accessible. AI-based image generators and deepfaking technology are used for a wide range of purposes, from students to cheat on tests to fraudsters spreading disinformation about wars, elections and natural disasters.

Researchers have identified several approaches to multimodality, a brief description of which is presented are presented on Fig. 3.

This indicates a significant growth of AI in last year.The most famous of them is Google, which announced a competitor to ChatGPT Bard. In April, thanks to the merger of Google Research and DeepMind teams, they are creating methods to improve the effectiveness of reinforcement learning. Such systems are able to perform extremely complex tasks through thousands of training iterations. In last year, Sundar Pichai, CEO of Google, together with Demis Hassabis of Google DeepMind presented Gemini, a multimodal AI that not only has the ability to understand text or images, but also combines different types of information in a way that is much closer to how humans perceive AI and other IT giants such as Microsoft Bing were also presented (renamed COPILOT in two thousand twenty-four year), Meta announces its open source LLM model, Anthropic releases Claude 2 and receives investment from Amazone. Working with the above systems on mobile devices requires a stable, fast and accessible connection.

This is precisely what can be used to show all possible combinations of high-resolution image synthesis, which helps to work on a whole series of different sets of solutions to the assigned problems. Usually, to create a given image, a special colab is used. In particular, Google Colab is a free service that presents everything you need for machine learning, divided into many separate cells.

Conclusions

The development of neural networks is a multi-component system, where each subsequent component is connected with the previous one, being its basis. First, language models appeared, tailored to work with text, and then, layer by layer, other modalities began to be added, such as photos, video and audio. Therefore, the primary source and basis of the research currently being conducted concerns multimodal AI, which is directly related to classical text neural networks.

AI identification tool	Characteristics of AI identification tool
Winston AI	the tool determines if the image was created by artificial intelligence, the results include image information such as C2PA, IPTC and Exif data
Illuminarty	a tool for verifying the origin of an image on the Internet, provides an estimate of the likelihood of the involvement of artificial intelligence
Hive Moderation	an AI detection tool, specifically for detecting AI in images and videos, provides API services for processing and tagging images, videos, GIFs, web pages, audio, and live streams for content moderation
Is It AI?	a tool designed to identify an image or text generated by artificial intelligence offers a free version for basic use with the option to subscribe to additional features or integrate it into your AI content review platform
Originality.ai	offers AI text recognition services for writers, marketers and publishers, has three modes – Lite, Standard and Turbo
GPTZero	AI text detector for teachers, writers, cyber security professionals and recruiters
Copyleaks	copyleaks' AI text detector is designed to detect human writing patterns and flags content as potentially AI only when it detects deviations from these patterns

Fig. 2. The tools for recognizing artificial intelligence

Method	Characteristics of the method
Tool-augmented LLM	combines several independent models in one product
End-to-end multimodal LLM	instead of using separate models for text and images, such a model is trained on all necessary types of data at once, within a single structure
Modality bridging with pretrained models	eliminating the gap between modalities, combining them, where models exchange data through't a text request, but using mathematical vectors

Fig. 3. Characteristics of approaches to multimodality

Separately, it is necessary to looks the problems of large multimodal models: inclusion of more data modalities; availability of diverse data sets; generation of multimodal outputs; list of instructions (LMMs face the challenge of mastering dialogue and following instructions beyond simple completion); multimodal reasoning (seamless integration of multimodal data for complex reasoning tasks); LMM compression (the resource-intensive nature of LMMs is a major obstacle, making them impractical for compute-constrained peripherals). Compressing LMMs to improve efficiency and make them deployable on resource-constrained devices is a critical area of current research.

REFERENCES

1. Chengyi Wang, Sanyuan Chen, Yu Wu, Zi-Hua Zhang, Long Zhou, Shujie Liu, Zhuo Chen, Yanqing Liu, Huaming Wang, Jinyu Li, Lei He, Sheng Zhao, and Furu Wei. Neural codec language models are zero-shot text to speech synthesizers. ArXiv preprint, abs/2301.02111, 2023. URL: https://arxiv.org/abs/2301.02111.
2. Zineng Tang, Ziyi Yang, Mahmoud Khademi, Yang Liu, Chenguang Zhu, and Mohit Bansal. Codi-2: In-context, interleaved, and interactive any-to-any generation. ArXiv preprint, abs/2311.18775, 2023a. URL: https://arxiv.org/abs/2311.18775.
3. Y. Wang, Y. Kordi, S. Mishra, A. Liu, N. A. Smith, D. Khashabi, and H. Hajishirzi. Self-instruct: Aligning language model with self generated instructions. ArXiv preprint, abs/2212.10560, 2022. URL: https://arxiv.org/abs/2212.10560.
4. Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm. ArXiv preprint, abs/2309.05519, 2023. URL; https://arxiv.org/abs/2309.05519.
5. Yusong Wu, K. Chen, Tianyu Zhang, Yuchen Hui, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5, 2022. URL: https://api.semanticscholar.org/CorpusID: 253510826.
6. Neil Zeghidour, Alejandro Luebs, Ahmed Omran, Jan Skoglund, and Marco Tagliasacchi. Soundstream: An end-to-end neural audio codec. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30:495–507, 2021. URL: https://api.semanticscholar.org/CorpusID: 236149944.
7. Tyshchenko D., Franchuk T., Zakharov R., Moskalenko V. Підтримка динамічних потреб безпеки засобами VPN *Системи управління, навігації та зв'язку*. Полтава: ПНТУ, 2024. Т. 3 (77). 149-152.
8. Курилех А., Капітон А. Використання штучного інтелекту для розвитку CRM-систем. *Стан, досягнення та перспективи інформаційних систем і технологій*. Одеса: ОНТУ, 2024. 357-359.
9. Капітон А., Сухоребрий О., Ненич Д. Використання мультимодального штучного інтелекту в економіці, освіті, науці та транспорті. *Інформаційні технології та цифрова економіка*. Київ: ДУІТ, 2024. 83-85.
10. Капітон А, Гладкий С., Пророк М. Практичні застосування інтеграції штучного інтелекту в процес освіти. *Стан, досягнення та перспективи інформаційних систем і технологій* Одеса: ОНТУ, 2024. 348-349.
11. PwC's 2023 Emerging Technology Survey. URL: https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html
12. Gemini. URL: https://blog.google/technology/ai/google-gemini-ai/#sundar-note
13. Bing. URL: https://www.microsoft.com/en-us/edge/features/the-newbing?form=MA13FJ
14. Introducing LLaMA. URL: https://ai.meta.com/blog/large-language-model-llamameta-ai/
15. Chat With RTX. URL: https://www.nvidia.com/en-us/ai-on-rtx/chat-with-rtxgenerative-ai/
16. Verner S. IBM adds AI-enhanced data resilience capabilities to help combat ransomware and other threats with enhanced storage solutions, 2024. URL: newsroom.ibm.com/
17. AnyGPT: Unified Multimodal LLM with Discrete Sequence Modeling URL: https://arxiv.org/pdf/2402.12226
18. Laion-aesthetics. URL: https://laion.ai/blog/laion-aesthetics/, 2022a.
19. Laion coco: 600m synthetic captions from laion2b-en. URL: https://laion.ai/blog/laion-coco/, 2022b.
20. AI identification tools URL: https://thetransmitted.com/ai/instrumenty-identyfikacziyi-shi-zhovten-2024/

Received (Надійшла) 26.06.2024
Accepted for publication (Прийнята до друку) 02.10.2024

Еволюція та аналіз розвитку мультимодальних систем штучного інтелекту

А. Капітон, Д. Тищенко, А. Десятко, В. Лазоренко

Анотація. У статті розглянуто основні аспекти еволюції та проаналізовано розвиток мультимодальних систем штучного інтелекту. Визначено, що в сучасних реаліях штучний інтелект зазнав трансформаційного зсуву в бік охоплення мультимодальності у великих мовних моделях. Праналізовано шляхи вдосконалення великих мовних моделей за допомогою здатності обробляти і генерувати великий обсяг даних. Метою цього дослідження є аналіз вимог до розробки та впровадження мультимодальних систем штучного інтелекту. Дослджено етапи трансформації штучного інтелекту у напрямку мультимодальності іноваційного розвитку у великих мовних моделях. Розглянуто питання верифікації та взаємодії інформаційних систем з навколишнім світом. Визначено, що вони за своєю суттю є мультимодальними, багатокомпонентними. Проаналізлвано шляхи вдосконалення великих мовних моделей за допомогою здатності обробляти і генерувати різні модальності даних. Досліджено, що сучасні мультимодальні системи штучного інтелекту ефективно використовуються в різних галузях науки, освіти, економіки та потребують подальшого розвитку та вдосконалення. Визначено, що внаслідок бурхливого розвитку інформаційних технологій та систем в різних спектрах життєдіяльності, ШІ переживає бурхливої модифікації, де особливої уваги заслуговують генеративні моделі, які стають все більш досконалими. Виконано огляд архітектури моделі AnyGPT, де модальності токенізуються в дискретні токени, на основі яких LLM виконує мультимодальне сприйняття та генерування в авторегресії. Визначено, що методологія, що лежить в основі AnyGPT, є багатокомпонентною, модель якої демонструє можливості на рівні зі спеціалізованими моделями в усіх протестованих модальностях оцінювання. Встановлено, що інструменти, призначені для виявлення об'єктів, згенерованих штучним інтелектом, перебувають у стані розвитку, та постійно модифікуються.

Ключові слова: штучний інтелект, біоінженерія, генеративні моделі, мультимодальність.