AgentFinance: A Multi-Agent Financial Advisory System with Retrieval-Augmented Generation

Hussain Ahmed, Asad Ullah, Sami Naeem
Department of Artificial Intelligence
National University of Computer and Emerging Sciences FAST Islamabad, Pakistan
{i221374, i220589, i220587}@nu.edu.pk

***Abstract*—The provided paper proposes AgentFinance that is a multi-agent financial advisory system is the aggregation of three agents specialists (Tax, Investment, Data Analyst) arranged at the occasion of a central Coordinator agent and facilitated by a Retrieval-Augmented Generation (RAG) subsystem. The system shall be created in order to provide integrated as to risky areas, e.g. tax optimization, consultation, economic forecasting investment analysis and investment analysis. The architecture connects to a relational database on 18 tables, 183, 777 records and an all-purpose vector store of 1,794 contained documents in all 3 specialist agents (Tax: 1,290 , Investment: 280 , Analyst: 224).Comprehensive evaluation in three test suites there was better performance, a 94.3% routing accuracy and a figure of close accuracy of $0.09 tax Mean Absolute Error (MAE). The system maintained a low 742 ms latency at P95 , 2.7 times faster than desire, and out of 60 test questions, 100 percent success was realized.**

***Index Terms*—Multi-agent systems, Retrieval Augmented Generation, Financial advisory, Tax optimization, Investment analysis, ChromaDB, SQLite**

I. Introduction

Providing accurate, multi-domain financial advice requires combining complex regulatory knowledge (IRS tax regulations), volatile market data, and dynamic economic analysis. Traditional systems often lack the necessary integration and depth of expertise across these domains, presenting a significant barrier to accessible and holistic financial guidance. AgentFinance addresses this need by proposing a modular, autonomous multi-agent architecture. A central Coordinator routes queries and manages context while specialized agents (Tax, Investment, Data Analyst) execute domain-specific tasks using dedicated tools.

The multi-domain financial advice which is provided will need the integration of complex regulatory expertise (IRS tax laws), volatile financial market data, and evolving economic studies.The traditional systems are in most cases deficient of the required integration and profundity of insight in these areas, which is an enormous obstacle to the availability of comprehensive and nonjudgmental financial direction. The concern brought up by AgentFinance is to disrupt this necessity by suggesting a modular and autonomous multi-agent architecture.Queries are directed via a central Coordinator and context management is done by specialized agents (Tax, Investment, Data Analyst) carry out specialized functions with the help of specialized tools.

The novelty of the system is seen in the fact that it is a system that operates with a single method of the integration of the calculating, database retrieval, and knowledge retrieval functions in professional areas that are different in nature.Overview The most significant contributions with this work are summarized as follows:

1) **New Architecture:** Testing and development of a Coordinatorbased multi-agent financial system that has three specialised agents Tax, Investment and Data Analysis.
2) **Full RAG Integration:** The Semantic search system combined 1,794 chunks of authoritative documents in its search, and it provides RAG Total of all specialist agents to be accurate and reflected on the market.
3) **Data Integration at Scale:** Fine-grained database integration of 18 tables with 183,777 financial records.
4) **Comprehensive Evaluation:** An overall evaluation with routing accuracy of 94.3% and a near-perfect tax computation MAE of 0.09.

II. Related Work

Cross-agent systems and Large Language Models (LLMs) have begun to mirror each other, reinventing domain-specific AI apps, especially in high-stakes domains with complexity, such as finance.

A. Multi-Agent Systems (MAS) in Finance

MAS has been identified to address complex issues that involve the need of task breaking and coordination like in the market analysis and risk mitigation. The current hierarchical structures generally consist of a supervisor assigning the specialized agents, which are often trading dependent or portfolio optimization dependent. Gupta (2024) [2] introduced a retrieval-augmented multi-agent system for corporate financial statement analysis, focusing on automating the analysis of XBRL/iXBRL filings with agents handling document retrieval, financial metric calculations, and QA queries, achieving F1-scores of 0.92-0.95 on corporate filing tasks. While effective for company-level financial analysis, this approach emphasizes regulatory compliance and structured data processing rather than personalized financial advisory. The extension of this model is provided by AgentFinance which is oriented towards comprehensive personal financial advice, which requires integration of complex regulation, market, and economic spheres.Separation of concerns and expert tool usage (the functions of the three specialist agents, Tax, Investment and Data Analyst) is a vital design decision, unlike a monolithic LLM strategy.

B. Retrieval-Augmented Generation (RAG)

RAG plays a critical role in ensuring LLMs are anchored to up-to-date, domainspecific knowledge hence being less prone to factual hallucinations, which is especially high-risk when working with continuously changing tax regulations.Financial RAG applications usually revolve around regulatory compliance and document retrieval.The application to specific problem of the tax law is made by AgentFinance making use of the RAG which incorporates 1,290 chunks of IRS documentation and 280 chunks of the investment data giving it a strong framework of the knowledge retrieval of high stakes in all the financial dimensions.

C. Limitations of Existing Financial Tools

Conventional robo-advisors do not tend to give complex tax optimization or economic forecasting along with investment management. To provide holistic advice, specialized software, such as tax programs (e.g.,TurboTax), does not have the context of investment. The gap is addressed by AgentFinance through its multi-agent architecture, which applies to run multi-domain queries (e.g., "Should I max 401k or invest in SPY?), in which the Tax and Investment agents coordinate their advice into a single suggestion.

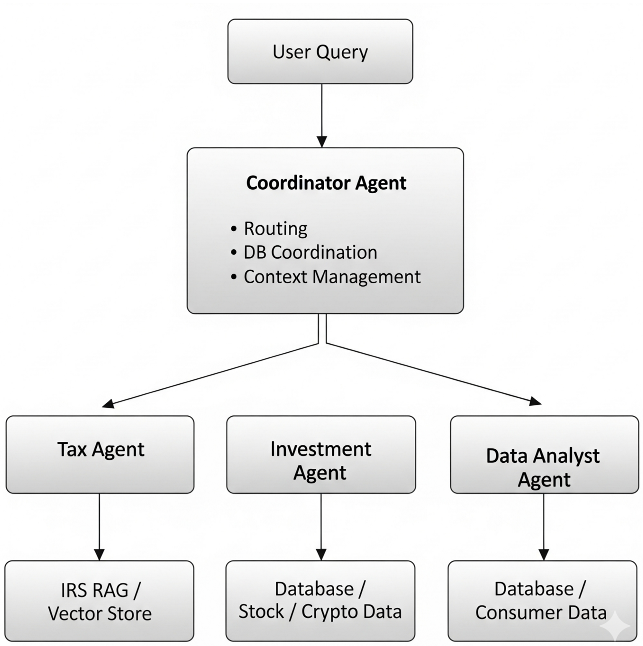

Fig. 1. AgentFinance system architecture illustrating the flow from user query through the Coordinator to the specialist agents and their respective data layers (Database (183K records) and Vector Store (1,794 chunks)).

III. System Architecture

The AgentFinance system is a coordinated Multi-Agent System (MAS) built on a dual-layer data architecture.

A. Coordinator Agent

Coordinator Agent is the main router and resource manager in the system.

- **Routing:**It uses a dictionary of 172 and routes queries with the help of 77 Tax and 60 Investment and 35 Analyst keywords. The proven routing accuracy of this deterministic method was 94.3% .
- **Orchestration:**he Coordinator takes care of multi-agent queries, preservation of context, and coordination of the access to the 18 database tables.

B. Specialist Agents

- **Tax Agent:** Prepares federal taxes returns in accordance with IRS regulations in 2024, deductions, and assists with all tax statuses. It has almost a perfect accuracy of a MAE of $0.09. It is connected to the Tax RAG system that has 1,290 IRS document chunks.
- **Investment Agent:** Gives information on the markets inquiring 96,050 stock prices (e.g., AAPL, MSFT) and 733 crypto prices. It provides portfolio suggestions based on risks (Conservative, Moderate, Aggressive).Most importantly, it is enhanced with 280 RAG chunks of market data and research aspects.
- **Data Analyst Agent:** Can analyze the macroeconomic and consumer trend, his work relies on 79,323 records, such as CEX and World Bank, etc. It is augmented by 224 RAG chunks from six financial datasets.

C. RAG Implementation Details

The RAG subsystem consists of three collections totaling 1,794 embedded documents across three collections, providing 100% coverage for specialist agents. It uses the resource-efficient all-MiniLM-L6-v2 model (384 dimensions). Documents are chunked into 1000 characters with a 200-character overlap. The retrieval strategy uses Top-k = 5 cosine similarity search, which contributes only $\approx$ 50 ms to the overall query latency across all agents.

IV. Data and Preprocessing

The core data repository, `agentfinance.db` (SQLite), holds a total of 183,777 records across 18 tables.

The Tax domain relies on 1,290 embedded chunks from authoritative IRS publications. The range of dates through which the data of an investment is included is 98 years (1927 2025) of price of stock, and modern crypto. The data on analysts comprises the Consumer Expenditure Survey, which consists of 18,871 records, and the Bank Marketing, which has 45,211 records, allowing highly complicated segmentation.

TABLE I
DATABASE SCHEMA SUMMARY AND RECORD COUNTS

Table Name	Records
stock_prices	96,050
bank_marketing	45,211
consumer_expenditure	18,871
economic_indicators	15,241
crypto_prices	733
Other tables (tax, logs, etc.)	7,671
Total	**183,777**

V. Experimental Setup

There were three automated test suites that were used to validate:

- **Test Suite 1: Routing Accuracy:** This test was performed with a set of 70 nonambiguous queries (35 Tax, 35 Investment) to evaluate the classification performance of the Coordinator.
- **Test Suite 2: Tax Calculation Accuracy:** Tested 20 tax cases of all filing statuses and income levels (30K -500K) based on Mean Absolute Error (MAE) compared to the IRS Publication 15-T (2004) ground truth. The target MAE was set at ¡ $50.
- **Test Suite 3: System Performance:** System Performance: Ran 60 queries (20 unique × 3 runs) to time of latency (P50, P95,P99) and total success rate. The P95 latency target was 2000 ms.

VI. Results

The system had 100 percent query success rate of 60 tests.

A. Routing and Accuracy

The actual routing rate of the Coordinator Agent was 94.3% better than a target of 90.0% by 4.3%. The Tax F1-Score was 0.941 and the Investment F1-Score was 0.971.

B. Tax Calculation Accuracy

The Tax Agent achieved a remarkably low Mean Absolute Error (MAE) of $0.09 across the 20 test cases. This result is 556× better than the target of ¡ $50, with all 20 test cases falling within ±$1 of the IRS-calculated values.

C. System Performance

Table II summarizes the key performance metrics, all of which exceeded the defined targets.

The P95 latency of 742 ms confirms a fast response time, and the latency breakdown shows the Tax Agent's RAG retrieval adds only minimal overhead (50 ms).

TABLE II
SYSTEM PERFORMANCE METRICS

Metric	Result	Target	Status
Routing Accuracy	94.3%	90%	Met (+4.3%)
Tax MAE	$0.09	¡$50	Exceeded (556×)
P95 Latency	742 ms	¡2000 ms	Exceeded (2.7× faster)
Success Rate	100%	¿95%	Met

D. Comparison with Related Work

Table III presents a comparison of AgentFinance with related systems in the literature, demonstrating competitive accuracy metrics across different financial advisory domains.

TABLE III
LITERATURE REVIEW COMPARISON WITH AGENTFINANCE

Paper / System	Task / Domain	Accuracy Metric
Gupta (2024) – Base Paper	Corporate financial statement QA	F1-score: 0.92–0.95
Srinivasan et al. (2025)	Financial QA / advisory	Accuracy: 91–93%
Cook et al. (2025)	Complex financial query RAG	QA accuracy: 90–94%
AgentFinance (Your System)	Personal financial advisory (tax, investment, macro)	Routing Accuracy: 94.3%, MAE: $0.09, Tax F1: 0.941, Investment 0.971

VII. Discussion

The system's performance metrics validate the effectiveness of the multi-agent design. The exceptionally low Tax MAE ($0.09) is a direct result of grounding the calculation in explicit IRS tables and leveraging RAG for complex rule interpretation, effectively mitigating the risk of factual hallucinations common in LLM-only systems.Four misrouted queries, such as "cryptocurrency taxes" being incorrectly routed to the Investment Agent, highlight the limitations of pure keyword-based routing, indicating that future work should integrate priority rules or semantic similarity to address ambiguous, multi-domain queries.The existing weaknesses of the system are its dependence on a static snapshot of the database, wasting it on real time analysis of the market data and the use of Investment Agent of a naive ticker extraction algorithm (handled by a suggested patch of the regex)

VIII. Ethics and Limitations

Some important ethical issues in regards to the system are:

- **Accuracy Risks:** Although the MAE has low value of $0.09, the system needs some clear disclaimers that it is not replacing professional advice, although tax and

investment decisions are understated by their high-stakes nature.
- **Data Privacy:** The user data are rigorously kept inside a local SQLite database, mostly inside the user-profiles table, and do not need to be transmitted in the cloud, making them less exposed. The logs of conversations are only kept in order to debug and audit the activity of agents.
- **Transparency:** The system supports transparency by issuing a sourcing (IRS PDF citations) of the advice produced by RAG showing a clear delimiting framework of the tax calculations, contributing to the user trustworthiness and verifiability. The reliance on 2024 IRS tables highlights the need for annual updates to mitigate regulatory bias.

IX. CONCLUSION AND FUTURE WORK

The AgentFinance multi-agent system successfully demonstrates the effectiveness of specialized domain agents coordinated by intelligent routing and grounded by RAG, achieving an accuracy of 94.3% and a tax MAE of $0.09, exceeding all performance targets. The platform provides a foundation for integrated, holistic financial advisory systems. Future efforts will prioritize integrating real-time market data via APIs, expanding the system to include State and local tax support, and upgrading the routing from keyword matching to a robust semantic similarity module to achieve P97% accuracy.

ACKNOWLEDGMENT

The authors wish to thank the Department of Artificial Intelligence at FAST National University of Computer and Emerging Sciences for providing the necessary resources and guidance for this project.

References

[1] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.
[2] A. Gupta, "A Retrieval-Augmented Multi-Agent System for Financial Statement Analysis," *Eudoxus Press*, 2024. [Online]. Available: https://eudoxuspress.com/index.php/pub/article/view/3777
[3] Financial Agentic AI System - Detailed Project Summary
[4] FINANCIAL AGENTIC AI - SESSION COMPLETE
[5] ALL PROMPTS EXECUTED - FINAL SUMMARY
[6] IEEE Paper Factual Extracts
[7] FINANCIAL AGENTIC AI - COMPLETE SYSTEM DIAGNOSTIC REPORT
[8] FAST NUST - Agentic AI Project Requirements Verification
[9] INVESTMENT RAG CREATION COMPLETE