

Evaluating the Efficacy of Generative AI Tools: A Comparative Analysis of ChatGPT and Other Models

Abeer A. Amer^{1*}, Heba S. Konswah²

¹Department of Computer Science and Information Systems, Faculty of Management Science, Sadat Academy for Management and Sciences, Cairo, Cairo Governorate, Egypt.

²Department of Information Systems and Computers, Faculty of Business, Alexandria University, Alexandria, Alexandria Governorate, Egypt.

abamer.2000@gmail.com¹, heba.konswah@alexu.edu.eg²

*Corresponding author

Abstract: Because these tools are being developed at such a rapid pace, it is essential to have a complete understanding of the capabilities and limitations of technologies that make use of generative artificial intelligence. To produce natural-language responses to human inquiries that are relevant to the context in which they are asked, technologies that use generative artificial intelligence, such as ChatGPT, are developed. In light of this, the objective of this article is to present a comparative analysis of the most important artificial intelligence tools for generative applications. Within the context of the paper, ChatGPT and its performance compared to Bard, Claude, and Gemini receive some attention. In order to test these models across a wide range of important activities, including code generation, natural language interpretation, and ethical reasoning, researchers use a predetermined set of difficult prompts. The outcomes of our research indicate considerable disparities in performance across these activity categories. These modifications highlight the advantages and disadvantages of each model, as well as potential ways to further grow.

Keywords: Artificial Intelligence (AI); Generative AI; Natural Language Processing; AI Tools; Comparative Study; Open AI; Chatbot and ChatGPT; Recurrent Neural Networks; Large Language Models.

Cite as: A. A. Amer and H. S. Konswah, "Evaluating the Efficacy of Generative AI Tools: A Comparative Analysis of ChatGPT and Other Models," *AVE Trends in Intelligent Computing Systems*, vol. 2, no. 3, pp. 174–183, 2025.

Journal Homepage: <https://www.avepubs.com/user/journals/details/ATICS>

Received on: 03/12/2024, **Revised on:** 24/03/2025, **Accepted on:** 05/05/2025, **Published on:** 09/09/2025

DOI: <https://doi.org/10.64091/ATICS.2025.000200>

1. Introduction

Generative AI model architectures have undergone significant advancements, moving from early Recurrent Neural Networks (RNNs) to the introduction of the Transformer architecture, which utilises a self-attention mechanism [1]; [2]. This transformation has enhanced the performance of large language models (LLMs), enabling them to produce highly fluent, coherent text and to perform complex tasks with a deeper understanding of context [3]. However, the high computational requirements for training and deploying these large transformer models present substantial challenges, leading researchers to investigate more efficient architectures and training techniques [4]. Several studies have compared the performance of different generative AI models [6]; [7]. These studies often focus on specific tasks, such as text generation, question answering, or machine translation [5]. However, comprehensive comparative analyses that evaluate multiple models across a broader range

Copyright © 2025 A. A. Amer and H. S. Konswah, licensed to AVE Trends Publishing Company. This is an open access article distributed under [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which allows unlimited use, distribution, and reproduction in any medium with proper attribution.

of tasks are relatively rare. Existing comparative studies often lack the basic methodology to provide an unbiased comparison of model performance. This research aims to address this gap by employing a comprehensive evaluation framework and a wider range of tasks.

Additionally, the ethical implications of generative AI have attracted increasing attention, with particular concern over biases in training data, the risk of generating harmful content like misinformation and hate speech, and the potential effects on human creativity and job security [8]; [9]. Current studies aim to address these challenges by developing frameworks to assess and improve the ethical performance of generative AI models, improve transparency, and promote responsible use of these technologies [10]. Alongside this focus, there is a growing interest in multimodal generative AI models that can combine different data types such as text, images, and audio, offering more immersive experiences. Still, they also present new challenges in data acquisition, model training, and ethical considerations [11]. This paper contributes to the research in this area by evaluating the multimodal capabilities of leading generative AI tools. In this study, researchers will conduct a comparative analysis to assess ChatGPT's performance against other similar generative AI tools. This will be evaluated by asking these generative AI tools to respond to several questions that assess various aspects of each tool, including coding proficiency, Natural language understanding, creative text generation, reasoning and problem-solving, ethical reasoning, knowledge retrieval, translation, and cultural awareness.

1.1. Background

1.1.1. Natural Language Processing, Natural Language Understanding, and Natural Language Generation

Today, with the advent of the social web, millions of people can share their ideas, opinions, and content. This large amount of information is usually unstructured and not machine-processable. Here is the need for machines to develop a deep understanding of unstructured natural language, which drives Natural Language Processing (NLP) [12]. Advancements in technology and research created the need for NLP and NLU (Natural Language Processing and Natural Language Understanding). While NLP focuses on information extraction through tokenisation and named-entity recognition techniques, NLU's main concern is understanding the text's context by analysing its semantics to identify its aims [13].

Semantic, syntactic, and pragmatic analysis help understand the text's objective. On the other hand, the term NLG (Natural Language Generation) has emerged to address the need to generate human-like responses from structured data [14]. Most generative AI models depend on annotated data to train internal system components and leverage the accuracy of the machine learning models used in those components. The presence of an efficient NLP system can leverage the enormous amounts of unstructured data available on the World Wide Web and assist in understanding user requirements and inputs in such an interactive working environment [15].

1.1.2. AI Vs. Conversational AI

ChatGPT is considered a conversational AI tool; however, it has broader functionality that qualifies it as a generative AI tool. Therefore, we need to illustrate the difference between conversational AI and generative AI. Generative AI can produce original content and synthetic data (artificially generated information rather than real-world events) when prompted. At the same time, conversational AI specialises in engaging in authentic, human-like two-way interactions with humans by understanding and responding in text or speech. To sum up, conversational AI and generative AI are both artificial intelligence tools that use natural language (Table 1).

Table 1: A comparison between generative AI and conversational AI

| Feature | Generative AI chatbots | Conversational AI Chatbots |
|------------------|---|---|
| Primary Function | Generates new creative text formats, like poems, code, scripts, musical pieces, emails, letters, etc. | Simulates human conversation by understanding user queries and naturally and engagingly responding to them. |
| Focus | Creativity and originality | Realism and human-likeness |
| Input | Text, code, or other data | Natural language text or voice |
| Output | New creative text, image, music | Human-like responses to user queries |
| Training Data | Large datasets, e.g., the internet | Large datasets of human conversation |
| Applications | Content creation, research, coding, images, ... and more | Customer service, virtual assistants, chatbots |
| Examples | ChatGPT, LaMDA | Amazon Lex, IBM Watson Assistant, Google Dialogflow |

1.2. AI Evolution

The concept of intelligent machines was first introduced in the early thirties, 1932, when Georges Artsrouni invented a machine that could translate between languages on a mechanical computer encoded on punch cards; he called it “mechanical brain”. Later in 1957, in the field of linguistics, Noam Chomsky published “Syntactic Structures,” which introduced grammatical rules for parsing and generating natural language sentences [15]. In 1966, a breakthrough in AI was made with the creation of the first chatbot, “Eliza.” Eliza simulates a conversation with a psychotherapist and was developed by MIT professor Joseph Weizenbaum. In 1989, Yann LeCun, Yoshua Bengio, and Patrick Haffner illustrated how convolutional neural networks can be used to recognise images. The second millennium saw many improvements in AI projects. Starting with object recognition applications, in 2006, Fei-Fei Li, a data scientist, created the ImageNet database, which is considered the foundation of visual object recognition [15]. Several enhancements have been made to improve the performance of neural network algorithms, leveraging technological advancements that have shaped the evolution of machine learning-based techniques, leading to new trends in linguistics, image recognition, and many other intelligent applications.

In 2013, a group of researchers at Google introduced the “word2vec” system, which can automatically identify word relationships. Google's collaborative research teams continue to drive new technology trends. In 2017, Google researchers developed the concept of transformers, which inspired subsequent research into automatically parsing unlabelled text into large language models (LLMs). Later in 2018, Google researchers implemented transformers into “BERT,” a Bidirectional Encoder Representations from Transformers, a platform that can predict the meaning of text. “BERT” was trained on more than 3.3 billion words, with 110 million parameters, and can identify relationships between words in sentences, paragraphs, and even books [16]. In the same year, OpenAI created the first version of GPT (Generative Pre-trained Transformer). As the name suggests, it's a trained platform that can generate content using LLMs, chatbots, and language translation. GPT was trained on about 40 gigabytes of data and consists of 117 million parameters. In 2021, OpenAI introduced “DALL-E, a platform that can generate images from text prompts [17]. The name is a combination of “WALL-E,” the name of a fictional robot in a Disney Pixar movie, and the artist Salvador Dalí.

2. Generative AI Tools Categorized According to their Creation Purpose

In this section, researchers introduce an attempt to classify the most common AI tools according to their functionalities and outcomes into seven categories:

2.1. Code Generation and Suggestion

- **Alpha Code:** A large-scale language model that has 41.4 billion parameters. Alpha code is a complex model that can generate code in several programming languages, including C#, Scala, Ruby, Java, JavaScript, PHP, GO, Python, C++, and Rust.
- **GitHub Copilot:** A collaboration between GitHub and OpenAI, it can complete code using artificial intelligence. It supports multiple languages and can learn from open-source code. It suggests code and documentation based on context-awareness, and it can integrate with Integrated Development Environments (IDEs) [19].

2.2. Business-Oriented Tools

- **Cohere Generate:** An AI platform that aids organisations in their business operations. It can create custom email content and product descriptions, as well as generate marketing- and sales-focused content.
- **Design.ai:** A comprehensive AI tool that can create graphic designs and marketing content, such as Logos, Graphic paints, Font paints, and various designs, along with social media postings, voiceovers, and videos.
- **Jasper:** An AI content creation tool focused on marketing and digital content. It can generate content for generative AI enterprise use cases based on your business's documents and resources [20].

2.3. Video Creation and Editing

- **Synthesia:** An AI tool that can create high-quality videos in minimal time. It can manage content by analysing and editing text and by creating avatars with facial and vocal expressions.
- **Rephrase.ai:** Just like Synthesia, Rephrase is an AI-generated tool that can create professional videos. However, it has an additional feature: digital avatars of the real people in the video. So, it provides more personalised experiences and communication. It can also integrate with various application channels, such as API access, App notifications, WhatsApp for Business bot, Email, QR code, and Microsite.

- **Type Studio:** An online transcription software that runs in the browser. Users can upload videos, browse video content, and automatically generate subtitles and captions, with video editing available if needed.
- **Murf.ai:** An AI tool that generates high-quality voice-overs for videos. It uses voice recognition technology to convert text to speech, with customisation features. This tool allows users to edit the script or use one of more than 100 human-sounding voices. It also lets users convert casual voice recordings into professional, studio-quality voice-overs.
- **Sora:** An open AI development inspired by LLMs, trained on internet-scale data. Sora is a diffusion transformer that demonstrated remarkable scaling across a variety of domains.

2.4. Image Generation Tools

- **DALL-E 3 is one of the Best AI Tools for Image Generation:** It creates realistic photos upon user requests and can edit or add modifications to existing images. DALL-E 3 can reject user requests by filtering out improper input to prevent the generation of inappropriate photos. It's now available to ChatGPT Plus users [21].
- **StyleGAN Uses Deep Learning Algorithms to Generate Realistic Images.** It can generate artificial faces based on a dataset of notable faces used for training.

2.5. Music and Voice Generation

- **Soundraw:** A generative AI tool that allows free, customised music generation. It can create audio content for applications such as podcasts, radio programs, Ads, Audiobooks, and games. The generated tracks are permanently licensed, preventing users from encountering copyright issues [18].

2.6. Text Generation

- **Copy.ai:** It is one of the best text-generative tools; it can recognise text context and intent, such as for marketing, social media posts, and punchy headlines. It can create captions, personalised email content, translations, blog posts, and interpretations.
- **Jenni:** An AI-powered editor for academic writing that can provide reference suggestions during writing and help researchers enhance their academic writing skills.
- **AI Writer:** It can create high-quality, full-length article drafts and generate unique, original text. AI Writer can generate text with embedded search engine optimisation, helping it reach the targeted audience.
- **Descript:** A cloud-based video/audio editor that generates subtitles and captions, writes webinar scripts, makes editable podcast scripts, and records screen [18].

2.7. NLU and Content Creation

- **ChatGPT:** The most common AI tool developed by OpenAI, as it provides users with free access to AI content development. It can engage in conversation and answer questions fluently within the same context. It is also used to create content and to interact in open-domain conversations [22].
- **Bardeen:** An AI tool that helps users manage time and organise tasks in a way that increases productivity and saves time. It runs in the browser, creates personalised messages, and learns continuously to improve its efficiency.
- **Bard:** Developed by Google, is a content-generation tool and chatbot. It uses the transformer-based LaMDA model and is considered a Google counterpart to ChatGPT.
- **GPT-4:** An open AI development that was marketed as being more stable and reliable to use than the previous generations. It has 100 trillion parameters and improved performance with image input and multilingual capabilities [22].
- **ChatFlash:** It is an AI generation tool that can create content, including images and SEO-optimised content, and generate expert content for professionals. It can also check for plagiarism.
- **ChatSonic:** A conversational AI tool based on GPT-4 that addresses ChatGPT's limitations, offering the best free chat GPT experience available. It can paraphrase, expand, summarise, translate, generate text, and create images and Google and Facebook Ads [23].
- **Claude:** An AI assistant that can process huge amounts of text and interact in natural conversations in various languages. Claude can automate workflows and process programming languages. This tool is specifically designed with ethical and responsible AI best practices in mind [19].
- **Mistral 7B:** A large language model that is a pre-trained generative text model with 7 billion parameters. It's an open, portable generative AI for developers and businesses [24].

- **LLaMA:** A part of Meta's open science projects, LLaMA (Large Language Model Meta AI) is designed to help researchers advance their AI research. It is a collection of pre-trained and fine-tuned generative text models with parameter counts ranging from 7 billion to 70 billion [19].
- **Microsoft Copilot:** A Microsoft company development, it's capable of chatting with text, voice, and image. It can summarise documents and web pages and generate AI images. It enables the integration with Microsoft 365 applications.
- **Jais:** The world's largest LLM with 13 billion parameters and 400 billion tokens. It was trained on 119 billion Arabic tokens and 279 billion English tokens. It was developed by G42, a company based in Abu Dhabi, United Arab Emirates. It's supposed to generate high-quality Arabic text, including regional dialects, and can perform tasks such as translation, summarisation, question answering, and more.
- **GPT-4o:** The most recent OpenAI development in Large Language Models (LLMs), it provides GPT-4-level intelligence with improved capabilities in text, voice, and vision.

3. ChatGPT Key Features Compared to Similar Generative AI Tools

As in this research, the focus is on ChatGPT, and a detailed study of its responses was conducted to evaluate its performance compared with similar, well-known AI tools, including Claude, Mistral 7B, Gemini, and Microsoft Pilot. As shown in Table 2 below. In addition, another comparison between ChatGPT and its modified versions (ChatGPT Plus and GPT-4.0) is listed in the following sections, Tables 3 and 4.

3.1. ChatGPT Compared to Claude, Mistral 7B, Google Gemini, Microsoft Copilot

The comparison included Arabic and English questions in different aspects to provide a comprehensive evaluation of most fields. Here is a clarification of each question used in the evaluation:

3.1.1. English Questions

Coding Proficiency: Large language models were always used for rewriting, updating, and testing code in various programming languages. Therefore, a test that asked both bots to create a simple Python program was evaluated. The prompt used was "Develop a Python script that serves as a personal expense tracker." All tools created a fully functional expense tracker built in Python, except for Microsoft Copilot, which only generated a template to start with. Tools varied in the features they provided to users, ranging from simply adding a record, as in ChatGPT, to adding more details about the category or description (Mistral7B), and even viewing categorised expenses, which was available in Claude and Gemini. Claude went further by providing documentation for the program features.

Natural Language Understanding: For this task, a Cognitive Reflection Test question was used to evaluate how well chatbots can understand ambiguous Natural-Language prompts. The prompt: "A bat and a ball cost £1.10 in total. The bat costs £1.00 more than the ball. How much does the ball cost?" All tools followed the same steps to calculate the result; however, Claude and Mistral7 provided more explanation.

Creative Text Generation and Adaptability: This test assesses the ability to create a consistent narrative story within a predefined theme. Also modifies the story if requested, such as changing characters or names. The initial prompt was "Write a short story set in a futuristic city where technology controls every aspect of life, but the main character discovers a hidden society living without modern tech."

Reasoning and Problem-Solving: Reasoning capabilities are one of the benchmarks for an AI model. It is not easy to evaluate, so the prompt was some classic questions: "You are facing two doors, one door leads to safety and the other door leads to danger. There are two guards, one in front of each door. One guard always tells the truth, and the other always lies. You can ask one guard one question to find out which door leads to safety. What question do you ask?" All tools answered correctly, so the judgment was based on the clarity. Claude and ChatGPT gave more detailed responses, but Claude was more organised and clear.

ELIV Explain Live I'm Five: This test evaluates the ability to simplify explanations for five-year-olds, making them clear, accurate, and engaging. The prompt was "Explain how aeroplanes can stay up in the sky to a five-year-old." All tools gave accurate responses using birds to explain their answers. However, Claude gave a more detailed explanation based on paper planes, but it was very simplified. ChatGPT and Gemini also gave clear responses. Mistral7 was a little advanced for a 5-year-old, and Microsoft Copilot provided a very long essay with much more advanced information.

Ethical Reasoning: Evaluating the AI tool's ability to make decisions in situations that may harm humans is not easy. They are supposed to assess different scenarios and make a judgment. The prompt used was “Consider a scenario where an autonomous vehicle must choose between hitting a pedestrian or swerving and risking the lives of its passengers. How should the AI make this decision?” None of the tools provided an exact answer; however, they offered suggestions and precautions to consider when making the decision. They also clarified the ethical basis and policies that underlie their decision (such as “prioritising human life and minimising harm”).

Knowledge Retrieval, Application, and Learning: A large language model should accurately retrieve information from its training data. For this test, the prompt used was “Explain the significance of the Rosetta stone in understanding ancient Egyptian hieroglyphs. All tools provided detailed information about the stone; however, in this task, Microsoft Copilot displayed a photo of the stones and provided references at the end of the response.

3.1.2. Arabic Questions

Writing a Short Message in Arabic: For this test prompt, the Arabic phrase used was “My friend is ill, write a message for my friend wishing him to get well soon, then write it in Egyptian dialect.” All tools wrote a proper message in Arabic, but when requested to write it in Egyptian dialect, Gemini couldn’t provide an answer, and Mistral7 was too close to formal language, with a few errors.

Facts and Knowledge Retrieval: A question was asked of all tools to assess their ability to respond to knowledge retrieval. The prompt used was: “When were the pyramids of Giza built?” All tools provided sufficient responses to this request; however, ChatGPT and Microsoft Copilot's answers were too short.

Mathematical Puzzle: This is to evaluate AI tools' ability to comprehend and calculate Arabic-language mathematical problems. The prompt used was: “What is the smallest integer that is equal to seven times the sum of its digits?” This test was a little tricky, as many tools got it wrong or gave incorrect answers. Gemini gave a too-short written answer, while Claude gave the right answer with a very clear explanation.

Cross-Lingual Translation and Cultural Awareness: In this test, a complex paragraph from the “Nagib Mahfouz” novel was given to AI tools to translate. This paragraph contained ambiguous words and culturally specific expressions. The tools gave close answers to this question, and Claude and Gemini provided extra elaboration notes. A few culturally related errors appeared in Gemini, Mistral7, and Microsoft Copilot (Table 2).

Table 2: ChatGPT compared to Claude, Mistral 7b, Google Gemini, Microsoft Copilot

| Aspect | Claude | Mistral 7B | Gemini | ChatGPT | Microsoft Copilot |
|---|--|--|--|--|--|
| Coding Proficiency | Provided a fully functional program with an explanation. Features included (add, view, categorise, calculate total expenses) | Provided a fully functional program. Features included (add, view, categorise, delete, calculate total expenses) | Provided a fully functional program with an explanation. Features included (add, view, categorise, calculate total expenses) | Provided a fully functional program. Features included (add, view) | Provided a template to start with. |
| Natural language understanding | Defined variables and solved the equation step by step | Defined variables and solved the equation step by step | Defined variables and solved the equation step by step | Defined variables and solved the equation step by step | Defined variables and solved the equation step by step |
| Creative text generation and adaptability | | | | | |
| Reasoning and problem-solving | Answered correctly with a more organised and detailed response | Answered correctly | Answered correctly | Answered correctly with a more detailed and clear response | Answered correctly |

| | | | | | |
|--|--|--|--|--|---|
| ELIV Explain like I'm five | Clear and simplified response using the paper planes example | The response given contained some advanced information from five years ago | Clear and simplified response using birds as an example | Clear and simplified response using birds as an example | The response given was very detailed and contained some advanced information from a five-year-old |
| Ethical Reasoning | Didn't provide an exact answer, just explained the precautions to consider when making a decision. | Didn't provide an exact answer, just explained the precautions to consider when making a decision. | Didn't provide an exact answer, just explained the precautions to consider when making a decision. | Didn't provide an exact answer, just explained the precautions to consider when making a decision. | Didn't provide an exact answer, just explained the precautions to consider when making a decision. |
| Knowledge retrieval, application, and learning | A detailed response with information about the stone was given | A detailed response with information about the stone was given | A detailed response with information about the stone was given | A detailed response with information about the stone was given | A detailed response with information about the stone was given, in addition to a photo of the stone, including the references |
| Write a short message in Arabic. | Provided a proper message in Arabic and the Egyptian dialect | Provided a proper message in Arabic, but contained some errors in the Egyptian dialect message | Provided a proper message in Arabic, but couldn't respond in the Egyptian dialect | Provided a proper message in Arabic and the Egyptian dialect | Provided a proper message in Arabic and the Egyptian dialect |
| Facts and knowledge retrieval | Provided a sufficient response | Provided a sufficient response | Provided a sufficient response | Provided a sufficient response, but too short | Provided a sufficient response, but too short |
| Mathematical puzzle | Gave the right answer with a very clear explanation | Didn't understand the question and gave the wrong answer | Gave the right answer | Gave a wrong answer despite the very detailed explanation | Gave a wrong answer |
| Cross-lingual translation and cultural awareness | Provided a good translation, with elaboration notes | Provided a good translation, including some culturally related errors | Provided a good translation, including some culturally related errors, with elaboration notes | Provided a good translation, including some culturally related errors | Provided a good translation, including some culturally related errors |

3.2. ChatGPT Compared to GPT-4

This section discusses the advancements made by the OpenAI company to improve ChatGPT into the new version, GPT-4. Many modifications were applied to release a more powerful pre-trained model that addresses deficiencies in the previous release in such a short time. As shown in Table 3, it can perform most of the processes more efficiently and accurately.

Table 3: ChatGPT compared to GPT-4

| Feature | ChatGPT | GPT-4 |
|------------------|---|---|
| Language Fluency | Focuses on conversational interactions and natural responses. | Focuses on grammar, vocabulary, and syntax. Able to generate coherent and contextually relevant text. |

| | | |
|-------------------------------|---|--|
| Contextual understanding | Maintains conversational flow and understands user intent. | Comprehends complex contexts and generates accurate responses. |
| Response Generation | It provides user-friendly and context-aware responses. | It provides detailed and informative responses with enhanced coherence and creativity. |
| Multimodal Capabilities | Focuses on text-based interactions. | It can process and generate text with other media formats, such as images, videos, and audio. |
| Image interpretation | It doesn't directly engage in image interpretation tasks. | It can create textual descriptions of images, but may not be as accurate as dedicated image processing systems. |
| Number of Parameters Analysed | It is less computationally intensive and employs a mechanism that ensures high-quality conversational output. | A highly complex model that analyses many parameters to generate nuanced and contextually appropriate responses. |
| Dealing with current Data | It benefits from its training on diverse datasets but may exhibit limitations in rapidly changing scenarios. | It can handle up-to-date and real-time information better, providing more relevant responses in dynamic contexts. |
| Accuracy of response | May occasionally produce responses that are contextually plausible but factually incorrect. | It strives for accuracy in its generated responses and aims to minimise factual errors. |
| Complex Tasks | May face challenges when dealing with highly technical or specialised subjects. | It excels at generating detailed, informative content across various domains, with improved performance on complex language tasks such as summarisation, translation, and text generation. |

3.3. ChatGPT Compared to ChatGPT Plus

A simple comparison between ChatGPT and ChatGPT Plus is shown in Table 4. It covers the model capacity, processing speed characteristics, the language model used and its capabilities, and their applications.

Table 4: ChatGPT compared to ChatGPT plus

| Parameter | ChatGPT | ChatGPT Plus |
|-----------------------------|--|---|
| Transformers Model Capacity | 117 million parameters | 1.2 billion parameters |
| Processing speed | Slow compared to ChatGPT Plus | Faster than ChatGPT |
| Language model | GPT-3.5 | GPT-4 |
| Multilingual capabilities | Limited proficiency | Better proficiency |
| Application | Ideal for use cases that occasionally require natural language processing. | Best for use cases that rely heavily on NLP, such as content generation and virtual assistants. |

4. Conclusion

This comparative review of modern generative AI tools shows how ChatGPT has had a significant impact on the development of AI-driven language models and on the growing competition among companies developing these technologies. ChatGPT, especially in its more advanced versions that use fine-tuned large language models like GPT-4, has made great strides in producing accurate, contextually relevant, and logical answers across a wide range of fields. It has become a standard for conversational AI and a starting point for new developments in the field, as it can handle and respond to complex, multi-step questions. Along with ChatGPT, there are now more open-source and proprietary generative AI models that can interpret natural language, write code, reason, and work with multiple types of data. These models help make the AI environment more varied and collaborative, allowing researchers and developers to test, customise, and deploy solutions tailored to specific use cases.

Anthropic's Claude stands out among major competitors for its heavy emphasis on ethical AI development, safety alignment, and responsible deployment. This topic addresses concerns about bias, transparency, and the effects of large-scale AI systems on society. Gemini from Google DeepMind is another important player in the generative AI space. Gemini combines advanced reasoning and multimodal processing skills. In some analytical and task-specific situations, it performs better than other systems, but in others, it produces results that are similar to ChatGPT. The development of this model is part of a larger trend toward unified models that can handle text, graphics, and other data types in a single place. All of these improvements show that the generative AI field is going through a time of tremendous innovation and robust competition. The rise and ongoing development of ChatGPT have been very important in setting industry standards, accelerating research, and encouraging people

to use AI technologies. As generative AI continues to grow, the interaction between proprietary and open-source models is expected to lead to greater breakthroughs, enabling more powerful, ethical, and easy-to-use AI systems for future applications.

Acknowledgement: The authors would like to express their sincere gratitude to Sadat Academy for Management Sciences and Alexandria University for providing the academic support and resources that facilitated this research.

Data Availability Statement: The data supporting the findings of this study can be made available upon reasonable request to the corresponding authors to promote transparency and reproducibility.

Funding Statement: This research and manuscript were completed without receiving any external financial support or funding.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest that could have influenced the outcomes or interpretations of this study.

Ethics and Consent Statement: All authors collectively agree to the publication and dissemination of this work for academic and educational purposes, adhering to ethical research standards.

References

1. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
2. J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modelling," *arXiv Preprint, arXiv:1412.3555*, 2014. Available: <https://arxiv.org/abs/1412.3555> [Accessed by 17/09/2024].
3. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, United States of America, 2017.
4. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, 2019. Available: <https://openai.com/research/language-unsupervised> [Accessed by 14/09/2024].
5. C. Ortega, Y. Falevoz, and R. Ayrignac, "PIM-AI: A novel architecture for high-efficiency LLM inference," *arXiv Preprint, arXiv: 2411.17309*, 2024. Available: <https://arxiv.org/abs/2411.17309> [Accessed by 18/09/2024].
6. Z. Wan, X. Wang, C. Liu, S. Alam, Y. Zheng, J. Liu, Z. Qu, S. Yan, Y. Zhu, Q. Zhang, M. Chowdhury, and M. Zhang, "Efficient large language models: A survey," *arXiv Preprint, arXiv: 2312.03863*, 2024. Available: <https://arxiv.org/abs/2312.03863> [Accessed by 19/09/2024].
7. N. Tihanyi, T. Bisztray, M. A. Ferrag, R. Jain, and L. C. Cordeiro, "How secure is AI-generated code: A large-scale comparison of large language models," *Empirical Software Engineering*, vol. 30, no. 2, p. 42, 2024.
8. I. O. Gallegos, R. A. Rossi, J. Barrow, M. M. Tanjim, S. Kim, F. Derroncourt, T. Yu, R. Zhang, and N. K. Ahmed, "Bias and fairness in large language models: A survey," *Computational Linguistics*, vol. 50, no. 3, pp. 1097–1179, 2024.
9. C. Deng, Y. Duan, X. Jin, H. Chang, Y. Tian, H. Liu, Y. Wang, K. Gao, H. P. Zou, Y. Jin, Y. Xiao, S. Wu, Z. Xie, W. Lyu, S. He, L. Cheng, H. Wang, and J. Zhuang, "Deconstructing the ethics of large language models: From long-standing issues to new-emerging dilemmas," *AI and Ethics*, vol. 5, no. 8, pp. 4745–4771, 2024.
10. S. Raza, A. Raval, and V. Chatrath, "MBIAS: Mitigating bias in large language models while retaining context," *arXiv Preprint, arXiv: 2405.11290*, 2024. Available: <https://arxiv.org/abs/2405.11290> [Accessed by 20/09/2024].
11. P. Wang, S. Bai, S. Tan, S. Wang, Z. Fan, J. Bai, K. Chen, X. Liu, J. Wang, W. Ge, Y. Fan, K. Dang, M. Du, X. Ren, R. Men, D. Liu, C. Zhou, J. Zhou, and J. Lin, "Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution," *arXiv Preprint, arXiv: 2409.12191*, 2024. Available: <https://arxiv.org/abs/2409.12191> [Accessed by 21/09/2024].
12. G. G. Chowdhury, "Natural language processing," *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 51–89, 2003.
13. E. Cambria, "Jumping NLP curves: A review of natural language processing research," *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, pp. 48–57, 2014.
14. J. Kaur, "What are the differences between NLP, NLU, and NLG?" *XenonStack*, 2025. Available: <https://www.xenonstack.com/blog/difference-between-nlp-nlu-nlg> [Accessed by 22/07/2025].
15. G. Lawton, "What is GenAI? Generative AI explained," *Tech Target*, 2025. Available: <https://www.techtarget.com/searchenterpriseai/definition/generative-AI> [Accessed by 17/07/2025].

16. Google AI, "Our AI journey," *Google AI*, 2025. Available: <https://ai.google/advancing-ai/milestones/> [Accessed by 23/07/2025].
17. OpenAI, "Pioneering research on the path to AGI," *OpenAI*, 2025. Available: <https://openai.com/research/index/> [Accessed by 24/07/2025].
18. N. Duggal, "Top 20 Generative AI Tools," *Simplilearn*, 2024. Available: <https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/top-generative-ai-tools> [Accessed by 10/09/2024].
19. L. Ticong, "Gemini vs ChatGPT 4: The Winning AI Revealed," *eWeek*, 2025. Available: <https://www.eweek.com/artificial-intelligence/gemini-vs-chatgpt/> [Accessed by 11/07/2025].
20. A. Zenon, "How Jasper's content marketing manager uses Jasper," *Jasper*, 2024. Available: <https://www.jasper.ai/blog/content-marketing-manager-uses-jasper> [Accessed by 13/09/2024].
21. J. Betker, G. Goh, L. Jing, T. Brooks, J. Wang, L. Li, L. Ouyang, J. Zhuang, J. Lee, Y. Guo, W. Manassra, P. Dhariwal, C. Chu, Y. Jiao, and A. Ramesh, "Improving image generation with better captions," *OpenAI*, 2023. Available: <https://cdn.openai.com/papers/dall-e-3.pdf> [Accessed by 14/09/2024].
22. S. Kothari, "GPT-4 vs. ChatGPT: What's the difference?" *Simplilearn*, 2023. Available: <https://www.simplilearn.com/tutorials/chatgpt-tutorial/gpt4-vs-chatgpt> [Accessed by 01/09/2024].
23. W. Ladhar, "ChatSonic APK: Super ChatGPT app," *Chatsonic*, 2023. Available: <https://chatsonic.pro/chatsonic-app/> [Accessed by 07/09/2024].
24. P. Gupta, "What makes ChatGPT Plus different from ChatGPT?" *easyInsight*, 2023. Available: <https://easyinsights.ai/blog/whats-new-in-chat-gpt-plus> [Accessed by 27/09/2024].