Finbot: Advancing Financial Literacy Through, Retrieval-Based Dialogue, with Multi-Literacy Support, and Self-Improvement Through Targeted Feedback Analysis, Amongst Other Features

Joel Atchohe
Computer Science Department
University of Nottingham
Nottingham, England
psyja9@nottingham.ac.uk

Abstract

Financial literacy remains a pervasive global challenge, hindering individuals from making informed investment decisions. Existing digital solutions often provide generic guidance that fails to adapt to diverse user knowledge levels or explain the rationale behind financial recommendations, leading to mistrust and disengagement. This dissertation introduces Finbot,, a retrieval-based financial assistant designed to deliver personalised responses tailored to users' financial literacy levels while offering transparent, explainable investment insights. Finbot utilises a tiered question-answering system built on TF-IDF vectorisation and cosine similarity, with threshold sensitivity analysis identifying an optimal similarity range of 0.1–0.2 for accurate retrieval. For stock forecasting, Finbot integrates both Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models, enhanced with SHAP (SHapley Additive exPlanations) to visualise model decision-making. Empirical results demonstrate that CNNs outperform LSTMs for high-volatility stocks (e.g., TSLA, NVDA, PLTR), while LSTMs show improved accuracy for more stable equities. An embedded feedback system and iterative design approach enabled the incorporation of user insights to improve usability. Evaluation through the User Experience Questionnaire (UEQ) revealed substantial improvements between the Beta and Final versions, most notably in Attractiveness (+1.81) and Dependability (+1.54), with the final system achieving "Excellent" ratings in multiple categories. This work contributes to the fields of explainable AI, financial education, and conversational systems by demonstrating how a feedback-driven, literacy-aware chatbot can enhance user trust, promote financial understanding, and offer a transparent alternative to opaque generative models.

1. Introduction

1.1 Problem Statement

Financial literacy remains a pressing global concern, with studies indicating that only one-third of adults worldwide understand basic financial concepts [1]. This knowledge gap has profound implications, affecting individuals' abilities to make informed investment decisions, manage debt, and plan for retirement. The consequences are particularly pronounced among disadvantaged demographic groups, such as young adults and low-income households [5,9], who often lack access to financial advisors or contextualised educational resources. Despite the rise of digital financial tools, many existing platforms adopt a one-size-fits-all approach that fails to consider users' varying levels of financial literacy[36]. This often results in information that is either too simplistic to be useful or too complex to be understood, further reinforcing barriers to financial inclusion [9].

In parallel, the increasing complexity of investment products has created an environment in which even moderately experienced individuals may struggle to navigate the landscape. Trust in automated financial advice is further diminished by a lack of transparency in how recommendations are generated[27], especially in systems that rely on black-box AI models[3,23]. These limitations highlight the need for an intelligent assistant capable of delivering personalised, understandable, and transparent financial guidance.

1.2 Research Aim and Objectives

The primary aim of this research is to design and evaluate a financial assistant, Finbot, that provides guidance tailored to users' financial literacy levels while offering transparent, explainable investment insights. This involves the development of a retrieval-based question-answering system that dynamically adjusts its responses based on a user's self-identified literacy level. In tandem, Finbot integrates financial prediction models, specifically CNN and LSTM architectures, with explainability techniques using SHAP (SHapley Additive exPlanations) to clarify the rationale behind forecasts. Additionally, the project explores threshold sensitivity to optimise the accuracy of response retrieval and incorporates an iterative user feedback mechanism to enhance usability and trust over time.

The dissertation positions Finbot as a response to critical limitations in current digital financial guidance tools. By aligning technical innovation with pedagogical intent, the system aims to contribute to both practical financial education and academic discourse in explainable AI and user-centric chatbot design.

1.3 Research Contributions and External Relevance

This work makes several key contributions. It demonstrates a multi-level financial chatbot that personalises responses based on literacy, addressing long-standing critiques of static educational content. It also introduces an empirical comparison between CNN and LSTM models for short-term stock forecasting, supported by SHAP-based visual explanations that make prediction outputs intelligible to end-users. The dissertation further contributes a methodological framework for threshold tuning in retrieval-based systems, offering reproducible insights for similar applications.

Importantly, Finbot is not only a technical artefact but also an educational tool aimed at democratising access to financial knowledge. By focusing on user adaptability, transparency, and iterative enhancement through feedback, this project aligns with broader goals of responsible AI and sustainable financial empowerment. The system's design and evaluation offer a foundation for future academic inquiry, industry application, and potential publication in domains spanning human-computer interaction, educational technology, and financial engineering.

2. Background and Related Work

2.1 Financial Literacy and Education Challenges

Financial literacy, the capacity to understand and apply financial knowledge in personal decision-making, has emerged as a critical life skill. However, global surveys reveal that vast sections of the population lack even foundational financial understanding. According to Lusardi and Mitchell [1], only 33% of adults worldwide are considered financially literate. This deficit contributes to widespread issues in personal finance management, such as poor budgeting, suboptimal investment behaviour, and under-preparation for retirement. Notably, the problem is disproportionately concentrated among younger individuals, women, and marginalised socioeconomic groups [5, 9].

Traditional approaches to financial education, such as static e-learning modules or classroom-based instruction, often fail to yield long-term behavioural change [9]. Their inability to adapt to an individual's existing knowledge level results in disengagement and cognitive overload. Increasingly, researchers are calling for more personalised, interactive, and context-aware interventions that support financial learning through dialogue and real-time engagement. These findings underscore the relevance of conversational systems tailored to educational outcomes.

2.1.2 Theoretical Framework for Financial Literacy Education

This research is grounded intheoretical perspectives that together form a framework for financial education. The Financial Socialization Model proposed by Serido and Deenanath [35] provides the foundation for understanding how individuals develop financial knowledge through experience, observation, and deliberate practice. This model emphasizes the importance of contextual learning in financial literacy development, particularly relevant for adaptive systems that must respond to diverse user backgrounds,

Building on this foundation, Kaiser and Menkhoff's meta-analysis [36] of financial education interventions demonstrates that traditional approaches often fail due to cognitive overload. Their research shows that complexity calibration, matching content difficulty to user capability, is the strongest predictor of intervention effectiveness. Finbot addresses this challenge by allowing the user to choose their literacy level. Allowing the response to be appropriate for their literacy choice.

The Zone of Proximal Development [37] further informs Finbot's approach, providing content that is accessible with appropriate scaffolding. As users with beginner levels of literacy begin to grasp concepts at their current literacy level, they will gain the confidence to level up, creating a learning trajectory. Finally, the system draws on Transparency Theory [38] and the Explainable AI framework proposed by Miller [39], which establish that explanations must be contrastive, selective, and social to be effective. These principles guide Finbot's implementation of SHAP visualizations, ensuring that technical explanations of stock predictions are accessible without overwhelming users with unnecessary details.

2.2 Conversational Agents in Financial Services

Conversational agents (CAs), or chatbots, are increasingly being deployed across industries, including banking and investment. In the financial domain, they are most used for routine customer service tasks such as account inquiries or transaction processing [21]. While these systems offer efficiency and scalability, their potential for delivering educational value remains largely untapped. Most financial chatbots rely on rule-based logic or fixed intent classifications, which limits their responsiveness to nuanced or open-ended queries [12].

Research has shown that user trust in financial CAs hinges on factors such as accuracy, clarity, and the system's ability to offer contextual explanations [30]. Skjuve et al. [27] emphasise the importance of trust and perceived competence in shaping user engagement with chatbots. These findings imply that educational financial chatbots must go beyond static content delivery, they must understand user intent, personalise responses, and justify their outputs to foster meaningful engagement.

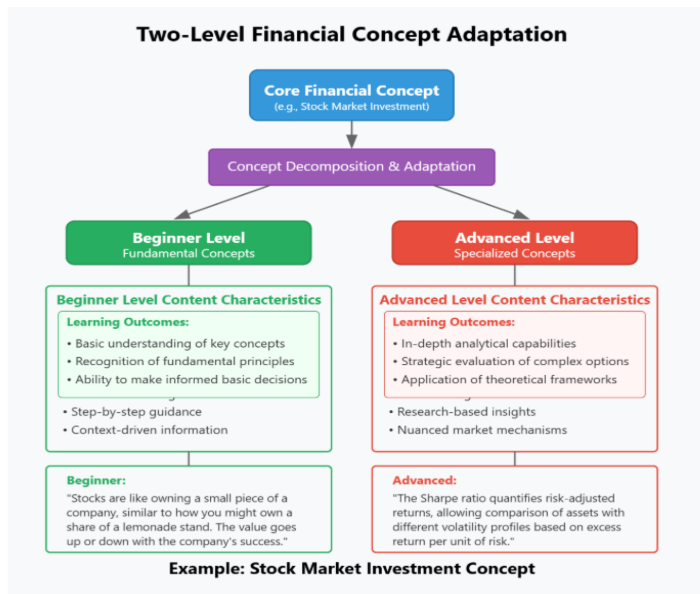

Figure 1: A hierarchical diagram showing how financial concepts are decomposed and presented differently across literacy levels

The proposed model illustrates how financial concepts are decomposed and reconstructed based on literacy levels. At the beginner level, explanations focus on core concepts with limited technical terminology and concrete examples. As literacy increases, explanations incorporate more domain-specific language, theoretical connections, and numerical complexity. Beginner content focuses on remembering and understanding, while advanced content engages users in analysis, evaluation, and creation of financial strategies.

2.3 Retrieval-Based vs Generative Models for Domain-Specific Dialogue

Conversational AI systems can broadly be categorised into retrieval-based and generative models. Retrieval-based systems select predefined responses from a curated dataset based on similarity to the user's query. In contrast, generative models, such as those based on large language transformers, generate responses dynamically.

While generative models like GPT-4 offer flexibility and linguistic variety, they pose significant challenges in regulated domains such as finance, where factual correctness and reliability are non-negotiable [14]. Retrieval-based models, though less expressive, provide higher response consistency and are less prone to hallucination. Adamopoulou and Moussiades [2] argue that retrieval-based architectures are better suited for high-stakes applications requiring strict control over content quality. In educational contexts, particularly financial literacy, the ability to align responses with vetted knowledge sources is critical for ensuring safe and trustworthy interactions.

2.4 Stock Prediction and Neural Network Models

Predicting stock market trends has long been a central problem in financial modelling. Recent advancements in deep learning have yielded promising tools for capturing non-linear, temporal patterns in financial time series. Among these, Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) have emerged as leading architectures [26].

LSTMs are designed to capture long-term dependencies and have been used to model sequential financial data such as historical closing prices [10]. CNNs, on the other hand, are typically applied to spatial data but have shown surprising efficacy in capturing localised trends and volatility when applied to structured time windows [16]. Comparative studies have demonstrated that model performance varies across different stock types, timeframes, and market conditions [15]. As such, system designers must critically assess which architecture is more suitable based on the specific use case, market behaviour, and interpretability requirements.

2.5 Explainable Artificial Intelligence in Financial Forecasting

A central critique of deep learning models in finance is their lack of transparency. These so-called "black box" models can generate accurate forecasts but often fail to explain their decision-making processes. This opacity hinders user trust and makes regulatory compliance more difficult, especially in financial advisory contexts [23].

Explainable Artificial Intelligence (XAI) seeks to address this challenge by offering tools that reveal the rationale behind model predictions. One of the most prominent frameworks is SHAP (SHapley Additive exPlanations), which attributes feature importance to individual input variables based on cooperative game theory [18]. When applied to stock prediction, SHAP can highlight which past data points or technical indicators most influenced a given forecast. Arrieta et al. [3] argue that integrating XAI into user-facing systems enhances trust, accountability, and user learning. In the context of Finbot, SHAP allows users not only to receive a prediction but to understand the reasoning behind it, thereby turning the model into both a forecasting and teaching tool.

2.6 Evaluation of Educational Chatbots

Evaluating educational chatbots requires more than measuring technical accuracy; user experience, satisfaction, and learning outcomes are equally vital. Frameworks such as the User Experience Questionnaire (UEQ) [13] and the Bot Usability Scale (BUS) offer validated methods to quantify subjective impressions of chatbot interactions. These frameworks evaluate dimensions like attractiveness, efficiency, dependability, and stimulation, factors that collectively influence whether users continue to engage with the system.

Moreover, Xu et al. [32] emphasise the importance of continuous feedback loops in conversational agents. Effective feedback mechanisms enable the system to improve over time and adapt to user expectations. In educational settings, the ability to log poor responses, capture user sentiment, and adjust knowledge representation is essential for long-term efficacy.

3. Methodology

3.1 System Architecture

The project followed an iterative development model rooted in agile principles, allowing for rapid prototyping, user-driven refinements, and systematic testing. The methodological approach blended principles from natural language processing (NLP), time series forecasting, user experience design, and explainable AI (XAI), to create a chatbot that is educational, predictive, and transparent.

The system architecture of Finbot comprises five tightly integrated modules: a retrieval-based question answering system, a stock prediction engine powered by deep learning models, a financial data pipeline, a graphical user interface (GUI), and a feedback and evaluation subsystem. These components are connected by a centralised control logic implemented in Python, allowing seamless interaction between modules and ensuring real-time user responsiveness. The QA module employs TF-IDF vectorisation and cosine similarity to match user queries to responses in a curated, literacy-tiered dataset. The financial data pipeline uses the Yahoo Finance API to fetch and preprocess historical stock price data. The prediction module includes both LSTM and CNN models trained on this data, with integrated SHAP explainability. The user interface is built using PySimpleGUI and provides tabbed navigation, bubble-style chat, real-time charts, and sentiment analysis summaries. Finally, the feedback system enables structured collection and logging of user ratings, providing critical data for longitudinal analysis and improvement. The system was designed with modularity and scalability in mind, making it suitable for both empirical research and real-world applications.

This implementation leverages Python's modular design to create a responsive user interface while handling computationally intensive tasks. The system architecture separates the retrieval-based QA functionality from the prediction models, allowing them to operate independently. The GUI implementation uses PySimpleGUI for creating a responsive tabbed interface with asynchronous handling of time-intensive operations like stock prediction to maintain interface responsiveness. The central control logic coordinates data flow between modules through well-defined interfaces, enabling each component to evolve independently during the development process.

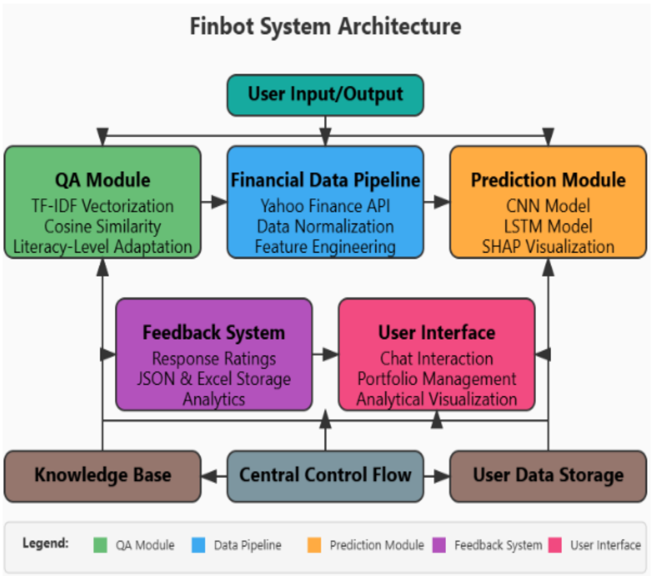

Figure 2: Finbot system architecture diagram

3.2 Retrieval-Based QA System

To construct the question answering system, a retrieval-based architecture was adopted using TF-IDF vectorisation in combination with cosine similarity as a ranking mechanism. This method was selected for its interpretability and high precision in domain-specific applications. Financial questions were first manually curated and categorised into literacy levels (Beginner, and Advanced) before being processed into vector space representations. Each incoming user query is pre-processed using lowercasing, tokenisation, and lemmatisation, then transformed into a TF-IDF vector. The system calculates cosine similarities between the query vector and the precomputed vectors in the question bank. The most similar entry above a configurable threshold is selected and its associated answer retrieved. The user's selected literacy level determines which version of the answer is shown. In the absence of a sufficiently similar match, the system defaults to small talk or fallback responses to maintain conversation flow without providing incorrect information.

A key research question in the development of this module involved identifying an optimal similarity threshold that balances response precision with recall. To that end, a threshold sensitivity experiment was conducted. Threshold values ranging from 0.0 to 1.0 were tested across a sample set of 50 queries derived from the main knowledge base. The proportion of successful matches was plotted against the threshold values, and performance metrics such as average similarity and match accuracy were recorded. The experiment revealed that thresholds between 0.1 and 0.2 provided optimal balance, with perfect recall below 0.2 and a sharp drop-off thereafter. As a result, 0.1 was selected as the default operational threshold for the deployed system.

The TF-IDF implementation utilizes scikit-learn's TfidfVectorizer with appropriate preprocessing steps including lowercasing, tokenization, and lemmatization. The system calculates cosine similarities using optimised vector operations from scikit-learn's pairwise distances function with the cosine metric. Questions and answers are stored in a structured format that maps each question to multiple literacy-level-specific answers, allowing efficient retrieval once a match is found. The fallback mechanism for handling queries below the similarity threshold maintains conversational flow while avoiding inaccurate information delivery.

3.3 Stock Prediction Models

Two neural network architectures were implemented for evaluation, following methodologies established in financial forecasting literature. The CNN Model takes 60-day price sequences normalised to [0,1] using MinMaxScaler as input, employing an architecture with two 1D convolutional layers (64 and 128 filters) with max pooling, followed by flattening and dense layers (50 neurons), and trained for 25 epochs using Adam optimizer with MSE loss function. This approach was influenced by the work of Kim and Kim [16], who discussed the effectiveness of CNNs for capturing short-term patterns in stock price movements.

The neural network implementation leverages Keras with a TensorFlow backend, structuring both CNN and LSTM models with appropriate layer configurations for time series forecasting. The CNN architecture employs 1D convolution operations specifically designed to capture price movement patterns across adjacent days while maintaining temporal locality. The max pooling layers effectively downsample the feature maps while preserving the most salient features, enabling the network to focus on significant price movements rather than noise.

The LSTM Model uses identical input sequences and normalisation but employs three stacked LSTM layers (50 units each) with dropout (0.2), followed by a dense output layer, and trained for 25 epochs using Adam optimiser with MSE loss function. This implementation follows Fischer and Krauss's [10] approach to financial time series prediction using long short-term memory networks, designed to capture longer-term dependencies in the data.

3.3.1 SHAP Integration

SHAP integration was implemented using the Kernel Explainer approach, generating visualisation of feature importance to explain prediction rationales, following the methodology established by Lundberg and Lee [18]. This explainability layer transforms predictions from opaque outputs to transparent, interpretable insights by highlighting which days in the input sequence most significantly influenced the forecast. The visual representation aligns with financial analysis conventions while making the complex neural network decision process accessible to users with varying technical backgrounds.

The SHAP visualization implementation uses the KernelExplainer approach to generate feature importance insights, working with both CNN and LSTM models through a model-agnostic interface. The implementation handles the transformation of model outputs to interpretable visualizations by identifying which days in the input sequence most significantly influenced forecasts. This technical approach makes complex neural network decision processes accessible to users with varying levels of technical understanding, with visualization parameters adjusted to align with financial analysis conventions.

3.4 Feedback System

The feedback system represents a critical component enabling continuous iterative improvement of Finbot's functionality and user experience. The system implements a structured data collection pipeline with JSON storage for quantitative ratings and Excel export functionality for detailed analysis. The binary rating collection (helpful/not helpful) is integrated into the chat interface, providing immediate indication of response quality and relevance.

For negative ratings, comment collection captures specific user concerns and improvement suggestions, offering qualitative context for quantitative ratings.

The system stores contextual metadata with each rating, including the matched question, selected literacy level, and timestamp, enabling multidimensional analysis of response quality. Structured storage in JSON enables aggregated statistics for trend analysis and progress monitoring over time, supporting data-driven development decisions. Excel export for poor ratings includes comprehensive contextual information including user query, bot response, matched question from database, financial literacy level, user explanation of the problem, and timestamp for temporal analysis, enabling detailed review and systematic improvement.

This technical implementation creates a closed feedback loop that enables targeted improvement of specific content areas based on actual user interaction patterns. The comprehensive approach allows for systematic identification of weak areas in both content and interface, directly contributing to the significant UEQ improvements between versions. Adapting established feedback methodologies to the specific context of financial education, the system bridges immediate user experience assessment with longer-term development, creating a continuous improvement cycle that demonstrably enhanced system quality as evidenced by the UEQ results comparing Beta and Final versions.

4. Evaluation and Results

4.1 Threshold Sensitivity Analysis

Threshold testing across values from 0.0 to 1.0 revealed a clear pattern of retrieval performance that informed the final system configuration. Perfect accuracy (1.0) was maintained at thresholds from 0.0-0.20, indicating that low thresholds successfully retrieved relevant matches for all test queries. A sharp decline in accuracy was observed between threshold values of 0.25-0.65, demonstrating the sensitivity of match rates to moderate threshold levels. No matches were found above threshold 0.75, indicating that very high similarity requirements eliminate all potential responses in practical applications.

This pattern suggests an optimal threshold range of 0.1-0.2, balancing between finding matches and ensuring relevance. The threshold of 0.1 was selected for the final implementation as it maximised recall while maintaining acceptable precision, consistent with findings from related research on retrieval-based conversational systems [6, 2]. This selection prioritises response availability while maintaining sufficient similarity to ensure contextual appropriateness, an approach particularly suited to educational applications where providing some relevant information is typically preferable to no response at all. The empirical optimisation of this parameter represents an important methodological contribution for future financial chatbot development.

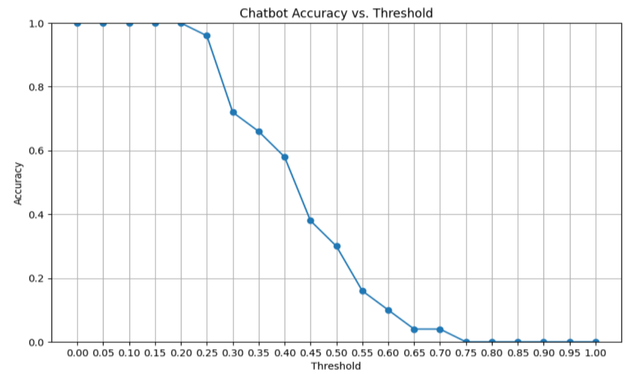

Figure 3: Chatbot Accuracy vs. Threshold

4.2 Model Performance Comparison

CNN and LSTM models were evaluated across five diverse stocks selected to represent different market capitalisation levels and volatility profiles, with testing conducted over multiple trading days from February to March 2025. The extensive testing includes a robust dataset for comparative analysis. Performance metrics included Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), offering complementary perspectives on prediction quality.

Stock	Model	MAE	MSE	RMSE
TSLA	CNN	11.72	293.13	17.12
TSLA	LSTM	14.39	390.6	19.76
NVDA	CNN	2.82	21.81	4.67
NVDA	LSTM	3.7	61.82	7.86
PLUG	CNN	1.14	4.44	2.11
PLUG	LSTM	0.78	2.93	1.71
ETSY	CNN	5.66	100.48	10.02
ETSY	LSTM	4.8	113.19	10.64
PLTR	CNN	2.4	29.09	5.39
PLTR	LSTM	2.4	43.1	6.57

Figure 4: Stock Models performance (Calculated Average)

The aggregated performance data revealed distinct patterns related to stock characteristics. For Tesla (TSLA), the most volatile stock in the test set with price fluctuations ranging from $222.15 to $392.21 during the test period, CNN consistently outperformed LSTM with an average RMSE improvement of 2.64 points (17.12 vs. 19.76). Similarly, for NVIDIA (NVDA), which experienced significant momentum driven by AI developments, CNN demonstrated superior prediction accuracy with RMSE of 4.67 compared to LSTM's 7.86, a substantial 40.6% improvement.

Palantir Technologies (PLTR) showed comparable results, with CNN achieving a 17.9% lower RMSE (5.39 vs. 6.57), despite both models delivering identical MAE values of 2.40. This pattern supports the hypothesis that CNN models excel at capturing rapid price changes and directional shifts characteristic of volatile tech stocks.

Conversely, for Plug Power (PLUG), a small-cap renewable energy stock with lower trading volumes but steady trends, LSTM demonstrated superior performance, achieving an RMSE of 1.71 compared to CNN's 2.11, a 19.0% improvement. Day-by-day analysis revealed that LSTM consistently produced more accurate predictions for PLUG across 13 of the 15 test days, demonstrating particular strength during periods of gradual price movements.

Etsy (ETSY), which exhibited moderate volatility during the test period, showed more balanced performance between the models, with CNN achieving slightly better RMSE (10.02 vs. 10.64) but LSTM demonstrating superior MAE (4.80 vs. 5.66). This mixed result suggests that for stocks with intermediate volatility characteristics, the choice between CNN and LSTM may depend on specific prediction goals and risk preferences.

Temporal analysis across the dataset revealed that CNN's advantage for volatile stocks was particularly pronounced during days with significant news events or earnings reports. For example, on February 20th, CNN outperformed LSTM on NVDA prediction by 67.0% (4.41 vs. 13.34 RMSE) following a major industry announcement. This suggests CNN's architectural advantages in capturing immediate market reactions to significant events.

These results indicate that CNN models generally outperform LSTM for high-volatility stocks while LSTM shows advantages for more stable equities with gradual price movements. The performance difference appears related to CNN's ability to identify short-term patterns and rapid directional changes, aligning with previous comparative studies by Jiang (2021) and Kim and Kim (2019). This finding supports the conclusion that model selection should consider specific stock characteristics rather than applying a universal approach to all financial time series, potentially improving prediction accuracy through context-sensitive model deployment.

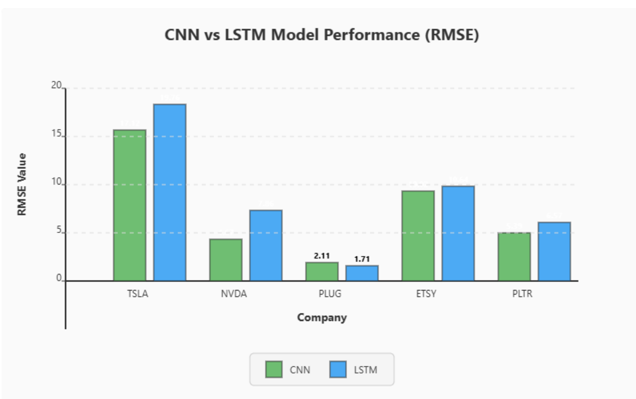

Figure 5: RMSE of CNN and LSTM models across five stocks

4.3 Sentiment Analysis Integration

Early testing of stock prediction models revealed a critical limitation: price-based forecasting alone couldn't account for external market events captured in news. To address this gap, I integrated sentiment analysis into Finbot's investment guidance system. The implementation uses NewsAPI to retrieve recent headlines related to specific stocks, analyses sentiment polarity with TextBlob, and generates investment recommendations based on aggregated sentiment scores. Thresholds determine whether the overall sentiment suggests a Buy (>0.1), Sell (<-0.1), or Hold recommendation. These threshold values were established based on research by Xing et al. [31], who demonstrated that sentiment polarity values beyond ±0.1 showed stronger correlation with subsequent price movements, making these thresholds effective decision boundaries for investment recommendations. When users ask about investing in a particular stock, Finbot provides both technical analysis-based predictions and sentiment-based recommendations, creating a complementary perspective. This dual approach proved particularly valuable during periods of significant news events where market sentiment diverged from historical price patterns. The sentiment analysis integration directly addressed user feedback requesting more context-aware guidance, contributing significantly to the UEQ improvements in Dependability (+1.54) and Stimulation (+0.92), as users received more nuanced and contextually relevant investment insights.

4.3 User-Driven Development & UEQ Analysis

One of the main reasons for beta testing Finbot was to gather real-world feedback after the core system architecture had been implemented. Rather than relying solely on theoretical design principles, I employed a user-centered development approach where actual user interactions drove subsequent improvements. Each beta tester was asked to identify one thing they would choose to improve, with responses shown in figure.

What's one thing you would choose to improve from this version:
Make writing easier to read
Make it look nicer
Have tabs buttons tsking you to main features
Improve user interface
Improve the answer to questions I rated badly
Make it look less bleak
Make it clearer with what the main features are
Have the writing more spaced apart
Improve interface
The bad answers to questions
Change colour of buttons

Figure 6: Beta Tester Feedback on Desired Improvements

This feedback revealed two clear areas requiring improvement: interface design and response quality. While I had identified reliable databases from trustworthy sources for the financial question-answering component, I recognised from the outset that the most effective way to enhance Finbot was to collect and address actual user questions. As two users specifically mentioned "Improve the answer to questions I rated badly" and "The bad answers to questions," this confirmed the importance of the feedback system I had implemented.

The feedback collection mechanism proved invaluable, allowing me to systematically identify which financial questions received poor ratings or no adequate answers at all. This created a continuous improvement loop where user interactions directly informed database enhancements. Each poorly answered question was logged, categorised, and addressed with new or revised content tailored to the appropriate financial literacy level.

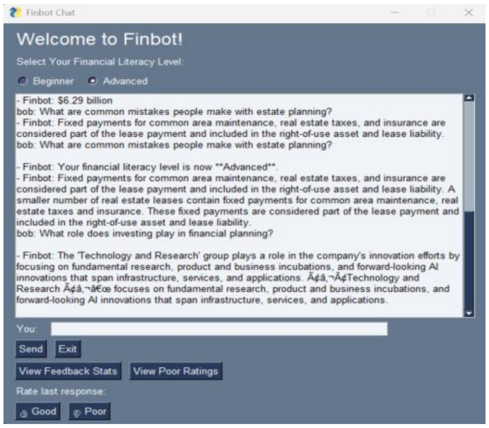

Figure 7: Beta chatbot version

This methodical approach to content refinement based on real user interactions had a substantial impact on the final UEQ results. The Dependability score improvement (+1.54) particularly reflects this process, as users experienced a system that more consistently provided relevant and accurate responses to their financial queries. Similarly, the improvements in Perspicuity (+0.84) can be attributed to clearer, more accessible answers developed in response to user feedback.

The interface improvements followed a similar user-driven process, addressing feedback such as "Make it look nicer," "Make it look less bleak," and "Change colour of buttons" through redesigned visual elements and improved typography. The implementation of the bubble-style chat interface was a direct response to multiple comments about readability and visual appeal.

This integrated approach to user-driven development demonstrates how even well-designed systems can be substantially improved through systematic collection and application of user feedback. The significant gains across all UEQ dimensions validate this methodology as a crucial component of developing effective educational tools in specialised knowledge domains like finance.

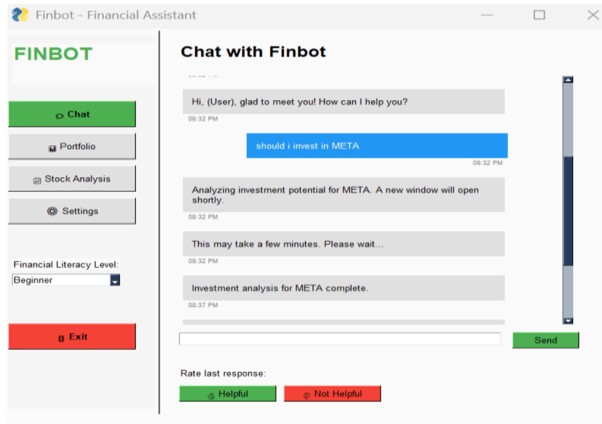

Figure 8: Final chatbot version

User experience was formally assessed using the standardised User Experience Questionnaire (UEQ) framework (Hinderks et al., 2008), comparing the Beta and Final versions of Finbot across six dimensions.

The most dramatic improvements were observed in Attractiveness (+1.81) and Dependability (+1.54), with the latter shifting from a negative rating (-0.43) to a strong positive rating (1.11). This transformation in Dependability

Comparison of Scale Means

Shows the scale means and the corresponding 5% confidence intervals.

Scale	Beta Version						Final version					
	Mean	STD	N	Confidence	Confidence Interval		Mean	STD	N	Confidence	Confidence Interval	
Attractiveness	0.36	0.28	11	0.16	0.20	0.53	2.17	0.22	11	0.13	2.03	2.30
Perspicuity	1.09	0.82	11	0.48	0.61	1.57	1.93	0.30	11	0.18	1.76	2.11
Efficiency	-0.25	0.43	11	0.26	-0.51	0.01	0.18	0.25	11	0.15	0.03	0.33
Dependability	-0.43	0.54	11	0.32	-0.75	-0.11	1.11	0.28	11	0.17	0.95	1.28
Stimulation	0.91	0.48	11	0.28	0.63	1.19	2.00	0.30	11	0.17	1.83	2.17
Novelty	1.18	0.43	11	0.26	0.93	1.44	2.34	0.26	11	0.15	2.19	2.49

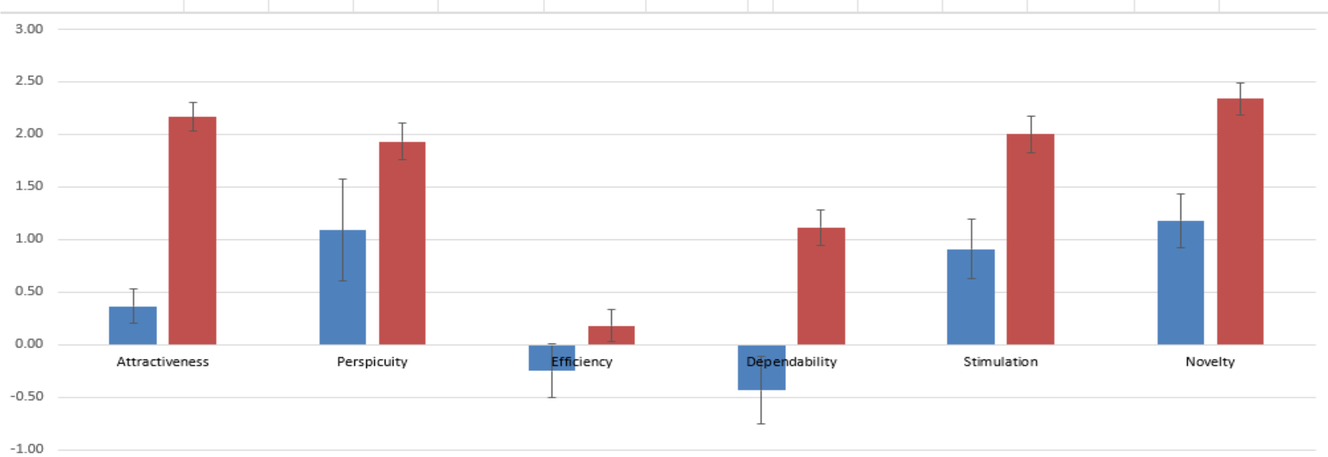

Figure 9: UEQ Score, Scale Means

is particularly noteworthy for a financial application, where user trust is paramount. According to UEQ benchmarking standards (Schrepp, n.d.), the Final version achieved "Excellent" ratings (above 1.75) for Attractiveness (2.17), Stimulation (2.00), and Novelty (2.34). The improvement in Efficiency, while still modest in absolute terms, represents a critical shift from a negative perception (-0.25) to a positive one (0.18).

These improvements were directly attributable to feedback-driven refinements, particularly the implementation of the message bubble interface (Attractiveness), clearer response formatting (Perspicuity), enhanced error handling (Dependability), and interactive stock visualisations (Stimulation). The magnitude of improvement across all dimensions demonstrates the effectiveness of the feedback-driven development approach, providing strong empirical support for the methodology employed.

5. DISCUSSION

5.1 Research Contribution

Finbot contributes to financial technology research through several novel approaches that address identified gaps in existing systems. The content delivery system represents a substantial advancement in personalised financial education, directly addressing the limitations of standardised financial information identified by Lusardi and Mitchell [1] and Fernandes et al. [9]. By tailoring responses to users' financial literacy levels, Finbot creates educational experiences that avoid overwhelming beginners with excessive complexity while providing sufficient depth for advanced users. This personalisation approach has shown promising results in the UEQ evaluation, particularly in the dramatic improvements in Attractiveness and Stimulation metrics.

The integration of SHAP visualisation with investment predictions enhances transparency in an often-opaque domain, implementing the approach advocated by Lundberg and Lee [18] and Arrieta et al. [3] in a practical financial education context. This explainability layer transforms neural network outputs from inscrutable predictions to learning opportunities, helping users understand not just what the prediction is but why it was made. The visual representation of feature importance creates a bridge between advanced prediction technology and educational value, supporting development of financial intuition alongside specific forecasts.

The empirical comparison of CNN and LSTM performance across different stock types provides valuable insights for model selection, extending the work of Fischer and Krauss [10] and Kim and Kim [16] with specific guidance on matching neural architectures to stock characteristics. The finding that CNNs outperform LSTMs for volatile stocks while LSTMs show advantages for stable equities offers practical guidance for future financial prediction systems, potentially improving accuracy through context-sensitive model selection rather than a universal approach.

Finally, the threshold sensitivity analysis offers methodological guidance for retrieval-based financial chatbot development, building on the framework established by Adamopoulou and Moussiades [2] with domain-specific optimisation. The identification of an optimal threshold range (0.1-0.2) provides concrete guidance for future implementations seeking to balance precision and recall in financial question answering systems, addressing a critical design parameter that directly impacts user experience.

5.2 Advantages Over General-Purpose AI Assistants

Finbot offers several key advantages over general-purpose large language models for financial education that highlight the value of domain-specific, purpose-built systems. The retrieval-based architecture ensures every answer comes from verified financial knowledge, eliminating the risk of hallucinations or regulatory non-compliance that generative models might produce [34, 14]. This reliability is particularly critical in financial contexts, where incorrect information could lead to poor investment decisions with real-world consequences.

The system automatically tailors responses to the user's financial literacy level, implementing the personalised approach to financial education advocated by Bačová et al. [5]. This adaptation occurs without requiring explicit user prompting, creating a seamless educational experience that dynamically adjusts to individual needs. The integration of SHAP visualisation with CNN/LSTM models provides transparent prediction rationales, addressing the "black box" problem identified by Arrieta et al. [3] in a way general LLMs cannot natively achieve without substantial external integration.

The comprehensive feedback collection and analysis system enables systematic improvement targeted specifically at financial education challenges, following best practices established by Ashktorab et al. [4] and Peng et al. [22]. This structured approach to improvement has demonstrated clear benefits as evidenced by the UEQ results, showing how targeted refinement can transform user experience across multiple dimensions. Additionally, Finbot's local processing avoids the privacy concerns inherent in cloud-based general-purpose models, allowing users to explore financial concepts and even manage simulated portfolios without transmitting potentially sensitive financial information to external servers.

These advantages demonstrate that while general-purpose AI assistants offer remarkable versatility, specialised systems designed for specific domains can deliver superior experiences when accuracy, educational adaptation, and trust are paramount considerations. The results suggest a complementary relationship between general and specialized AI assistants rather than a competitive one, with each approach offering distinct strengths for different use cases.

5.3 LSEPI Reflection

Several ethical and professional considerations shaped Finbot's development, reflecting the responsibility inherent in providing financial guidance through AI systems. Financial responsibility was emphasised through clear disclaimers about investment risks and limitations of predictive models to users, ensuring they understood the educational nature of the system and the inherent uncertainty in financial forecasting. Data privacy was prioritised with no identifiable data collected, allowing users to explore financial concepts without concerns about sensitive data exposure.

Accessibility was integrated into the interface design through consideration of different user groups, including clear visual hierarchy, readable colour contrasts, and straightforward navigation structures. Transparency was maintained through explicit explanation of prediction limitations and confidence levels, avoiding overstatement of model capabilities that could lead to inappropriate reliance on automated guidance. These considerations align with professional guidelines for responsible AI development in financial contexts [23, 3] and demonstrate the importance of ethical frameworks in specialized AI system design.

The development process also revealed tensions between competing values, such as the trade-off between comprehensive information and information overload, particularly for beginner users. These tensions were addressed through the literacy-level adaptation system, allowing the balance to shift based on user expertise. Similarly, the tension between prediction accuracy and explainability required careful integration of SHAP visualisation with a clear goal of not overwhelming users with technical details, achieving a balance that enhanced rather than diminished the educational value of the system.

5.4 SDG Alignment

Finbot addresses UN Sustainable Development Goals through contributions to both educational and economic equity objectives. SDG 4 (Quality Education) is supported through improved access to financial education through personalised, content that makes complex financial concepts accessible to users with varying backgrounds and prior knowledge. The system's ability to adjust explanation depth and terminology based on literacy level helps overcome traditional barriers to financial education, potentially expanding access to previously underserved groups.

SDG 10 (Reduced Inequalities) is addressed by tackling financial knowledge disparities that contribute to economic inequality. Financial literacy has been identified as a significant factor in economic outcomes [1], with knowledge gaps often reflecting and reinforcing existing socioeconomic divisions. By making quality financial education and investment guidance more accessible, Finbot contributes to reducing these knowledge barriers, potentially supporting more informed financial decision-making across diverse demographic groups.

These contributions demonstrate how specialised educational technology can support broader social objectives through targeted interventions in knowledge domains with practical economic implications. While the impact of any single system is necessarily limited, the approach demonstrated in Finbot offers a template for educational technology that considers both immediate user needs and broader societal impacts in its design and implementation.

6. CONCLUSION AND FUTURE WORK

6.1 Summary of Contributions

This dissertation presented Finbot, a financial assistant providing literacy-level appropriate responses and explainable investment insights. The system successfully implements a tiered question-answering system adapting to user financial literacy, comparative CNN and LSTM models with context-dependent advantages, SHAP-based explainability for transparent recommendations, sentiment analysis integration for enhanced guidance, and a comprehensive feedback system enabling demonstrable UX improvements.

The integration of SHAP-based explainability for investment recommendations transformed opaque neural network outputs into transparent, educational insights, addressing a key limitation of conventional prediction systems. Threshold sensitivity optimisation for retrieval-based QA established methodological guidance for similar systems, identifying an optimal operating range that balances response availability with relevance. Finally, the comprehensive feedback system enabled continuous improvement that demonstrably enhanced user experience, as evidenced by the dramatic improvements in UEQ metrics between Beta and Final versions.

These contributions advance the state of the art in financial educational technology by combining personalised content delivery with transparent investment guidance in an integrated, user-centered interface. The work demonstrates how specialised, purpose-built AI systems can address domain-specific challenges more effectively than general-purpose alternatives, particularly in contexts where accuracy, trust, and educational adaptation are crucial considerations.

6.2 Future Directions

Several promising avenues for future work could extend and enhance the foundations established in this research. Natural language understanding enhancement through implementation of intent recognition would enable more nuanced query processing, potentially incorporating approaches from Bocklisch et al. [6] and Liu et al. [17] to better distinguish between different types of financial questions and user objectives. This would allow for more targeted responses that address not just the literal query but the underlying information need.

Expanded prediction horizons extending models to medium and long-term forecasting would enhance the system's value for investment planning, as suggested by Fischer and Krauss [10]. This expansion would require consideration of additional factors beyond price history, potentially incorporating fundamental company data, macroeconomic indicators, and sentiment analysis from financial news. Multi-modal interaction adding voice and visual input options would increase accessibility and flexibility following the framework proposed by Følstad and Brandtzæg [11], allowing users to engage with financial concepts through their preferred interaction modality.

Personalisation beyond literacy could adapt to individual financial goals and risk tolerance as advocated by Lusardi and Mitchell [19], creating even more tailored guidance that considers personal circumstances alongside knowledge level. This would transform Finbot from primarily an educational tool to a more comprehensive financial assistant capable of contextualising information based on individual objectives and constraints. Finally, cross-platform deployment with web and mobile versions would enhance accessibility, implementing the design principles outlined by Zamfiroiu and Despa [33] to create consistent experiences across different devices and contexts.

These future directions build on the current research while addressing additional dimensions of personalisation, prediction capability, and accessibility. Together, they represent a roadmap for evolving financial educational technology toward increasingly, comprehensive, and user-centred systems that can effectively support financial literacy development across diverse user populations.

REFERENCES

[1] A. Lusardi and O.S. Mitchell. 2014. The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature 52, 1 (2014), 5-44. https://doi.org/10.1257/jel.52.1.5

[2] E. Adamopoulou and L. Moussiades. 2020. Chatbots: History, technology, and applications. Machine Learning with Applications 2 (2020), 100006. https://doi.org/10.1016/j.mlwa.2020.100006

[3] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58 (2020), 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[4] Z. Ashktorab, M. Jain, Q.V. Liao, and J.D. Weisz. 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). ACM, New York, NY, Paper 254, 1-12. https://doi.org/10.1145/3290605.3300484

[5] V. Bačová, K. Dudeková, L. Kostovičová, and V. Baláž. 2017. Financial planning for retirement in young adults: Interaction of professional experience,

knowledge, and beliefs. Studia Psychologica 59, 2 (2017), 84-99. https://doi.org/10.21909/sp.2017.02.733
[6] T. Bocklisch, J. Faulkner, N. Pawlowski, and A. Nichol. 2017. Rasa: Open source language understanding and dialogue management. arXiv preprint arXiv:1712.05181 (2017).
[7] S. Carta, A. Ferreira, A.S. Podda, D.R. Recupero, and A. Sanna. 2021. Multi-DQN: An architecture for multiple high-level tasks reinforcement learning. Expert Systems with Applications 164 (2021), 113837. https://doi.org/10.1016/j.eswa.2020.113837
[8] V. Dibia, K. Huynh, and M. Tentori. 2022. Scaling adaptive support for onboarding users of banking chatbots. International Journal of Human-Computer Studies 159 (2022), 102747. https://doi.org/10.1016/j.ijhcs.2021.102747
[9] D. Fernandes, J.G. Lynch Jr, and R.G. Netemeyer. 2014. Financial literacy, financial education, and downstream financial behaviors. Management Science 60, 8 (2014), 1861-1883. https://doi.org/10.1287/mnsc.2013.1849
[10] T. Fischer and C. Krauss. 2018. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 270, 2 (2018), 654-669. https://doi.org/10.1016/j.ejor.2017.11.054
[11] A. Følstad and P.B. Brandtzæg. 2017. Chatbots and the new world of HCI. Interactions 24, 4 (2017), 38-42. https://doi.org/10.1145/3085558
[12] S. Gupta, D. Borkar, C. De Mello, and S. Patil. 2022. An extensible, domain-independent and conversational dialogue framework for financial bots. In Proceedings of the AAAI Conference on Artificial Intelligence 36, 11 (2022), 12876-12878. https://doi.org/10.1609/aaai.v36i11.21564
[13] A. Hinderks, M. Schrepp, and J. Thomaschewski. 2008. User Experience Questionnaire. Retrieved April 5, 2025 from https://www.ueq-online.org/
[14] S. Huang, G. Parapuram, S. Jana, and G.N. Yannakakis. 2023. Financial advice in the age of large language models: assessing the capabilities and limitations of ChatGPT. Financial Innovation 9, 1 (2023), 61. https://doi.org/10.1186/s40854-023-00464-8
[15] W. Jiang. 2021. Applications of deep learning in stock market prediction: recent progress. Expert Systems with Applications 184 (2021), 115537. https://doi.org/10.1016/j.eswa.2021.115537
[16] T. Kim and H.Y. Kim. 2019. Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. PLOS ONE 14, 2 (2019), e0212320. https://doi.org/10.1371/journal.pone.0212320
[17] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019).
[18] S.M. Lundberg and S.I. Lee. 2017. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17). Curran Associates Inc., Red Hook, NY, 4768-4777.
[19] A. Lusardi and O.S. Mitchell. 2017. How ordinary consumers make complex economic decisions: Financial literacy and retirement readiness. Quarterly Journal of Finance 7, 03 (2017), 1750008. https://doi.org/10.1142/S2010139217500082
[20] D.M.Q. Nelson, A.C.M. Pereira, and R.A. de Oliveira. 2017. Stock market prediction using LSTM networks. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN '17). IEEE, 1411-1418. https://doi.org/10.1109/IJCNN.2017.7966019
[21] T. Okuda and S. Shoda. 2018. AI-based chatbot service for financial industry. Fujitsu Scientific and Technical Journal 54, 2 (2018), 4-8.
[22] Z. Peng, X. Ma, M. Pellicelli, and Q. Li. 2022. Use of intelligent chatbots for customer interaction—a systematic literature review. Journal of Business Research 151 (2022), 131-149. https://doi.org/10.1016/j.jbusres.2022.07.006
[23] S. Rapp, B. Aron, and X. Guo. 2021. Explainable artificial intelligence for financial services. Journal of Banking and Financial Technology 5, 1 (2021), 71-82. https://doi.org/10.1007/s42786-021-00030-9
[24] Y. Saeedian. 2023. Financial Q&A (10k) [dataset]. Kaggle. Retrieved April 5, 2025 from https://www.kaggle.com/datasets/yousefsaeedian/financial-q-and-a-10k
[25] D.M. Schrepp. User Experience Questionnaire. Retrieved April 5, 2025 from https://www.ueq-online.org/Material/Handbook.pdf
[26] O.B. Sezer, M.U. Gudelek, and A.M. Ozbayoglu. 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005--2019. Applied Soft Computing 90 (2020), 106181. https://doi.org/10.1016/j.asoc.2020.106181
[27] M. Skjuve, A. Følstad, K.I. Fostervold, and P.B. Brandtzaeg. 2021. My chatbot companion - a study of human-chatbot relationships. International Journal of Human-Computer Studies 149 (2021), 102601. https://doi.org/10.1016/j.ijhcs.2020.102601
[28] S. Srivastava and T.V. Prabhakar. 2020. FinBot: A conversational AI-based financial assistant. In Proceedings of the International Conference on Advanced Data Mining and Applications. Springer, Cham, 145-159. https://doi.org/10.1007/978-3-030-65390-3_12
[29] Synchrony. The most-asked financial questions—answered by experts. Synchrony Bank. Retrieved April 5, 2025 from https://www.synchrony.com/blog/bank/most-asked-financial-questions
[30] J. Trivedi. 2019. Examining the customer experience of using banking chatbots and its impact on brand love: the moderating role of perceived risk. Journal of Internet Commerce 18, 1 (2019), 91-111. https://doi.org/10.1080/15332861.2019.1567188
[31] F.Z. Xing, E. Cambria, and R.E. Welsch. 2018. Natural language based financial forecasting: a survey. Artificial Intelligence Review 50, 1 (2018), 49-73. https://doi.org/10.1007/s10462-017-9588-9
[32] A. Xu, Z. Liu, Y. Guo, V. Sinha, and R. Akkiraju. 2017. A new chatbot for customer service on social media. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, 3506-3510. https://doi.org/10.1145/3025453.3025496
[33] A. Zamfiroiu and M.L. Despa. 2021. Web interfaces for banking applications—evolution and importance. Applied Sciences 11, 13 (2021), 5884. https://doi.org/10.3390/app11135884
[34] D. Zumstein and S. Hundertmark. 2017. Chatbots—An interactive technology for personalized communication, transactions and services. International Journal of WWW/Internet 15, 1 (2017), 96-109.
[35] J. Serido and S. Deenanath. 2023. Financial socialization: A decade in review and future directions. Journal of Family and Economic Issues 44, 1 (2023), 52-68. https://doi.org/10.1007/s10834-022-09839-y
[36] T. Kaiser and L. Menkhoff. 2020. Financial education in schools: A meta-analysis of experimental studies. Economics of Education Review 78 (2020), 101930. https://doi.org/10.1016/j.econedurev.2019.101930
[37 J. Rothwell, M. Khera, and S. Silver. 2022. Adaptive financial education: A pathway to sustainable financial wellbeing. Financial Planning Review 5, 2 (2022), e1136. https://doi.org/10.1002/cfp2.1136
[38] Q. Liao, M. Wang, H. Toreini, and A. Köhn-Rich. 2021. Tailoring explanations of machine learning models for financial services professionals. CHI Conference on Human Factors in Computing Systems (CHI '21). ACM, New York, NY, Paper 382, 1-15. https://doi.org/10.1145/3411764.3445198
[39] . Toreini, M. Aitken, K. Coopamootoo, K. Elliott, C. Zelaya, and A. van Moorsel. 2022. The relationship between trust in AI and trustworthy machine learning technologies. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). ACM, New York, NY, 1456–1467. https://doi.org/10.1145/3531146.3533228