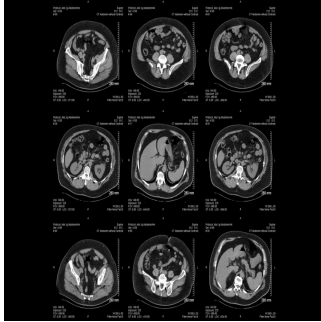


---

## Comparing GPT-4's and Gemini's Performance in Oncological Reporting



---

Integrating artificial intelligence (AI) into healthcare transforms traditional practices, offering innovative solutions to long-standing challenges. Among these, large language models (LLMs) such as GPT-4 and Gemini are emerging as pivotal tools in oncology. Their ability to analyse vast amounts of unstructured medical data and generate actionable insights is expected to alleviate clinicians' cognitive and logistical burdens. A recent comparative study highlights the potential of these LLMs to enhance oncological surveillance through the evaluation of radiological reports.

### The Role of LLMs in Oncology

Oncology relies heavily on serial imaging studies, such as computed tomography (CT) scans, to monitor disease progression, assess therapeutic efficacy and detect recurrences. These studies generate radiological reports, often presented in free-text formats. However, the lack of standardisation in these reports, coupled with the extensive data they encompass, can hinder efficient clinical decision-making. Oncologists frequently face challenges in extracting critical information from these reports due to inconsistencies in language, structure and terminology, especially when reports are produced by non-native English-speaking radiologists.

LLMs such as GPT-4 and Gemini offer a solution to these challenges. Structuring unstructured radiological reports enables clinicians to navigate complex data with greater ease. These models utilise natural language processing (NLP) to interpret and organise findings into actionable summaries, significantly reducing the time required for manual analysis. Moreover, they can identify subtle changes in tumour-related findings across serial reports, a task that requires meticulous attention to detail. Automating such tasks enhances efficiency and ensures that critical findings are not overlooked, potentially improving patient outcomes.

### Comparative Performance of GPT-4 and Gemini

The comparative study between GPT-4 and Gemini highlights the strengths and limitations of these LLMs in oncological applications. The study involved analysing consecutive radiological reports from patients undergoing abdominal CT scans. Key tasks included matching findings between sequential reports, identifying tumour-related issues and classifying tumour status into categories such as improved, stable or aggravated. Performance metrics such as accuracy, precision, recall and F1 scores were used to evaluate the models.

GPT-4 demonstrated superior performance, correctly matching 96.2% of findings between reports, compared to 91.7% for Gemini. Its precision and recall scores for identifying tumour-related findings were also higher, at 0.68 and 0.91, respectively, compared to Gemini's 0.63 and 0.78. This indicates GPT-4's greater reliability in pinpointing oncological issues. Furthermore, GPT-4 excelled in assessing tumour status, achieving accurate classifications in 87.6% of cases, compared to Gemini's 73%. This capability is critical for determining the progression or regression of a patient's condition, which directly informs treatment decisions.

Despite their strengths, both models exhibited limitations. A notable challenge was the high rate of false positives, where benign findings were misclassified as malignant. This issue stems from the models' inability to differentiate between benign and malignant medical terminology consistently. Additionally, false negatives were observed, particularly when low-probability malignancies were excluded from the analysis. While GPT-4 performed better overall, these findings underscore the need for further refinement of both models to enhance their accuracy in clinical contexts.

### Enhancing Clinical Integration

Several challenges must be addressed for LLMs to be fully integrated into clinical workflows. First, prompt engineering is crucial in guiding these models to generate accurate outputs. The study employed a customised prompt that broke down complex tasks into sequential steps, improving the models' performance. Future iterations of LLMs should focus on refining prompt designs to better capture the nuances of medical language, particularly in differentiating between benign and malignant findings.

Second, medical datasets are essential for training LLMs to handle domain-specific terminology. Currently, these models rely on general NLP training, which may not adequately cover the complexities of medical language. Fine-tuning LLMs with diverse and comprehensive datasets can reduce errors and improve their reliability. Additionally, incorporating multilingual datasets could address the language biases observed in this study, where reports were exclusively written in English by non-native speakers.

Lastly, the security and ethical implications of using AI in healthcare cannot be overlooked. Anonymising patient data and implementing robust data protection measures are critical for ensuring patient privacy. Developing AI systems that adhere to stringent regulatory standards will build trust among clinicians and patients, paving the way for broader adoption of these technologies.

The study comparing GPT-4 and Gemini illustrates the transformative potential of LLMs in oncology. By optimising the analysis of radiological reports, these models can reduce the cognitive load on clinicians and improve the accuracy of oncological surveillance. GPT-4, in particular, has demonstrated superior performance in matching findings, identifying tumour-related issues and assessing tumour status. However, the challenges of false positives and false negatives, as well as the need for domain-specific training, highlight areas for improvement.

LLMs' integration into clinical workflows holds immense promise. These models can revolutionise oncology and other medical fields by addressing existing limitations through refined prompt engineering, enhanced training datasets, and robust security measures. Embracing these advancements will enable healthcare professionals to deliver more efficient, precise and patient-centred care, ultimately improving outcomes for those facing the challenges of cancer.

**Source:** [Academic Radiology](#)

**Image Credit:** [iStock](#)

Published on : Mon, 16 Dec 2024