C7-ASM: A Multi-Sensory Tree-Based Cognitive Architecture with Adaptive Shortcut Mode

Author: Bahram (2025)

Preprint v1.0

Abstract

We introduce C7-ASM, a new cognitive architecture designed to overcome fundamental limitations of current large multimodal models (LMMs), which treat language, vision, and audio as a single undifferentiated computational stream. In contrast, C7-ASM adopts a biologically-inspired design: sensory pathways are fully separated, while higher-level cognition is performed in a shared, modular core. The architecture consists of: three independent sensory front-ends for audio, visual, and textual inputs; a unified three-layer embedding interface (Emb-A/B/C); a Prism-7 Router that dynamically activates a subset of cognitive modules; a seven-array cognitive core (C7): Meaning, Intent, World-Model, Reasoning, Memory, Behavior, Integration; and an Adaptive Shortcut Mode (ASM)—a functional state, not a structure—enabling parallel activation, selective suppression, and path skipping for efficient inference.

1. Introduction

Modern multimodal models attempt to integrate audio, visual, and textual signals inside a single unified computational pipeline. This monolithic architecture leads to modality interference, high computational cost, unstable reasoning, and inefficient scaling. Biological systems keep sensory pathways separated while higher cognition emerges only after independent feature extraction. Inspired by this, we propose C7-ASM.

2. Motivation and Contributions

Limitations of Current AI Architectures:

- Vision/Audio/Text encoders fused early
- Reasoning and memory not modular
- Intuition emerges as hallucination
- High cost and unstable behavior

Our Contributions:

1. Three independent sensory pipelines
2. Shared embedding interface
3. Branching cognitive tree
4. Seven interpretable cognitive arrays
5. Adaptive Shortcut Mode (ASM)

6. Practical implementation potential

3. Sensory Front-Ends (Independent Pipelines)

Audio Front-End processes raw waveform into spectral features, phonetic units, and prosody. Visual Front-End extracts spatial, semantic, and motion features. Text Front-End captures token embeddings, syntax, and local semantics. Sensory data remain strictly separated.

4. Unified Embedding Interface (Emb-A/B/C)

Emb-A: Form (structure)

Emb-B: Features (distinguishable traits)

Emb-C: Abstract semantic cues

All modalities map into a unified representational space.

5. Prism-7 Router

The router selects active cognitive arrays, assigns weights, and forwards payloads. It chooses processing paths but does not perform reasoning.

6. The C7 Cognitive Core

A1 – Meaning: concept extraction

A2 – Intent: user goal modeling

A3 – World Model: situation graphs

A4 – Reasoning: logical inference

A5 – Memory: long-term access

A6 – Behavior: tone and policy

A7 – Integration: unified cognitive state

7. Adaptive Shortcut Mode (ASM)

ASM is a processing state enabling:

- Parallel activation

- Selective suppression

- Path skipping

- Pattern-based inference

ASM provides intuitive-like reasoning without structural complexity.

8. Output Decoder

The decoder renders language or action outputs. Cognition occurs inside C7.

9. Discussion

C7-ASM offers biological plausibility, modularity, efficiency, reduced interference, and stable integration of reasoning, memory, and intent. It forms a foundation for next-generation cognitive systems.

10. Future Work

- Prototype implementation (10–50M params)
- ASM efficiency benchmarks
- Latency profiling
- Memory optimization
- Agentic extensions

11. Conclusion

C7-ASM proposes a separation of sensory processing and unification of cognitive operations. This yields robust reasoning, efficient inference, and natural intuitive behavior, aligning more closely with human cognition than existing multimodal transformers.