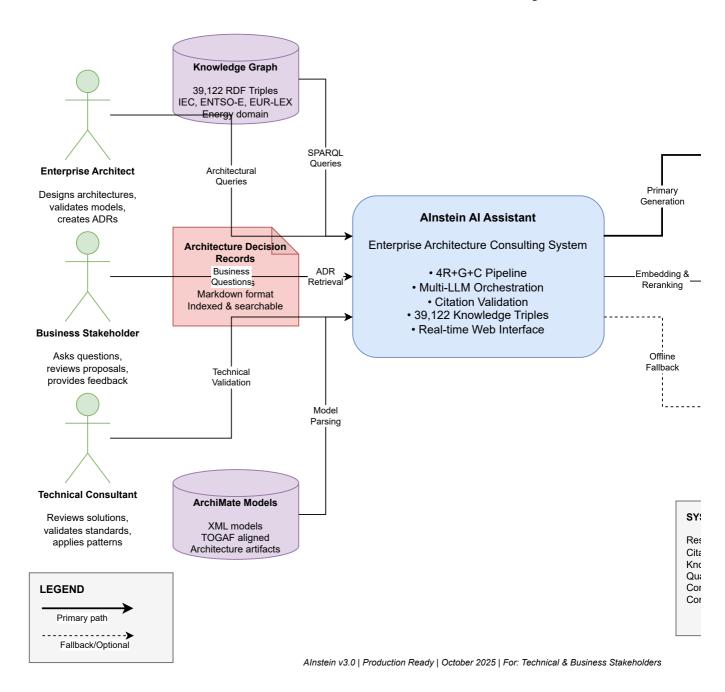
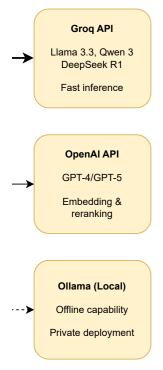
AInstein AI Assistant - System Context





STEM METRICS

sponse Time: < 3s P50

ation Accuracy: 100%
owledge Coverage: 39,122 triples
ality Improvement: +15-20% (API reranking)
nfidence Threshold: 0.75

ncurrent Users: 50+

Alnstein 4R+G+C Pipeline Architecture

User Prompt

PHASE 1: REFLECT Query Analysis & Intent Detection

Query Parser

- Tokenization
- · Entity extraction
- Intent classification

Comparison Detector

- Pattern matching
- Extract terms
- Validate structure

Context Analyzer

- Session history (3 turns)
- Extract concepts
- Build context

PHASE 2: ROUTE LLM-Based Intelligent Query Routing

Homonym Detector

- · Check ambiguous terms
- Domain context
- Disambiguation

Domain Router

- · Identify domains
- Select sources
- Priority ranking

ADR Matcher

- Pattern matching
- Number extraction
- Content search

PHASE 3: RETRIEVE

Multi-Source Knowledge Retrieval

Knowledge Graph

SPARQL queries 39,122 triples IEC definitions

ArchiMate

XML parsing Model elements TOGAF mapping

ADRs

13 decisions Markdown Indexed

Embeddings

Semantic search (pkl) OpenAl textembedding-3-small

API Rerank

OpenAl eml +15-20% qu text-embedo

PHASE 4: REFINE Multi-LLM Response Generation

Prompt Builder

- EA-specific templates
- Inject candidates
- Citation constraints

LLM Provider

Primary: Groq (Llama 3.3) Fallback: OpenAI (GPT-5) Local: Ollama

LLM Council

- Consensus building
- Quality voting
- Conflict resolution

PHASE 5: GROUND

Citation Validation (Zero Tolerance for Fake Citations)

Grounding Check

- Extract all citations
- Mandatory validation
- No ungrounded text

Citation Validator

- Pre-loaded pools (3,970+)
- · Authentic sources only
- Pattern matching

Fingerprint Check

- Vector validation
- Semantic accuracy
- Content match



king

beddings uality ding-3-small



PHASE 6: CRITIC Quality Assessment & Confidence Scoring

Quality Assessor

- Completeness check
- · Relevance scoring
- Accuracy validation

Confidence Calculator

- Source quality weighting
- Coverage analysisUncertainty quantification

Human Review Trigger

- Threshold: 0.75
- Auto-flag low confidence
 Audit trail logging

✓ Cited answer with authentic sources
✓ High confidence score
✓ Low response time (SLAs)
✓ Minimal human input

Minimal human input needed Full audit trail recorded

Alnstein v3.0 | 4R+G+C Pipeline | Production Implementation | October 2025

