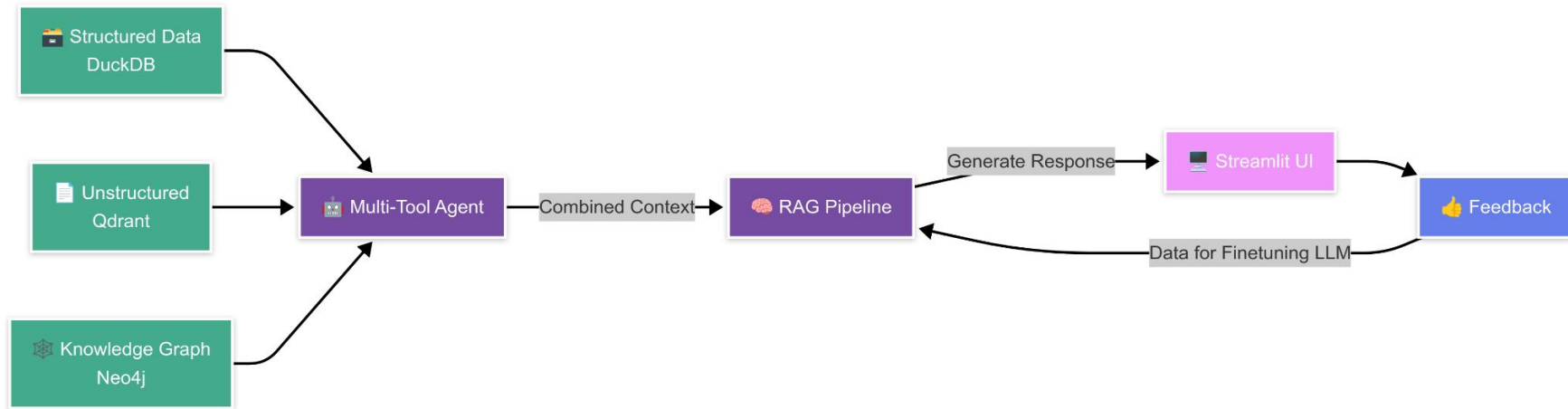


Allyin Compass

Multi-Tool AI Agent that searches across structured databases, unstructured documents, and knowledge graphs to deliver comprehensive, source-attributed answers.

System Architecture

- Ingest CSVs, PDFs, emails, graphs
- Tool-chaining agent picks best search tool(s)
- Retrieval-Augmented Generation (RAG) for LLM-based answers
- Modern UI for queries, feedback, and analytics



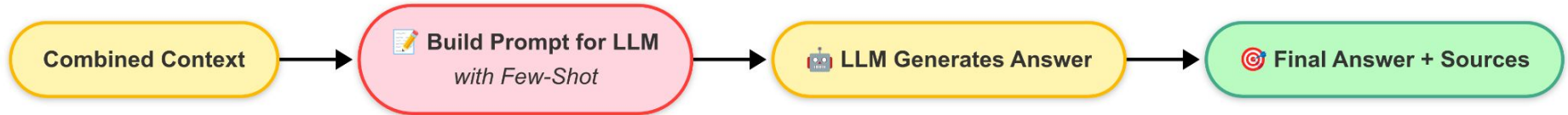
Multi-Tool Agent Logic

- Parses queries and selects best tool(s): SQL, vector search, graph
- Executes searches, combines results
- Enriches with domain-specific templates (finance, biotech, energy)
- Sends combined context to RAG for LLM answer

```
self.domain_focus = {  
    "Finance": "Focus on financial metrics, revenue, risk analysis, and regulatory compliance.",  
    "Biotech": "Emphasize research findings, clinical data, safety protocols, and lab compliance.",  
    "Energy": "Highlight environmental impact, emissions data, facility operations, and sustainability."  
}
```

Retrieval-Augmented Generation (RAG) Flow

- Retrieves context from SQL, docs, and graphs
- Built prompt for LLM with citations, few shot examples to improve answer quality
- Generates answer, along with sources



Compliance & Security

- PII detection (email, SSN, phone) with regex
- Risk/compliance keywords automatically flagged
- All actions logged for traceability



PII Detection

SSN, emails,
phone
numbers
automatically
masked



Risk Flagging

Identifies
terms like
"restatement",
"violation"

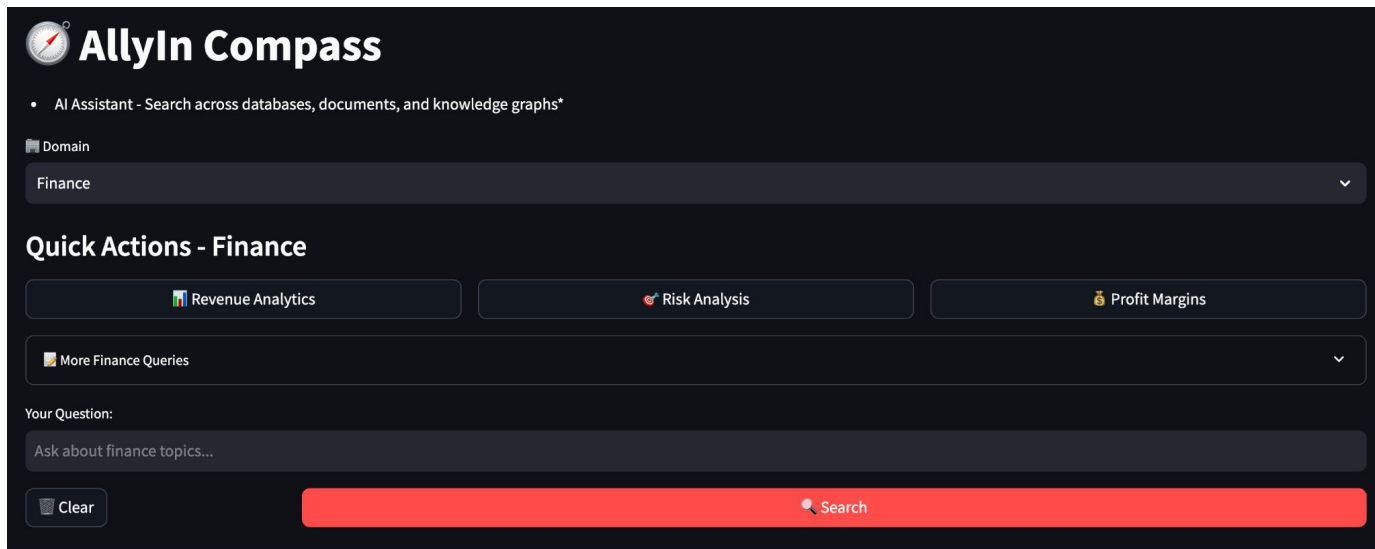


Audit Trail

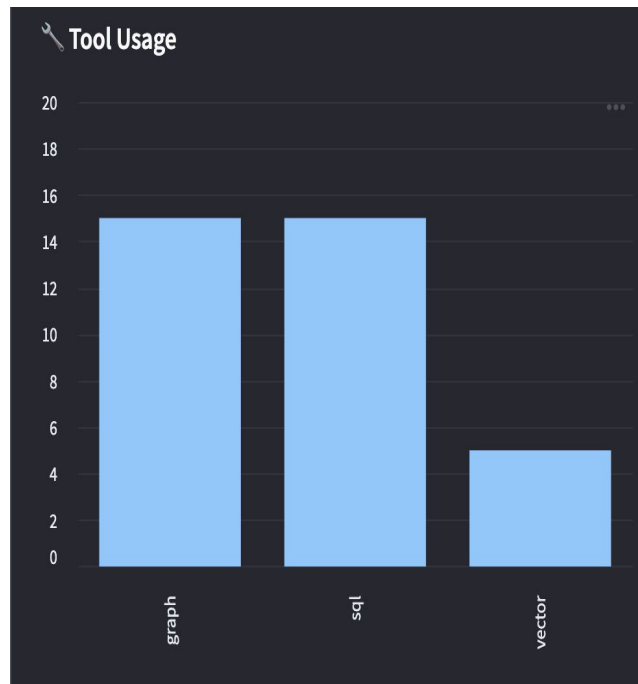
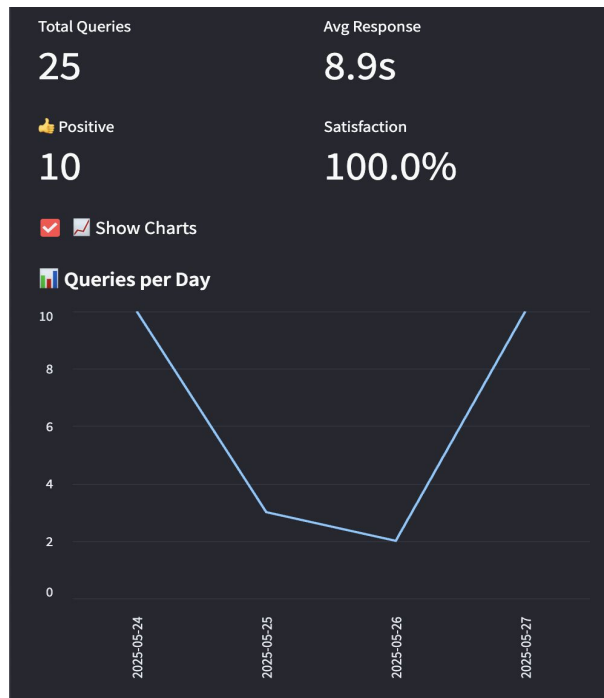
Complete
logging of
queries
and data
access

UI, Feedback, and Observability

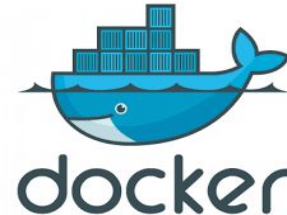
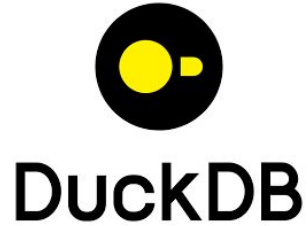
- **Streamlit app:** Query bar, domain selection, live answer view and sources
- **Feedback:** Thumbs up/down system collects training data for continuous improvement
- **Sidebar dashboard:** Usage stats, response time, tool trace



Sidebar Dashboard



Key Technologies & Modules



Sample query

Your Question:

Adverse outcomes for molecule X

Context

Sources & Context

****Document Search:****

****Found 3 relevant documents:****

****1. Unknown** [doc] (Score: 0.696)**

Subject: SAFETY ALERT: Molecule X (BGX-4521) - Immed

****2. Unknown** [doc] (Score: 0.624)**

ert: Molecule X (Development Code: BGX-4521) Our com

****3. Unknown** [doc] (Score: 0.605)**

on: 8 cases o Myocardial infarction: 3 cases (1 fata

****Knowledge Graph:****

Available graph queries: compliance violations, custome

Answer

Answer

Adverse Effects for Molecule X (BGX-4521)

- Total Adverse Events Reported: 127
- Serious Adverse Events (SAEs): 34 (26.8%)
- Deaths Possibly Related to Molecule X: 3
- Hospitalizations: 18

Specific Adverse Outcomes:

- Myocardial Infarction: 3 cases (1 fatal)
- Neurological Complications:
 - Severe Headaches: 45 patients
 - Seizures: 5 patients (no prior history)
 - Peripheral Neuropathy: 12 cases
 - Cognitive Impairment Reports: 9 patients

Site-Specific Adverse Outcome Patterns:

- Genomics Plus (Cambridge): Highest rate at 31% of patients
- BioMed Research: Specific data not provided in the context

Note: Immediate review and attention required based on aggregated adverse event data from all trial sites.

Details

Domain: Biotech

Tools Used: Vector + Graph

Time: 3.44s

Rate Answer



In real world scenarios

Finance

"Top flagged clients in last quarter"

Searches risk scores, compliance violations, and financial metrics

Biotech

"Adverse outcomes for molecule X"

Analyzes clinical trials, safety reports, and research documents

Energy

"CO2 violations near San Jose since Q1"

Queries emission data, facility locations, and compliance records

CHALLENGES FACED

- **Database Compatibility:** LangChain didn't support DuckDB dialect, built SQLAlchemy bridge
- **Knowledge Graph Integration:** Neo4j retriever returned hardcoded strings instead of live queries
- **Tool Routing:** No domain-aware routing initially, refined tool selection logic
- **Response Quality Hallucination Crisis:** System ignored SQL results, fabricated responses instead of using real data, lost specific customer names and revenue figures in generic corporate responses
- **Multi-Domain Sample Data:** Diverse datasets across Finance, Biotech, Energy sectors
- **Domain-Specific Knowledge Graphs:** Relationship networks tailored to each business vertical
- **UI/UX (Session State Management):** Query results and user sessions not persisting correctly

THANK YOU