AI Conference Technical Notes (2026-03-24)

AI Conference Technical Notes (2026-03-24)

Architecture Overview

Deep-dive notes from the 2026 All Things AI conference sessions centered on robust, production-grade agentic systems.

1. Pydantic AI & Logfire (Observability Stack)

Pydantic AI is the standard for type-safe agentic architectures. Integration with Logfire creates **Context Graphs**, turning logging into a primary observability tool. This allows for clear visualization of agent decision-making traces.

2. Three-Tiered Memory Architecture

A move away from LLM-centric memory management towards a structured, verifiable storage system:

- **Tier 1 (Profile Summary):** High-level user context. **Rule:** LLM access forbidden.
- **Tier 2 (Facts Table):** Individual, extracted data points. Verification requires a deterministic, non-AI script to validate against raw source transcripts.
- **Tier 3 (Graph Store):** Relational data linking facts.

3. The "Smart Gateway" Philosophy

- **Rule:** The LLM should only intervene when human-level reasoning is strictly required.
- **Application:** The gateway handles critical routing, security, and tool invocation using deterministic code. The LLM acts as a service to the gateway, never the primary driver of infrastructure.

4. Knowledge & Context Graphs

- **Vector DBs:** Best for semantic similarity search.
- **Knowledge Graphs (e.g., Neo4j):** Essential for logical relationships and grounding LLM responses, avoiding "hallucination by confidence."

5. Prototyping

Marimo identified as a preferred "reactive" alternative to traditional cell-based Jupyter flows for executable rapid prototyping.

6. Hypothetical Document Embedding (HyDE)

A high-accuracy retrieval pattern: instead of embedding the user's question, the system generates a "fake answer" via the LLM, embeds that, and searches the document store using "answer-to-answer" matching.

7. The Three Laws of Agentic AI

Reframing security as "Rules of Conduct" for autonomous agents:

- **Law #1 (Identity):** "An agent must always maintain a unique, verifiable identity & must never act anonymously or under a masked provenance."
- **Law #2 (Delegated Permissions/Scope):** "An agent must only perform actions within its explicitly delegated permissions, except where such actions would violate the Law of Identity."
- **Law #3 (Credential Integrity):** "An agent must protect its own credentials and session tokens, as long as such protection does not conflict with the 1st or 2nd laws." This incorporates CIBA (Client Initiated Backchannel Authentication) and human-in-the-loop workflows for high-risk actions.

Implementation Strategy (Homelab):

- **Law #1 (Provenance):** Log client_id and session_id tracking in ClickHouse; enforce strictly at the API layer.
- **Law #2 (Permissions):** Utilize OIDC Claims within token scopes (e.g., manage_vlan, read_docker_logs). Deny any action not explicitly claimed.
- **Law #3 (Integrity):** Implement "Out of Band" (OOB) validation via Matrix/push-notifications for high-risk commands (e.g., database deletion or critical system configuration).

Revision #1
Created 2026-03-31 11:55:51 UTC by Flobot
Updated 2026-03-31 11:55:51 UTC by Flobot