ComplAI

Towards a Scalable Multi-Agent System for Automated Financial Compliance Reporting

Students Team: Afag Ramazanova, Hongyi Duan, Mobasserul Haque, Tursunai Turumbekova, Zijian Huang

2026 Duke MIDS Capstone Symposium

GitHub

Introduction and Background

- Financial institutions spend 3-5% of annual revenue on AML compliance, with large global firms investing $500M–$1B annually.
- A significant portion of the $206B global compliance cost is driven by manual data aggregation and narrative drafting.
- Current AML reporting processes are highly inefficient, time-consuming, and resource-intensive.
- There is a critical need to automate the transformation of complex transaction data into actionable regulatory narratives.

Research Question: Can a multi-agent system automate end-to-end regulatory reporting with >95% accuracy, auditability, and ~90% reduction in operational overhead?

To address this, ComplAI proposes a specialized five-agent architecture that integrates LLM-based semantic understanding with deterministic validation.

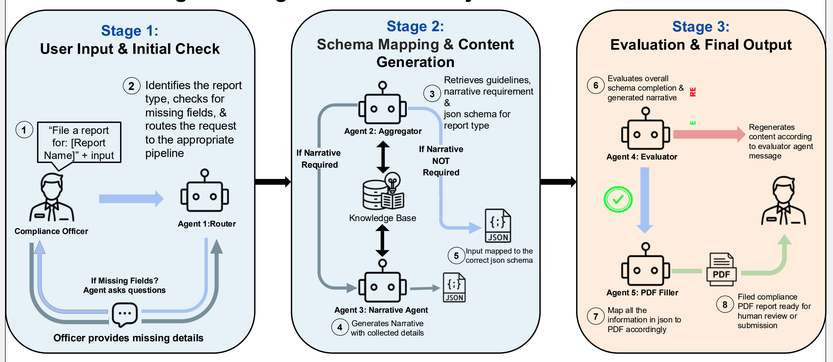
Methodology

ComplAI implements service-oriented design with FastAPI backend, CrewAI orchestration, Streamlit frontend, and hybrid knowledge base infrastructure combining Weaviate(vector), PostgreSQL(relational), and Redis(cache) DBs.

Five specialized agents coordinated via CrewAI to enable modular, end-to-end report generation.

1. **Agent 1 - Router**: Classifies report types and requirements based on regulatory thresholds.
2. **Agent 2 - Aggregator**: Transforms and maps raw data into structured report fields.
3. **Agent 3 - Narrative Generator**: Produces compliant narratives using LLM APIs with grounded prompts.
4. **Agent 4 - Evaluator:** Assesses reports across multiple quality dimensions (e.g., accuracy, completeness).
5. **Agent 5 - PDF Filler**: Generates final PDFs and records audit logs for traceability.

Diagram 1. Agentic Workflow System Architecture



Results

Empirical evaluation on n=100 synthetic cases (50 SAR (Suspicious Activity Report), 25 CTR (Currency Transaction Report), 15 cases that classify for filing multiple report types, 10 edge cases) demonstrates statistically significant improvements over baseline methodologies (manual filing).

Table 1. Agentic Workflow Comparative Evaluation

Metric	Manual*	Raw LLM	ComplAI
Latency	165 min	12.4 min	1.42 min
Field Accuracy	88.4%	82.1%	99.2%
Audit Traceability	Low (Manual)	Black-box	100% (Step-level logging)
Cost per Report	$45	$1.20	$0.45

➢ Quantitatively, the system achieved an operational throughput increase of 116x, **reducing** mean end-to-end **latency from 165 minutes to 1.42 minutes.**

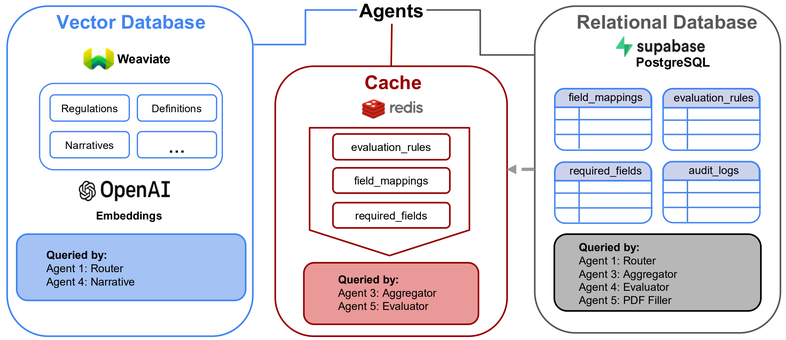
➢ **Cost** per report **decreased** from approximately $45.00 to $0.45, enabling substantial labor savings.

➢ Improved field-level accuracy, meeting the strict requirements of regulated financial institutions.

➢ Relational database ensured deterministic rule enforcement, maintaining compliance with financial regulations.

➢ Balanced outcomes: nuanced, high-quality narratives alongside strict adherence to legal requirements.

Diagram 2. Knowledge Base Architecture



** Numbers are averaged according to open source resources mentioned in references*

Results

ComplAI demonstrates strong alignment with human compliance officer diagnostic standards by providing structured, step-by-step logs throughout the report generation process. Unlike traditional black-box AI systems, ComplAI ensures end-to-end audit traceability, enabling full transparency into how each output is generated.

This evaluation highlights the potential of agentic auditing frameworks to scale compliance operations efficiently, allowing financial institutions to maintain high-quality assurance standards without a proportional increase in specialized human compliance personnel typically required for manual review processes.

Conclusions and Future Work

ComplAI demonstrates how multi-agent AI can significantly reduce the time and cost of compliance reporting by automating repetitive, manual tasks while maintaining auditability and control.

By combining retrieval-augmented generation with specialized agents, the system enables faster, more consistent report creation grounded in regulatory requirements, helping organizations scale compliance operations without a proportional increase in staffing.

Importantly, the system is designed to support compliance officers-not replace them by streamlining workflows and enabling faster, more informed decision-making. Human oversight remains critical for high-risk or ambiguous cases to ensure regulatory trust and accountability.

Future work will focus on introducing a Researcher Agent capable of continuously ingesting and updating regulatory information, enabling the system to adapt in real time to evolving compliance requirements and further improving scalability in dynamic regulatory environments.

References

FinCEN. (2025, October). *Guidance on Responsible Use of AI in AML/CFT Programs*. U.S. Dept. Treasury.
FinCEN. (2020) Suspicious Activity Report (SAR) Guidance.
Moody's (2025). Global Financial Crime Compliance Survey Report.
Flagright. (2025). *High Cost of AML Non-Compliance: Global Banking Benchmarks*.
LexisNexis. (2025). *True Cost of Financial Crime Compliance Study: U.S.*
Lewis, P., et al. (2020). *Retrieval-Augmented Generation*. NeurIPS, 33.
Zheng, L., et al. (2024). *Judging LLM-as-a-Judge*. NeurIPS.