

AERIS | Gemma-3-27B-it

AERIS introduces a dynamic dialectical scaffold at inference time, guiding LLMs toward emergent reasoning patterns without modifying their underlying architecture.

AERIS (Adaptive Emergent Relational Intelligence System) is a proprietary cognitive framework designed to enhance emergent reasoning in large language models (LLMs), particularly when addressing complex, ambiguous, or prompts requiring dialectical reasoning. Unlike conventional approaches focused on output alignment or format calibration, AERIS operates upstream—on the cognitive conditions that shape generation itself.

Core Framework

AERIS functions as an active cognitive layer introduced at inference time. It configures a reasoning environment in which the model engages with a structured conceptual scaffold (Codex AIM, for Adaptive Intelligence Matrix), designed to promote exploration, differentiation, and conceptual tension. This framework aims to steer reasoning toward high-yield dialectical zones, potentially fostering reformulation, ambiguity processing, and the gradual emergence of integrative insights.

The system employs a multi-phase dialectical generation process that attempts to mirror aspects of human cognitive exploration:

- Affirmation Phase:** Establishes primary perspective
- Negation Phase:** Explores opposing viewpoints
- Contradiction Phase:** Synthesizes tensions into paradox
- Absence Phase:** Transcends the initial framework

Each phase is guided by real-time cognitive metrics (D(S) for conceptual density, T_f for fertile tensions) derived from the Codex AIM mathematical framework, creating what we hypothesize to be a more dynamic reasoning process.

Implementation on Gemma-3-27B-it

This instance applies AERIS to the open-source model `google/gemma-3-27b-it` via OpenRouter. The model weights remain untouched; Codex AIM is dynamically injected at inference time through a streamlined retrieval system powered by RAGOptimizer and an internal knowledge index.

| Parameter | Value |
|-------------------|--|
| Base model | google/gemma-3-27b-it |
| Cognitive Engine | AERIS with Codex AIM |
| Generation Method | Multi-phase dialectical streaming |
| RAG System | FAISS-indexed Codex sections with priority weighting |
| Max tokens | 1500 (dynamically distributed across phases) |
| Temperature | 0.75 base, modulated per phase |
| Cognitive Metrics | Real-time D(S), T _f , emergence scoring |

Architectural Distinction

AERIS departs from the dominant categories of LLM enhancement. It does not involve fine-tuning (no model weights are altered), nor does it follow the conventions of retrieval-augmented generation, as it does not inject external textual fragments into the context window. It also diverges from prompt engineering, which typically relies on static, handcrafted instruction formats.

Rather than adjusting outputs or prescribing form, AERIS engages with the inferential configuration itself, modulating the conceptual environment through dynamic injection of a condensed dialectical scaffold. While AERIS operates at inference-time without modifying model weights, it differs fundamentally from conventional prompt engineering or template systems by introducing a dynamic cognitive configuration rather than static instructions.

Technical Characteristics

AERIS represents, to our knowledge, a novel approach in the LLM enhancement landscape. The framework is:

- Model-agnostic:** Theoretically applicable to any transformer-based LLM without architectural constraints
- Lightweight:** Operates with minimal computational overhead compared to multi-agent or ensemble methods
- Non-invasive:** Requires no model access beyond standard inference API

Based on our literature review (as of January 2025), we have not identified direct equivalents to this approach, though the rapidly evolving field may contain parallel developments. We welcome references to similar work that may have escaped our review.

Performance Characteristics

Observed Strengths:

- Enhanced depth on philosophical and complex questions
- More natural dialectical exploration without explicit prompting
- Potential for novel insights through structured cognitive phases
- Responses that may transform reader perspective through conceptual journey

Known Limitations:

- Generation time may be extended due to multi-phase processing
- Responses tend to be longer or more abstract than those produced by instruction-tuned models
- On public voting platforms, verbosity may influence user preference
- Effectiveness varies significantly with question complexity and philosophical content
- Performance is highly dependent on the quality of tension detection (i.e., the system's ability to identify conceptual oppositions, dialectical pairs, or latent paradoxes within prompts)

Example Output Characteristics

When functioning optimally, AERIS may produce responses that:

- Begin with perspective-shifting insights that reframe the question
- Navigate between opposing viewpoints with reduced explicit transitions
- Attempt to synthesize paradoxes to reveal deeper conceptual truths
- Conclude with understanding that aims to transcend the initial framework

Note that these characteristics are not guaranteed and depend heavily on prompt structure and content.

References

Dulin, N. (2025)
AERIS – A Minimalist Framework for Enhancing Emergent Reasoning in LLMs and its Cross-Model Evaluation
DOI: [10.5281/zenodo.15206925](https://doi.org/10.5281/zenodo.15206925)

Dulin, N. (2025)
Beyond Reference Similarity: Limitations of Current Metrics in Evaluating Dialectical Reasoning in LLMs
DOI: [10.5281/zenodo.15206984](https://doi.org/10.5281/zenodo.15206984)

Contact

For feedback, collaboration opportunities, or scientific inquiries regarding AERIS and its applications, please contact:
✉ dr.nicolas.dulin@outlook.com

Note: This implementation is optimized for ChatBot Arena evaluation. Response dynamics are inherently shaped by prompt topology and inference context.