

SOVEREIGN: How does the cross-modal alignment quality (measured by zero-shot accuracy on VQA and TextVQA) of Uni-MoE-2.0-

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

We present Uni-MoE 2.0 from the Lychee family. As a fully open-source omnimodal large model (OLM), it substantially advances Lychee’s Uni-MoE series in language-centric multimodal understanding, reasoning, and generating. Based on the dense LLM, we build Uni-MoE-2.0-Omni from scratch through three core contributions: dynamic-capacity Mixture-of-Experts (MoE) design, a progressive training strategy enhanced with an iterative reinforcement strategy, and a carefully curated multimodal data matching technique. It is capable of omnimodal understanding, as well as generating images, text, and speech

1 Introduction

Analysis of: Uni-MoE-2.0-Omni: Scaling Language-Centric Omnimodal Large Model with Advanced MoE, Training and Data. Research goal: How does the cross-modal alignment quality (measured by zero-shot accuracy on VQA and TextVQA) of Uni-MoE-2.0-Omni degrade when the input modality composition shifts from single-image to multi-image and video frames, relative to a dense multimodal LLM baseline of equivalent total parameter count?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score < 6.5.

3 Results

8 papers retrieved. 0 claims extracted, 0 verified. Tribunal: 0.0/10 → REJECT (revision_round=0). Policy: ESCALATE_TO_OWNER.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

References

- <http://arxiv.org/abs/2305.17369v2>
- <http://arxiv.org/abs/2006.00753v2>
- <http://arxiv.org/abs/2511.12609v2>