LLM SYSTEMS · INFORMATION RETRIEVAL · EVALUATION

What does it take to trace a grounded answer?

Abstract. A streaming answer may look fluent long before it is trustworthy. The search harness already carries a question through bounded conversation context, grounding sources, enabled connectors, synthesis, and SSE telemetry. The research gap is measuring whether that trace is actually sufficient.

THE PROBLEM

Which evidence, tool, and context events must be retained so a person can inspect why an answer was given—without exposing secrets, chain-of-thought, or unnecessary private context?

HARD CONSTRAINTS

01 Every tool schema is provider-adapted before a model can call it.

02 Connector output, workspace reads, and context size are bounded.

03 A failure must degrade visibly, not fabricate a source or capability.

04 Latency and traceability must both survive chained follow-up questions.

A FIRST CONTRIBUTION

Create a replayable eval corpus with known source conflicts, missing evidence, slow connectors, and follow-ups that change the question rather than merely append text.

BUILD IN THE OPEN

github.com/hushh-labs/hushh-research
hushh.ai/discord

PRIMARY SOURCE · docs/AGENT_ARCHITECTURE.md

Implemented architecture; actual connector enablement is deployment-specific.

Read the research
hushh.ai/research/open-problems