An Agentic Multimodal Architecture for Personal Knowledge Compilation, Epistemic Evolution, and Expert-Level Reasoning

Alberto Espinosa

January 16, 2026

Abstract

Personal knowledge collections increasingly span heterogeneous sources and modalities, yet remain epistemically static, fragmented, and prone to hallucination when used in contemporary AI systems. This work proposes a rigorous agentic architecture for personal knowledge compilation and expert-level reasoning, extended with a constrained deep external research agent to support epistemic evolution over time. The system integrates advanced multimodal knowledge acquisition, including structured video analysis, a command-line interaction interface for local-first operation, and optional diagrammatic explanation generation. A principled distinction between shallow operational agents and deep epistemic agents is enforced to preserve epistemic integrity, manage uncertainty, and avoid uncontrolled knowledge drift.

1 Motivation and Problem Statement

The central challenge addressed by this research is epistemic integrity rather than information access. Personal knowledge systems derived from saved references, academic papers, courses, and multimedia lectures are incomplete, temporally bounded, and vulnerable to hallucination when treated as closed worlds.

Existing retrieval-augmented systems assume static corpora and lack mechanisms for epistemic uncertainty detection, controlled external inquiry, and structured multimodal understanding. This project reframes personal AI systems as epistemic infrastructures rather than information tools.

2 Research Objectives

1. Design a universal ingestion and compilation architecture for heterogeneous personal knowledge artifacts.

2. Construct a multi-resolution internal knowledge representation supporting expert reasoning and pedagogy.

3. Integrate multimodal acquisition pipelines, including explicit video-level semantic analysis.

4. Explicitly model epistemic boundaries to minimize hallucination.

5. Introduce a principled deep external research mechanism for controlled knowledge evolution.

6. Provide a local-first, scriptable command-line interface for interaction and system integration.

3 Conceptual Distinction: Shallow vs Deep Agents

3.1 Shallow Agents

Shallow agents operate within fixed epistemic boundaries. They perform bounded transformations, do not introduce external knowledge, and do not reason under open uncertainty. These agents are responsible for ingestion, expansion, scraping, multimodal analysis, structuring, indexing, and presentation.

3.2 Deep Agents

Deep agents are permitted to cross epistemic boundaries. They formulate research-grade questions, reason under uncertainty, compare competing claims, and explicitly represent confidence and evidential strength.

Design Principle: Only agents that cross epistemic boundaries are permitted to be deep.

4 High-Level Architecture

The system consists of layered shallow agent pipelines orchestrated via a graph-based execution framework, complemented by a single constrained deep external research agent. User interaction is mediated through a command-line interface enabling composability with other local systems.

5 Layer 0: Universal Ingestion (Shallow)

5.1 Epistemic Question

"What reference has the user saved, independent of format or source?"

5.2 Canonical Source Descriptor

```
{
  "source_id": "uuid",
  "origin": "linkedin | browser | manual | other",
  "raw_reference": "original input",
  "raw_text": "optional extracted text",
  "links": ["url_1", "url_2"],
  "content_hints": {
    "contains_pdf": false,
```

```
    "contains_video": true,
    "contains_external_links": true
  },
  "timestamp": "ISO-8601"
}
```

6 Layer 1: Exploration and Expansion (Shallow)

6.1 Epistemic Question

"What actual knowledge artifacts exist behind this reference?"

This agent identifies referenced papers, videos, courses, or datasets and dispatches specialized acquisition pipelines accordingly.

7 Layer 2: Knowledge Acquisition and Multimodal Scraping (Shallow)

This layer acquires primary knowledge materials and performs modality-specific extraction. For video content, the system integrates a dedicated analysis pipeline based on the open-source project **video-analyzer** (`https://github.com/byjlw/video-analyzer`), which provides:

- Shot and scene segmentation
- Transcript alignment
- Visual-semantic frame extraction
- Multimodal chunk generation

This pipeline transforms long-form video lectures and courses into structured knowledge units suitable for downstream representation and reasoning.

8 Layer 3: Multi-Resolution Knowledge Representation (Shallow)

Knowledge is stored explicitly at four abstraction levels:

- Level 0: Raw content chunks (text, frames, transcripts)
- Level 1: Section-level structured representations
- Level 2: Document- or lecture-level synthesis
- Level 3: Cross-document thematic synthesis

Lower levels provide factual grounding; higher levels support reasoning, comparison, and pedagogy.

9 Layer 4: Classification and Ontology Construction (Shallow)

This layer constructs topic hierarchies, prerequisite relationships, and redundancy mappings, enabling selective reasoning without exhaustive semantic search.

10 Layer 5: Expert Reasoning and Teaching (Shallow-Composite)

This agent composes explanations strictly from internal representations. It dynamically selects abstraction levels but is epistemically constrained to the compiled knowledge base.

11 Layer 6: Diagrammatic Explanation Agent (Shallow)

When requested or pedagogically beneficial, this agent generates schematic diagrams, conceptual graphs, or explanatory visuals to accompany textual explanations. It operates exclusively on internal representations and produces images as explanatory artifacts rather than sources of new knowledge.

12 Layer 7: Deep External Research Agent (Deep)

12.1 Purpose

The Deep External Research Agent (DERA) prevents epistemic stagnation and hallucination under temporal or conceptual uncertainty.

12.2 Epistemic Question

> **"Does authoritative external knowledge exist that materially alters or invalidates the current internal understanding?"**

12.3 Operation

DERA is activated only under explicit trigger conditions and produces comparative epistemic reports. It annotates and versions internal knowledge but does not overwrite it. Its design is informed by the Deep Agent paradigm described in recent literature (`https://huggingface.co/papers/2510.21618`).

13 User Interaction Layer

User interaction is mediated through a command-line interface built upon the `llm` framework by Simon Willison (`https://github.com/simonw/llm`). This interface supports:

- Scriptable queries and reproducible workflows
- Integration with other local tools and pipelines
- Multimodal output, including text and generated diagrams

The CLI is treated as a first-class architectural component rather than a peripheral interface.

14 Technology Stack

- Agent orchestration: LangGraph (`https://github.com/langchain-ai/langgraph`)
- Local model execution: Ollama (`https://github.com/ollama/ollama`)
- Text models: LLaMA 3.x (`https://ai.meta.com/llama/`), Qwen (`https://github.com/QwenLM`), Gemma (`https://ai.google.dev/gemma`)
- Multimodal models: LLaMA 3.2 Vision
- Video analysis: `video-analyzer` (`https://github.com/byjlw/video-analyzer`)
- CLI interface: `llm` (`https://github.com/simonw/llm`)
- Vector storage: FAISS (`https://github.com/facebookresearch/faiss`) or Qdrant (`https://qdrant.tech`)
- Structured storage: relational and graph databases

15 Project Structure

```
project_root/
  ingestion/
  exploration/
  acquisition/
    video_analysis/
  representation/
  ontology/
  reasoning/
  diagramming/
  deep_research/
  interface/
  orchestration/
  storage/
```

16 Conclusion

By integrating explicit multimodal video analysis, diagrammatic explanation, a command-line interaction layer, and a constrained deep external research agent within a rigorously layered architecture, this system advances personal AI from retrieval-oriented tools toward a robust epistemic infrastructure. The explicit separation between shallow operational agents and deep epistemic agents preserves reliability while enabling principled knowledge evolution and expert-level pedagogy.