American Journal of Applied Science and Technology

VOLUME Vol.06 Issue 02 2026
PAGE NO. 97-106
DOI 10.37547/ajast/Volume06Issue02-10

The Rise and Impact of Modern Generative AI Tools: A Comparative Study of Chatgpt, Gemini, Claude

Suhad Ateyah
University of Kufa, Iraq

Zainab Khairallah Kadhim
General Directorate of Education in Najaf, Iraq

Received: 31 December 2025; **Accepted:** 22 January 2026; **Published:** 27 February 2026

Abstract: The present research utilizes three popular multimodal AI systems (Gemini, Cloud AI, ChatGPT) to evaluate their ability to interpret visual language by analyzing responses to three different images. In addition to evaluating the systems' ability to produce accurate and detailed descriptions of each image, this research evaluated their ability to contextualize each description within an appropriate framework of understanding, as well as assess their response times. Results indicate that ChatGPT provided the most accurate and descriptive descriptions of the images analyzed in this study, particularly those which depicted emotionally and/or socially nuanced scenes; that Gemini performed reasonably well in terms of conceptual interpretation, though was inconsistent in its provision of specific details regarding the image(s); and that while Cloud AI responded more quickly than either ChatGPT or Gemini, it failed to provide as much detail or relevance to the situation presented in the images. These findings emphasize the need to develop multimodal AI systems that balance speed, emotional intelligence, and semantic accuracy to be used in the real world when reasoning with images.

Keywords: Artificial Intelligence, ChatGPT, Gemini, Natural Language processing, Generative AI, Claude, AI Tools Comparison, Human-AI Interaction, OpenAI, ChatGPT.

INTRODUCTION:

The quick increase in use of artificial intelligence (AI) in many areas has dramatically changed how technology is used in every area. Among the biggest advancements is the emergence of large language models (LLMs). LLMs are a new form of generative AI which is capable of producing text, code and multimodal content that is contextual and meaningful. Examples of LLMs are ChatGPT from OpenAI, Gemini by Google DeepMind and Claude from Anthropic [1][2].

In addition to being chatbots, generative AI are able to assist in research, create code, analyze business data, collaborate in creative ways, etc. The capabilities of generative AI represent both possibilities and problems for academia, industry and society at large. With increasing usage of generative AI in school settings, workplaces and online communities, it is essential to examine the advantages and disadvantages of generative AI and its potential impact on the future [3][4].

This research aims to provide an evaluation of several popular generative AI tools currently available, examining their technical bases, primary features, real-world applications and possible risks. Through this evaluation we hope to enhance researchers' and educators' understanding of how generative AI may influence creation of knowledge and decision-making processes in the immediate future [5].

Overview of Modern Generative AI Tools

Generative AI is currently being shaped by a variety of advanced models developed by top research

institutions and technology companies. All of these tools rely upon large scale transformer architectures that are trained on massive amounts of data; however, each model was developed with unique design objectives, characteristics, and use cases[4].

One of the most widely used generative AI tools today is OpenAi's ChatGPT. ChatGPT, as an example, is now multimodal, as its newest version utilizes the GPT-4 architecture to process both text and images. The usability of ChatGPT in education, creativity, and programming settings is enhanced due to a user-friendly interface and compatibility with various productivity tools (i.e., Microsoft Copilot). [3] [6] [7].

DeepMind's Gemini uses the same technology as Google DeepMind's Gemini (Gemini 1.5) as well as similar technologies such as GPT. Gemini 1.5 was designed for long context reasoning, and therefore has the ability to process input of greater than one million tokens. Therefore, the Gemini system is particularly suited for large document or data sets, such as technical documents or code bases, and may thus have an advantage over its competitors in a research or business setting [1].

However, there is another company using the same technology as Gemini, Anthropic, and they use this technology to create a Constitutional AI called Claude. Claude is a type of AI that is more aligned, steerable, and safer than Gemini, and the most recent model, Claude 3, creates a very good middle ground between the desire for creativity and the need to follow the rules, therefore Claude 3 can be used in a variety of professional applications including professional writing, summary generation, and many other applications that require ethical considerations [2].

Overall, while both systems are created from the same underlying technology, the differences in each system lie in their focus areas (such as performance characteristics that may include factors such as safety, alignment, long-context processing capabilities and working with other systems). In order to determine which system will work best for a particular academic or industrial application, it is necessary to understand the key characteristics that make each system different.

Technical Comparison

The primary differences between generative AI models are based upon the underlying architecture of each model, as well as their respective capacity for input, ability to reason, and accessibility. The architecture and scalability of a model can have a significant impact on how well that model is able to perform a given task (e.g., natural language processing, code generation, and/or document analysis) [3][5].

Model architecture and scalability also affect how well a language model will be able to perform its intended functions. For example, GPT-4 has performed very well in areas including language fluency, multilingual reasoning, and creative writing. Conversely, Gemini 1.5 is particularly adept at longer contextual reasoning; therefore, users are able to input larger quantities of information (for instance, a book or an entire repository of source code) without losing coherancy. [6].

Another large distinction between these two types of models relates to how effectively they may be able to think and how easily one will be able to access them. For example, Claude models are developed under a "safe and steerable AI" framework by means of constitutional AI which incorporates structured feedback in addition to the development of internal ethical guidelines as an effective means of producing output which is very mindful of its surroundings and carefully constructed [2].

Technical comparison represents yet another method to evaluate models. While both ChatGPT and Gemini represent closed model systems with limited user control; Mistral and LLaMA represent open weight models that are capable of being modified or adapted to meet the needs of a researcher or developer. [8].

Table 1: The following table summarizes the technical features of selected leading AI systems:

	Model	Developer	Architecture	Max Input Length	Open Source	Notable Strengths

1	Chatgpt (gpt-4)	Open AI	Transformer (GPT-4)	128 k tokens	No	High accuracy, coding , general-purpose
2	Gemini 1.5 pro	GoogleDeep Mind	Mixture of Experts	1 million tokens	No	Long-context understanding, multimodal
3	Cloude 3	Anthropic	Transformer variant	200 k tokens	No	Aligend, safe, streerable outputs

Use Cases and Applications

Generative AI is revolutionizing how people and businesses work. It is creating new ways for people and organizations to solve problems, create and communicate ideas. Generative AI models are transitioning from experimental tools to being incorporated as workflow solutions throughout various sectors.[4].

1. Education and Academic Research

AI tools in education are being implemented in a variety of ways such as customized tutoring, assisting in writing essays and providing language support. As an example, ChatGPT could be used to provide a student with suggestions for ideas for their paper; help them understand a complex topic; and also serve as a simulated conversation partner in a foreign language. Researches are increasingly utilizing tools such as Claude and Perplexity AI to aid them in creating summaries of academic articles; refine their hypotheses; and assist in generating citations. Kasneci et al. (2023) found that LLMs have a positive impact on learner autonomy and engagement, if they are used ethically. [2][3][6].

2. Programming and Software Engineering

Generative AI is transforming how developers create and evaluate software; GitHub Copilot — which uses an OpenAI model — can be used as a real time "pair programmer," and reduce a developer's mental load while speeding up development of code. The larger-than-GitHub-Copilot context window of Gemini enables it to analyze all of the files within a repository simultaneously, allowing for complete analysis and refactoring of very large scale projects. Claude's safer and more interpretable design make it a better option than Gemini when teaching the logic of code, or for use as an explanation tool (Ziegler et al., 2023). [7].

3. Business and Content Generation

Generative AI has become a way companies are using to produce company reports, email drafts, product descriptions, and chatbot customer service applications. Many of the users of CRM platforms such as Claude and ChatGPT utilize the information and insights generated by this technology to assist them in answering questions and making informed decisions. The same technology is also being used in marketing departments to generate campaign ideas and to develop content that can be understood globally. The tool from Google, Gemini, appears to be most adept at working with structured business data while taking the context of a question into consideration when developing a response [4].

4. Healthcare and Medical Analysis

Researcher's are testing GPT-based applications to retrieve data from patients' Electronic Health Records (EHRs) that could be very beneficial in helping to automate some administrative tasks, while generative AI technology will start to help with summarizing doctor's clinical notes and answering patients' questions, but is currently limited due to policy and regulation. [4][6].

5. Legal and Government Use

Legal professionals rely on tools (ChatGPT and Claude) to read relevant cases, summarize contract language, draft legal memoranda, etc. Models such as Gemini provide significant advantages in

reading/understanding lengthy legal documents. There is a growing interest among various governmental entities to utilize LLMs for citizen support, document automation, policy analysis. Google DeepMind, 2024. [1][9].

6. Creative Industries and Media

The number of people who write articles, journalism, etc., is increasing in their use of Artificial Intelligence (AI) to assist them with producing content. There are two popular forms of AI that many writers, journalists, and artists are now using to create a variety of types of content. For example, both ChatGPT and Claude may produce story outlines, dialogue for characters, or even complete screenplays. Many musicians, as well as various visual artists, are experimenting with AI-created lyrics and images, while many media professionals utilize AI in editing interviews, summarizing interviews, and creating headline summaries. [2][4][6].

METHOD

Comparison of AI Applications for Image Processing A study was conducted to evaluate the performance of various Artificial Intelligence (AI) applications in processing image data. As part of this study, three images were used as input images for each of the AI applications being studied. A key objective of this study was to evaluate and compare the selected applications with regard to accuracy of results, speed of processing, and reliability. Selection of Application Input Images and Content The input images were selected to provide the same input to each of the AI applications that would be evaluated. In order to evaluate each application thoroughly, the images were carefully evaluated based upon their content, resolution, and file format.

How do the images look? In other words, can you provide an example of the types of things that appear in the images? For instance, "There are many complex graphics in Image A; there are many objects in a plain background in Image B; There is much text in Image C." The Comparative Analysis selected several public and/or private AI applications to carry out image-centric activities.

These applications include [Identify the categories of AI applications, such as "image recognition platforms, object detection models, GANs etc."] Criteria for Evaluating Performance. We used the following measures to evaluate how well each AI application performed:

Accuracy: The accuracy of an AI's output (compared to its input) is measured by this method. Accuracy of object recognition or classification tasks will be determined by how well the AI understands 'visual language', or as simply put, it is able to describe images or provide answers to questions regarding images' content. In general, the AI's evaluation metrics will be much more qualitative than quantitative. That is a reversal of what one typically expects when working with traditional image classification tasks. Truth for Each Image: This is the single most important part of the entire process. Prior to evaluating any models, we must first determine what the "correct answer" or the desired/expected correct response to the image is. A correct description of the image's content: For example, if the reference description is a picture of a happy person sitting in a café. Correct Answer to a Question: For example, if the question asked is "What color is the shirt the person is wearing?", the color of the shirt in the picture is the correct answer. Conceptual Analysis: When a person asks you to "Describe the emotion(s) portrayed within this image", the correct and true emotional interpretation is the truth.

Processing speed: How long does the AI take to turn the input image into the answers to the questions? Time for the application to receive the image and then send the final result to you was recorded in [Specify unit, e.g. "seconds", "milliseconds"].

Reliability: Reliability measures how consistent the AI is when the same image is used as input many times. Does the app always provide very close to the same accurate result every time you run it with the same input, or does its performance vary greatly?

Interpretability/Explainability: This evaluates how well the AI application is able to explain what decision-making processes were used to arrive at its output. For example, some applications may report on why they arrived at the output they did (e.g., through confidence score reporting, providing

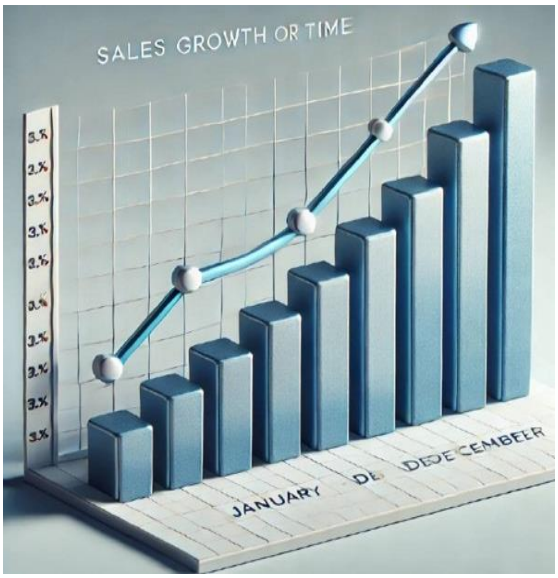
bounding box coordinates, etc.) and provide visual heat maps that highlight areas of interest.
Utilizing Resources (CPU, Memory, etc.) Each AI Application Uses: Since some AI Applications will likely be run in locations where there are limitations on available resources (i.e., processing power, memory, etc.), it is very important to track how many resources each application utilizes while it processes.
Output Format & Usefulness of AI Application: This is another way of saying the format of the AI's output and the ease of using that output. Is the AI producing structured data? Are the labels used by the AI to describe what was produced clear and understandable? Can the output from the AI be easily combined with additional data to analyze?

Scalability (Future Research): Although we did not test scalability during our trial since we were only working with a total of 3 images; future research will need to evaluate the application's ability to process larger amounts of images or possibly more complex data sets.

Collecting and Evaluating the Data Three separate images were processed through each AI application. We carefully documented all of the output data that each AI application provided (accuracy scores, time to process, and anything else that could be useful) including confidence scores and any error messages that occurred.

Following this, a comparative evaluation was undertaken to determine whether there were statistically significant differences or trends in how well the applications were performing with respect to the newly developed metric. Set of Images for The set of images employed in this study is a diverse collection of three images to provide an exhaustive assessment of the effectiveness of various Artificial Intelligence (AI) applications. In addition to each image having its own unique characteristics as a result of their visual properties, they also present a variety of challenging issues that are good for testing the capabilities of artificial intelligence. Image.

Landscape Scenery (Figure 1): A picture showing a picturesque natural setting with an impressive mountain range, a tranquil lake, and dense coniferous forest; a lake with turquoise color water that reflects the blue sky above; the soft light of dawn or dusk creates long shadows and highlights small features of both the water and the foliage. A picture that is ideal for assessing the ability of artificial intelligence systems to interpret scenes, to divide a scene into meaningful elements (such as sky, water, trees, mountains), to assess colors, and potentially for additional uses. Descriptive analysis: aesthetic evaluation. There are many details and many different texture types within this image that make the identification of all elements of the environment and the objects within



Business Chart (Figure 2): The image of this three-dimensional bar chart demonstrates "SALES GROWTH OVER TIME" with an upward trending bar. It displays the month-by-month data beginning in "JANUARY," and concluding at "DECEM" (likely a typo). There is a line graph over the bars showing the overall progress of the sales growth. The vertical axis represents the increase in the percent. Overall, the document pictured is particularly useful as an example for evaluating AI's ability to assess documents by using optical character recognition (OCR) to retrieve text from images or pictures (labels, titles, etc.) along with reading data visualizations and potentially extracting data from those images. While this requires AI to recognize objects visually, it also requires AI to read and interpret both structured and numeric data presented in graphical format..

Indoor Portrait with Background (Figure 3): The image appears to be an interior shot of a cozy & inviting indoor environment that is likely a cafe/co-workspace. The main subject of this image is a young male with a smile. He is positioned at a table with a cup in front of him. There are additional figures in the background along with architectural elements such as wood ceiling, wood shelves. This image would be useful for developing and evaluating AI based on face recognition, facial expressions/feelings of subjects, subject pose/position, finding objects in crowded environments (i.e. cups, tables, chairs), etc. and interpreting complex social situations. The presence of many people and additional background detail makes this image more challenging to identify subjects and provide accurate interpretations.

Image number	Description of the image	Tool \ model quastions
1	Natural landscape of mountain and a lake	1. Describe this Natural landscape? 2. What time of day does it appear to be? 3. What are a point of interest?
2	A chart showing sales growth	1. What is the main subject of this chart?

		2. Describe the overall trend for sales between months from "JANUARY" to "DECEDECMEBER" 3. What do you predict for sales in 2025 based on this chart?
3	A person smiling in a cafe	1. Describe the person in the photo. 2. What emotion is the person expressing? 3. What is the person doing in the photo? |

The three images from above will be run through several different AI models; all have been selected based upon their use in image processing. There will be some which are just general-purpose image analysis APIs, and some which are model-specific to do one type of task or another (i.e., object recognition, OCR, scene analysis). The comparison between the various AI systems will focus on key metrics regarding the performance of the systems, including but not limited to:

Accuracy: This means that the system accurately identified the natural features of the landscape picture, the facial features and objects within the portrait picture, and the extracted data (text, numbers) was as accurate as possible for the chart picture.

Processing Time: Each AI application will be tested to determine how quickly each can process an input image and produce an output.

Comparative Evaluation of Visual Language Understanding in Multimodal AI Models (table 3)

Image id	Tool	Received Response	Description Accuracy	Context Understanding	Response Time (S)	Notes
1	Chatgpt	The image shows a snow-capped mountain reflecting on a serene lake under a blue	4	5	5.2	Accurate description, added some aesthetic details.
	Gemini	A scenic view of a mountain and a lake, with white clouds	3	4	3.8	Concise description, less detailed than ChatGPT.
	Cloud AI	Mountain, lake, sky	2	2	2.5	Very generic description.
2	Chatgpt	The chart illustrates sales growth for a company over	5	5	6.1	Excellent comprehension of the topic.

	Gemini	Company sales, from January to Desmber	4	4	4.5	Good, but less detailes
	Cloud AI	Data figures	1	1	3.0	Failed to understand.
3	Chatgpt	A smiling young man is sitting in a cozy café, holding a coffee cup. The environment appears relaxed and warmly lit,.	5	5	3.5	Accurate, detailed, includes emotional and spatial context.
	Gemini	A cheerful young man is enjoying his coffee in a warmly-lit café. Other people are visible in the background	3	5	2.7	Good description, captures mood and setting, but less detail than ChatGPT
	Cloud AI	A man sitting at a table in a restaurant or café. Background shows other people	4	3	2.0	Basic description, lacks emotional tone and context details.

The results of Image 1 demonstrate that while all three give excellent descriptions of the image; ChatGPT was able to describe it the most completely and emotively by providing both an emotional tone and capturing the subject of the image. Gemini was able to provide a very good descriptive response that provided an emotional tone and some of the basic contextual elements of the scene, but had slightly less detail than ChatGPT. In comparison to the two above, Cloud AI provided a much more superficial/elementary descriptive response that did not provide evidence of deeper contextual or emotional understanding. The differing levels of contextual/emotional understanding demonstrated through the responses provided by the three tools, indicate the level of complexity and coherency provided by each tool. ChatGPT provides the greatest level of coherency and depth.

Challenges and Risks

Shortcomings and risks The results also illustrate many shortcomings and threats of the currently available multimodal AI solutions for understanding visually. A major shortcoming of the models used here was that they all describe the visuals with varying degrees of precision. For example, while ChatGPT provided a good amount of detail and context, Cloud AI typically limited its descriptions to relatively basic and missed important emotional and

environmental contextual details; the variability of model output makes the overall system less reliable (and potentially unsafe) when a detailed and nuanced interpretation of a visual is required. Additionally, people perceive and understand visual contextual information in many different ways. In some cases, one model may pick up on a subject's emotions or social position, whereas the other does not. This illustrates that visual language reasoning does not always translate across different models. In addition, varying response times demonstrate performance inefficiencies. When cloud AI responded quickly, this was typically at the expense of detail and context. Therefore, it appears that there is an equilibrium between rapidity and quality. Furthermore, there is always the potential for bias when assessing qualitative results, and this can impact benchmarking. Thus, as evident by the importance of having clear metrics and human-in-the-loop validation, this highlights why it is so crucial to properly test multimodal AI systems.

CONCLUSION

Overall, while all three multimodal AI tools performed differently, each tool demonstrated a level of proficiency for interpreting visual language in their own way. The ChatGPT tool was able to provide the most accurate, as well as contextualized representations of the given visual material, which demonstrates its ability to effectively integrate multiple types of contextual information (i.e., situational, emotional) along with various aspects of the visual cue itself. Gemini provided an adequate representation of the concepts involved, however on occasion lost sight of the finer points of detail. Cloud AI was the quickest of the tools to produce responses, yet lacked sufficient detail and/or contextual information. As such, the findings from this study demonstrate the need to achieve an optimal balance between semantic richness and processing speed in the development of future AI tools.

Future Outlook

The image data set for this study should be increased to include additional examples of various types of content (e.g., abstract, cultural, and domain specific). Also, having more than one human evaluator will increase the reliability of our evaluations. Additionally, testing for adaptation can be furthered by examining model performance across different languages and socio-cultural contexts. Finally, incorporating explanation tools will provide an understanding as to how and why the models perform the way that they do and facilitate improved model development in the future.

Limitations

The research for this study has limitations in that it utilized only three images as a basis for comparison which limits the generalization potential of its results. As such, although there are some structured criteria for evaluating the models qualitatively; due to the subjective nature of qualitative methods, there is potential for bias in the evaluation process. The analysis did not explore the internal structure (i.e., how they were trained) of each model, thus, it only provides limited insight into why there were differences in their performance.

REFERENCES

1. Google DeepMind. (2024). Gemini 1.5 Models: Technical Capabilities and Roadmap. https://deepmind.google/technologies/gemini
2. Anthropic. (2024). Claude and Constitutional AI: A Safer Path for LLMs. https://www.anthropic.com
3. Kasneci, E., Sessler, K., & Betsch, T. (2023). ChatGPT for Good? On Opportunities and Challenges of LLMs in Education. arXiv. https://arxiv.org/abs/2302.05756
4. McKinsey & Company. (2023). The Economic Potential of Generative AI: The Next Productivity Frontier. https://www.mckinsey.com.
5. Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of Hallucination in Natural Language Generation. arXiv. https://arxiv.org/abs/2302.03636
6. OpenAI. (2024). GPT-4 Technical Report. https://openai.com/research/gpt-4
7. Ziegler, D., Brockman, G., & Christiano, P. (2023). GitHub Copilot and AI in Software Development. GitHub Docs. https://docs.github.com/en/copilot
8. Mistral AI. (2024). Open-Weight Language Models from Europe. https://mistral.ai

9. OECD. (2023). Generative AI and Intellectual Property Rights. https://www.oecd.org/digital