Midterm Project Report
Financial Multi■Agent Analysis System with Comprehensive Evaluation Framework

1. Project Essence and Objective

This project presents a production■oriented multi■agent AI system for financial analysis and decision support. The system is designed to transform large volumes of market data, fundamentals, and risk indicators into structured investment insights. Unlike a generic chatbot, the system emphasizes controllability, auditability, and evaluation at scale.

2. System Architecture Overview

The architecture follows an Agentic design with a central orchestrator coordinating four specialized agents. A retrieval layer supplies verified financial data to the agents, ensuring that all reasoning is grounded in explicit sources. Agent outputs are aggregated into a final, user■facing recommendation.

3. Agent Descriptions

Market Data Agent: Retrieves structured market data including prices, returns, benchmarks, and historical performance.
Fundamental & News Agent: Analyzes company fundamentals and relevant financial news to identify material signals.
Portfolio & Risk Agent: Computes portfolio■level metrics, risk exposure, and scenario implications.
Summarizer Agent: Synthesizes all agent outputs into a concise, evidence■based investment recommendation.

4. Methodology: ReAct and Agentic Reasoning

Each agent follows the ReAct paradigm, explicitly reasoning over retrieved financial context before producing an output. This separation between reasoning and action reduces hallucinations and enables transparent inspection of intermediate decisions.

5. Evaluation Framework

Evaluation is a first■class component of the system. The project includes more than 175 test cases covering multiple dimensions of system behavior. Four complementary evaluation types are applied to ensure robustness and reliability.

Evaluation Type	Purpose	Applied Scope
Hard / Deterministic	Verify correctness against expected outputs	Critical financial logic
Retrieval (RAG)	Validate relevance and coverage of retrieved data	All data■dependent agents
LLM■based	Assess reasoning quality and coherence	Analytical and summarization tasks
Human Review	Qualitative validation of recommendations	High■impact scenarios

6. Testing Coverage and CI/CD Integration

The evaluation suite contains over 175 automated test cases integrated into a CI/CD pipeline. Each system update triggers regression evaluation to ensure that changes do not degrade existing capabilities. This approach aligns the system with production■grade engineering standards.

7. Production Readiness

The system is designed with deployment in mind, including modular components, clear interfaces, and monitoring hooks. Extensive documentation accompanies the implementation, enabling onboarding and future extension.

8. Conclusion

This midterm project demonstrates that a financial multi■agent AI system can be both analytically powerful and rigorously evaluated. By combining agentic reasoning, retrieval grounding, and a comprehensive evaluation framework, the project illustrates how LLM■based systems can be made reliable, auditable, and suitable for real■world financial applications.