

Project Specifications Report

1. Introduction

This project presents the design and development of an AI-powered personalized financial assistant aimed at supporting retail investors in making informed investment decisions. The system integrates multiple AI modules, including news-based sentiment analysis, individual stock and sector index forecasting, a reinforcement learning-based portfolio optimizer, and a Retrieval-Augmented Generation (RAG) module for document question answering.

All functionalities are accessible through an intuitive, natural language chat interface powered by a central AI agent that orchestrates data retrieval, model inference, and response generation. By combining diverse data streams, advanced predictive modeling, and explainable AI mechanisms, this platform bridges the accessibility, integration, transparency, and personalization gaps in current financial advisory tools, making sophisticated financial analysis more interpretable and actionable for individual investors.

1.1 Description

At its core, the AI Financial Assistant aggregates insights from five interrelated modules to provide holistic investment guidance:

- **Sentiment Analysis Module:** This component scrapes financial news, applies a model to assign positive/negative/neutral scores, and stores results in the relational database. Real-time scraping and inference support on-demand sentiment requests, while batch processing at market close ensures up-to-date historical sentiment data for downstream modules
- **RL Portfolio Optimizer Module:** Leveraging user portfolio and risk-preference inputs alongside market data, a RL based agent learns adaptive asset allocations. Routine policy updates occur after each market closes. This module reads historical price and sentiment features from the database to inform reward calculations and action proposals
- **Stock Price Prediction:** A Temporal Convolutional Network ingests OHLCV time series (and optional technical indicators) to forecast short-term price movements. Daily data is appended post-market, and the model output—“up” or “down”—is written back to the database for display or further processing by the portfolio or sectoral modules
- **Sectoral Index Prediction:** Similar to stock-level forecasting but focused on sector indices hence using bigger models, the MAMBA multi-scale attention model predicts longer-term sector trends. Monthly weight updates keep the model attuned to shifting market regimes. Its predictions inform both the portfolio optimizer and user-facing sectoral information.
- **RAG Document Assistant:** Users upload documents (PDF, Word, TXT), which are chunked and embedded via an ai model into a vector store. On question, top-similar chunks are retrieved and passed to an LLM to generate context-aware answers. This module operates in parallel with market-driven modules.

1.1.1 Related Work

AI Agents in Financial Advisory

Recent studies [1,2] have investigated LLM-based agents in financial advisory systems, revealing both opportunities and challenges. While LLM-advisors can match human advisor performance in eliciting user preferences, they struggle with conflicting user needs and complex investor profiles. Notably, users often prefer extroverted AI personalities even when they provide inferior recommendations, highlighting a critical gap between user satisfaction and actual advisory quality. These findings also point to a broader lack of effective personalization strategies that balance individual preferences with sound financial principles, which remains a key need for retail investors.

Multi-Agent Systems for Financial Services

Multi-agent systems [2] have shown effectiveness in institutional financial tasks including fraud detection, credit approval, and portfolio risk modeling. These systems typically employ hierarchical structures with specialized agents for tasks like model training and compliance checking. However, their scope largely centers on institutional-level applications, leaving retail investors underserved and without accessible solutions that integrate multiple aspects of financial decision-making in one platform.

Specialized Financial Market Applications

A recent study [3] investigated various specialized applications, including TradingGPT with hierarchical memory systems, StockAgent for modeling investor behavior, and Alpha-GPT for automated quantitative research. These systems primarily target sophisticated investors or institutional applications, often requiring significant technical expertise, and rarely incorporate transparent mechanisms that allow users to understand and trust the decision-making process.

Stock Price Movement Prediction with Machine Learning

A recent study [4] evaluated the effectiveness of machine learning models including logistic regression, artificial neural networks, and support vector machines (SVM) for predicting stock price movement direction using VN30 index data. The results showed that SVM outperformed the other models in short-term prediction tasks, suggesting its strong potential for retail-level financial forecasting. Still, most implementations remain narrow in focus, addressing prediction accuracy without integrating other essential components of holistic investor support.

Financial Risk Forecasting with Edge Computing and Knowledge Graphs

A study [6] proposes an intelligent risk prediction framework for financial investments by leveraging mobile edge computing and knowledge graph-enhanced CNN-LSTM models. The system processes operational data from investment platforms to forecast risks across the investment lifecycle, achieving high accuracy across multiple datasets. While demonstrating the power of deep learning and structured knowledge representations, such approaches often overlook accessible, personalized delivery for non-expert investors, leaving a gap for solutions that combine technical sophistication with usability and transparency.

1.2 Constraints

Economics Constraints

- **Budgetary Limits:** The project operates under a strict financial budget. The total allocated budget for GPU usage (via cloud services or similar) is fixed at \$500. Additionally, a separate budget of \$300 has been designated for deployment and maintenance processes.
- **Deployment Costs:** Deployment is recognized as a potentially costly and challenging process. The system design must strictly adhere to these financial limits to ensure the project's feasibility.

Environmental Constraints

- **Energy Consumption:** Training large AI models (such as TCN, MAMBA, and RL agents) is energy-intensive. To mitigate the environmental impact, the project employs efficient training methods (e.g., fine-tuning) to minimize unnecessary GPU usage and energy waste.
- **Resource Efficiency:** The system utilizes distributed training across available hardware to optimize resource usage, ensuring that computational power is not wasted.

Social Constraints

- **Accessibility for Novices:** A primary social constraint is the target audience's varying level of financial literacy. Since many expected users are novices, the system's interface must be robust and intuitive, abstracting the logical complexity to the backend.
- **Digital Divide:** The system is designed to be accessible via standard web browsers to ensure broad access regardless of the user's device, helping to bridge the gap in financial information accessibility.

Political and Legal Constraints

- **Regulatory Compliance (SPK):** To comply with SPK (Capital Markets Board) regulations, the system acts strictly as an informational tool. It is constrained from providing official "investment advice" to prevent legal liabilities.
- **Data Protection Laws:** The project is strictly bound by KVKK and GDPR regulations. All personal and portfolio data must be encrypted at rest and in transit.
- **Content Usage:** Web scraping functionality is legally restricted to publicly licensed websites. Paywalled or copyright-restricted sites are explicitly excluded from data collection.

Ethical Constraints

- **Algorithmic Bias:** Financial data may contain inherent biases. The AI models operate under the constraint of monitoring and mitigating these biases to prevent the systematic disadvantage of specific user profiles.
- **Transparency:** The system must adhere to "Explainable AI" principles. Users must be clearly informed that they are interacting with an AI agent, and all predictions must

be accompanied by disclaimers.

Health and Safety Constraints

- **Financial Health:** While physical safety is less applicable to software, the system safeguards the user's "financial health." It utilizes risk profiling to prevent the recommendation of high-risk assets to conservative investors, which could otherwise lead to financial distress.
- **Data Security:** To ensure user safety in the digital realm, stringent security measures (encryption, secure storage) are enforced for all uploaded files and portfolio data.

Manufacturability Constraints

- **Hardware Limits:** The "manufacturing" or training environment is constrained to the university-provided infrastructure: two H100 GPUs with a combined storage capacity of 100 GB. The training processes are designed specifically not to exceed this storage limit.
- **DevOps and Infrastructure:** Managing multiple separate AI models introduces significant operational complexity. The system requires a robust DevOps strategy to handle containerization and deployment within the allocated resources.

Sustainability Constraints

- **Data Lifecycle Management:** To ensure the system remains sustainable despite storage limitations, a weighted time decay functionality is implemented. This optimizes database usage by prioritizing fresh data while managing the accumulation of historical records.
- **Continuous Learning:** The system's relevance depends on up-to-date models. The weekly fine-tuning schedule is a hard constraint; it must be completed before the stock market opens to ensure seamless service continuity for the end user.

1.3 Professional and Ethical Issues

- **Data Privacy and Confidentiality (KVKK/GDPR) [7,8]:** We acknowledge our professional responsibility to protect user data. The system is built to be fully compliant with KVKK (Personal Data Protection Law) and GDPR. User portfolios and uploaded RAG documents are secured with strict access controls, guaranteeing that files are accessible only to the uploading user.
- **Integrity of Information:** The system aggregates data from various sources. Ethically, we ensure that the source of news and financial data is credited where appropriate and that the AI does not fabricate financial facts. The "Hallucination" problem in LLMs is mitigated by grounding the RAG module in retrieved documents rather than generating text from valid-sounding but incorrect patterns.
- **Responsibility to the User (Non-Advisory Role):** Professionally, we must clearly define the boundaries of the software. The system serves as an analytical tool, not a certified financial advisor. It is our ethical duty to implement prominent disclaimers preventing users from interpreting AI probabilities as guaranteed market outcomes.

- **Fairness and Inclusivity:** The system is designed to serve a broad spectrum of users, from novices to experts. We are ethically bound to ensure the system does not discriminate based on the size of the user's portfolio, providing high-quality insights regardless of investment capital.

2. Requirements

External Interface Requirements:

These requirements define the interfaces between the system and users, as well as external hardware and software components.

- The system shall provide a web-based user interface accessible via standard web browsers.
- The interface shall allow users to upload portfolios, view summarized news, see sentiment scores, and access AI-based predictions.
- The system shall feature a chat interface that enables users to ask natural language investment-related questions and receive specific stock predictions.
- The interface design shall be simple and intuitive, optimized for users with varying levels of technical and financial expertise.
- The system shall integrate with external News APIs (e.g., BloombergHT, Investing.com) and Financial Data APIs (e.g., Yahoo Finance) to retrieve real-time market data.

Functional Requirements:

These requirements specify the fundamental actions the system must perform.

- The system shall allow users to upload their investment portfolios in supported formats (e.g., CSV, Excel).
- The system shall automatically identify and retrieve financial news specifically relevant to the assets listed in the user's portfolio.
- The system shall summarize retrieved news and apply sentiment analysis to classify the content as positive, neutral, or negative, identifying the impact scope (global, national, or sector-based).
- The system shall process user queries such as "Will stock X go up or down?" or "What is the latest news about stock Y?".
- The system shall generate and display stock predictions based on a synthesis of recent market trends, news data, and sentiment analysis scores.
- The system shall generate investment strategies and suggestions tailored to the user's specific risk level (e.g., conservative, aggressive).
- The system shall utilize a Retrieval-Augmented Generation (RAG) model to provide natural language responses and context-aware explanations from uploaded documents.
- The system shall provide an option for users to select between "technical-only" or "sentiment-enhanced" prediction modes.

Performance Requirements:

These requirements define the specific performance metrics the system must meet.

- The system shall respond to user questions within 10 seconds to maintain a conversational flow.

- The model inference time for sentiment analysis and prediction modules shall not exceed 10 seconds per request.
- Model training or fine-tuning sessions for datasets under 100,000 records shall be completed within 0.5 to 1 hour.
- The system architecture shall support a minimum of 100 active concurrent users without service degradation.
- The technical prediction module shall support batch processing to minimize latency during periods of high user load.
- The system shall implement caching mechanisms (e.g., Redis) for frequently requested assets and repeated queries to reduce redundant model inference load.

Software System Attributes:

These requirements address non-functional quality attributes such as reliability, availability, and security.

- The system shall target an uptime of 99.5%, ensuring continuous availability for users.
- The system shall implement retry-on-failure mechanisms for API calls and fallback responses for critical functions.
- All data transfers shall be encrypted using TLS/SSL to ensure confidentiality and data integrity.
- User authentication shall be managed via secure tokens, and uploaded files must be isolated to prevent unauthorized access.
- The codebase shall follow a modular design (separating preprocessing, inference, and UI) to facilitate future updates and debugging.

3. References

- [1]Takayanagi, T., Izumi, K., Sanz-Cruzado, J., McCreadie, R., & Ounis, I. (2025, April). Are generative AI agents effective personalized financial advisors? arXiv. <https://arxiv.org/abs/2504.05862>
- [2]Yang, H., Lin, L., She, Y., Liao, X., Wang, J., Zhang, R., Mo, Y., & Wang, C. D. (2025, June). FinRobot: Generative business process AI agents for enterprise resource planning in finance. arXiv. <https://arxiv.org/abs/2506.01423>
- [3]Okpala, I., Golgoon, A., & Kannan, A. R. (2025, April). Agentic AI systems applied to tasks in financial services: Modeling and model risk management crews. arXiv. <https://arxiv.org/abs/2502.05439>
- [4]Bui, T. K., & Huynh, T. (2022, December). Forecasting stock price movement direction by machine learning algorithm. International Journal of Electrical and Computer Engineering, 12(6), 6625–6634. <https://doi.org/10.11591/ijee.v12i6.pp6625-6634>
- [5]Cheng, C., & Huang, H. (2023, December). Smart financial investor's risk prediction system using mobile edge computing. Journal of Grid Computing, 21, Article 27. <https://doi.org/10.1007/s10723-023-09710-w>
- [6] Global Industry Classification Standard (GICS). MSCI Inc. and S&P Dow Jones Indices. Available: <https://www.msci.com/our-solutions/indexes/gics>

[7] General Data Protection Regulation (GDPR) - Regulation (EU) 2016/679. European Parliament and Council, 2016.

[8] Kişisel Verilerin Korunması Kanunu (KVKK) - Law No. 6698. Republic of Turkey, 2016.