GPT-4o: The Cutting-Edge Advancement in Multimodal LLM

Raisa Islam and Owana Marzia Moushi

July 2, 2024

GPT-4o: The Cutting-Edge Advancement in Multimodal LLM

Raisa Islam
Computer Science
New Mexico Institute of Mining and Technology
Socorro, NM, USA
raisa.islam@student.nmt.edu

Owana Marzia Moushi
Electrical and Computer Engineering
University of Nebraska - Lincoln
Omaha, NE, USA
omoushi2@huskers.unl.edu

***Abstract*—GPT-4o marks a significant advancement in AI technology, enhancing multimodal capabilities. OpenAI has launched several GPT models over the years, with GPT-4o being the latest. This paper provides a concise overview of these models, focusing on their key features and technological advancements. The main objective is to present a brief overview of GPT-4o, including its technological innovations. GPT-4o offers substantial improvements over its predecessors by introducing multimodal capabilities, larger context windows, efficient tokenization, and faster processing speeds, achieving state-of-the-art performance in text, audio, video, and image generation and understanding. We have compared GPT-4o with ten top LLMs using metrics such as throughput, response time, and latency, where GPT-4o demonstrated clear superiority. Additionally, this paper explores various application domains, highlighting GPT-4o's versatility and potential to modernize multiple aspects of human life.**

***Index Terms*—GPT-4o, LLM, OpenAI, AI, ChatGPT, multimodal**

I. Introduction

Artificial Intelligence (AI) is one of the most popular cutting-edge technology in this era which is gaining attention because of its revolutionizing aspects in our life [1]. Natural Language Processing (NLP), a subset of AI, focuses on the interaction between computers and human languages, aiming to enable machines to understand, interpret, and generate human language meaningfully. The advent of Large Language Models (LLMs) has initiated a paradigm shift and revolutionized the field of NLP by leveraging deep learning techniques to process and generate natural language with unprecedented accuracy and fluency [2]. These models are trained on vast amounts of textual data, allowing them to capture intricate linguistic patterns and contextual nuances. As a result, LLMs have significantly enhanced applications in various domains, including chatbots, language translation, sentiment analysis, and content creation, driving forward the capabilities of AI in understanding and utilizing human language [3] [4].

OpenAI's Generative Pre-trained Transformer (GPT) models have fundamentally transformed the fields of AI and NLP over the past ten years. GPT models are built upon the transformer architecture, which has demonstrated exceptional effectiveness, particularly in developing LLMs. The first version, GPT-1, was launched in 2017. Subsequent versions, including GPT-2 and GPT-3, culminated in the groundbreaking ChatGPT. Since its release in November 2022, ChatGPT has gained immense popularity, reaching 1 million users within a week and 100 million users within two months [5]. ChatGPT aims to simplify various tasks, such as coding, math solving, and problem-solving, by using a combination of pre-trained unsupervised models and fine-tuned supervised models to provide human-like responses [3]. However, the last version, GPT-4, had limitations in video chatting, image processing, and audio processing. To address these shortcomings and advance toward a more digital world, OpenAI has released a new version of ChatGPT, called GPT-4o.

The recent version of ChatGPT is based on GPT-4o architecture which is gaining success on the previous chatbots. It was released on May 13, 2024, called GPT-4omni (or GPT-4o) the latest multimodal LLM from OpenAI [6]. The term ”omni,” derived from the Latin word ”omnis,” meaning ”all” or ”every,” highlights the model's omni-modal capabilities [7]. GPT-4o can process and understand multimodal inputs, including text, images, audio, and video, making it a significant advancement in AI technology. It is the first LLM capable of interpreting emotions from videos, enhancing user experience by analyzing various data types. Although the model is already available for use, OpenAI continues to improve its capabilities, with updates expected soon [8]. The model's efficiency in handling different types of data results in high success rates and reduced processing times and costs [9]. For users seeking unlimited access and enhanced features, OpenAI offers a paid version, such as ChatGPT Plus, which includes a higher message limit and upcoming macOS desktop app support [10] [11]. GPT-4o is twice as fast as GPT-4 Turbo (released after GPT-4) and represents a milestone in the digital interaction between machines and humans, promising to significantly impact the future of digital communication [7] [12] [13].

Tech giant Apple and OpenAI are collaborating to integrate ChatGPT into Apple's ecosystem, focusing on the upcoming iOS 18 operating system [14]. This partnership aims to enhance the user experience by embedding ChatGPT's advanced language capabilities into Apple services like Siri, making them more intuitive and responsive. Features include real-time conversation synchronization, and voice input through Whisper. These enhancements will allow Siri to process

	GPT-4o	GPT-4	GPT-3.5	GPT-3
Initial Release Date	May 13, 2024	March 14, 2023	March 15, 2022	June 11, 2020
Modality	Audio, video, text, images	Text and images	Text	Text
Context window [2] [15]	128000	8192	4096	2048
Parameters [2] [16] [17]	Yet-to-Disclose (YTD)	1.76 trillion	175 billion	175 billion
Cost per 1M token [18] [19]	Input: $5 Output: $15	Input: $30 Output: $60	Input: $1.5 Output: $2	$0.4 to $20
Decoder layers [16] [17] [20]	YTD	120	96	96
Prompting Method [17] [21]	YTD	Chain-of-Thought, n-shot[1]	n-shot	n-shot
Training Data [17] [22]	YTD	$\sim$ 13T tokens which includes both text data and code data	Over 570GB of text data	About 45TB text data
Response Time [23] [24]	30% faster than GPT-4	94ms per token	35ms per token	Unknown
Performance	State-of-the-art performance	Advanced in handling complex tasks, slower response time than GPT-4o	Improved version of GPT-3, lacks depth of understanding	Poor on complex tasks

TABLE I: Summary of latest GPT models

complex queries more effectively and provide functionalities such as instant answers, tailored advice, creative inspiration, professional input, and educational support. This integration aligns with Apple's broader strategy to incorporate AI-driven features while ensuring privacy by powering most features on-device.

A. Evolution of GPT

Table I provides an overview of the latest GPT models to understand the improvement OpenAI is making for the updated GPT models. GPT-4o significantly enhances the capabilities of its predecessors by incorporating multimodal functionalities, including text, audio, image, and video processing. It offers faster response times and reduced costs compared to GPT-4, while maintaining high accuracy and performance, surpassing the improvements seen in GPT-3.5 and GPT-3.

Contribution: The main contributions of this paper are as follows:

- We have provided an in-depth overview of GPT-4o, detailing the technologies that contribute to its advancement over previous models. This study also summarizes the key features and advantages of GPT-4o, highlighting the significant improvements it offers compared to its predecessors.
- We have analyzed the text and vision performance comparisons provided by the OpenAI team. Additionally, a comparison of GPT-4o with other leading LLMs using metrics such as throughput, response time, and latency has been discussed. The results demonstrate the superior performance of GPT-4o.
- Furthermore, we have presented the various fields where GPT-4o can be applied, showcasing its versatility and potential impact.

The structure of this paper is as follows: Section II provides an overview of the newly launched GPT-4o, including its technologies, features, advantages, and challenges. Section III analyzes the performance of GPT-4o by comparing it with popular LLM models, and Section IV explores the application domains where GPT-4o can be utilized. Finally, Section V concludes the paper.

II. GPT-4o Overview

This section provides an overview of GPT-4o, detailing the technologies employed in its development, its features, and its advantages. While GPT-4o incorporates numerous updated technologies, it also faces some challenges that need to be addressed. These aspects, including both the strengths and areas for improvement, are comprehensively discussed here.

A. Technology

GPT-4o builds upon the architecture of its predecessors, integrating enhancements in context window size, tokenization efficiency, and multimodal capabilities, which include processing text, audio, video, and images. This single model is trained end-to-end across text, vision, and audio, ensuring all inputs and outputs are managed by the same neural network [8]. It employs refined reinforcement learning with human feedback (RLHF) [25], significantly improving its alignment with human values and ethical standards. GPT-4o also features an advanced transformer architecture with enhanced self-attention mechanisms, allowing for better comprehension and generation of nuanced, contextually relevant responses. Utilizing Nvidia's most advanced GPUs [26], known for their parallel processing capabilities, GPT-4o efficiently handles the massive computations required by its sophisticated architecture. This model gives better performance for scalability since it uses larger training sessions [25]. GPT-4o's rate limits are 5x higher than GPT-4 Turbo—up to 10 million tokens per minute. These advancements make GPT-4o a versatile tool for applications ranging from conversational AI to content generation and data analysis. The following subsection discusses the technologies that are being used in GPT-4o.

1) o200k_base Tokenizer: GPT-4o introduced a new $o200k\_base$ tokenizer algorithm, marking a shift from the $cl100k\_base$ tokenizer used by GPT-4, GPT-4 Turbo, and GPT-3.5 Turbo[2]. Tokenization, which breaks down text into smaller units called tokens, is critical in NLP. The $o200k\_base$ tokenizer improves upon previous methods by being faster and more efficient, allowing GPT-4o to process and generate language at unprecedented speeds. It enhances semantic coherence in generated text and improves the handling of multiple languages [27], expanding GPT-4o's applicability across various linguistic contexts.

[1] the value of n starts from 0

[2] https://github.com/openai/tiktoken

2) RAG-GPT: RAG-GPT[3] is an advanced implementation of Retrieval-Augmented Generation (RAG) technology, designed to enhance the capabilities of LLMs by integrating them with efficient document retrieval systems. This integration allows RAG-GPT to provide more accurate and contextually relevant responses by fetching and incorporating information from extensive knowledge bases during query processing. Integration of RAG-GPT in GPT-4o enables delivering precise and grounded answers, making it particularly useful for applications requiring detailed and up-to-date information. This approach not only improves the quality of generated content but also enables the handling of specialized and complex queries with greater reliability.

3) Context Window: GPT-4o features an impressive context window size of 128k tokens [6]. The larger context window allows the model to maintain and process a much larger amount of information within a single interaction, enhancing its ability to understand and generate responses that are contextually relevant over extended conversations. This large context window is particularly beneficial for complex tasks requiring extensive context retention, tracking multiple threads of a conversation, and integration of multiple pieces of information.

4) Cloud Infrastructure and API Access: GPT-4o is designed to be deployed on scalable cloud infrastructures (Microsoft Azure) [28] [29], offering flexible API access for diverse applications. This cloud-based deployment model allows for seamless scaling to meet varying user demands and workloads.

B. Features

`OpenAI` is improving its models to make the world digital by adding some additional features which are discussed below:

1) Multimodality: GPT-4o is the newest multimodal LLM. It can understand and generate spoken languages, recognizing speech to transcribe text and using text-to-speech to generate speech [8]. Its vision capabilities allow it to interpret and generate visual content, such as recognizing images, generating new images, and solving problems by analyzing uploaded images. In text interaction, GPT-4o excels in NLP, enabling it to write essays, answer questions, provide summaries, and create stories or poetry based on user input.

2) Enhanced Interaction: Users can interact with ChatGPT more dynamically, interrupting and receiving responses in real-time. The model can detect nuances in users' emotions and respond in various emotive tones, making conversations more natural and engaging [30]. Additionally, it can handle interactive Q&A sessions which require extensive knowledge [6].

3) Data Analysis: Users can create interactive tables and charts from uploaded data in a variety of file formats. ChatGPT automatically generates an interactive table view, enabling users to scroll through all rows and columns. Users can create interactive charts by specifying chart type or automated selection, customize the graphics of these charts, and generate summaries to explain their findings [31].

4) Multilingual: GPT-4o is capable of responding to 50 different non-English languages [6] [12]. This proficiency ensures effective communication and content generation for a global audience, making it a valuable tool for diverse linguistic applications.

5) Memory: ChatGPT introduced memory capabilities to save users from repeating information across conversations and to enhance future interactions. Users have control over the memory; they can instruct it to remember or forget specific details, inquire about what it remembers, and manage these settings or disable memory entirely [32].

C. Advantages

The following section explains the important advantages that GPT-4o has made over the previous chatbots.

1) Faster Response: GPT-4o can respond to audio inputs in as little as 232 milliseconds, with an average of 320 milliseconds, which is similar to human response time in a conversation [8].

2) Cost-effective: This GPT-4o model is cost-effective since we are using one model for various data types of input. This model will help analyze the video, security purposes, sports, and also for content analysis [13]. Most importantly, this chatbot has around 50% reduced processing cost than the previous model GPT-4 Turbo [27].

3) Safe and Reliable: GPT-4o is safe and reliable compared to the previous model. This model is based on the feedback of humans with a backbone of reinforcement learning which makes it more reliable. They also reduce the amount of misleading content generation by using the feedback [25]

D. Challenges

Despite the numerous advancements of GPT-4o, it still has noteworthy limitations. The outage on June 4th, 2024, caused by a major system issue, affected all users of ChatGPT-related services and lasted several hours[6]. The exact technical causes were not detailed, but such incidents typically involve a mix of software bugs, infrastructure issues, or configuration errors. Furthermore, GPT-4o's audio models are limited to preset voices, and the model's pronunciation or explanations can sometimes be incorrect [8]. Additionally, data breaches remain a significant concern in the digital world, emphasizing the need to protect user data and comply with data protection regulations to maintain trust and legal compliance. Ensuring responsible AI use, avoiding biases, and adhering to ethical practices are crucial considerations [29].

III. Performance Evaluation

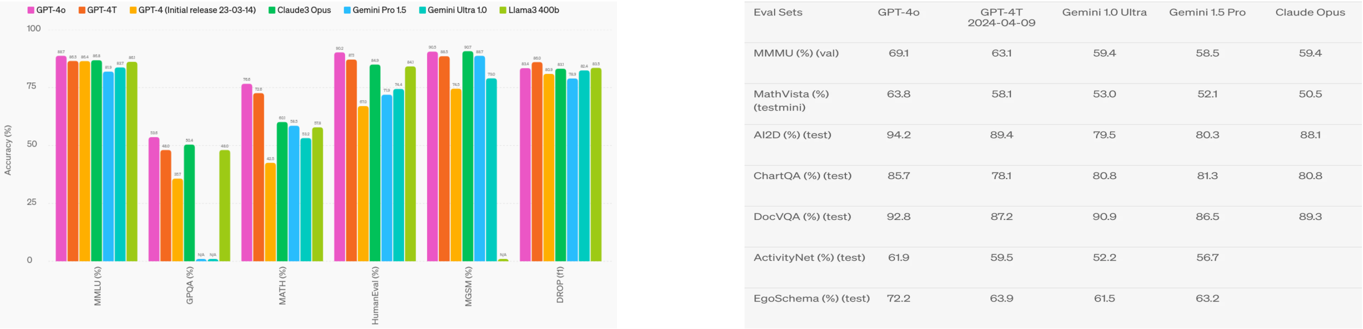
`OpenAI` [8] provided a performance comparison between its models and their counterparts. Figure 1a illustrates the text evaluation results, showing that GPT-4o achieved state-of-the-art (SOTA) performance across the MMLU (88.7%),

[3] https://github.com/gpt-open/rag-gpt

[5] All vision evals are 0-shot, with MMMU, MathVista, and ChartQA as 0-shot CoT

[5] recommanded to check original images for better view

[6] https://status.openai.com/



Eval Sets	GPT-4o	GPT-4T 2024-04-09	Gemini 1.0 Ultra	Gemini 1.5 Pro	Claude Opus
MMMU (%) (val)	69.1	63.1	59.4	58.5	59.4
MathVista (%) (testmini)	63.8	58.1	53.0	52.1	50.5
AI2D (%) (test)	94.2	89.4	79.5	80.3	88.1
ChartQA (%) (test)	85.7	78.1	80.8	81.3	80.8
DocVQA (%) (test)	92.8	87.2	90.9	86.5	89.3
ActivityNet (%) (test)	61.9	59.5	52.2	56.7	
EgoSchema (%) (test)	72.2	63.9	61.5	63.2	

(a) Text Evaluation

(b) Vision Understanding Evals[4]

Fig. 1: Performance comparison provided by OpenAI [8][5]

GPQA (53.6%), Math (76.6%), and HumanEval (90.2%) benchmarks. For the MGSM benchmark, Claude 3 Opus performs best (90.7%), followed by GPT-4o (90.5%). For the DROP benchmark, GPT-4T performed the best (86.0), followed by Gemini Ultra 1.0 (83.5) and GPT-4o (83.4). Figure 1b presents a performance comparison across various vision evaluation sets. Similar to the text evaluation set, GPT-4o consistently outperforms other models in most metrics, achieving SOTA in MMLU (69.1%), MathVista (63.8%), AI2D (94.2%), ChartQA (85.7%), DocVQA (92.8%), ActivityNet (61.9%), and EgoSchema (72.2%). GPT-4T follows closely but slightly lags behind GPT-4o. For the selected evaluation sets, it is clear from Figure 1 that GPT-4o demonstrates strong capabilities in these tasks, while Gemini 1.0 Ultra, Gemini 1.5 Pro, and Claude Opus show competitive but generally lower performance, with some exceptions in specific tasks.

OpenAI offers an evaluation framework called Evals[7] on GitHub. This framework provides tools for assessing LLMs and includes an open-source registry of benchmarks. It enables users to create and run evaluations using datasets to generate prompts, assess the quality of model outputs, and compare performance across various datasets and models.

GPT-4o	GPT-4T
Claude 3 Opus	Claude 3 Haiku
Gemini 1.5 Pro	Gemini 1.5 Flash
Mixtral 8x22B	Mixtral 8x7B
DBRX	Command-R+

TABLE II: LLMs for performance comparison

Artificial Analysis [33] provides tools to compare selected LLMs by calling APIs and generating comparison graphs. To further analyze the performance of the GPT-4o model, we selected 10 popular LLMs (listed in Table II) using these tools. Figure 2 illustrates a relative performance analysis of the listed LLMs based on throughput, response time, and latency for a single query and a 10k token prompt length. These graphs collectively highlight the strengths and weaknesses of each model.

Figure 2a shows the latency comparison, defined as the time to the first token received (in seconds) after an API request is sent. Mixtral 8x7B and Mixtral 8x22B exhibit the lowest latencies, at 0.59 and 0.81 seconds, respectively. GPT-4o has a lower-than-average latency, receiving the first token in 1.04 seconds.

Figure 2b depicts throughput, measured in tokens per second received while the model generates tokens (i.e. after the first chunk has been received from the API). Gemini 1.5 Flash and Claude 3 Haiku have the highest throughput, scoring 89 each, followed by Mixtral 8x7B with a score of 73. GPT-4o ranks fourth, with a throughput of 64 tokens per second, outperforming the average.

Finally, Figure 2c evaluates the total response time, defined as the time to receive a 100-token response, estimated based on latency and throughput. Mixtral 8x7B and Claude 3 Haiku have the fastest response times, at 2.2 and 2.3 seconds, respectively. GPT-4o has a response time of 2.7 seconds.

Overall, the analysis in Figure 2 indicates that GPT-4o performs better than most models but lags behind Mixtral 8x7B and Claude 3 Haiku.

IV. Application Domain

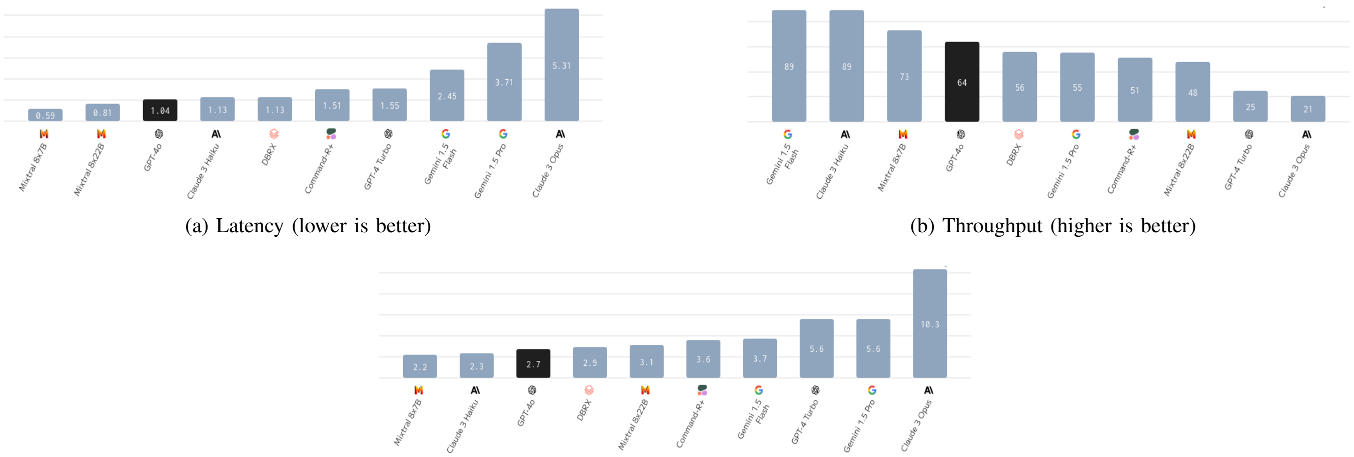
GPT-4o has the potential to significantly modernize and digitize the world by simplifying various aspects of human life. Its versatile applications span numerous fields, as detailed below, showcasing its substantial impact on a wide range of sectors.

a) Education: GPT-4o can guide students step-by-step to solve any math problems omitting the necessity of extra tutor [8] [30]. Additionally, it can be utilized in academic research by generating summaries of research papers, suggesting research topics, and providing insights from large datasets. These capabilities accelerate the research process and contribute to academic advancements by enabling researchers to quickly understand and explore extensive amounts of information.

b) Medical: GPT-4o can analyze medical images and patient data, aiding doctors in diagnosing diseases more accurately and quickly [29]. It enhances patient interaction by handling inquiries, providing information on medical conditions,

[7] https://github.com/openai/evals

[8] Generated from https://artificialanalysis.ai/models/gpt-4o/prompt-options/single/long#performance



(a) Latency (lower is better)

(b) Throughput (higher is better)

(c) Response time (lower is better)

Fig. 2: Performance comparison[8]

and managing appointment scheduling through chatbots. It can also support learners with visual impairments by offering both speech-to-text and text-to-speech solutions [34].

c) Customer Support: GPT-4o can be tailored to specific business needs, enabling AI-powered chatbots to handle banking queries, transaction details, and account management $24/7$, reducing the need for human intervention [29]. It supports advanced virtual assistants capable of effective responses, including visual data [25]. With real-time emotion analysis and multilingual capabilities, GPT-4o enhances global reach and customer satisfaction.

d) Finance: GPT-4o can analyze financial data and forecast market trends, enabling institutions to manage risks and make informed investment decisions [29].

e) Content Generation: GPT-4o can be used for creative and analytical purposes, i.e., making posters, creating visual representations from text input, designing characters, and producing outputs in different styles. It can also design logos, create artwork of people or objects, print text in different fonts, and generate 3D images.

f) Code Review: GPT-4o, can also review code effectively. It can identify the appropriate notes and comments while analyzing the code [6] [35]. Moreover, if provided with a picture of a desktop displaying the code, GPT-4o can detect issues or problems within the code [12].

V. Conclusion

GPT-4o represents a significant advancement in the field of AI, offering substantial improvements over its predecessors in terms of multimodal capabilities, context window size, tokenization efficiency, and processing speed. By integrating advanced technologies like refined RLHF and utilizing powerful hardware such as `Nvidia`'s GPUs, GPT-4o achieves remarkable performance in understanding and generating text, audio, video, and images. Its applications span various domains, including education, medicine, customer support, finance, and content generation, showcasing its versatility and potential to modernize and digitize numerous aspects of human life.

However, despite these advancements, GPT-4o still faces challenges such as ensuring ethical AI use, protecting user data, and overcoming technical issues like system outages. Addressing these challenges will be crucial for maximizing the model's benefits and maintaining user trust.

References

[1] T. Wu, S. He, J. Liu, *et al.*, "A brief overview of chatgpt: The history, status quo and potential future development," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp. 1122–1136, 2023. DOI: 10.1109/JAS.2023.123618.

[2] Y. Xu, L. Hu, J. Zhao, Z. Qiu, Y. Ye, and H. Gu, *A survey on multilingual large language models: Corpora, alignment, and bias*, 2024. DOI: 10.48550/arXiv.2404.00929.

[3] Y. K. Dwivedi, N. Kshetri, L. Hughes, *et al.*, ""So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *International Journal of Information Management*, vol. 71, p. 102 642, 2023.

[4] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, 2023.

[5] H.-W. Cheng, "Challenges and limitations of ChatGPT and artificial intelligence for scientific research: a perspective from organic materials," *AI*, vol. 4, no. 2, pp. 401–405, 2023.

[6] S. M. Kerner, *GPT-4o explained: Everything you need to know*, https://www.techtarget.com/whatis/feature/GPT-4o-explained-Everything-you-need-to-know, [Accessed 26-05-2024], May 2024.

[7] GeeksforGeeks, *OpenAI Launches GPT-4o Omni*, https://www.geeksforgeeks.org/openai-announces-gpt-4o-omni, [Accessed 25-05-2024], May 2024.

[8] OpenAI, *Hello GPT-4o*, https://openai.com/index/hello-gpt-4o/, [Accessed 10-06-2024], May 2024.

[9] K. Doyle, *'The "o" is for omni' and other things you should know about GPT-4o*, https://www.jasper.ai/blog/what-is-gpt-4o, [Accessed 26-05-2024], May 2024.

[10] R. Montti, *OpenAI Announces GPT-4o Omni*, https://www.searchenginejournal.com/openai-announces-chatgpt-4o-omni/516189/, [Accessed 25-05-2024], May 2024.

[11] M. Zeff and Gizmodo, *OpenAI's new ChatGPT sounds more human than ever*, https://qz.com/openai-new-chatgpt-gpt4-omni-voice-human-ai-1851475246, [Accessed 26-05-2024], May 2024.

[12] K. Wiggers, *OpenAI debuts GPT-4o 'omni' model now powering ChatGPT*, https://techcrunch.com/2024/05/13/openais-newest-model-is-gpt-4o/, [Accessed 26-05-2024], May 2024.

[13] K. Gomez, *Harnessing the Power of GPT-4 Omni for Multimodal Processing: A Comprehensive Guide*, https://medium.com/@kyeg/harnessing-the-power-of-gpt-4-omni-for-multimodal-processing-a-comprehensive-guide-9301ae001576, [Accessed 24-05-2024], May 2024.

[14] *Introducing Apple Intelligence, the personal intelligence system that puts powerful generative models at the core of iPhone, iPad, and Mac*, https://www.apple.com/newsroom/2024/06/introducing-apple-intelligence-for-iphone-ipad-and-mac, [Accessed 11-06-2024], Jun. 2024.

[15] *Introducing GPT-4o: our fastest and most affordable flagship model*, https://platform.openai.com/docs/models/gpt-4o, [Accessed 06-06-2024], 2024.

[16] M. Schreiner, *GPT-4 architecture, datasets, costs and more leaked*, https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked, [Accessed 07-06-2024], Jul. 2023.

[17] T. B. Brown, B. Mann, N. Ryder, *et al.*, *Language models are few-shot learners*, 2020. arXiv: 2005.14165 [cs.CL].

[18] OpenAI, *Pricing*, https://openai.com/api/pricing, [Accessed 06-06-2024], 2024.

[19] D. Sheremetov and A. Bitkina, *OpenAI API Pricing 2024: Understanding GPT-3 Pricing In-Depth*, https://onix-systems.com/blog/how-much-does-it-cost-to-use-gpt-models, [Accessed 07-06-2024], Jun. 2023.

[20] C. Lang, *ChatGPT's Architecture - Decoder Only? Or Encoder-Decoder?* https://datascience.stackexchange.com/questions/118260/chatgpts-architecture-decoder-only-or-encoder-decoder, [Accessed 07-06-2024], Dec. 2023.

[21] OpenAI and J. A. et al., *GPT-4 Technical Report*, 2024. arXiv: 2303.08774 [cs.CL].

[22] N. Bijani, *What are the differences between GPT, GPT3, GPT 3.5, GPT turbo GPT 4?* https://www.codiste.com/what-difference-between-gpt-gpt3-gpt-3-5-gptturbo-gpt-4, [Accessed 08-06-2024], Oct. 2023.

[23] T. Pungas, *GPT-3.5 and GPT-4 response times*, https://www.taivo.ai/__gpt-3-5-and-gpt-4-response-times, [Accessed 09-06-2024], May 2023.

[24] E. Eckert, *GPT-4 vs GPT-4o: The Ultimate AI Smackdown!* https://supernormal.com/blog/gpt-4-vs-gpt-4o, [Accessed 09-06-2024], Jun. 2024.

[25] C. D. Under, *OpenAI GPT-4o: The Next Generation of Omni-Multimodal AI*, https://medium.com/@cognidownunder/openai-gpt-4o-the-next-generation-of-omni-multimodal-ai-e8d64d211a2c, [Accessed 26-05-2024], May 2024.

[26] R. LeFebvre and J. Ledford, *OpenAI Reveals New GPT 4o: AI For Everyone*, https://www.lifewire.com/openai-reveals-gpt-4o-8647637, [Accessed 06-06-2024], May 2024.

[27] M. Rajguru, *Exploring the New Frontier of AI: OpenAI's GPT-4-o For Indic Languages*, https://techcommunity.microsoft.com/t5/ai-azure-ai-services-blog/exploring-the-new-frontier-of-ai-openai-s-gpt-4-o-for-indic/ba-p/4142383, [Accessed 25-05-2024], May 2024.

[28] *Introducing GPT-4o: OpenAI's new flagship multimodal model now in preview on Azure*, https://azure.microsoft.com/en-us/updates/new-openai-model-on-azure/, [Accessed 06-06-2024], May 2024.

[29] V. K. Upadhyay, *OpenAI : GPT-4o (Comprehensive Guide)*, https://vivekupadhyay1.medium.com/openai-gpt-4o-comprehensive-guide-ccf15fd93870, [Accessed 06-06-2024], May 2024.

[30] A. Toolz, *15 Abilities of GPT4o that you Won't Believe*, https://aitoolzai.medium.com/15-abilities-of-gpt4o-that-you-wont-believe-8cba07c1cf1f, [Accessed 31-05-2024], May 2024.

[31] OpenAI, *Data analysis with ChatGPT*, https://help.openai.com/en/articles/8437071-data-analysis-with-chatgpt, [Accessed 04-06-2024], Jun. 2024.

[32] OpenAI, *Memory and new controls for ChatGPT*, https://openai.com/index/memory-and-new-controls-for-chatgpt, [Accessed 04-06-2024], Feb. 2024.

[33] *Artificial Analysis*, https://artificialanalysis.ai, [Accessed 05-06-2024], 2024.

[34] A. Kuzdeuov, O. Mukayev, S. Nurgaliyev, A. Kunbolsyn, and H. A. Varol, "Chatgpt for visually impaired and blind," in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2024, pp. 722–727. DOI: 10.1109/ICAIIC60209.2024.10463430.

[35] D. Eastman, *Reviewing Code With GPT-4o, OpenAI's New 'Omni' LLM*, https://thenewstack.io/reviewing-code-with-gpt-4o-openais-new-omni-llm/, [Accessed 26-05-2024], May 2024.