# Advancements in AI's ability to analyse visual data



In a recent exploration of the capabilities of artificial intelligence (AI), particularly in multimodal systems, the potential for AI to understand and analyse various forms of visual data has been highlighted. This investigation centres on the application of OpenAI's GPT-4o models, which have demonstrated an ability to process images alongside text, propelling advancements in fields such as data analysis, robotics, and practical everyday tasks.

To investigate the understanding of AI concerning visual data, various tests were conducted involving the GPT-4o and GPT-4o-mini models. These models can interpret and provide insights on a range of images, including photographs, maps, molecular structures, and more complex data plots. The tests were aimed at evaluating how effectively these models could analyse visual information and produce coherent summaries or analyses.

One of the primary tools used in this investigation is ChatGPT’s free version, which allows users to make up to four image analyses each day. Users can upload images directly within a conversational prompt, enabling a blend of visual and textual data processing. Alternatively, for those with programming skills, the OpenAI API provides access to process images programmatically, allowing developers to create applications that automate image processing tasks.

The first significant test leveraged a Google Maps screenshot to assess the model's ability to extract relevant geographical information. Upon querying the model regarding the contents of the image, it successfully identified essential navigation details, including landmarks and directions.

Following this, driving scenarios were tested. By uploading photographs of driving situations, the model evaluated road signs and conditions. For instance, in one analysis, GPT-4o-mini provided a substantial overview of what drivers should look for, including traffic signs and geographical particulars. The model advised to “pay attention to the following signs and features in the image,” demonstrating its capacity to assist drivers with safety information while on the road.

Additional testing ventured into the realm of robotics, where the model was prompted to guide a robotic arm to interact with objects based solely on visual input. The model accurately described the required movements for the arm to grab objects, incorporating details such as rotations and the estimated angles needed for successful manipulation.

In terms of data interpretation, the investigation also included analysing various types of data plots. For instance, in one example involving a tide chart, the model identified peak high tides with precision, demonstrating a logical understanding of the significance of the data. The experiment also delved into bioinformatics, with assessments made on common analysis plots within the field, where GPT-4o provided detailed interpretations of gene expression data, showcasing its adaptability across fields of scientific inquiry.

Notably, the experimental results indicated that the AI models occasionally struggled with specific details, particularly when images contained dense data or poorly resolved labels. Nonetheless, the overall findings suggest a promising trajectory for future applications of AI in visual data interpretation across various sectors, potentially reducing the burden on human analysts and experts.

Moreover, insights from Google's Gemini 2.0 Flash model revealed even greater refinement in visual analysis, often outperforming the GPT-4o models in specific evaluations. This exploration has engendered excitement around the continuing evolution of AI capabilities in multimodal contexts, signalling a burgeoning era in which visual data can be more readily understood and used as a powerful tool across disciplines.

The advancements presented in these findings reflect the innovative strides being made within the realm of artificial intelligence, encompassing a wide array of fields from transportation to complex scientific research, and pose intriguing questions regarding the future integration of AI in everyday tasks.

Source: [Noah Wire Services](https://www.noahwire.com)

## References

* <https://blog.roboflow.com/gpt-4o-vision-use-cases/> - This article provides insights into GPT-4o's capabilities, particularly its ability to process images and audio, highlighting its use in tasks such as OCR, object detection, and visual question answering.
* <https://www.techtarget.com/whatis/feature/GPT-4o-explained-Everything-you-need-to-know> - This source explains GPT-4o as a multimodal model that can understand and analyze visual content, including images and videos, offering advanced capabilities for integrating visual data into AI applications.
* <https://blog.mlq.ai/gpt-4-vision-data-analysis/> - This article discusses GPT-4 Vision's utility in analyzing and interpreting data from images, such as charts and graphs, highlighting its potential in data analysis and scientific research.
* <https://www.openai.com/> - OpenAI's official website provides general information on their AI models, including GPT-4 and its vision capabilities, which are central to the exploration of multimodal AI systems.
* <https://www.streamlit.io/> - Streamlit is a platform used for building interactive web applications. It can be used to integrate GPT-4 Vision for data analysis, facilitating the development of tools that analyze visual data.