International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
www.computersciencejournals.com/ijcai
IJCAI 2025; 6(2): 115-122
Received: 03-06-2025
Accepted: 08-07-2025

Indraneel Parthasarathy
Department of Computer Science, California State University East Bay Hayward, California 94542, USA

Tanu Priya
Department of Computer Science, California State University East Bay Hayward, California 94542, USA

Moayed Daneshyari
Department of Computer Science, California State University East Bay Hayward, California 94542, USA

Corresponding Author:
Indraneel Parthasarathy
Department of Computer Science, California State University East Bay Hayward, California 94542, USA

Effects of chain-of-thought on large language model in designing and development of a financial assistant application

Indraneel Parthasarathy, Tanu Priya and Moayed Daneshyari

DOI: https://doi.org/10.33545/27076571.2025.v6.i2b.185

Abstract
In a time when daily expenses frequently compound in the back-ground, sabotaging financial goals and independence, the Financial Decision Helper App is a game changer for personal finance. To tackle the common issue of unaccounted-for daily money dissipations and inconsistent account tracks, this model makes use of the state-of-the-art AI model, DeepSeek-R1-Distill-Llama-8B large language model fine-tuned with Low-Rank Adaptation (LoRA). The app offers personalized, step-by-step guidance on how users should spend their money, which includes clear instructions on whether something is "APPROVED," "DELAYED," "DENIED," "FORCEFULLY APPROVED," or "RECONSIDERED"-all with detailed explanations behind each decision to help people understand what impact it may have on their finances. User profile, budget cycle and transaction details are overlaid, producing financial awareness and sustainable habits. A well-designed manually constructed data set based on real-world financial situations guarantees that the model is accurate and relevant. The application is equipped with advanced strategies, including parameter-efficient fine-tuning, to effectively align the language model with the application requirements. Successive evaluation metrics such as the Precision, Recall, F1-score and BERTScore prove the effectiveness of the model in terms of predicting recommendations and generating comprehensible reasoning output. The application is designed to simplify financial management and to educate individuals by linking daily expenditures with long-term financial well-being.

Keywords: Financial decision helper, chain-of-thought, large language models, personal finance, Low-Rank Adaptation (LoRA)

1. Introduction
In contemporary times, it is increasingly arduous to monitor and manage personal finances as individuals accumulate numerous small payments. The impact on the stability or balance of one's budget may remain entirely unnoticed until it evolves into a problem that is difficult to resolve. Budgeting applications can assist users in exerting control over their spending and periodically comparing it to their income. However, most of these applications fail to account for the fact that minor purchases can significantly affect a user's financial well-being by the end of the day, as small expenditures tend to accumulate rapidly and effortlessly. While budgeting applications can facilitate the tracking of substantial spending, they do not effectively manage minor financial irregularities. This oversight leaves users vulnerable in their financial situation, often fostering the illusion—encouraged by the application—that limiting spending through the app equates to covering part of their expenses, when in reality, it addresses only the specific item monitored. Among the plethora of applications available in today's market, a significant majority do not perform adequately in terms of user engagement. This inadequacy primarily stems from the paucity of a personalized interface that enables individual users to feel as though they possess a unique input interface, along with the liberty and structure necessary to navigate the complex spending scenarios encountered daily.
Essentially, the majority of budgeting apps are capable of portraying the user's financial picture by zooming in at larger purchases. They often fall short of enabling individuals to track their daily micro-spending in a convenient fashion, and in turn, lack the most valuable capability crucial for helping them to construct a rationale behind their spending habits on a daily basis and its effect on their finances and future.

The Financial Decision Helper App fills the void left by conventional budgeting tools. Pen and paper and niche apps do not offer a technological solution that is accurate, personalized and easily-accessible to a user in mind. The Financial Decision Helper App is designed to mitigate these issues through AI where personalized financial conduct is set. The application stands out due to its Chain of Thought (CoT) reasoning method implemented with DeepSeek-R1-Distill-Llama-8B model. It is expected to grant users dynamic and coherent behavior regarding their expenses within also non-standardized limits.

The Chain-of-Thought reasoning approach implemented in the DeepSeek-R1-Distill-Llama-8B model serves as a pivotal development in personal finance management, where the integration transcends beyond conventional static responses to offer tailored financial guidance. Through leveraging the nuanced pathways of CoT, the app advances users' interaction with financial datasets, transitioning from mere data presentation to engaging in simulated reasoning that assists users in understanding their unique financial habits. As Qian *et al*. (2025) discuss, this alignment introduces a robust frame-work that enhances transferability of reasoning capabilities from general AI to real-world finance applications, capturing individual patterns and delivering customized advice [1]. With this innovative framework, users receive dynamically generated suggestions that mirror personalized financial landscapes, overcoming the previous limitations seen in rule-based systems. This focus on adaptive and predictive analytics signifies a paradigm shift, empowering users to implement informed decision-making processes with precision and foresight, effectively bridging the gap between AI potential and user-specific financial demands.

To advance the capabilities of financial decision-making tools, the application capitalizes on the integration of CoT reasoning within the context of financial operations, leveraging the DeepSeek-R1-Distill-Llama-8B model for enhanced personalization. By utilizing the LoRA fine-tuning technique, the app transcends the constraints of rule-based systems, enabling it to tailor recommendations more precisely according to individual spending patterns and objectives. This methodology allows for a nuanced understanding and anticipation of user behavior, thereby facilitating proactive and adaptive financial advice that adjusts as circumstances evolve. Moreover, the application's architecture is designed with interoperability in mind, allowing it to communicate with other applications and incorporate extensive datasets, thus broadening its analytical depth and accuracy [2]. Consequently, this strategic fusion of CoT principles and LLM-based technology not only elevates the quality of financial guidance but also positions the app as a cornerstone in the burgeoning land-scape of intelligent personal finance management.

2. Related Work

LLMs have created a revolution in various fields and their application of the Chain-of-Thought (CoT) reasoning brought significant changes to the landscape of their operation. However, there are on-going applications of LLMs, especially in the field of finance that are still focused on their generic reasoning capabilities rather than on their user-specific applications [3]. Despite the significant performance improvements, CoT prompting remains to be underutilized in the field of personal finance and financial management [4]. While, systems today are empowered by the LLM to perform generic applications like arithmetic operation or predicting stocks and the market [3], they still do not account for the uniqueness of individual users. There still exists a challenge in the development of applications that are concerning user-specific financial items, such as the Financial Decision Helper App that uses CoT-enhanced reasoning applications for more personalized financial assistance.

In addition, past studies on AI-based applications within the fi-nance industry show significant technological progress while also pointing out important shortcomings in their ability to cater to personal finance. Evidence shows how the application of AI and machine-learning technology in the financial technology (fintech) industry has the ability to produce cutting-edge applications that support better financial decisions [5]. However, most of these avail-able applications are limited in terms of their ability to provide a more personalized and responsive intervention that can adequately address individual financial behaviors. Researchers find that existing applications focus mainly on common financial activities and not on such tasks that require custom interventions which are able to absorb user financial behavior. As such, these findings present a significant gap that the Financial Decision Helper App intends to address through the use of AI-based technologies such as Chain-of-Thought reasoning that promotes more user-relevant financial recommendations.

A considerable amount of work has also been done to embed artificial intelligence into budgeting programs; however, in this case, one can also observe significant limitations. The earlier AI-inclusive solutions concentrated on rule-based approaches, and while being considered a breakthrough at that point, they soon revealed their inability to cater for their use case individualization. In successful examples explained by Talasila (2024), that included the possibility to integrate with a variety of financial institutions, such solutions still remained limited in dynamic advice provision. Their inability to implement higher order reasoning architectures made them incapable of addressing the challenges brought about by that use case individualization, and hence their advice remained almost universal and impersonal. As such, our application is unique for combining CoT reasoning within the DeepSeek-R1-Distill-Llama-8B algorithm. This combination in employment offers a first-of-its-kind form of personalized financial support through extensive user spending analysis.

Furthermore, there are currently existing technologies with shared goals to our proposed application but fail to provide a holistic decision-support system for day-to-day financial decisions. For ex-ample, while apps such as Mint and PocketGuard allow for budget adherence and guidance at a basic analytical level, they only allow for higher-level categorization of spending and do not incorporate CoT or cognitive reasoning that is required for mid-level decision-making. Qian *et al*. (2025) also suggest that smaller model, such as DeepSeek-R1-Distill-Llama-8B have poor performance with dynamically changing financial models, thereby indicating that CoT implementation can improve upon this area in a financial application [1]. Other similar systems eliminate the ability to use dynamic decision support as their advice remains stagnant over time, rendering financial advice un-adaptable to changing user traits and habits besides calculated dollars and cents reasoning. Our proposed application aims to provide a solution to these problems through the use of CoT reasoning in advanced ways and to use this framework to allow for endorsed user decision-making and financial development in accordance with user-specific situations.

3. Problem Statement and Data

One of the most critical obstacles for managing finances in today's consumer world is assuming that small expenses are negligible. Even though budgeting applications are constantly growing, they are often abandoned by users who find them incapable of fitting their specific spending patterns. On a different note, current financial applications are yet to support the integration of reasoned, cognitive comprehension of daily expenses. Our app aims to solve these issues through

a Chain-of-Thought reasoning based conversational system which will allow accurate classification and identification of transaction habits. Besides, the creation and evaluation of a synthetic dataset aimed at imitating user profiles and expense patterns guarantees that our application delivers reasoned financial advice tailored to specific conditions.

3.1 Problem Statement

The Financial Decision Helper App represents a novel approach in the realm of budgeting tools by leveraging large language models (LLMs) to meet the nuanced financial needs of individual users. Unlike traditional budgeting applications, this app uses Chain-of-Thought reasoning in combination with custom datasets to tailor its financial advice, thus resonating more intimately with user-specific financial behaviors. Such customization is critical to enhancing the user experience since it ensures the AI comprehensively coaches the user through personalized financial journeys while verbalizing the logical steps behind each recommendation. By doing so, it not only improves user engagement but also addresses the over-looked issue of accumulative low-cost item expenditure which, when unchecked, can significantly derail personal financial goals. Through this integration of advanced AI, the app encourages user involvement in their budgeting processes and fosters a more disciplined management of finances.

The application of CoT reasoning in the Financial Decision Helper App serves as its backbone, enabling it to navigate complex financial decisions by deploying this technique to simplify intricate financial problems. By breaking down spending patterns into manageable steps through CoT prompting, the app guides users in dissecting their financial habits to foster better understanding and control, addressing the disparity between user-specific spending behaviors and existing generic budgeting tools. This approach ensures that recommendations are crafted not just on a surface analysis of transactions but are deeply rooted in the rationale behind each spending decision, akin to the advancements in AI-driven financial advisement platforms noted in emerging financial technology research. Such nuanced handling of financial data offers users a more immersive and educational experience in understanding their financial standing, contrasting starkly with the simplistic methodologies of traditional financial applications. As users interact with this enriched recommendation system, they are empowered to refine their spending choices, paving the way for both immediate and future financial well-being through a tailored, user-centric approach.

Conventional data assessment by existing applications analyze inputs with no regard for user particulars, and this results in generic recommendations that lack functionality. The Financial Decision Helper App adopts the use of a synthetic user profile, spending categories and transaction patterns dataset drawn from JSON objects to create a recommendation system communicative to the user. The analysis of prepared dataset allows the provision of recommendations that are tailored according to user-specific financial trends and consumption behaviors.

3.2 Data

Creating the dataset, which reflects different user types and spending scenarios, had multiple stages. It began with specifying user types. Based on the overview of existing demographic data and spending behaviors, relevant user types and scenarios to cover them in detail in the dataset were defined. In addition to typical spending patterns, specified types of financial activities were also covered in the dataset, such as user transactions or savings target for various financial ac-counts. After that, the data collection process proceeded. Due to the variety of financial activities, advanced simulation algorithms were used to generate synthetic data based on realistic patterns of trans-action sequences. The resulted dataset can be represented in JSON and JSONL formats for app/system integration. The final dataset can be used as a part of various interactions, including user spending and account management scenarios. Dataset verification required cross-verifying it with actual user type spending scenarios.

Moreover, the synthetic data on which the application operates is a carefully designed dataset representative of user segments with a wide range of spending habits and preferences. The dataset includes specific and general spending categories, covering needs and wants, which facilitates comparison with results from similar apps and previously published statistics. The dataset includes user transactions that reflect general patterns of financial activity. The data framework allows for advanced AI techniques to be applied to provide personalized recommendations, shifting the paradigm of decision-making in comparison with existing solutions. Financial applications designed using synthetic data provide a means to address unmet financial requirements within a framework that recognizes the limitations of existing applications, which typically rely on generalized spending patterns rather than concentrating on the needs of individual users.

The dataset is designed to fine-tune a financial decision helper model by simulating realistic user financial scenarios and decision-making processes. Each record in the dataset adheres to a detailed and consistent structure composed of seven core sections: User Profile, Current Budget Cycle, Category Details, Historical Insights, Transaction Details, Decision Question, and Model Output. These sections collectively provide a comprehensive snapshot of the user's financial context, behavior, and transaction justification.

The User Profile captures demographic and behavioral traits such as age, income range, family structure, time zone, and psychological spending habits. The Current Budget Cycle outlines the active financial plan, including total allocation, savings targets, category-wise allocations, and current spending. Importantly, the sum of category allocations must exactly match the total budget, ensuring mathematical consistency. Category Details zoom in on a specific spending category, comparing allocated amounts to current and proposed expenditures.

Historical Insights provide data-driven context based on past bud-get cycles, including savings performance, transaction patterns, and summarized behavioral insights. Each record includes a Transaction Description that proposes a specific new expense, prompting the model to answer a structured Decision Question asking whether the transaction should be APPROVED, DELAYED, RECONSIDERED, DENIED or FORCEFULLY APPROVED. The Model Output must include a clear recommendation and detailed reasoning with numerical justifications and cost-effective alternatives.

The dataset emphasizes diversity, requiring coverage of different user types, financial behaviors, and decision outcomes. It enforces realism, ensuring that figures and logic are grounded in plausible financial scenarios. Strict validation rules enforce internal consistency across numerical values and decision logic. Together, this structure enables the model to learn how to analyze context-rich financial data, reason through trade-offs, and generate well-supported financial decisions.

In addition, the synthetic dataset is constructed in JSON format, which serves as a standardized medium for organizing user information. User information, categorized into user profile, which includes personal data, spending patterns, and historical transactions, will be integrated into the application to provide appropriate recommendations to facilitate users for effective financial suggestions. Moreover, synthetic data is provided in both training and validation format. The process begins with validating each record in the dataset to ensure all required fields are present for accurate training and evaluation. This ensures data completeness and prevents issues during later stages. Once validation is complete, attention shifts to addressing class im-balance within the training data. Underrepresented categories are identified based on the distribution of recommendation labels, and additional records are generated for these categories. To maintain diversity and avoid repetition, the transaction descriptions in these new samples are paraphrased using a T5 language model. These augmented records are then combined with the original training data, resulting in a more balanced and varied dataset that enhances model robustness and generalization. The training data is systematically organized to comprehensively reflect user behavior, which is then divided into training and validation sets in order to evaluate the model's performance, especially its capability to generalize to novel data. Initially, data is fine-tuned using LoRA to adapt, optimize, and modify the layer weights and configurations.

4. Financial Assistant Application

To address the needs identified, the financial decision helper app is divided into five components that together are able to provide a complete solution. During onboarding, user must fill-in a personalized budget which helps them to set their own real achievement target value. Then, during transaction logging, they will enter their transactions in real-time which allows recommend engine to suggest a spending recommendation between value ranges by inputting the logged transactions data [3]. The model utilizing DeepSeek-R1-Distill-Llama-8B and then LoRA, allows the app to be fine-tuned to give the recommendation. Finally, spending overview allows a user to have a summary of their spending behavior in order to better plan future transactions. It also allows users to understand where their money goes in order to map out plan to improve their spending behavior.

Building on the distinct sequence of actions outlined in the app design, incorporating Chain-of-Thought reasoning provides an advanced layer of analytical depth that significantly enhances user engagement through more precisely tailored financial advice. CoT reasoning is particularly effective for breaking down complex financial decisions into manageable parts, allowing the app to mimic a step-by-step decision-making approach that resonates with the users' thought processes. This approach not only facilitates more accurate financial recommendations but also improves user confidence in their decision-making abilities by making underlying analysis clear and comprehensible. By leveraging insights from platforms like FinRobot, the app can integrate sophisticated CoT prompting to elicit reasoning that aligns with the users' unique financial needs and behaviors. Thus, the Financial Decision Helper App not only informs users about their spending but also educates and empowers them to make informed financial decisions through an intuitive, user-centric AI framework.

4.1 App Components

User onboarding in financial decision helper enables user to meld into daily activities with minimum friction. It is a guided journey for the user which begins at the user onboarding stage. A personalized onboarding prioritizes user needs and helps in defining a financial milestone. Onboarding adds to the usability of budget setting mod-ule which is highly adaptable and fits various financial aims and spending habits [1]. User can define budget parameters and choose limits on spending and expectations for savings as per his/her wish. This high level of personalization appeals to the user and aids in adhering to the purpose as it is evolving according to user needs.

Second, the transaction logging tool will have a secondary function; the user will have the ability to log the transaction with de-scription and amount but the user will also be able to instantly view the recommendation along with reasoning for each transaction. This recommendation will also help the user to determine whether to complete the transaction or not.

Furthermore, the Financial Decision Helper App benefits from recent advancements in AI-driven financial platforms, such as the insights demonstrated by FinRobot, which uses Chain-of-Thought techniques to break down complex financial queries into man-ageable parts. This methodology not only enhances user engagement by facilitating AI-driven financial analysis but also improves the clarity and effectiveness of financial guidance offered to users. By leveraging CoT prompting, the app elicits sophisticated reasoning from large language models, making it adept at offering personalized financial advice tailored to the user's unique spending patterns and goals. Moreover, the integration of CoT within the app's recommendation engine enables dynamic adaptation to real-time financial scenarios, thereby refining user support beyond static budget planning. Such innovative applications underscore the app's potential to redefine financial discipline by providing concrete, actionable insights derived from a nuanced understanding of individual user behaviors.

4.2 App Functionality

The design model of the Financial Decision Helper App includes important constructs that help to deliver relevant and useful financial advice to the app users in an effective manner. First, the user onboarding construct helps to orient new users to the functionalities of the app in an effective and natural manner. The budget setup construct helps in framing the users' budgetary goals and limits and is flexible enough to accommodate user-specific profiles. The trans-action logging construct helps in capturing each user's transactions

in an orderly manner, thus helping to define the transaction history that will be later used as basis for the transaction log analysis. The recommendation engine construct helps to deliver financial decision recommendations to the user, based on the expense history and avail-able profile information. The spending analytics construct helps to analyze the users' expenses to help them understand their spending behavior and the ways that they deviate from the defined budgetary goals and limits. Overall, the design model includes a series of coherent and logical actions of constructs that define a framework for adaptive user-specific financial advice. The underlying concept is already incorporated in the latest AI-powered development frame-works.

Secondly, user onboarding of the Financial Decision Helper App is an important process. During it, the application may provide the user with personalized recommendations in the sphere of finance. At the onboarding stage, the user's data and preferences are collected in a structured way. In the onboarding process, the user's financial needs and goals are comprehensively analyzed to allow the application to provide recommendations and to customize the application's features. In order to meet financial literacy goals, it is important to provide personalization of the application at the onboarding stage because personalization is the crucial factor of the user engagement. Provided personalization increases the application efficiency in delivering financial recommendations significantly. Since the content will be personalized, the user will be able to make a step towards their financial literacy because they will be able to obtain the content that is necessary for them.

The app's financial decision helper budget-setting feature aims to create cleft budgetary goals and objectives for each user. It follows that setting cleft budgetary goals and limits on the user's spending and financial decisions will make it easier to adapt and change them according to the user-specific financial profile. Based on the knowledge acquired from the onboarding stage of the users' spending behavior, the set budgetary goals can be altered accordingly [5]. Personalization of this budgeting aspect also improves the decision-making process by relying on specific adjustments to the users' financial profile in addition to the historical profile. All in all, these elements are likely to increase user engagement owing to the high accuracy and personalization of the employed financial recommendations.

Likewise, the Financial Decision Helper App's recommendation feature, which offers users feedback relating to their current spending behavior, also serves to further improve user engagement through the provision of real-time feedback relating to the user's financial position. The app utilizes the feature to systematically log each transaction made by the user in relation to their chosen spending categories. The recorded data then helps the user gain insights into spending behavior in such categories, thus allowing the app to pro-vide the user with feedback regarding their current spending behavior. Engaging the user in such a manner, by providing real-time feedback regarding their spending habits, helps them to manage their finances more effectively and makes better decisions regarding their finances [3]. The real-time feedback provided by the recommendation engine is important as it immediately notifies the user of any discrepancies in their progress towards their financial goals and thereby allows them to correct their spending accordingly. The transaction logging functionality offered by the app serves to promote user engagement

by maintaining a precise and real-time view of the user's financial position. While, it helps the model provide accurate and relevant recommendations, which can be modified in accordance with any changes made to the user's financial position.

5. Model and Technique

For fine-tuning, DeepSeek-R1-Distill-LLaMA-8B is chosen: a dis-tilled version of LLaMA (Large Language Model Meta AI) architecture that has shown state-of-the-art results. The size of 8-billion parameters achieves a reasonable trade-off between performance and resource consumption. Distill means that a bigger model was distilled, which reduced its memory and compute footprint without losing the majority of its capabilities. DeepSeek-R1 has been de-signed to deliver optimized context and throughput, enabling long sequence processing needed for financial reasoning tasks with structured data.

The model optimization is further improved through LoRA (Low-Rank Adaptation), a recently popularized and efficient fine-tuning technique. Instead of updating the complete set of model parameters, LoRA adds small trainable rank-decomposed matrices to attention and feedforward layers of the model where the matrix rank (r) is 32 in our fine-tuned model. This configuration allows for the effective representation of adaptation signals without the additional cost of full parameter updates. The value of lora_alpha is set to 32, a critical parameter in the training process that plays a significant role in scaling matrix values. This scaling is important because it helps maintain proportional relationships among the matrix elements, thereby ensuring the stability of the training process. Such stability is crucial because it prevents the model from diverging or producing erratic results, which can otherwise occur if the matrix values fluctuate un-predictably. The configurations involved in LoRA includes dropout regularization turned on for certain components and targeted major components of transformers such as query projection, key projection, value projection, output projection, up projection, and down projection. The target architecture covers all attention-based and feedforward paths along the neural network.

Other training hyperparameters are selected to improve the training efficiency and resource-savings. Gradient accumulation steps are set to 8 to increase batch-size with no extra GPU memory overhead. Learning rate is set to 1e-4, which is a commonly effective learning rate in adapter-based fine-tuning and a linear scheduler is selected to decrease the learning rate gradually during the training. Weight decay is utilized with 0.01 to regularize the model not to overfit. Mixed-precision training is selected with fp16 or bf16 based on augment hardware support to improve training efficiency without losing accuracy. Checkpointing, evaluation steps and seed-based determinism are configured to have traceable and reproducible training results.

This training workflow combines careful data curation, semantic augmentation, and parameter-efficient fine-tuning to build a capable financial decision model that balances interpretability, diversity, and generalization. Utilizing this technique would allow the app to break down spending habits and decisions into smaller components, thus producing more understandable conclusions for customers. Such an approach would accommodate unique patterns in individual expenditure along with desired goals. Since CoT reasoning is, by itself, personalized, it would increase user loyalty because people tend to perceive the app's advisory behavior as more intuitive (attractive) and logical. This automatically positions the application as superior to other existing solutions for financial decision-making. As an additional promising result of this process, the educational effect may be understood in terms of the improvement of two strongly interconnected and truly important components related to financial decision-making-financial literacy and financial discipline.

Overall, the integration of DeepSeek-R1-Distill-LLaMA-8B with LoRA and finely tuned hyperparameters results in a robust, scalable training strategy. It combines state-of-the-art architecture with efficient training practices to produce a model well-suited for nuanced, structured decision-making in financial contexts.

6. Performance Evaluation

Evaluation metrics are essential for any machine learning system as well as the Financial Decision Helper Model proposed in this work. This model pursues two main goals: 1) producing a recommendation (e.g., APPROVED, DENIED) to the user; 2) producing rationale or a sequence of reasoning statements to support the recommendation. The first goal corresponds to a multi-class classification problem and the second one relates to the natural language generation task, which requires producing semantically and contextually congruent statements according to domain knowledge. In this section, we propose a performance evaluation framework for both tasks. The developed framework aims to integrate standard classification performance metrics with the state-of-the-art metrics for the evaluation of natural language processing.

In the case of recommendation task, the model is trained to as-sign one of the five discrete classes to financial transaction data: APPROVED, DELAYED, RECONSIDERED, DENIED, FORCE-
FULLY APPROVED. Classification metrics such as standard model evaluation metrics such as precision, recall, F1-score, and normalized confusion matrix are used to assess the predictive abilities of the model. For APPROVED class, the model achieves near perfect performance in all key metrics (≈1.0), signaling its high identification performance with very few false positives and negatives. For FORCEFULLY APPROVED and DELAYED classes, the model achieves high precision (≈1.0), and high recall (≈0.90 and 0.88), respectively, resulting in F1-scores of ≈0.95. In the case of RECON-SIDERED and DENIED classes, the model yields slightly lower values of recall (0.80-0.85). However, the precision value for these classes remains high (≈0.90). This indicates a heightened conservativeness of the model when assigning these classes to data samples, possibly due to their delicate class border definitions within the financial domain.

The normalized confusion matrix highlights the distribution of original classes over the predicted label for individual classes in terms of proportions. The correctly classified instances (diagonal) vary from 83.33% (DELAYED) to 96.52% (FORCEFULLY AP-
PROVED), indicating a strong performance for the most part. No label is strongly misclassified compared to others, and misclassified instances are adjacent in terms of decision logic, such as RECON-SIDERED → APPROVED (4.35%) or DELAYED → DENIED
(7.04%) indicating that the decision was a logical one perhaps due

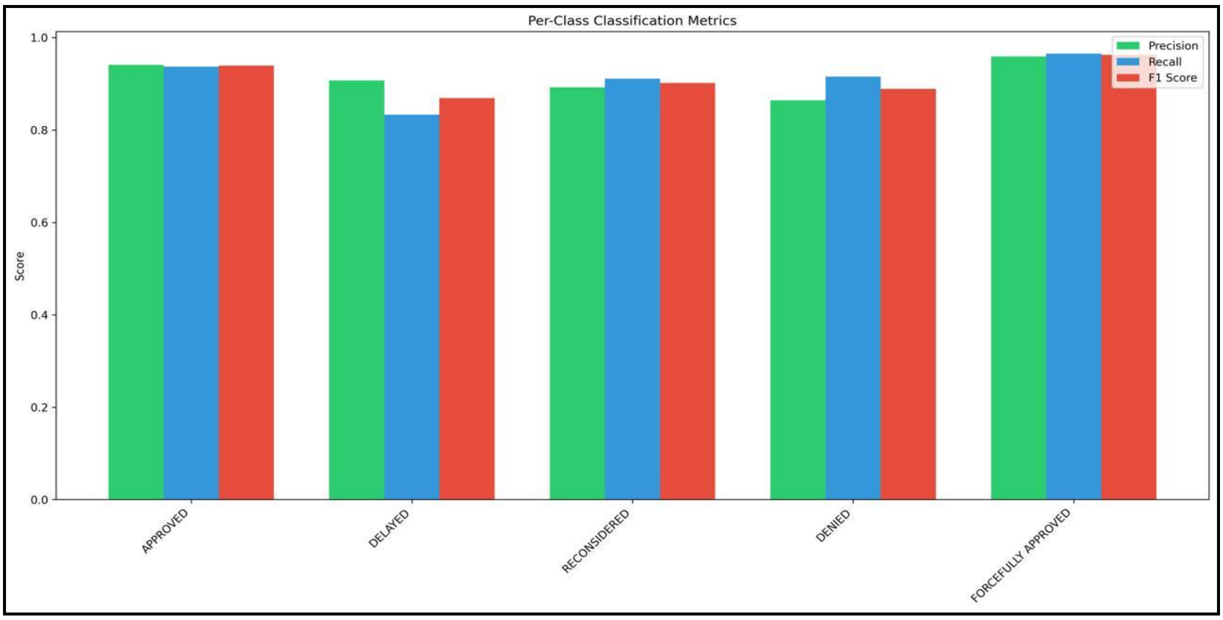

Fig 1: Precision, Recall, and F1-Score for the recommendation task across all classes.

to some overlap in the features but not a faulty decision made by the model. These considerations also indicate model reliability and a thorough fitting of the decision boundary for apparently complex financial decisions.

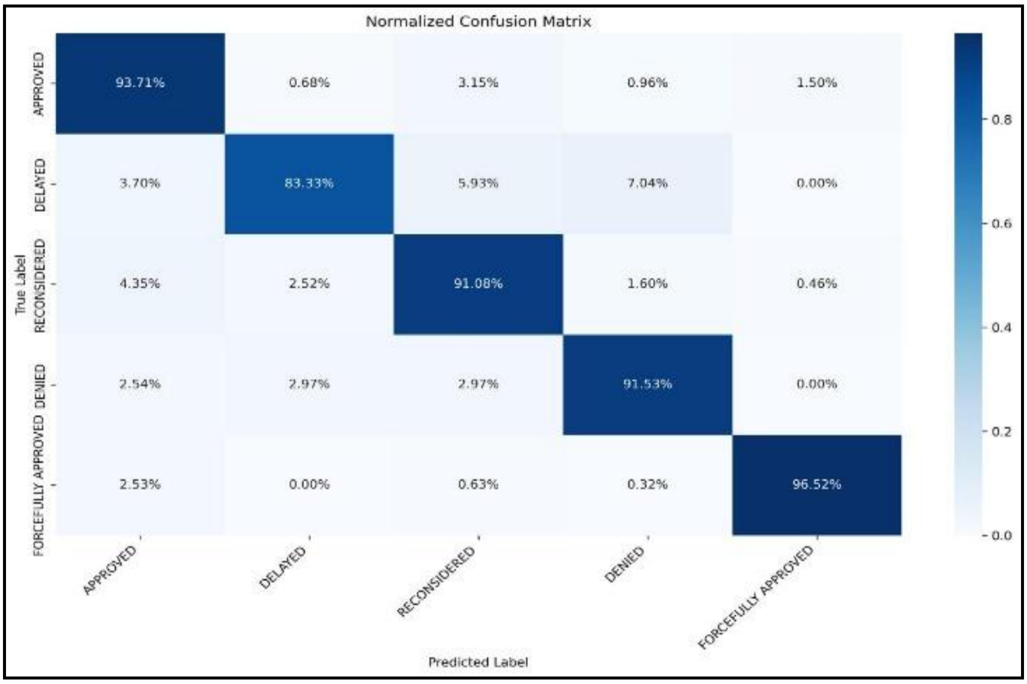

Fig 2: Normalized confusion matrix showing the distribution of predicted versus actual labels for the recommendation task.

A traditional lexical overlap measure is not appropriate for the reasoning task that involves outputting the coherent, logistically valid rationale for the recommendation. Consequently, BERTScore is used to compare the semantic similarity of the model's hypothesis rationale and expert-sourced human ground truth explanations. BERTScore computes the sentence similarity using contextual embeddings extracted from BERT to capture the meaning of phrases and words beyond mere overlap. This method is particularly suited to paraphrase identification and equivalence of meaning identification, exemplified by the similar statements like "this expense fits the budget" and "the budget fits this expense" that will be treated as a highly similar pair. This sensitivity to semantics is critical for the financial domain in comparison to the general domain due to the frequent changes in terminology and the need for a high precision in intent despite varying terminology.

The BERTScore statistics for the fine-tuned DeepSeek-R1-Distill-Llama-8B model indicate its remarkable capacity: 0.80-0.82 for precision; 0.88-0.90 for recall; and 0.85-0.87 for F1-score. High recall implies that most reference explanations critical information gets covered by the model. Lower precision, on the other hand, indi-cates the possibilities of stylistically inflated or slightly related token inclusion, common for jargon-rich generative models. Nevertheless, this precision-recall set made DeepSeek-R1-Distill-Llama-8B a powerful candidate for specialized reasoning generation challenges. Since the BERTScore independently measures semantic preserve, the proposed analysis places emphasis on its precision-recall score rather than general metric in order to assess reasoning quality with-out regard to classification accuracy. This dual metric's impact can thus characterize the decision and the reasoning to maximize the trustworthy automation of financial decisions.

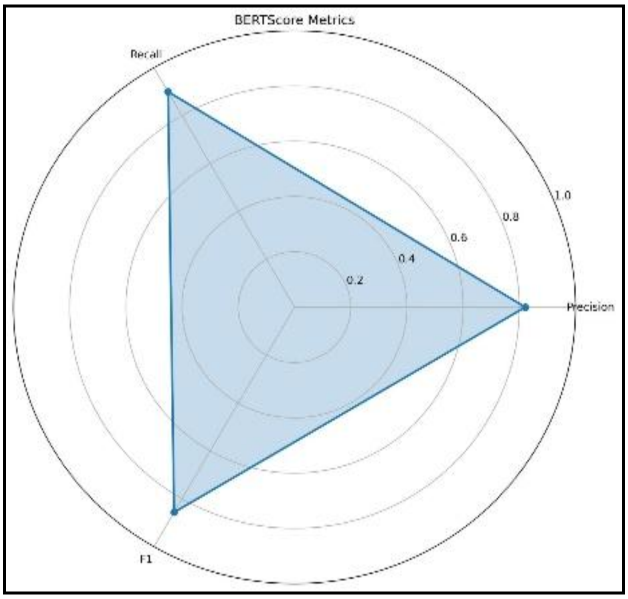

Fig 3: Distribution of BERTScore metrics (precision, recall, F1-score) for the reasoning task.

7. Conclusion and Future Scope

The Financial Decision Helper App features an efficiency in employing Chain-of-Thought reasoning technique that emulates LLM architecture to achieve individualized financial recommendations. It utilizes a deep AI model and technique to deliver precise reasoning and recommendations in response to users' unique transactions, financial habit and conditions.

The Financial Decision Helper App's utilization of CoT reasoning is reminiscent of its application in other fields, such as healthcare, which underscores its versatility and potential for broad impact. By incorporating step-by-step reasoning, the app provides users with transparent insights into their financial behaviors, thereby enhancing their understanding and control over personal spending. This technique, as evidenced in healthcare research, demonstrates how effective CoT is in breaking down com-plex decision-making processes into manageable steps, allowing users to capitalize on this logic for refined financial decision-making. The app's capabilities are further bolstered by dynamic feed-back loops that assess and adjust to user actions in real-time, thus strengthening its cognitive adaptability and ensuring tailored financial advisory services. Through this innovative adaptation, the Financial Decision Helper App not only exemplifies a leap forward in fintech applications but also redefines the potential for AI-driven personalized financial guidance.

As for the future innovations of the Financial Decision Helper App, there are already several avenues that can be explored in helping further enhance the engagement of the users. One possible opportunity lies in the implementation of voice-assisted features that would enable a hands-free interaction of the users with the application. Having this feature also helps to promote the accessibility of the application towards the user especially to the people with disabilities. Another possible innovation opportunity that can be developed in the future regarding the application is the ability to additional integration of financial data sources that users will have access to and be consolidated within the application. Investments accounts and savings accounts are example of sources that if integrated, the user will have a deeper understanding of their overall financial health instead of just having control over their spending habits but the actual state of their net worth. Another possible future innovation regarding the application would be through the development of predictive analytics capabilities that can study the user's behavior and suggest actions that they can take based on patterns that can be observed in their spending and saving habits. This enables the users to have an effective planning method on how to adjust their current behaviors based on projected trends on their future financial situation. These possible future innovations showcase the future of the application and its potential innovations based on the current ongoing trends under research with AI-based financial planning and its impact in creating an enhanced user engagement and financial literacy.

Secondly, the future scope to overcome the verbosity in reasoning part is to follow the chain of draft. Chain of draft means creating the drafts in a sequence and enhancing the previous drafts at each stage. The draft should focus on: Iterative Refinement: Content should be sharp and unambiguous by eliminating the unimportant details. Clarity and Conciseness: Reasoning should always be clear and straight forward. Quality: Final output should be effective and self-explanatory.

Moreover, the integration of sophisticated addictive AI-based features into the Financial Decision Helper App can advance the level of individualization even more. The algorithmic learning will enable the app to modify the specifics of recommendations and financial guidance according to the evolving tendencies of users' behavior. The algorithmic interaction will contribute to a more complicated realization of the user's financial conduct which will also elevate the level of the user's loyalty to the app and contentment with the financial recommendations. For instance, through the app implementation of behavioral analytics and sentiment analysis will construct a visual image of the user attitudes towards their finances. In its turn, it will enable the app to foresee users' possible endeavors and deliver them the financial recommendation even before they will require it. Such sophisticated AI-based methodologies align with the overarching fintech trend to adopt innovative approaches to ameliorate the quality of app usability and reactivity and assist the users with their financial decision-making.

References

1. Qian H, Zhang L, Chen Y, Li P. Enhancing financial decision-making with chain-of-thought reasoning. 2025. p. 1-15.
2. Easin M, Rahman T, Chowdhury A, Karim M. Interoperable AI systems for financial applications. 2024. p. 45-63.
3. Alonso NI. Large language models in finance: reasoning [Internet]. 2024 [cited 2025 Sep 6]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5048316. p. 1-24.
4. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, *et al*. Chain-of-thought prompting elicits reasoning in large language models. Adv Neural Inf Process Syst. 2022;35:24824-37. Available from: https://proceedings.neurips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html.
5. Saiyed A. AI-driven innovations in fintech: applications, challenges, and future trends. Int J Electr Comput Eng. 2024;14(2):1001-1014.