Optimizing LangChain Pipelines for Trust, Observability, and Compliance

Audit-Ready RAG for Banking and Regulated Domains

Sivakumar Mahalingam

Why Trust, Compliance, and Observability Matter?

TRUST

Prevent hallucinations & enforce citations

COMPLIANCE

Mask PII, follow jurisdiction rules, avoid policy violations

OBSERVABILITY

Full traceability for audits and debugging

Optimization Importance

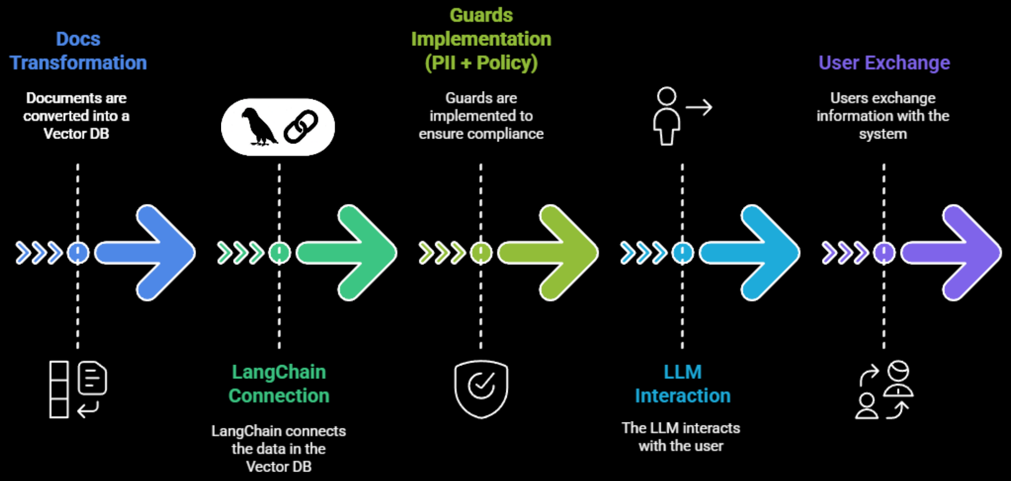
The Banking Challenge

- Analysts and agents need fast answers (KYC, AML, fees)
- Knowledge is scattered: PDFs, policies, SOPs, product docs
- Wrong or non-compliant answers = regulatory risk

Naive pipelines suffer from **latency**, hallucinations, and lack of auditability, impacting reliability and user confidence.

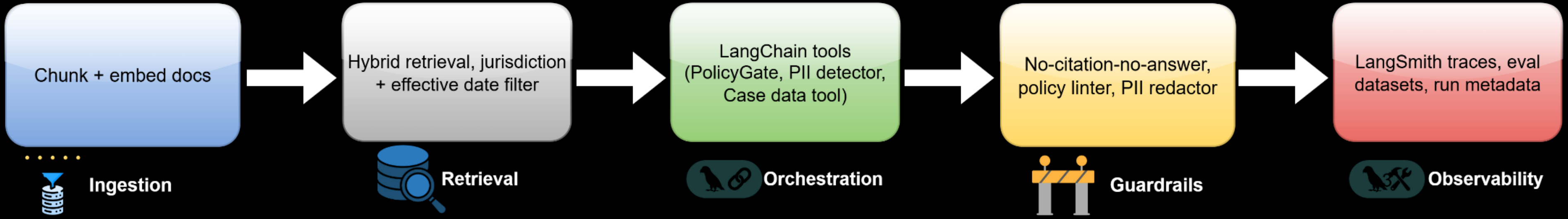
Audit-Ready RAG Copilot

- Retrieval-Augmented Generation with guardrails
- Jurisdiction-aware policy enforcement
- Full LangSmith traceability & replay


Docs Transformation
Documents are converted into a Vector DB
LangChain Connection
LangChain connects the data in the Vector DB
Guards Implementation (PII + Policy)
Guards are implemented to ensure compliance
LLM Interaction
The LLM interacts with the user
User Exchange
Users exchange information with the system

This is not just RAG, it's **RAG with governance**
Ensures every answer is backed by citations

Architecture Overview


Chunk + embed docs
Ingestion
Hybrid retrieval, jurisdiction + effective date filter
Retrieval
LangChain tools (PolicyGate, PII detector, Case data tool)
Orchestration
No-citation-no-answer, policy linter, PII redactor
Guardrails
LangSmith traces, eval datasets, run metadata
Observability

Code Repository Layout

```
app/
├── chains/      # LangChain graphs & tools
├── guards/      # PII, policy linter
├── retrieval/   # Indexer, retriever
├── policies/    # Policy packs
├── schemas/     # Pydantic models
├── server/      # FastAPI endpoints
└── telemetry/   # LangSmith client
```

How Guardrails Work?

- **PII Redaction:** Masks PAN, IBAN, SSN, EID pre + post generation.
- **Policy Linter:** Blocks banned phrases like “guaranteed approval”
- **No Citation, No Answer:** Refusal if no supporting docs

Example:

- Input: “4242 4242 4242 1234”
- Output: “**** **** **** 1234”

Guardrails — the backbone of trust and compliance in this system

Observability with LangSmith?

- Every run logged with:
  - Inputs, retrieved docs, tool calls, outputs
  - Metadata: jurisdiction, policy_pack_version, timestamp
- Replay past runs for audits

THANK YOU

Questions & Discussions