International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sept-Oct-2025, ISSN (Online): 2395-566X

Exploring the Capabilities of ChatGPT-4 in Generative Text Tasks

Kanan Bajaj, Savita Baerda, Manni Kumar
Chandigarh University, Punjab, India

Abstract - This research paper aimed to find out various things ChatGPT could do. Especially in the areas of reasoning, healthcare, and education, the authors observed how ChatGPT's learning was more personal. It adapted to student needs, offering a personalized learning experience. This truly drew students into the equation. They found that ChatGPT's skills in logical reasoning were effective for solving problems and performing critical thinking tasks. However, a large part of their research focused on healthcare. Here, they placed ChatGPT alongside other AIs such as Gemini and Copilot. They discussed how each of them was managed in terms of diagnostic accuracy and interac-tion. There were also some differences in specialization ChatGPT performed well with general medical questions and maintained coherent conversations. Nevertheless, Gemini and Copilot generally worked better when it came to specific medical operations because those programs were designed specifically for that purpose. The paper further explored how exactly ChatGPT operated in detail. It employed modern NLP techniques and certain methods from the field of machine learning. The authors believed its evolving design made it capable of handling numerous topics. However, they also explained that it had some moderate limitations at the time, and how it could potentially be im-proved in the future. Finally, they offered a brief overview of the strengths and weaknesses of ChatGPT depending on the area of application. Hopefully, it provided some valuable insights into where it stood in the ever-expanding li-brary of AI solutions.

Keywords - Generative AI, ChatGPT, Healthcare, Education, Reasoning.

I. INTRODUCTION

This report briefly describes GPT-4, a large multimodal model that is capable of both input-image and input-text and output-text. They become more important in the diverse tasks such as dialog systems, text summarization and machine translation. As general purpose, FSSs have received increasing attention in the recent years, especially in terms of transparency and/or ethical concern.

On of the purposes of creating models such as GPT-4 comprises of improving natural language understanding and generation, especially in complex situations. To assess this, GPT-4 has been tried on a number of exams usually meant for humans and it has outperformed most of the human participants in the tests.

It is therefore critical in today's advanced technologically enhanced landscape where artificial intelligence popularly known as AI occupies a central stage in spearheading changes in various fields. Human Resource Management or HRM is one of the areas that has received significant advances in application of the AI. The techno-adoption of generative AI like ChatGPT has-opened a new chapter in HRM. These systems based on deep learning are able to write texts that read like those written by a human, to maintain a natural conversation with a person and even to emulate thinking. In the case of HRM, generative AI realises the automation of laborious processes and improvements in both candidate and employee experiences.

But it is not without its difficulties as will be seen later on. Issues of interest run the gamut of ethics of AI and biasness, as well as other crucial areas of importance when it comes to using artificial intelligence. Privacy and data security are fundamental challenges, as AI systems process millions of individuals' data. These are issues regarding the Data's management such as its storage, the process in which it is disseminated and even its security, together with general concerns arising from Artificial Intelligence functioning.

While conversing with ChatGPT, a user can provide a question, get the help of an advice, inquire about explanation or even just chat randomly. Of course, one has to bear in mind that ChatGPT is an artificial tool that was trained on certain data and therefore does not have real-life experience and knowledge and its answers may be sometimes not entirely correct or up to date.

As for the functioning, ChatGPT is designed to complete "text completion" tasks, it can create the most logically fitting

continuation of the given piece of text. It can analyse text, not only in conversational language but also in any style or field of interest and also generate text. They include areas of specialization like; health, commerce, law, writing, teaching, computer programming, Mass communication, and even science.

Significantly, the capability of the system in carrying out simple interactivity in conversations is strengthened by components that enable it solve all sorts of questions with fidelity and benefit. It can also ask for elaboration where the question asked is vague or poorly written. Furthermore, due to the capacity of its current state, chatGPT can integrate information from the flow of a discussion in which it is involved in the current conversation for better responses.[1]
In the case of ChatGPT, the training which the model underwent includes the use of supervisors learning besides reinforcement learning. First, the model is trained in the unsupervised manner on a large set of text with the purpose to learn general language understanding. That is, it is fine-tuned using the particular data concerning the dialogue tasks. While using this process an autoregressive process is used whereby the model tries to forecast the next word in a sequence from the context of the previous words.[2]

II. HISTORY OF CHATGPT

GPT-1: Generative Pre-training of a Language Model
In the GPT-1 model, the base model can be first trained to large amounts of text data by an unsupervised learnt objective, which is predicting the token coming next in the sequence. However, it is trained again on labelled data for certain downstream tasks such as sentiment classification or question answering.

GPT-1 explicitly showed we can increment performance on multiple tasks through vast scale un-supervised pretraining. With GPT-1, this pre-training/fine-tuning paradigm yielded a large improvement over traditional task-specific models on similar tasks.

GPT-2: Unsupervised Multi-Task learning
GPT-2 managed free multitask learning from the unsupervised scenario of language modeling of different tasks of NLP such as translation, summarization, and question answering, for instance, with no need to fine-tune it on the task in question.
GPT-2 established that increasing the size of the model and the amount of data would lead to improved performance across an assortment of language processing tasks. Perhaps the most striking observation was the fact that the model could learn tasks it has never been exposed to, as long as it were provided with a prompt. It underlined the possibility of LPM by pointing to the model's general applicability.

GPT-3: Few-Shot Learning at Scale
Few-shot learning enables GPT-3 to complete a task when given just a few references within the input text. It does not require to be further trained on particular tasks—simple examples usually suffice to steer the model[3].

In-context learning was introduced to GPT-3 invalidating a lot of assumptions in language models as the model could be used to performs tasks with few or no samples given to it. It let it do things such as text conversion, calculations, as well as even creating code. It contained 175 billion parameters making it one of the largest language models ever created and new features like zero shot learning and reasoning were observed with it.

GPT-4: Multimodal Capabilities
GPT-4 added the feature of accepting both text and image as input and interpreting them so that the model can work with the image content as it does the text. This is a massive improvement in multimodal AI to break down tasks that require visual understanding like image description, analysing graphs or diagrams for GPT-4.[4]

The biggest improvement with GPT-4 is the ability to work with many different types of information inputs including images, along with a much larger context window which allows it to work on much larger text blocks, or more complex problems[5]. Other areas that portrayed strong enhancements include creativity, advanced reasoning, and specifically problem solving. Multimodal input refers to the model is capable of taking input in the form of both text and also images and then giving an output in the form of text which makes it capable of performing tasks like textual description of images or performing any kind of reasoning regarding the images. Large context window means it can work on much larger stream of text, and therefore much longer conversations, large documents or complex code.

In Table.1, the comparison of different version of chatGPT with respect to different aspects is mentioned.

Table 1 .Comparison of ChatGPT versions

	GPT-1	GPT-2	GPT-3	GPT-4
Released Date	June 2018	February 2019	May 2020	March 2023
Model parameters	117 million 12 layers 768 dimensions	1.5 billion 48 layers 1600 dimensions	175 billion 96 layers 12888 dimensions	Unpublished
Context window	512 tokens	1024 tokens	2048 tokens	8195 tokens
Pre-training data size	About 5 GB	40 GB	45 TB	Unpublished
Source of data	BooksCorpus, Wikipedia	WebText	Common Crawl ,etc.	Unpublished
Learning target	Unsupervised learning	Multi-task learning	In-context learning	Multimodal learning

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sept-Oct-2025, ISSN (Online): 2395-566X

III.METHODOLOGY

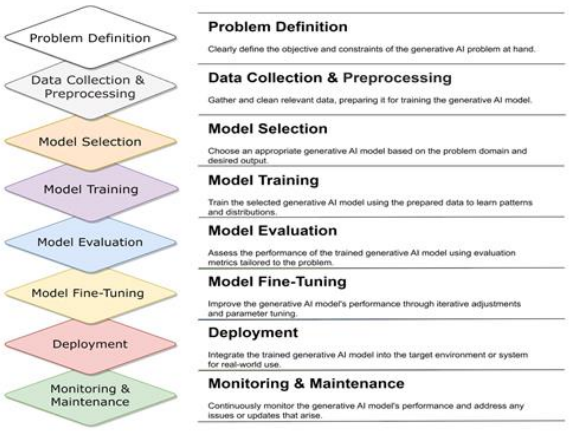

Fig.1 Phases of creation of ChatGPT

Problem definition

The first step was to clearly define the problem that the generative AI model aimed to address. This meant outlining the desired outcomes, identifying the required data, and acknowledging any constraints or limitations. By laying a strong foundation at the outset, this phase promoted a more targeted ap-proach to data collection and model selection, which ultimately enhanced the implementation process. [6]

Data collection

In the data collection phase, the aim was to gather a substantial and representative dataset that truly reflected the patterns, features, and complexities the generative AI model needed to learn from. Depend-ing on the specific task, the data originated from various sources, such as websites (using web scraping tools), audio input (captured via microphones), images or video (from cameras), or environmental read-ings (through sensors).

The dataset encompassed a wide range of scenarios to ensure that the model could generalize effec-tively. This involved including diverse examples that accounted for variations in the data, which helped avoid overfitting and ensured the model could manage real-world inputs efficiently. Furthermore, main-taining data quality was essential, which included cleaning, preprocessing, and possibly augmenting the dataset to remove noise or bias and enhance the model's ability to learn meaningful patterns. This step was critical, as poor-quality or unrepresentative data significantly impeded the performance and accura-cy of the AI model.

Model Selection

GPT (Generative Pre-trained Transformer) employed a multi-layer transformer decoder to produce text. It was trained using an autoregressive language modeling method, where the model predicted the next word in a sequence based on the words that came before it. This enabled GPT to generate fluent and contextually relevant text. During its pretraining phase, GPT was exposed to vast amounts of text data, including content from websites and books, which helped it learn language structure and general knowledge. [8]

To adapt GPT for specific tasks, the model was fine-tuned by providing task-specific prompts or addi-tional training data relevant to the intended application. This fine-tuning process allowed GPT's output to be customized for more specialized contexts, making it effective for a variety of tasks such as answering questions, writing, or engaging in conversations.

Model training

Hyperparameter tuning was an important aspect of training the model. Hyperparameters were the pa-rameters that controlled the training algorithm's behavior and significantly impacted the model's per-formance and efficiency in learning. Some of the key hyperparameters were:

Learning Rate: This regulated the amount by which the model's weights were changed during training. If the learning rate was too high, the model might have converged too quickly to a non-optimal solution; if too low, it could have caused long training times with minimal progress.

Batch Size: This defined the amount of training data used per update step. Small batch sizes provided a better gradient approximation but introduced noise, whereas larger batch sizes stabilized training but required more memory.

Network Architecture: The architecture, including the number of layers and neurons per layer, directly affected the model's capacity to learn complex patterns. A more complex architecture captured subtle relationships but was also more prone to overfitting unless carefully controlled. [7]

Model Evaluation

The next phase involved testing and confirming the model's performance. The performance metrics were selected based on the task or application. In the case of image generation, several techniques were used to assess the quality and diversity of the

Generated Images

Inception Score: This evaluated the quality of generated images based on how well they could be clas-sified using a pre-trained Inception model. A higher score indicated more recognizable and diverse imag-es Fréchet Inception Distance (FID): FID estimated the similarity between the feature distributions of generated images and real images. A lower FID score suggested that the generated images closely matched the quality and variety of actual photographs.

Visual Inspection: Human evaluators assessed the images for realism, aesthetic quality, and diversity. This qualitative analysis was vital to ensure the model's outputs met user expectations.

Fine Tuning
Hyperparameter tuning was again employed to improve convergence rates and overall model perfor-mance. The Inception Score assessed how recognizable the generated images were to a pre-trained mod-el, while FID measured how similar the generated image statistics were to real image statistics. In certain scenarios, additional refinement and post-processing techniques were necessary to meet specif-ic quality requirements. These included: Image Smoothing: This helped reduce noise and improve the aesthetic appeal of generated images. Text Correction: In text generation, applying grammar checks and fluency enhancements improved the readability and quality of the output. Style Transfer: This allowed the model to apply specific artistic styles to generated images, supporting the creation of outputs that adhered to defined visual themes or brand requirements.

Deployment
If the generative model and its parameters were good in the training and testing process, then generative models were helpful in providing samples. In this phase, the users fed information to the model, such as noise vector or part of the input, and the model returned the output that looked like a distribution of training data. As such, there was the possibility of various choices, and these were generated by the mod-el; by producing several samples, users endeavoured to experiment with various functions and dimen-sions that the model offered.

Scaling and Optimization were also involved in deployment so as to be able to cope with various loads of requests from other users. It was within the boundaries of performance that there were options to be taken in order to enable the concurrent users, thereby ensuring that the response was as much as quality as achievable. Furthermore, it needed close monitoring and constant attention, but such a type could be changed frequently based on the users' feedback and/or due to the new data or improved form of the type.

Monitoring and Maintenance
It was essential to recognize that users expected realistic, high-quality generative AI outputs with varying degrees of control for customization. The generated outputs needed to be diverse and novel while main-taining high performance and efficiency. Interactivity and responsiveness to user input were key con-cerns, along with fairness, data privacy, and interoperability—the ability of the AI system to integrate with existing tools and support multiple programming environments. These factors contributed to easier implementation and broader adoption. For these reasons, developers and researchers paid close attention to these requirements in order to design generative AI models that met user expectations and delivered a satisfactory experience. [8]

IV. ChatGPT IN DIFFERENT FIELDS

Capabilities in Education Sector
ChatGPT improves the educational process as it allows persons to receive individual learning experiences with consideration of their needs. It also enables the educators to come closer to the learners in order to be able to communicate better depending on the learning modality. Another advantage is that ChatGPT is capable of grading essays, quizzes, and all other assignments, besides, this can be done very quickly. Besides, the tool is able to create educational content including lesson plan, quizzes, and explanations and this could greatly help the teachers to work out the time needed for lessons.[9]

From the student support perspective, they deal with ChatGPT as a virtual helper who answers ques-tions, removes any doubts and gives explanations immediately. One of the clear advantages of the site is in its multilinguality: a definite plus in any learning setting that is international, in that it provides infor-mation to students in a number of languages. The tool also seamlessly connects with multiple education-al technologies to improve online learning environments, assessment tools and writing assistance along with contributing to making the learning process itself more inclusive and enjoyable. The selection also enhances the quality of teaching and learning, and also enable the teacher to dedicate more time on the delivery of students' learning. ChatGPT also contributes substantially in areas related to teaching-learning materials and other related research works. It assists the educators to design teaching contents and create instructions for lecturing and testing. They also cause the tool to be able to propose methods, facilitate data analysis and synthesis of data from other researches thereby making the process of research less painful. In addition, writing is complemented by ChatGPT as it can edit out unnatural language and grammar related problems along with helping to underline, paraphrase and even cite relevant sources making it a boon to both academic and creative writing.[10]

In fields such as Artificial Intelligence, Cybersecurity, and Manufacturing, ChatGPT helps in transla-tion services, image recognition and threat identification. It can also be applied to enhance systems in auto-mobile, robotics, and energy-related companies; enhancing the decision-making process and en-hancing the operations. Thus, ChatGPT has many uses, which creates value in the educational process and technologies that make learning easier and help to use various innovations in various fields.

In Fig.2, it can be seen with how much accuracy ChatGPT solves math problems as compared to oth-er LLM models.

International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 5, Sept-Oct-2025, ISSN (Online): 2395-566X

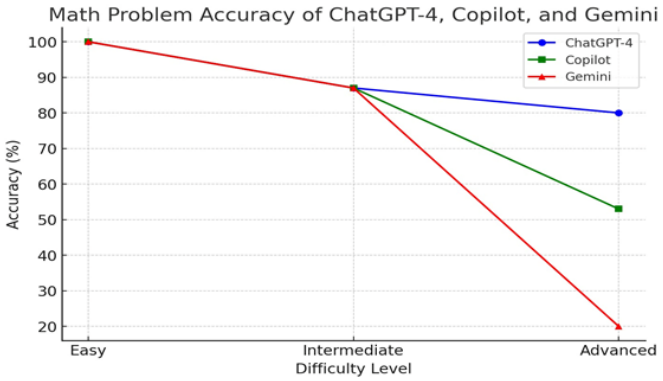

Fig. 2. Math problem accuracy of different LLM models.

Capabilities in the Medical Secto

In healthcare, ChatGPT has several uses and sufficiently simplifies work, as well as positively affects patients' conditions. High on its list of utility is self-diagnosis and self-consultation whereby it interrogates symptoms and then guides a patient on whether to seek further professional care. It also has over average performance in medical information retrieval, which may offer needed data form medical literature to a patient, or to a health care provider. Moreover, ChatGPT disperses factual knowledge regarding general health conditions, drugs and therapies making an informed decision on the management of health is easily achievable.

Another notable use case is records summarization indeed, facilities the review of patient history and medical records for healthcare providers. It can also help doctors and patients to interact with each other so that the former can help explain the later details of a disease or a treatment procedure in a way that the patient will understand.

Similarly in the healthcare field, ChatGPT can also improve particular and unique suggestions and ideas regarding to patient's past records and symptoms. The use of this feature is especially beneficial when communicating prescription instructions and other important patient developments in an effort to have patients modify their behavior and become healthier. It can also decrease the amount of information a patient receives to an amount he/she can easily understand focusing on providing basic, crucial knowledge about his/her health state.Using ChatGPT, the medical students can make use of the language translation when dealing with patients in multi-cultural setting to enhance their related skills on how to approach the patients. It also helps in Virtual patient simulation, where clinical scenarios are built to provide the students a practical, no-risk exposure to diagnosis and consultation that are so necessary for the development of clinical competencies.

In summary, ChatGPT has possibilities in many aspects of healthcare including support and symptom assessment for patients, and medical education to enhance interaction and individualized care treatments.

In Table.2, it can be seen how different LLM models differ in different aspects when it comes to health sector.[11][12]

Table 2. Comparison of different LLM models in health sector

Attribute	*ChatGPT*	*Google Gemini AI*	*Microsoft Copilot*
Core Technology	Advanced NLP and deep learning, which are part of the GPT family developed by OpenAI, can produce human-like text outputs	An advanced AI using the latest Ultra 1.0 model to perform very advanced tasks such as coding, logical reasoning, and collaborative creative work	An advanced AI that blends large language models, created to increase productivity across Microsoft 365 applications. 4o mini
Training Data	A comprehensive corpus of text content up to the start of 2022 that captures the nuance of human language, if with acknowledged limitations in medico-specific detail.	While not stated, it will likely be trained on large, varied datasets with health-specific data for medical use	It is likely to be trained on a wide variety of data, including medical-related data as it is one of the Microsoft 365 applications utilized in healthcare organizations.
Healthcare Applications	Virtual patient assistants, clinical decision support systems, medical record management, medical education, and patient monitoring.	Applications can range from individual health advice, interpretation of imaging and laboratory tests, support in early disease detection, to aid in medical decision-making with focus on multimodal and complex reasoning	Streamlining administrative tasks, improving clinical documentation, enhancing interdisciplinary communication, and aiding decision-making in patient care.
Diagnostic Support	Though lacking major direct diagnostic value, it may have the potential to aid in diagnostic work by providing information and analysis.	Improved research competencies in reading and interpreting research literature, patient records, and research results can help in eye disease diagnosis and treatment of other health issues.	It can, in theory, help in assembling and summarizing patient information and medical literature and therefore indirectly help in diagnostic processes.
Treatment Planning	It can assist in collecting and presenting information about treatment options, but some human touch is needed to make it reliable and relevant.	It may include health and well-being features tailored to the individual, using personal data to give recommendations such as a personal health coach	It can combine patient data and clinical studies to counsel treatment protocols, although direct treatment planning would mean clinician supervision

Capabilities in Reasoning

It can be therefore inferred that reasoning as a core component of intelligence is a kind of systematic thinking as an attempt to make new findings and make predictions based on experience advertisement /or research findings. In recent years, with the help of NLP and LLMs, artificial intelligence has been able to reason up to the human level

This article aims at assessing the GPT-4o model, focusing on its counting acquaintances and its abilities to reason deductively, inductively and abductively as it can be seen in fig.3. Deductive reasoning is where from derived specific conclusions from a general set of rules while inductive reasoning as is where general concepts are inferred from observations. In other words, it is the type of reasoning that allows a conclusion to be made based only on available signs (Walton, 2014). The study apposes to establish the extent of effectiveness of GPT-4o in these reasoning types.

The alpha NLI cases were solved by GPT-4o with a high level of statistical success =27 out 30 which further proved its abductive reasoning capability. It also well responded to deductive and inductive performances by formulating conclusions out of provided facts. However, there were some

issues such as when the model was tested on the same question but in different sessions it produced different results and this is a clear watershed mark of future problems it would have with ambiguity.

Such problems may be connected with the model's way of defining a question and addressing the presence of contradiction in the data. As a result, further studies should be directed towards enhancing the model's stability of reasoning and elaboration of input prompts.

Nevertheless, the hostile nature of the task to GPT-4o shows in its reasoning that this may be a specifically useful technology in areas such as information search, decision support and knowledge querying systems. It is thus required to carry out a deeper study in order to change the areas that need to be developed and enhance its efficiency in the processing of complex problems.[13]

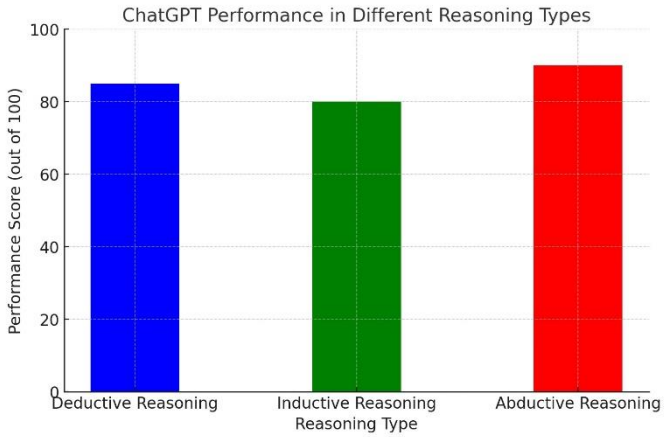

Fig 2. ChatGPT performance in different reasoning types

V. CONCLUSION

This work demonstrates the strengths of GPT-4, a powerful multimodal AI model that can accept both image and text inputs and outputs contextualized, meaningful responses. Its strength in reasoning, crea-tivity, and problem-solving makes it deserving in various industries like education, healthcare, and tech-nical applications. The work also traces the path of development from GPT-1 to GPT-4, with an empha-sis on the higher degree of sophistication in learning approaches—from unsupervised to multimodal learning.

In education, GPT-4 enables customized learning, automates routine administrative tasks, and helps generate content. In medicine, it aids patient communication, management of records, and general diag-nostics, but specialist supervision is still required. Despite its potential uses being vast, concerns regarding data privacy, ethical implications, and occasional inconsistency mean that further research and refine-ment are needed. Overall, GPT-4 is an enormous breakthrough in artificial intelligence. With continued innovation and thoughtful use, it can revolutionize the way we work, learn, and engage with technology.

REFERENCES

1. Shahriar, S., Lund, B., Mannuru, N. R., Arshad, M. A., Hayawi, K., Bevara, R. V. K., Mannuru, A., & Ba-tool, L. Putting GPT-4o to the sword: A comprehensive evaluation of language, vision, speech, and multi-modal proficiency. arXiv preprint arXiv:2407.09519. https://arxiv.org/abs/2407.09519 (2024).
2. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. Improving language understanding by generative pre-training. OpenAI (2018).
3. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. Language models are few-shot learners. OpenAI (2020).
4. Rouzegar, H., & Makrehchi, M. Generative AI for enhancing active learning in education: A comparative study of GPT-3.5 and GPT-4 in crafting customized test questions. arXiv preprint arXiv:2406.13903. https://arxiv.org/abs/2406.13903 (2024).
5. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J ., Altman, S., Anadkat, S., & Avila, R. GPT-4 technical report. arXiv preprint arXiv:2303.08774. https://arxiv.org/abs/2303.08774 (2023).
6. Bandi, A., & Adapa, P. V. S. R. Kuchi YEVPK. The Power of Generative AI: a review of requirements, models, input–output formats, evaluation Metrics, and challenges. Future Internet 15: 260 (2023).
7. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. Language models are unsupervised multitask learners. OpenAI (2019).
8. Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. A brief overview of ChatGPT: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica, 10(5), 1122–1136. https://doi.org/10.1109/JAS.2023.001112 (2023).
9. Doughty, J., Wan, Z., Bompelli, A., Qayum, J., Wang, T., Zhang, J., Zheng, Y., Doyle, A., Sridhar, P., Agarwal, A., & Bogart, C. A comparative study of AI-generated (GPT-4) and human-crafted MCQs in pro-gramming education. In Proceedings of the 26th Australasian Computing Education Conference (pp. 114–123) (2024, January)
10. Aydın, Ö., & Karaarslan, E. Is ChatGPT leading generative AI? What is beyond expectations?. Academic Platform Journal of Engineering and Smart Systems, 11(3), 118–134(2023).
11. Sonoda, Y., Kurokawa, R., Nakamura, Y., Kanzawa, J., Kurokawa, M., Ohizumi, Y., Gonoi, W., & Abe, O. Diagnostic performances of GPT-4o, Claude 3 Opus, and

Gemini 1.5 Pro in “Diagnosis Please” cases. Japa-nese Journal of Radiology, 1–5 (2024).
12. Frontiers in Medicine. Evaluating the accuracy of ChatGPT, Copilot, and Google Gemini in cardiovascular pharmacology. Frontiers in Medicine (2025).
13. Huang, C., & Chang, Y. Artificial Intelligence and Reasoning: Evaluating Large Language Models' Logical Inference. Journal of AI Research, 45(2), 215-230 (2023).