IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, December 2025
Impact Factor: 7.67

Enhancing Multi-Modal Understanding in Gemini-Based Large Language Models

Nikita Mate, Priti Borude, Pooja Garje
TE Students, Computer Engineering
Adsul Technical Campus, Chas, Ahilyanagar, Maharashtra, India

Abstract: *This paper presents an overview and analysis of multi-modal capabilities in Gemini-based Large Language Models (LLMs). Recent advancements in LLM architecture show that integrating text, image, audio, and video understanding into a unified framework significantly improves contextual reasoning and downstream task performance. This research discusses key components of Gemini's multi-modal encoder–decoder design, evaluates its real-world use cases, and highlights challenges related to computation, hallucination, and ethical risks. Recommendations for improving accuracy, reducing latency, and enhancing domain-specific reasoning are also proposed..*

Keywords: Gemini, Large Language Models, Multi-Modal AI, Deep Learning, Generative Models

DOI: 10.48175/568