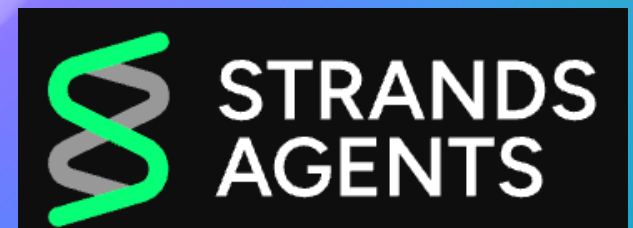


Lab 6: Observability with LangFuse and Evaluation with RAGAS



Why Observability and Evaluations Matters?

Observability

- X-ray vision into your AI agent
- See thought processes and tool usage
- Real-time debugging capabilities
- Trace execution paths

Evaluation

- Measure agent performance objectively
- Use specific metrics instead of guesswork
- Track improvements over time
- Ensure quality standards

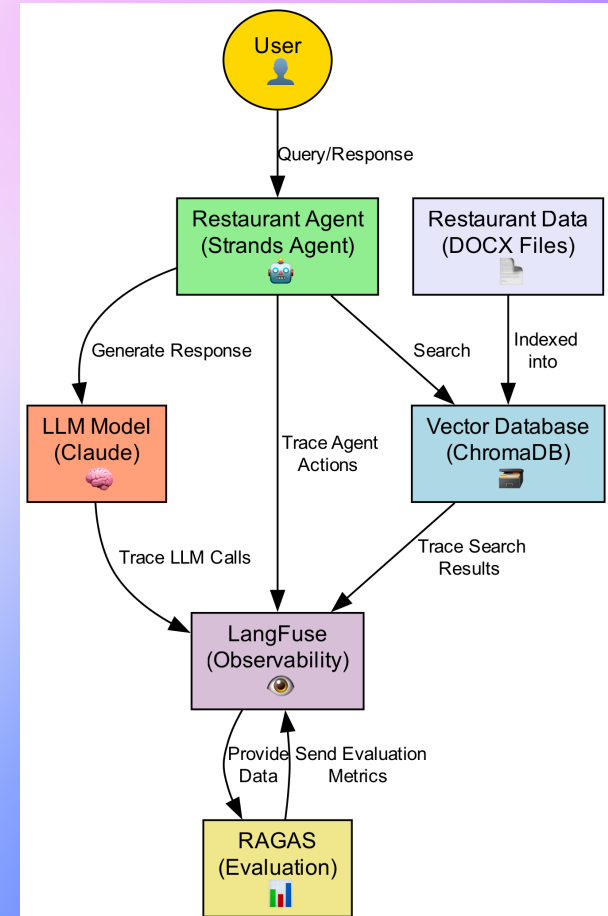
Observability in the Strands Agent

- **Metrics:** Performance and resource usage numbers
- **Traces:** Detailed execution path records
- **Logs:** Operation and event text records
- **Evaluation:** Performance measurement methods

Strands integrates with **OpenTelemetry**

Restaurant Agent Architecture

- A user **asks** a question about restaurants
- Our agent searches the **vector database** for relevant information
- The agent generates a response **based on the search results**
- Trace with **LangFuse**
- Evaluate the quality of the response using **RAGAS** metrics



RAGAS Metrics

- **Aspect Criteria metric:** a simple yes/no evaluation
- **Rubric Score metric:** a set of different scores, each with an explanation of what it means
- **To evaluate the RAG:**
 - ContextRelevance
 - ResponseGroundedness

By implementing observability and evaluation:

- Debug issues more effectively
- Measure and improve your agent's performance over time
- Ensure your agent meets quality standards
- Gain insights into how users interact with your agent

Thank You

