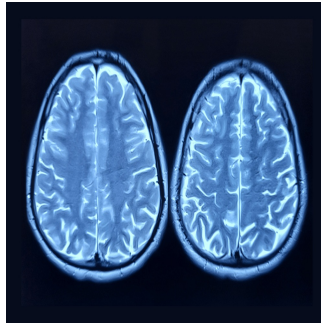

Optimising Multimodal Prompts for GPT-4V in Brain MRI Diagnosis



The rapid advancement of large language models (LLMs) has led to increasing interest in their application within medical diagnostics. While traditional text-based models have demonstrated potential in clinical decision-making, newer multimodal models such as OpenAI's GPT-4 with vision (GPT-4V) introduce the capability to process both textual and imaging data. This evolution raises expectations for improved diagnostic performance in radiology, where accurate interpretation of medical imaging is crucial. However, studies have indicated that the effectiveness of these models depends significantly on how input data is structured and presented. A recent study published in *Radiology* explored the impact of different multimodal prompt configurations on the accuracy of GPT-4V in diagnosing complex brain MRI cases. The findings highlight the importance of specific prompt elements, particularly textual descriptions of imaging findings, in enhancing the model's diagnostic accuracy. Understanding these insights is key to optimising AI-assisted radiology and ensuring its reliable application in clinical settings.

The Role of Multimodal Inputs in Diagnosis

GPT-4V's capability to analyse both images and text suggests that integrating multimodal inputs should enhance diagnostic accuracy. However, the study found that merely providing images—whether unmodified or annotated—resulted in extremely poor performance, with diagnostic accuracy as low as 2.2%. This suggests that, unlike human radiologists, who possess an innate ability to interpret visual patterns and detect abnormalities, LLMs require additional context to generate meaningful diagnostic insights. The inclusion of a patient's medical history provided some improvement, as it allowed the model to establish a clinical framework for its analysis. However, the greatest impact on diagnostic performance was observed when detailed textual descriptions of radiologic findings were included in the prompt. This finding highlights a fundamental limitation of LLMs in medical imaging analysis: they do not yet possess the ability to independently extract detailed pathological features from images with the accuracy and nuance required for differential diagnosis. Instead, they rely heavily on structured textual inputs to bridge the gap between raw visual data and their text-based reasoning processes.

Enhancing Prompt Structure for Improved Performance

The study evaluated seven distinct prompt configurations, each incorporating different combinations of four key elements: unmodified images, annotated images, medical history and textual image descriptions. The results demonstrated that the highest diagnostic accuracy (69%) was achieved when all four elements were provided as input. Notably, prompts that included an image description—whether combined with medical history or presented alone—consistently outperformed those that relied solely on visual data. The presence of a medical history contributed to a moderate increase in accuracy, but textual descriptions of imaging findings proved to be the single most influential factor.

Regression analysis confirmed this observation, showing that image descriptions had the strongest positive impact on diagnostic accuracy, followed by medical history, while image annotations provided no significant improvement. In contrast, providing only images—either unannotated or with annotations—resulted in very low accuracy, reinforcing the idea that GPT-4V lacks the capability to independently interpret medical imaging at a high level. These findings suggest that structuring prompts effectively is crucial for maximising the model's ability to process and contextualise radiologic data. They also imply that even in an AI-assisted diagnostic setting, the role of expert radiologists remains essential in generating high-quality input data.

Implications for AI-Assisted Radiology

The study's findings indicate that while GPT-4V has the potential to assist in differential diagnosis, its effectiveness depends largely on how input data is structured. The results emphasise that textual descriptions of imaging findings are vital for improving diagnostic accuracy, reinforcing the need for radiologists to play an active role in AI-assisted workflows. This underlines the continued necessity of expert human interpretation, as AI models alone cannot yet replace the analytical and contextual reasoning abilities of trained medical professionals.

appear to have been trained extensively on labelled radiologic datasets, which may explain its poor performance when analysing images without accompanying text. This suggests that fine-tuning LLMs on structured radiologic datasets, particularly those containing expert-generated image descriptions, could enhance their diagnostic accuracy. Additionally, integrating LLMs into clinical workflows requires consideration of their limitations, including susceptibility to errors and potential over-reliance by clinicians. While AI tools may offer valuable assistance, they must be used with caution and in conjunction with expert oversight to avoid misinterpretations that could affect patient care.

The study demonstrates that the accuracy of GPT-4V in brain MRI diagnosis is significantly influenced by the structure of multimodal inputs. Textual descriptions of radiologic findings play the most critical role, followed by medical history, while images alone contribute very little to diagnostic accuracy. These insights offer valuable guidance for optimising AI-assisted medical diagnostics and highlight the need for structured data and expert oversight. While GPT-4V and similar multimodal LLMs have the potential to support radiologists in differential diagnosis, their performance is highly dependent on how they are prompted. Future research should focus on refining AI training methodologies, evaluating domain-specific fine-tuning approaches and exploring ways to integrate multimodal LLMs into clinical practice in a way that complements, rather than replaces, human expertise.

Source: [Radiology](#)

Image Credit: [iStock](#)

Published on : Tue, 4 Feb 2025