

Updating a Local-First Multimodal Memory Architecture to an Active/Passive Multi-Core Agentic System with a Graph Layer Executive summary This report updates an existing local-first multimodal memory assistant architecture (EchoGarden) into an active/passive multi-core agentic, multi-model system and strengthens it with a graph layer that simultaneously supports (i) user memory and semantic recall, (ii) multi-model orchestration, and (iii) durable execution / traceability. EchoGarden's current baseline already includes an ingestion pipeline (documents/images/audio), Memory Cards, hybrid retrieval (full-text + vectors + recency), grounded answering with citations, and explicit abstention/verification patterns, all deployed locally via a small service set and an inbox queue for asynchronous processing.