Üretken Yapay Zekaya Dayalı Bireysel Emeklilik Bilgilendirme ve Öneri Sistemi

Araştırma Makalesi/Research Article

Ezgi AVCI[1]*, Mehmet Furkan ATİK[2], Nur Muazzez MEMİŞ[3]

[1]Department of Applied Data Science, TED University, Ankara, TURKEY
[2]Department of Digital Product Management, Strategy and System Development, Pension Monitoring Center, İstanbul, Türkiye
[3]Department of Advanced Analytics and Research Services, Pension Monitoring Center, İstanbul, Türkiye

ezgi.avci@tedu.edu.tr, furkan.atik@egm.com.tr , muazzez.memis@egm.org.tr
(Geliş/Received:29.05.2024; Kabul/Accepted:25.06.2024)
DOI: 10.17671/gazibtd.1475239

Özet— Bu makale, üretken yapay zeka (GenAI) ile güçlendirilmiş yenilikçi bir bireysel emeklilik bilgi ve tavsiye sisteminin tasarımını sunmaktadır. Sistem, kullanıcı verilerini analiz etmek ve kişiselleştirilmiş emeklilik planlama tavsiyeleri üretmek için gelişmiş AI algoritmalarını kullanacak şekilde özelleştirilmiştir. GenAI entegrasyonu ile sistem, kullanıcılar arasında finansal okuryazarlığı önemli ölçüde artırmayı, emeklilik planlaması ve finansal ürünler hakkında daha derin bir anlayış sağlamayı hedeflemektedir. GenAI destekli içgörüler, kullanıcıların uzun vadeli emeklilik hedefleri ve risk tercihleriyle uyumlu bilinçli kararlar alabilmelerini sağlayacak şekilde özelleştirilmiş yatırım stratejilerini kolaylaştıracaktır. Bu yaklaşım, sadece bireysel finansal sonuçları iyileştirmeyi amaçlamakla kalmayıp, geleneksel olarak yalnızca finansal danışmanlar aracılığıyla erişilebilen finansal tavsiyeye erişimi demokratikleştirmeyi de hedeflemektedir. Sistem geliştikçe, değişen ekonomik koşullara ve kişisel durumlara uyum sağlaması, kullanıcıların yaşam değişiklikleriyle uyumlu dinamik tavsiyeler sunması beklenmektedir. Bu sistemin amacı, emekliliğe yaklaşırken ve emekliliğe girerken kullanıcılarının finansal refahını ve güvenliğini artıracak şekilde proaktif bir emeklilik planlaması yaklaşımını teşvik etmektir.

Anahtar Kelimeler— bireysel emeklilik, üretken yapay zeka, yapay zeka, karar destek sistemi

GenAI-Based Private Pension Information and Recommendation System

Abstract— This paper presents the design of an innovative private pension information and recommendation system powered by generative artificial intelligence (GenAI). The system is tailored to leverage GenAI algorithms to analyze user data and generate personalized retirement planning advice. By integrating GenAI, the system seeks to significantly enhance financial literacy among users, providing them with a deeper understanding of retirement planning and financial products. The GenAI-driven insights will facilitate tailored investment strategies, enabling users to make informed decisions that align with their long-term retirement goals and risk preferences. This approach aims not only to improve individual financial outcomes but also to democratize access to financial advice that is traditionally available only through financial advisors. As the system evolves, it is expected to adapt to changing economic conditions and personal circumstances, offering dynamic advice that keeps pace with users' life changes. The goal of this system is to foster a proactive approach to retirement planning, thereby enhancing the financial well-being and security of its users as they approach and enter retirement.

Keywords— private pension, generative artificial intelligence, artificial intelligence, decision support system

1. INTRODUCTION

The importance of private pension planning cannot be overstated in today's financial landscape. In an era marked by rapid economic changes and increasing life expectancy, effective retirement planning has become crucial for ensuring long-term financial security. Retirement planning involves understanding and managing various financial elements, including savings, investments, pensions, and other income streams, to achieve a financially secure retirement. However, financial literacy—the ability to understand and effectively use various financial skills, including personal financial management, budgeting, and investing—is a critical component that influences the effectiveness of individual retirement planning. Financial literacy is not merely about understanding basic financial concepts; it's about applying that knowledge to make informed decisions about saving, investing, and managing money, which ultimately affects one's quality of life in retirement. Unfortunately, many individuals lack the financial literacy required to navigate the complex landscape of retirement options effectively. Studies have shown that a significant portion of the population lacks the knowledge necessary to make decisions that optimize their retirement outcomes. This lack of understanding can lead to poor financial choices, such as inadequate savings, inappropriate investments, and a general lack of preparation for retirement.

In response to these challenges, the emergence of generative artificial intelligence (GenAI) in financial services presents a transformative opportunity to enhance decision-making and improve individual outcomes. GenAI, particularly models like Generative Pre-trained Transformer (GPT), has revolutionized various sectors by providing advanced data processing capabilities and generating human-like text based on the input they receive. In the context of financial services, GenAI can analyze vast amounts of financial data, recognize patterns, and provide personalized advice based on an individual's financial status, goals, and risk tolerance. The application of GenAI in financial decision-making extends beyond mere automation. It involves understanding complex financial scenarios, adapting to new financial information, and evolving with the user's changing financial conditions and life stages. For retirement planning, this means that GenAI can assist users in crafting a comprehensive, dynamic financial plan that adjusts over time, providing recommendations on savings rates, investment choices, and other financial decisions that align with the user’s retirement goals and financial situation.

The primary objective of the proposed GenAI-driven system is to develop a tool that harnesses the power of GenAI to deliver personalized retirement planning advice directly to end-users, empowering them with tailored financial insights and strategies. While the system primarily targets individual users seeking to enhance their retirement planning, it also offers valuable support for financial advisors and intermediaries by providing them with advanced tools to better serve their clients. This system aims to bridge the gap between complex financial knowledge and the average individual's understanding. By integrating GenAI, the system can process detailed personal financial information provided by users, including their current financial status, retirement goals, risk tolerance, and other relevant data. Utilizing advanced algorithms, the system can generate customized advice that guides users through the myriad options available, suggesting strategies to increase their retirement savings, optimize their investment portfolio, and manage risks associated with their financial decisions. Moreover, the GenAI system is designed to be interactive, allowing users to receive explanations and clarifications on financial concepts and the reasoning behind specific advice. This educational component is crucial as it helps to enhance the user's financial literacy over time, empowering them to make informed decisions independently. The system’s ability to learn from user interactions and feedback also enables it to improve its advice, ensuring that it remains relevant as economic conditions change and as the user's personal financial situation evolves.

2. BACKGROUND and LITERATURE REVIEW

2.1. Rationale to enter the Individual Pension System (IPS)

2.1.1. *A Comfortable Retirement*

The IPS has been created as a supplement to the social security system to increase one’s wealth by providing them with an additional income during retirement. One can participate in the IPS by acquiring a pension product from their choice of pension company. Eligibility for retirement requires to have been stayed in the system for at least 10 years and attained the age of 56. There are three different options to be chosen from after becoming eligible for retirement. (1) One can simply request a lump-sum payment of some or all the accumulations in the individual pension and state contribution account (2) One choose to stay in the IPS and receive accumulations in part from the pension company according to the reimbursement plan they decide. (3) One can earn regular income by purchasing an annuity with some or all the accumulations.

2.1.2. *Transparency*

One can access their daily individual pension account information via two different platforms: (1) from the BES Mobile application, provided by the Pension Monitoring Center (EGM) (2) from the customer portal on one’s pension company's website (3) from the mobile app provided by the pension company. Since the individual pension accounts are kept by the Istanbul Settlement and Custody Bank (TakasBank), participants can also keep track of their daily fund shares from TakasBank's accounts.

2.1.3. *Right to Choose the Fund*

The pension company offers participants a pension plan suitable for their pension expectations, income level and age. It is up to the participants to decide on the funds in which they wish to invest their contributions. If one does not choose any funds, their contributions shall be invested in the funds stipulated by the legislation. Also, Pension Fund Trading Platform (BEFAS) enables participants to purchase funds from different companies without requiring them to change their pension companies. Fund distribution can be changed 12 times per year.

2.1.4. *State Contribution*

A state contribution is paid into the participant's account in the amount of up to 30 percent of the contributions they made to the IPS. A participant can benefit from the state contribution for the contribution payments up to the total amount of the gross minimum wage determined for the relevant year within a calendar year. Contribution payments exceeding the limit will be considered in the calculation of the state contribution as carryover contribution shares transferred to the company accounts on the first day of each year. Separate state contribution limits have been set for the contracts within the scope of Auto Enrollment System (AES) and IPS, respectively.

2.1.5. *Professional Fund Management*

Participant's accumulations are managed by the portfolio management companies overseen by the Capital Markets Board of Turkey (SPK).

2.1.6. *Purchasing Service from Licensed Intermediaries*

Only licensed intermediaries can engage in the IPS promotion, marketing and sales activities. The professional competency of these intermediaries is tested with an exam conducted by the Pension Monitoring Center.

2.1.7. *Flexible Contribution Payments*

In addition to the regular contributions, participants can also make initial or additional contribution payments in interim periods to the IPS; they can also suspend contributing anytime.

2.1.8. *Transfer to Another Pension Company*

If one wishes, savings can be transferred to another pension company at least two years after the effective date of the contract. The waiting period is one year for subsequent transfers.

2.1.9. *Tax Advantage for Employers*

Employers can enter employees into the IPS by drawing up a group pension contract. In this case, the contributions paid on behalf of employees are recorded as expenses when calculating business income without associating them with the salary. Employers may determine the vesting period for the accumulations arising from contributions paid for the employees and from their earnings on the condition of not exceeding seven years. This is a good way to achieve employee motivation and loyalty to the company.

2.1.10. *Reliability*

All elements of the system are under supervision. The operations of all the pension companies are supervised by the Insurance and Private Pension Regulation and Supervision Agency (SEDDK) while the operations of the funds, portfolio managers and custodians are supervised by the SPK. The company's accounts and transactions are subject to independent external audits at least once a year. EGM monitors all pension companies' operations electronically daily. Assets that belong to pension companies and to pension mutual funds are separate from each other. Fund assets cannot be pledged, used as collateral except for portfolio-related transactions, seized, or included in bankruptcy assets. SEDDK may impose sanctions against a pension company if the financial structure of the pension company is deemed to have been weakened to a level that may put the participants' rights and interests in danger and if it is impossible to remedy the situation. In the event of the death of the participant during the contract period, their accumulations and all of the state contributions in their account shall be paid to the beneficiaries stated in the pension contract or lawful heirs stated.

The procedures and principles regarding instant monitoring of the return performance of pension mutual funds, rewarding the managers of the funds with relatively high returns during the evaluation period, and imposing sanctions on the managers of relatively unsuccessful funds have been regulated. The task of operating the system has been entrusted to EGM stakeholders.

2.2. Challenges of entering the IPS

2.2.1. Complexity and Accessibility

One of the primary challenges is the inherent complexity of many retirement systems. The options available can be bewildering, with a range of investment choices, contribution levels, and rules regarding withdrawal. This complexity can act as a barrier to effective participation, particularly for those with lower levels of financial literacy.

2.2.2. Lack/Level of Financial Literacy

Financial literacy plays a critical role in effective retirement planning. It influences an individual's ability to make informed financial decisions, which in turn affects their financial well-being in retirement. However, financial literacy levels vary significantly across different demographics, impacting the effectiveness of retirement planning efforts. Studies have shown that financial literacy is not uniformly distributed across populations. Younger people, women, and those with lower income levels often exhibit lower levels of financial literacy. This demographic variation presents a challenge to universal retirement planning efforts, as these groups are often the most vulnerable to poverty in retirement. Lower levels of financial literacy are strongly correlated with poor retirement planning. Individuals with limited financial knowledge are less likely to calculate their retirement needs accurately, less likely to invest in diversified portfolios, and more likely to opt out of beneficial retirement plans.

3. THEORETICAL FRAMEWORK AND METHODOLOGY

3.1. Generative AI

GenAI signifies a major leap in AI, especially for its capacity to create new content. This technology leverages *large language models* (LLMs) that process extensive datasets of text and documents to produce coherent outputs in various formats, including text, images, and languages. Its capabilities have substantial implications across multiple sectors, particularly in finance. A pivotal advancement in LLMs was the introduction of the Transformer architecture by Vaswani et al. in their seminal paper "Attention is All You Need" in 2017 [1]. This architecture, using the *self-attention mechanism* [2], supports parallel processing and effectively manages long-range dependencies. This framework underpins models like OpenAI’s GPT series [3] and Google’s BERT [4], which have achieved remarkable results across various language tasks. These models are designed to interpret entire contexts, providing genuine contextual understanding [5]. Notable variations of the GPT architecture include models like ChatGPT, Llama, and Mistral [6] which share a consistent training approach: During *pre-training*, the models learn grammar, information, reasoning abilities, and some level of common-sense knowledge from a wide range of internet text [7][8]. This broad linguistic understanding equips them well for various tasks. *Fine-tuning* then focuses these models on specific applications, like conversational contexts for ChatGPT, making them suitable for applications such as chatbots and virtual assistants [9][10][11]. While models like Llama and Mistral are less well-known, they may represent advancements or specialized variants designed for specific research or practical uses. These models are at the forefront of NLP developments, facilitating more natural human-like interactions through advanced AI-driven language models [12][13][14].

3.2. Generative AI in Financial Advisory

Robo-advisors are digital platforms that provide financial investment services through algorithmic processes, modernizing and automating traditional financial advising and wealth management using advanced information technologies [15] [16]. These platforms enhance passive indexing strategies to reduce costs, risks, and the time required for portfolio rebalancing. They also mitigate behavioral biases and improve financial transaction efficiency [17]. The Covid-19 pandemic has boosted the popularity of robo-advisors, which currently manage over $1 trillion in assets and are expected to grow annually by 16.72% from 2022 to 2025 [18][19] noted that the pandemic highlighted the value of online investment services, prompting even those who preferred traditional in-person banking or human advisors to shift towards digital solutions. This shift has led to a significant increase in robo-advisor-related research, particularly over the past six years.

Research on robo-advisors covers several domains, including their impact on financial literacy, adoption and acceptance, investment performance, risk management, regulation and compliance, and human-robot interactions. [16] examined design principles, factors influencing adoption and acceptance, and portfolio recommendations. Additionally, the ongoing digitalization of the economy and the rise of fintech have accelerated the integration of digital experiences in financial services, transforming how services and channels converge [20]. The high-quality and effective services offered by robo-advisors ensure client satisfaction by providing tailored and personalized solutions [21][22]. *Personalization*, a key competitive edge, enhances sales by improving conversion rates, strengthening customer relationships, fostering loyalty, and boosting revenues. It involves tailoring offers to meet individual customer needs, but understanding these unique needs can be challenging. Personalization anticipates needs based on previously gathered data, whereas customization directly responds to specific client requests. AI and digital technology enable these personalized experiences by analyzing patterns from contextual and personal data [23] [24]. As technology and the industry continue to develop, the research focus within the robo-advisory services sector is expected to adapt and expand. Despite increased academic attention, there is still a notable research gap in understanding how robo-advisory services can be more effectively cater to the distinct preferences and requirements of individual investors.

3.3. Impact on Financial Literacy and Retirement Planning

GenAI and robo-advisors significantly influence financial literacy and retirement planning. Individuals need to possess sufficient financial knowledge about pension products and information provided by pension companies and be capable of making informed investment decisions. [25] developed a private pension literacy scale to determine the pension literacy levels of academic and administrative staff and found that 36.7% of respondents had medium and 8% had high levels of individual retirement literacy. The findings suggest that participants have a low level of private pension literacy. A significant difference was observed between individual retirement literacy levels and variables such as gender, marital status, and age. [26] studied the relationship between financial literacy levels of students and their awareness of the individual pension system. They found that students who attend financial courses more frequently have higher financial literacy scores and greater awareness of the private pension system.

A positive but weak correlation was found between students' financial literacy levels and their awareness of the private pension system. The influence of GenAI and robo-advisors extends into the realm of financial literacy and retirement planning. Literature suggests that AI tools can enhance users' understanding and engagement with their retirement plans, simplifying complex financial concepts and promoting proactive management of retirement savings.

3.4. Trust and Satisfaction in Retirement Systems

A critical factor in adopting new technologies in personal finance is user trust and satisfaction. [27] study on trust and satisfaction perceptions within individual retirement systems indicates that AI's personalization capabilities can lead to higher satisfaction and trust among participants, fostering more robust engagement with the retirement planning process. The Private Pension System has been promoted through recent regulations and legislative changes. Examining these applications reveals that the system has a solid legislative structure, which creates trust. This trust and the profitability provided by the system are significant factors in customer satisfaction. The study found that participants' perceptions of trust and satisfaction were slightly above average and that customer trust positively impacted customer satisfaction.

To establish trust in the Individual Retirement System, accurate and complete information must be provided to potential participants in a clear and understandable manner. Individuals are reluctant to commit to a system they do not fully understand. Therefore, it is recommended that investment advisors assigned by banks provide personalized and frequent updates rather than just general investment bulletins. This approach enhances the perception of helpfulness, which significantly impacts trust. Additionally, investment advisory services are a key advantage of individual pension funds. Through fund management, individuals can generate additional income beyond their contributions. In investment and savings systems, the perception of competence stands out more than accuracy/honesty and helpfulness, as the primary goal of these systems is to enable individuals to save in addition to their social security retirement income. Thus, increasing individuals' perception of competence can be achieved by ensuring they believe they are part of a profitable investment system [27].

4. DECISION SUPPORT SYSTEM DEVELOPMENT

The proposed GenAI-driven decision support system (DSS) is designed to address the complexities of retirement planning by integrating advanced GenAI technologies. This system aims to provide dynamic, personalized retirement planning advice tailored to individual user profiles, enhancing both the precision of financial guidance and the user's financial literacy over time.

4.1. System Architecture

The architecture of the proposed system (Figure 1) is structured around a core GenAI engine that leverages a GenAI model, similar in functionality to models like ChatGPT, which is fine-tuned for financial advising and retirement planning.

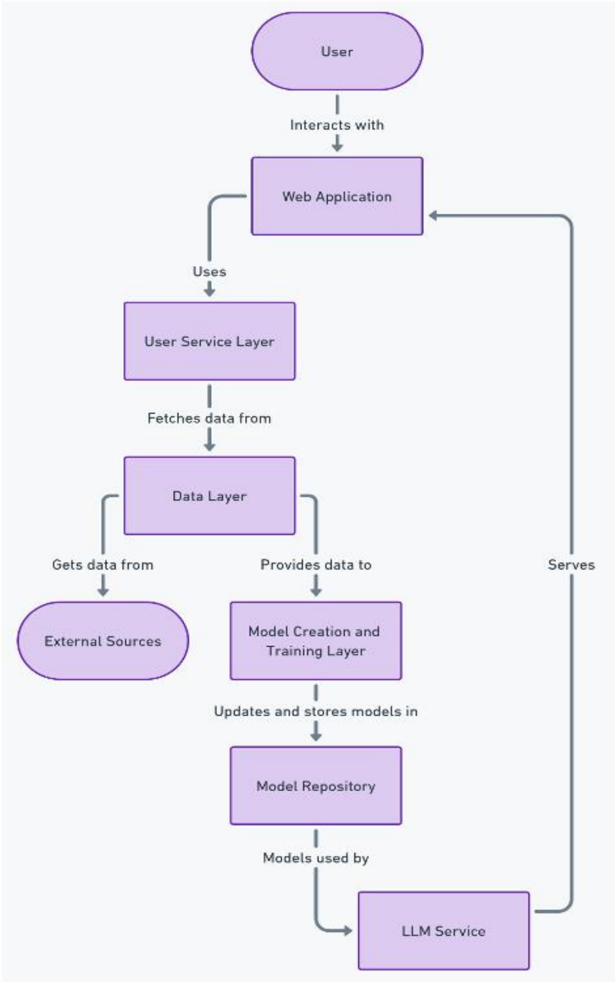

Figure 1. System architecture

The system architecture is composed of several key components:

- **User:** This component represents the end-user who interacts with the system. Users are the primary actors who initiate queries and receive responses, driving the operational flow of the application.

- **Web Application:** Serves as the interface through which users interact with the system. It is designed to be user-friendly and responsive, facilitating easy navigation and efficient access to services provided by the underlying architecture.

- **User Service Layer:** This layer acts as a mediator between the user and the data layer. It processes user requests, retrieves the necessary data from the data layer, and ensures that responses are tailored to the user's needs. This layer improves user experience by optimizing data retrieval and response generation.

- **Data Layer:**

Central to the system's operation, this layer manages the acquisition, storage, and provision of data. It fetches data from both internal and external sources and serves it to other parts of the system for further processing. The data layer is essential for ensuring that the GenAI system and its agents have access to high-quality, relevant data necessary for generating accurate and personalized retirement planning advice.

The Extract, Transform, Load (ETL) process is a crucial component of the data layer, converting raw data into a structured format suitable for machine learning models. The ETL process consists of three main stages:

Extract:
Data is extracted from various internal and external sources, including user input, financial databases, market data feeds, and other relevant financial information. The data layer connects to external databases or APIs to fetch data that enriches the application's responsiveness and functionality. This include integrating real-time data feeds or accessing large datasets relevant to the user's queries. The data can come in various formats such as CSV, JSON, XML, and SQL databases. Extraction tools and custom-built scripts are used to gather data efficiently from the aforementioned multiple sources.

Transform: Raw data is cleaned to remove inconsistencies, duplicates, and errors. This step ensures the quality and reliability of the data. Data is standardized to a consistent format, making it easier to integrate and analyze. This includes converting currencies, normalizing date formats, and ensuring consistent units of measurement. Additional relevant information is added to the data to enhance its value. This include deriving new features, aggregating data, or integrating supplementary data from external sources. Transformation tools (Apache Spark, AWS Glue) and Python libraries (Pandas, NumPy) are used to process and transform the data.

Load: Transformed data is loaded into a data warehouse or a data lake, where it can be accessed by the GenAI system and its agents for further analysis and processing. For our system, we chose a hybrid approach using both a data warehouse and a data lake.

We use data lake for storing raw data, log files, market data feeds, and other large datasets. This allows for flexible data exploration and transformation before loading into the data warehouse for detailed analysis. We chose Amazon S3 which provides a highly scalable and flexible storage solution for unstructured and semi-structured data. It allows for the storage of large volumes of data in its raw form.

We use data warehouse for storing structured data, such as user profiles, financial transactions, and historical performance metrics. We chose Amazon Redshift which offers high-performance, scalable, and cost-effective data warehousing solutions. It supports complex queries and integrates seamlessly with other AWS services.

Model Creation and Training Layer: This critical layer receives data from the data layer and utilizes it for training and refining language models. The layer incorporates machine learning algorithms and data processing techniques to develop models that understand and generate language effectively.

- **Model Repository:** Acts as a storage unit for all the language models created and refined by the model creation and training layer. This repository ensures that models are readily accessible for use, updates, and further refinements.

- **LLM Service:** This service utilizes the models stored in the model repository to process user queries and generate appropriate responses. It is an essential component that applies natural language understanding and generation capabilities to meet user needs effectively.

The architecture illustrated in Figure 1 underscores a structured approach to managing user interactions, data processing, and language model utilization. Each component is strategically placed to optimize data flow and service delivery, ensuring the system is efficient, scalable, and responsive to the dynamic needs of users. This system architecture supports continuous improvement and adaptation of language models, enhancing the overall quality and effectiveness of the web application.

4.2. User Interface Design

The UI is designed to be intuitive and accessible, catering to users with varying levels of financial literacy. It provides clear navigation, straightforward data input methods, and the presentation of advice in a digestible format.

4.2.1. Front end

Figure 2 details the process flow from a user query to the response delivery in a web application environment. Each step in the sequence diagram is explained below:

- **User:** Represents the individual or system initiating the interaction. The user inputs a query through the user interface (UI), initiating the sequence of operations. The user's role is crucial as the originator of requests that drive the application's functional responses.

- **UI:** The front-end component where the user inputs their query. This interface is designed to be intuitive and user-friendly, providing a seamless experience that efficiently captures user inputs and communicates them to the back-end systems.

User Service Layer

- **Forward Query:** Once the query is received from the UI, it is forwarded to the User Service Layer. This layer acts as a conduit, passing the query deeper into the system for processing.
- **Authentication & Authorization checks:** Critical for security, the User Service Layer performs checks to verify the user's identity and ensure they have the necessary permissions to perform the query. This step prevents unauthorized access and ensures compliance with security policies.

LLM Service

- **Forward Query:** After passing authentication and authorization, the query is forwarded to the LLM Service. This transition is crucial as it moves the query to the processing phase.
- **Process Query:** The LLM Service processes the query using pre-trained language models. This involves interpreting the query's intent and generating an appropriate response based on the information available to the model.
- **Send Response:** Once the response is generated, it is sent back through the User Service Layer to the UI. This marks the completion of the query processing phase.

UI (User Interface)

- **Display Response:** The final step in the sequence where the processed response is displayed to the user. This phase is critical as it provides the user with the information or feedback they sought, completing the interaction cycle.

This sequence diagram illustrates a streamlined interaction flow within a web application, emphasizing the critical steps of query processing and response generation. Each component—from the UI to the LLM Service—plays a pivotal role in ensuring the system's responsiveness and security. The process effectively demonstrates how user queries are handled, authenticated, processed, and responded to within a modern web application architecture, providing a clear and efficient pathway from user input to system output.

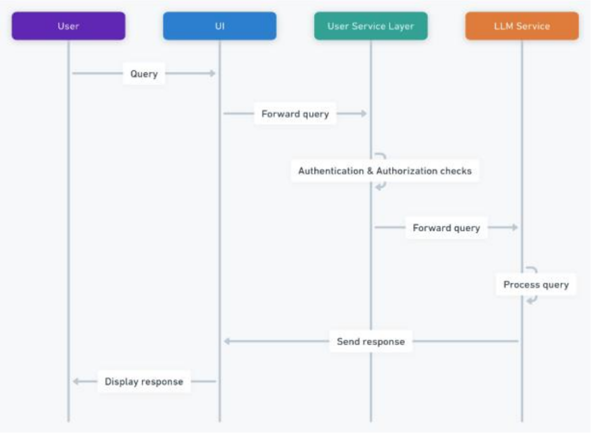

Figure 2. Front-end Sequence Diagram

4.2.2. Back-end

Figure 3 represents a back-end sequence diagram detailing the workflow from data collection to model utilization in a system that employs LLMs.

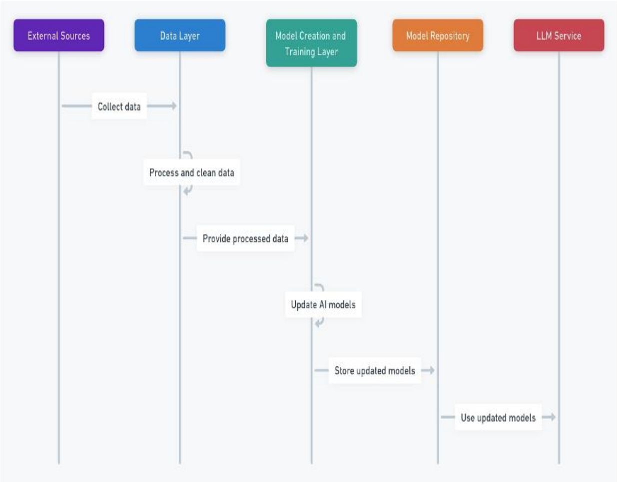

Figure 3. Back-end sequence diagram

External Sources

- **Collect Data:** This initial step involves gathering data from various external sources. These sources may include databases, web services, APIs, and other data repositories that provide the raw data necessary for training and refining language models. The efficiency and breadth of data collection are crucial for building robust and versatile models.

Data Layer

- **Process and Clean Data:** Once data is collected, it moves to the data layer, where it undergoes processing and cleaning. This step is essential to ensure the quality and usability of the data. Processing involves converting raw data into a more structured format, while cleaning involves removing inaccuracies, duplicates, and irrelevant information, which helps in enhancing the data's reliability for training purposes.

- **Provide Processed Data:** After processing and cleaning, the data is made available to the next layer. This provision ensures that the data is in a usable state, formatted correctly, and free from elements that could degrade the model's performance.

Model Creation and Training Layer

- **Update AI Models:** This layer receives the processed data and uses it to update or train new AI models. Updating models involves adjusting existing models based on new data, which can include retraining or fine-tuning. This step is crucial for keeping the models current and effective, allowing them to adapt to new information or changes in data trends.

Model Repository

- **Store Updated Models:** Once the models are updated, they are stored in the Model Repository. This storage is vital for version control, accessibility, and deployment. The repository ensures that all model versions are tracked and can be retrieved or rolled back if needed. This feature supports continuous improvement and testing of models before they are deployed into production.

LLM Service

- **Use Updated Models:** The final step in the sequence involves the LLM Service utilizing the updated models to perform tasks such as data analysis, prediction, or any other language-related processing required by the application. This step highlights the practical application of the models, where they serve the end-user or automated systems by providing intelligent responses based on learned patterns and data.

Each step in the bacj-end sequence is geared towards enhancing the data quality, model accuracy, and overall system efficiency, ensuring that the end-service remains relevant and effective in real-world applications. This workflow not only underscores the systematic approach to data-driven model training but also highlights the continuous cycle of improvement and adaptation that characterizes advanced AI-driven systems.

4.3. Integration of LLM for Dynamic, Personalized Pension Planning

The integration of GenAI into the system allows for dynamic and personalized pension planning in several ways. Firstly, the system can simulate various financial scenarios based on current trends and possible future events, such as economic downturns, inflation changes, or shifts in the stock market. This helps users understand potential risks and the impact of different retirement planning decisions. Secondly, as financial markets fluctuate and personal financial situations change, the AI updates its advice in real-time. This ensures that users always have access to the most current and relevant financial guidance.

4.3.1. User Query LifeCycle

Figure 4 represents the User Query Life Cycle, detailing the process and infrastructure required for managing user queries within the system. This comprehensive breakdown of the User Query Life Cycle illustrates how each component is crucial for ensuring the system not only responds accurately to user queries but also manages these interactions securely and efficiently. This cycle is pivotal for maintaining high user satisfaction and system reliability.

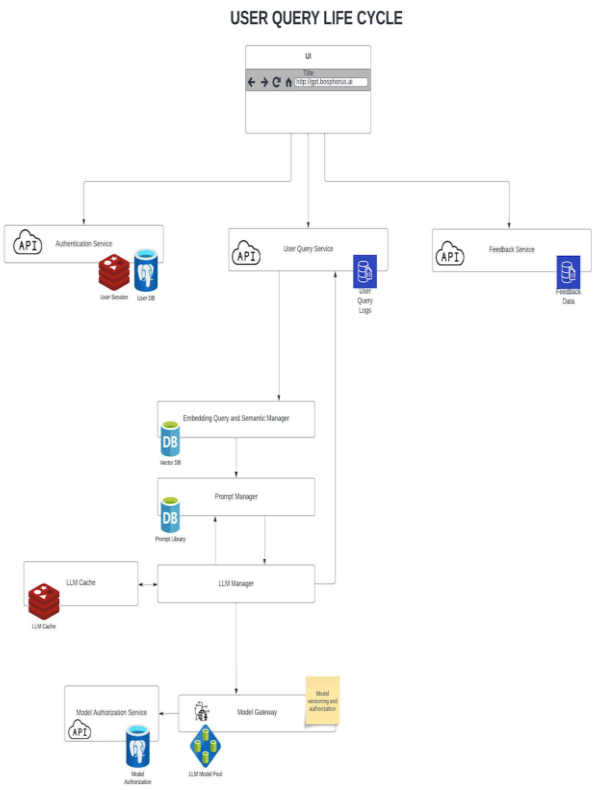

Figure 4. User Query Life Cycle

1. User Interface (UI): The interface serves as the primary interaction point between the user and the system. It's designed to be intuitive, enabling effortless navigation and query input. Innovation in UI design incorporates adaptive interfaces that adjust to user preferences and behaviors, enhancing engagement and satisfaction.

2. Authentication Service

- **API for Authentication Service:** This component is critical for security and personalization. It ensures that user interactions are secure and that data is only accessible to authenticated users. The innovative use of session management and advanced encryption enhances user trust and system integrity.

3. User Query Service

- **User Query Service API:** Central to the query processing framework, this service efficiently manages the distribution and logging of user queries. Its innovative aspect lies in its ability to handle massive volumes of queries simultaneously while maintaining high performance and reliability.

4. Feedback Service

- **Feedback Service API:** This service collects user feedback to continuously refine and improve the response quality. The innovative use of machine learning algorithms to analyze feedback helps in dynamically adjusting query responses based on real-time user input, greatly enhancing the learning capabilities of the system.

5. Data Management and Processing

- **Embedding Query and Semantic Manager:** By converting queries into semantic embeddings, this manager plays a crucial role in understanding user intent more deeply. The use of cutting-edge natural language processing technologies to generate and manage these embeddings represents a significant innovation, leading to more precise and context-aware responses.
- **Prompt Manager:** This component manages a library of query prompts to streamline response generation. The innovation here includes the use of advanced retrieval algorithms that can quickly match user queries with the most effective prompts, greatly speeding up response times and improving accuracy.

6. LLM Cache: The caching mechanism is vital for enhancing system efficiency. It reduces latency by storing frequently or recently accessed data. The innovative application of smart caching algorithms predicts which data is likely to be reused, thus optimizing resource allocation and speeding up query processing.

7. LLM Manager: It intelligently coordinates between cached and new data processing, ensuring optimal system performance. This manager's innovative feature includes the use of decision algorithms that automatically determine the most efficient way to process a query, considering current system load and query complexity.

8. Model Authorization and Access

- **Model Authorization Service API:** This ensures that all operations within the system comply with legal and ethical standards. The innovative use of policy-driven access controls and real-time monitoring tools safeguard sensitive data and ensure compliance with global data protection regulations.
- **Model Gateway:** Acts as a secure conduit between the query services and the LLMs, managing access permissions and resource allocation. Its innovative aspect is the dynamic scaling capability, allowing it to handle varying loads seamlessly and ensuring that resources are efficiently utilized.

9. LLM Model Pool: A sophisticated collection of language models tailored to various aspects of query processing. The innovation here lies in the models' ability to learn from interactions continuously and adapt to new information, improving their accuracy and relevance over time.

4.3.2. Integration of Agents within the GenAI System

Agents are employed to perform specific, well-defined tasks such as mathematical calculations, data preprocessing, and basic financial modeling. For instance, calculating the future value of retirement savings based on current contributions and expected interest rates can be handled by a simple machine learning model or even a deterministic algorithm. By delegating these tasks to smaller models, the system can operate more efficiently, reserving the computational power of the GenAI model for more complex, context-aware tasks such as generating personalized advice and interpreting user queries.

LangChain Agents facilitate the integration of multiple machine learning models and APIs into a cohesive system. This approach allows for modularity, where each agent performs a specific function, and their outputs are fed into the GenAI system for further processing. Using LangChain Agents, the system optimizes workflows by chaining together different models and tools, ensuring that each task is handled by the most appropriate component.

In our system, a user’s query about their retirement savings initiate the following sequence: (1) The user inputs their current savings, expected contributions, and retirement goals. (2) An agent model calculates the projected future value of the savings using a basic financial formula. (3) The result is passed to the GenAI system, which integrates this information with broader financial data and generates a comprehensive retirement plan. The GenAI model provides personalized advice, considering the user’s risk tolerance and investment preferences. (4) The user interacts with the GenAI system to refine their plan, ask questions, and explore different scenarios.

4.3.3. LLM Training Life Cycle

This life cycle (Figure 5) demonstrates the sequential and structured processes involved in preparing, training, and deploying LLMs:

- **Data Sources:** This is the initial phase where diverse data sets are collected to train the LLMs. The data encompasses:
    - **Documents:** Written content sourced from various formats and domains.
    - **User Query Logs:** Data collected from user interactions, providing insight into user intent and language usage.
    - **Feedback Data:** Information derived from user responses to the system, crucial for iterative model improvements.
    - **Oracle EGM Data Repo:** A specialized repository that might include structured data, relevant for domain-specific training.

- **Data Crawler Layer:**

This layer is designed to handle the systematic retrieval and processing of data. This layer encompass both web scraping (for real-time data extraction from the web) and the transfer of "cold data" or file-based data into a structured data catalogue.

1. **Web Scraping for Real-Time Data Retrieval:** Aims to extract real-time financial data, market trends, and news relevant to retirement planning from various online sources. There are three types of patterns: (1) Factory pattern is used to create different types of web crawlers based on the source and data structure. For example, one factory might create a crawler for financial news websites, while another creates a crawler for market data APIs (2) Iterator pattern allows the crawler to systematically iterate through the pages or API responses to retrieve data. (3) Visitor pattern applies specific operations to each element of the retrieved data, such as parsing HTML, extracting relevant information, and transforming it into a usable format.

   The process initializes by crawler' identifying the target websites or APIs and the data structures used. The crawler navigates through web pages or API endpoints, extracting the necessary data. This can include parsing HTML content, handling pagination, and managing API rate limits. The extracted data is processed to remove unnecessary elements, normalize formats, and ensure consistency. The cleaned and structured data is then stored in the data warehouse or data lake for further analysis and use by the GenAI system.

   Use Case: The system initializes a web scraper for a financial news website. The scraper navigates through the website, extracting headlines, summaries, and publication dates. The HTML content is parsed, and the extracted data is cleaned and structured into a JSON format. The structured data is stored in the data lake for real-time analysis and integration into user recommendations.

2. **Crawling for Cold Data Transfer:** Aims to transfer large volumes of historical or "cold" data from various sources such as databases, file systems, and data repositories into a structured data catalogue. There are three types of patterns: (1) Factory Pattern is used to create different types of crawlers for various data sources, such as SQL databases, CSV files, or cloud storage. (2) Iterator pattern enables the crawler to systematically process files or database records, ensuring no data is missed. (3) Visitor pattern applies specific operations to each data item, such as validating, cleaning, and transforming it before storage.

   The process initializes by crawler identifying the source locations (e.g., database connections, file directories). The crawler reads data from the source, handling different formats such as CSV, JSON, XML, or SQL tables. The retrieved data undergoes transformation processes, including cleaning, normalization, and enrichment. The processed data is organized into a data catalogue, stored in the data.

   Use Case: The system initializes a crawler for an on-premises SQL database containing historical financial transaction records. The crawler reads data from the database, processing one table at a time. The retrieved data is cleaned to remove duplicates, normalized to ensure consistent formats, and enriched with additional metadata. The processed data is uploaded to a data warehouse, indexed in a data catalogue, and made available for query and analysis.

   The data crawler layer plays a vital role in the GenAI-based private pension information and recommendation system by ensuring comprehensive and flexible data retrieval. Whether through web scraping for real-time data or transferring cold data into a structured data catalogue, this layer ensures that the system has access to high-quality, relevant data necessary for providing accurate and personalized retirement planning advice. By leveraging design patterns like Factory, Iterator, and Visitor, the data crawler layer is both robust and adaptable, capable of handling diverse data sources and structures effectively.

- **Data Storage:** The crawled data is stored in a robust system like Amazon S3 bucket, ensuring security, scalability, and accessibility. This setup supports various data types and is critical for large-scale data operations.

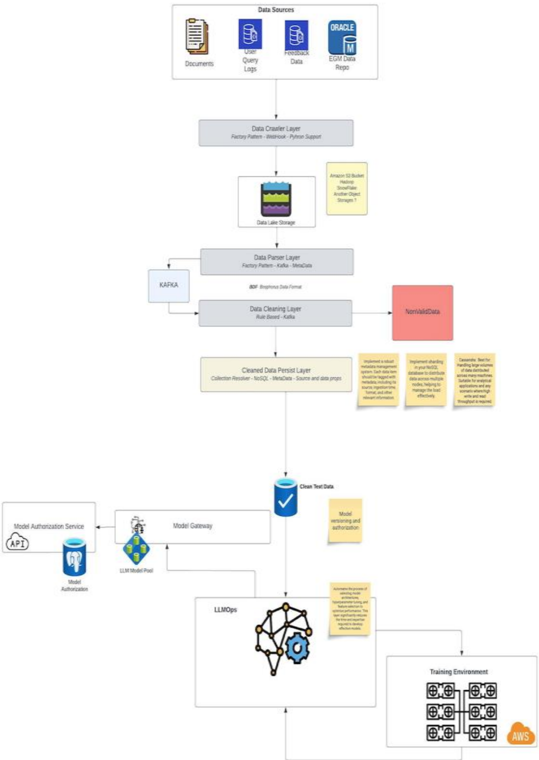

Figure 5. LLM Training Life-Cycle

- **Data Parser Layer:** Following storage, data is parsed using Factory and Lambda/Mediator patterns into structured formats like AVRO or Parquet. This transformation is crucial for standardizing data before it undergoes cleaning and further processing.

- **Data Cleaning Layer:** This layer focuses on refining the data using defined rules and real-time processing via Kafka to ensure quality and relevance. Non-valid data is identified, logged, and handled appropriately to maintain the integrity of the training set.

- **Cleaned Data Persist Layer:** Post-cleaning, the data is persisted in systems that support efficient retrieval and management, employing NoSQL databases with ACID compliance to ensure reliability and performance during the data retrieval phase.

- **Model Gateway:** Acts as a checkpoint that integrates with model authorization services to ensure all data used for training complies with regulatory and ethical standards, safeguarding against the misuse of sensitive information.

- **LLM Model Pool:** This is the core repository where various pre-trained and to-be-trained models are stored. The models available here are ready to be fine-tuned or trained with the newly cleaned and structured data.

- **LLM Ops:** This operational unit manages the continuous integration and deployment of language models, ensuring they are up-to-date and perform optimally. It incorporates advanced automation tools for monitoring, scaling, and managing the models within a production environment.

- **Training Environment:** The actual training of models occurs in a controlled environment, often cloud-based platforms like AWS, which provide the necessary computational power and scalability to handle extensive model training sessions efficiently.

4.3.4. Selection of LLMs for the Proposed System

The choice of LLMs for the GenAI-based private pension information and recommendation system is critical for its effectiveness and reliability. The selection should be based on several factors including the model's ability to understand and generate human-like text, its performance on financial data, and its adaptability to the specific needs of the system. Most used LLMs and their capabilities are:

GPT-4 (Generative Pre-trained Transformer 4) by OpenAI

GPT-4 has demonstrated superior capabilities in understanding and generating human-like text across various contexts. It can maintain context over long conversations, making it suitable for interactive financial advising. Fine-tuning capabilities allow the model to be adapted specifically for financial data and personalized advice. GPT-4 has been extensively tested and validated across numerous applications, ensuring reliability. Its ability to handle a wide range of queries and provide detailed explanations makes it ideal for a complex domain like financial planning. GPT-4 can handle large-scale deployments, making it suitable for use in a system intended for a broad user base.

BERT (Bidirectional Encoder Representations from Transformers) by Google

BERT processes text in both directions, allowing for a deeper understanding of context. It is trained on extensive text corpora, making it robust in various language tasks. It can be fine-tuned for specific tasks, such as financial data analysis and personalized recommendations. BERT's bidirectional nature enhances its ability to understand nuanced financial queries. Fine-tuning BERT on financial datasets improve its performance in providing accurate and relevant advice. For institutions already using Google's cloud services, BERT offers integration and optimization.

LLaMA (Large Language Model Meta AI) by Meta

LLaMA is designed to be efficient and effective in various NLP tasks, making it suitable for experimental applications. It delivers high accuracy and performance in generating human-like text. Can be fine-tuned for specific applications, including financial advising. LLaMA's design ensures it can provide accurate responses quickly, enhancing user experience. Its optimization for research allows continuous improvement and adaptation based on feedback. Suitable for integration into systems requiring high performance and adaptability.

For our system, GPT-4 by OpenAI is the most suitable LLM. Several factors contribute to this selection. Firstly, GPT-4's advanced natural language understanding capabilities enable it to comprehend and generate human-like text with high accuracy, essential for interpreting complex financial queries and providing detailed, personalized advice. Secondly, its ability to maintain contextual awareness over long conversations ensures that users receive consistent and coherent advice throughout their interaction with the system. Thirdly, GPT-4's customization potential allows it to be fine-tuned specifically for financial data and personalized recommendations, enhancing its relevance and accuracy for users. Additionally, GPT-4's proven performance across numerous applications demonstrates its reliability and effectiveness in delivering high-quality responses. Finally, its scalability makes it suitable for large-scale deployments, ensuring that the system can accommodate a broad user base without compromising performance. These attributes make GPT-4 the optimal choice for enhancing the user experience and effectiveness of the GenAI-based private pension information and recommendation system.

4.3.5. Language Model Operational Framework

Figure 6 represents a streamlined structure within a language model operational framework, emphasizing the integration of model management and prompt categorization to enhance response generation.

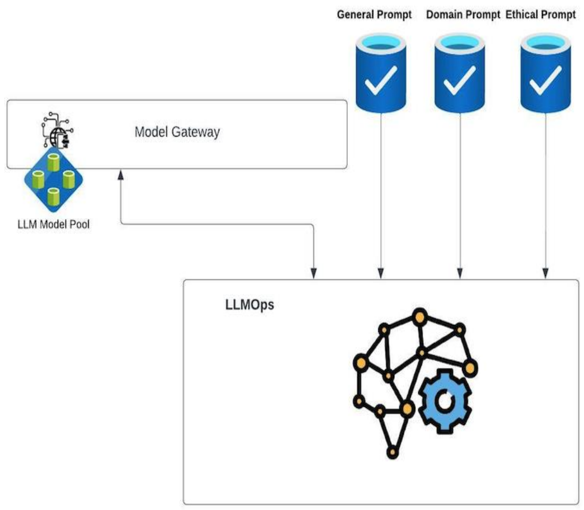

Figure 6. LLM Operational Framework

- **Model Gateway:** This is the primary interface between the stored language models and the operational environment. It functions as a regulatory node that ensures all requests for model utilization pass through a uniform gateway, providing a centralized point of control. This setup allows for efficient management of model resources, security protocols, and compliance checks. The gateway's innovation lies in its ability to dynamically route requests based on model availability, query complexity, and priority, ensuring optimal allocation of computational resources.

- **LLM Model Pool:** A robust repository of pre-trained language models that cater to a wide range of tasks and domains. This pool is fundamental in providing diversity in response capabilities and specialization in content generation. The innovative aspect of the model pool is its extensive coverage of different linguistic styles and domain-specific knowledge, enabling highly accurate and contextually appropriate outputs.

- **Prompt Categorization**

  - **General Prompt:** These prompts cater to a broad range of queries without specific domain constraints. They are designed to handle general knowledge questions and everyday informational needs, making them versatile and widely applicable.
  - **Domain Prompt:** Specifically tailored to address queries within a particular field or industry, these prompts enhance the accuracy and relevancy of responses by utilizing domain-specific language models. This allows for specialized knowledge processing, which is crucial in fields like medicine, law, or finance.
  - **Ethical Prompt:** Designed to ensure responses adhere to ethical guidelines and social norms. These prompts are crucial for maintaining the integrity and trustworthiness of the system, especially when dealing with sensitive topics. They incorporate filters and checks that prevent the generation of harmful or biased content.

- **LMM Operations (LLMOps):**

The LLMOps component is crucial for the ongoing maintenance, monitoring, and optimization of language models within the GenAI-based private pension information and recommendation system. This component encompasses a comprehensive suite of tools and processes designed to ensure that the models perform at their peak efficiency, adapting to changing usage patterns and requirements. Key processes in LLMOps include: (1) Automated Model Tuning: Continuously improve the language model's performance through regular fine-tuning based on new data and user feedback. The process begins with gathering new data from user interactions, financial markets, and other relevant sources. Then we periodically retrain the model with this updated data to enhance its accuracy and relevance. We utilize machine learning frameworks (TensorFlow, PyTorch, or Hugging Face's Transformers library) for efficient model tuning. (2) Real-Time Performance Analytics: Monitor the model's performance metrics in real-time, ensuring it meets the desired accuracy and efficiency standards. We use BLEU score to evaluate the quality of generated text against reference texts. We monitor the response time of the model (latency) to ensure it meets performance expectations. Then we implement our custom dashboards to visualize and track these metrics. (3) Adaptive Deployment Strategies: Deploy the language model in a manner that optimally balances performance and resource usage. The process initializes with distributing requests across multiple instances of the model to prevent overloading any single instance (Load Balancing). Then the system automatically scales the number of active model instances based on current demand. We use serverless computing solutions (Kubernetes and AWS Lambda) for scalable and adaptive deployment. (4) Monitoring Post-Training Performance: To ensure that the language model maintains its performance after retraining or fine-tuning. The process initializes with comparing the model’s performance metrics (accuracy, BLEU score) against previous versions to detect any regressions (Baseline Comparison). Then we Implement continuous integration/continuous deployment (CI/CD) pipelines to automatically test and deploy updates. We use tools MLflow for tracking and versioning model performance. (5) Debugging and Validation: To identify and resolve issues that may arise from model updates or changes in data patterns. The process initializes with analyzing incorrect predictions to understand the root causes and make necessary adjustments. Then we conduct thorough validation tests using a separate validation dataset to ensure the model’s robustness. We employ debugging frameworks (TensorBoard, integrated with Jupyter notebooks) for in-depth analysis. As an example, suppose the GenAI model is retrained after a year to incorporate the latest financial data and user feedback. Post-training, the model’s performance is tracked using metrics such as accuracy and BLEU scores. These metrics are compared to the baseline established before retraining. If any performance drop is detected, the system triggers an error analysis process to identify the issues. Continuous monitoring tools ensure that any significant deviations are promptly addressed, and adaptive deployment strategies help in balancing the load and scaling the resources as needed.

By integrating advanced prompt management with robust model control and operations, the system ensures that outputs not only meet the diverse needs of users but also adhere to high standards of accuracy, relevance, and ethical consideration. This framework is pivotal in deploying GenAI-driven applications that require nuanced understanding and generation of human-like text.

5. DECISION SUPPORT SYSTEM IMPLEMENTATION and USE CASE

5.1. Implementation Challenges and Strategies

5.1.1. Integration with Existing Financial Platforms

Integrating the GenAI-based private pension information and recommendation system with existing financial platforms presents several technical and operational challenges. One major challenge is ensuring compatibility with diverse systems that use different technologies and data formats. Financial platforms often have established infrastructures, proprietary software, and specific data standards that can complicate seamless integration. To deal with these challenges we will start with a phased (incremental) integration approach, beginning with less critical components and gradually expanding to more complex systems. This allows for the identification and resolution of integration issues without disrupting the entire system. We will then develop robust APIs that allow seamless communication between the GenAI system and existing financial platforms. These APIs are designed to handle different data formats and protocols, ensuring compatibility and ease of integration. We will also implement data standardization practices to ensure that the data exchanged between the GenAI system and financial platforms is consistent and accurate. This involves using industry-standard data formats and protocols. Moreover we ensure that robust security measures are in place to protect sensitive financial data during integration. This includes encryption, secure data transmission protocols, and regular security audits.

5.1.2. Ethical Implications of Using GenAI in Financial Advice

The first ethical risk is that: the system collects and processes sensitive financial data, raising concerns about data privacy and security. To mitigate this risk we use advanced encryption techniques to protect data both at rest and in transit. Also we collect only the data necessary for providing accurate recommendations, reducing the risk of data breaches. Moreover we clearly communicate data usage policies to users, ensuring they are informed about how their data is handled.

The second risk is; AI algorithms can inadvertently perpetuate biases present in training data, leading to unfair or discriminatory recommendations. To mitigate this risk we implement regular audits to identify and correct biases in the algorithms. We use diverse and representative datasets to train the AI models, reducing the risk of biased outcomes. Moreover we develop and monitor fairness metrics to ensure that recommendations do not disproportionately favor or disadvantage any user group.

Transparency and explainability is the third risk. Users need to understand how and why specific recommendations are made to trust the system. To mitigate this risk we develop models that can provide clear, understandable explanations for their recommendations. We implement feedback mechanisms that allow users to ask questions and receive explanations about the advice given.

Accountability is another risk that we should consider. Determining accountability for financial advice provided by an AI system can be complex. To mitigate this risk we establish clear policies outlining the roles and responsibilities of the AI system, financial advisors, and the institution. We ensure that human advisors are available to review and validate the recommendations provided by the AI system.

5.1.3. Ensuring User Adoption

User adoption is another significant challenge, as users may be hesitant to trust and rely on an AI-based system for their retirement planning. Factors such as lack of familiarity with AI, concerns about data privacy, and resistance to change can hinder adoption. To deal with these challenges we start with designing a user-friendly interface that is intuitive and easy to navigate. The system provides clear instructions, feedback, and support to help users feel comfortable and confident in using it. Then we will conduct educational campaigns to inform users about the benefits and functionalities of the GenAI system. This includes webinars, tutorials, and detailed user manuals that explain how the system works and how it can enhance their retirement planning. We will also implement and communicate strong data privacy measures to assure users that their personal and financial data is secure. This includes transparent data handling policies and compliance with relevant data protection regulations.

5.1.4. Real-World Testing

We will perform a pilot program involving a small group of users to test the system in a real-world setting. This pilot will target a diverse user base to gather comprehensive feedback. We will choose participants based on demographics, financial literacy levels, and retirement planning needs to ensure a representative sample. We will establish a robust feedback mechanism (survey and where necessary, interview) to collect user experiences, identify issues, and gather suggestions for improvement. We will define clear performance metrics to evaluate the system's effectiveness. These metrics include user satisfaction, accuracy of recommendations, ease of use, and impact on financial literacy. We will continuously collect data on these metrics during the pilot phase to assess performance and make necessary adjustments. We will use the feedback and performance data to iteratively improve the system. This involves refining algorithms, enhancing user interface design, and addressing any technical issues identified during the pilot phase.

5.1.5. Scaling

Collaborating with a range of partners facilitates broader adoption and integration thus we will establish partnerships with multiple financial institutions to expand the system's reach. Moreover we will invest in scalable infrastructure to support increased user loads and data processing demands. This includes cloud-based solutions that offer flexibility and scalability. Also we will ensure compliance with financial regulations and standards. This involves continuous monitoring of regulatory changes and adapting the system accordingly. Then we will launch a comprehensive marketing campaign to promote the system to a wider audience. This includes digital marketing, participation in financial industry conferences, and leveraging social media platforms. Finally we will provide ongoing support and updates to users to maintain engagement and satisfaction. This includes customer service, regular system updates, and continuous improvements based on user feedback.

5.1.6. Use Case

Objective of our use case is to enhance retirement planning services using GenAI technology to provide personalized advice and improve user engagement. There are three actors: (1) User: The end-user seeking personalized retirement advice. (2) System: GenAI-based private pension information and recommendation system. (3) Financial Advisor: A SecurePension financial advisor providing support and additional consultation. Our primary goals are: (1) To provide personalized retirement planning advice. (2) To improve financial literacy and user engagement. (3) To ensure adaptive recommendations. Steps and LLM utilization are explained below:

Step 1: User Profile Creation
Activity: User logs into the platform and creates a profile by providing personal and financial information (e.g., age, income, current savings, retirement goals).
LLM Utilization: The LLM analyzes user input to ensure completeness and clarity, providing prompts for missing or unclear information. Example output: "`It looks like you haven’t specified your current savings. Can you please provide this information to tailor your retirement plan better?`"

Step 2: Initial Analysis and Recommendation
Activity: The system uses the provided data to generate an initial retirement plan recommendation.
LLM Utilization: The LLM processes the user data, applying financial models and algorithms to generate personalized advice. The LLM provides a summary and detailed explanation of the recommended retirement plan. Example output: "`Based on your current savings and retirement goals, we recommend increasing your monthly savings by $200 to meet your target. Here’s a detailed breakdown of how this will impact your retirement fund.`"

Step 3: Scenario Simulation
Activity: The user can simulate various financial scenarios (e.g., changes in income, market fluctuations) to see how they impact the retirement plan.
LLM Utilization: The LLM generates and explains multiple financial scenarios, highlighting potential risks and benefits. Example output: "`If you experience a 10% increase in income next year, your retirement fund could grow by an additional $50,000 by the time you retire. Conversely, a 5% market downturn could reduce your fund by $20,000.`"

Step 4: Interactive Q&A
Activity: The user asks specific questions about the retirement plan and financial concepts they don't understand.
LLM Utilization: The LLM provides real-time, context-aware responses to user queries, enhancing financial literacy. Example prompt: "`What are the benefits of diversifying my investment portfolio?`" The LLM responds, example output: "`Diversifying your portfolio reduces risk by spreading investments across various asset classes, which can help protect your retirement fund from market volatility.`"

Step 5: Feedback and Continuous Improvement
Activity: The user provides feedback on the system's recommendations and overall experience.
LLM Utilization: The LLM collects and analyzes user feedback to improve its algorithms and user interactions continuously. Example: "`How satisfied are you with the clarity of the advice provided? Your feedback will help us enhance our recommendations.`"

This use case illustrates the practical application of a GenAI-based private pension information and recommendation system within a financial institution. By leveraging LLMs, the system can provide personalized, real-time, and adaptive retirement planning advice while enhancing user engagement and financial literacy. The LLM's capabilities in processing and analyzing data, generating detailed explanations, simulating scenarios, and interacting with users in a context-aware manner are crucial for delivering a comprehensive and user-friendly experience. This approach ensures that users receive the most relevant and effective financial advice tailored to their unique needs and circumstances.

6. CONCLUSION

The development and implementation of the GenAI -based private pension information and recommendation system outlined in this paper represents a significant step forward in individual retirement planning. The system's use of GenAI algorithms to process and analyze user data has demonstrated substantial potential in enhancing users' financial literacy and ability to make well-informed retirement planning decisions. Through personalized advice, the system facilitates a deeper understanding of various financial products and investment strategies tailored to individual risk preferences and retirement goals. One of the notable outcomes of this system is its ability to democratize financial advice, historically the domain of professional financial advisors. By making personalized financial guidance accessible and affordable, the system breaks down barriers to financial knowledge and planning tools, potentially leveling the playing field for a broader range of individuals. Furthermore, the system's dynamic adaptability to both macroeconomic conditions and personal life changes ensures that the advice remains relevant and timely, providing users with confidence in managing their retirement planning proactively. The continuous learning capability of the GenAI algorithms ensures that the system evolves with the latest financial trends and data, enhancing its reliability and accuracy over time.

Potential limitations of this study are: (1) Different users have diverse financial goals, risk tolerances, and investment preferences. A one-size-fits-all approach may not adequately address individual needs. To mitigate this limitation, we will utilize the system's personalization capabilities to tailor advice to individual profiles and preferences. (2) Economic conditions and cultural attitudes towards savings and investments vary across regions and demographics. Recommendations that are effective in one context may not be suitable in another. To mitigate this limitation, we will incorporate localized economic data and culturally relevant advice to ensure applicability across diverse contexts.

Future developments should focus on: (1) Exploring advanced personalization techniques using deeper machine learning models to cater to highly specific user needs. Since there is a potential for improved user satisfaction and engagement through more accurate and relevant recommendations. (2) Investigating methods for seamless integration with a wider range of financial platforms and services. Since providing a unified financial management experience will increase system adoption and user convenience. (3) Developing algorithms capable of processing and acting on real-time financial data more efficiently. This will enhance responsiveness and relevance of financial advice in dynamic market conditions.

TEŞEKKÜR (ACKNOWLEDGMENTS)

This study was supported with the project number 3235018 under the program of "TÜBİTAK 1707".

REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[2] Shaw, P., Uszkoreit, J., Vaswani, A.(2018). Self-attention with relative position representations. arXiv preprint arXiv:1803.02155.

[3] Ghojogh B., Ghodsi, A. (2020). Attention mechanism, transformers, bert, and gpt: tutorial and survey

[4] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv preprint arXiv:1810.04805.

[5] Liu, Q., Kusner, M. J., Blunsom, P. (2020). A survey on contextual Embeddings. arXiv preprint arXiv:2003.07278.

[6] Roumeliotis, K. I., Tselikas, N. D., Nasiopoulos, D. K. (2023). Llama 2: Early adopters' utilization of meta's new open-source pretrained model.

[7] Baladn, A., Sastre, I., Chiruzzo, L., Ros, A. (2023). Retuyt-inco at bea 2023 shared task: Tuning open-source llms for generating teacher responses. In: Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications

(BEA 2023), 756–765.

[8] Nay, J. J. (2023). Large language models as fiduciaries: A case study toward robustly communicating with artificial intelligence through legal standards.

[9] Sennrich, R., Haddow, B., Birch, A. (2015). Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909.

[10] Zhuo, T. Y., Li, Z., Huang, Y., Li, Y.-F., Wang, W., Haffari, G., Shiri, F. (2023). On robustness of prompt-based semantic parsing with large pretrained language model: An empirical study on codex. arXiv preprint arXiv:2301.12868.

[11] Chiang, W.-L., Li, Z., Lin, Z., Sheng, Y., Wu, Z., Zhang, H., Zheng, L., Zhuang, S., Zhuang, Y., Gonzalez, J.E., Stoica, I., Xing, E.P. (2023). Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. March.

[12] Yao, Z., Yazdani Aminabadi, R., Zhang, M., Wu, X., Li, C., He, Y. (2022). Zeroquant: Efficient and affordable post-training quantization for large-scale transformers. Advances in Neural Information Processing Systems, 35, 27168–27183.

[13] Zou, A., Wang, Z., Kolter, J. Z., Fredrikson, M. (2023). Universal and transferable adversarial attacks on aligned language models," arXiv preprint arXiv:2307.15043.

[14] Katz, D. M., Bommarito, M. J., Gao, S., Arredondo, P. (2023). GPT-4 Passes the Bar Exam. March.

[15] Jung, D., Dorner, V., Glaser, F. and Morana, S. (2018a). Robo-advisory: digitalization and automation of financial advisory. *Business and Information Systems Engineering*, 60(1),81-86. https://doi.org/10.1007/s12599-018-0521-9.

[16] Jung, D., Dorner, V., Weinhardt, C. and Pusmaz, H. (2018b).Designing a robo-advisor for risk-averse, low-budget consumers. Electronic Markets, 28(3), 367-380. https://doi.org/10.1007/s12525-017-0279-9.

[17] D'Acunto, F., Prabhala, N. and Rossi, A.G. (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies*, 32(5), 1983-2020. https://doi.org/10.1093/rfs/hhz014.

[18] Isaia, E. and Oggero, N. (2022). The potential use of RAs among the young generation: evidence from Italy. Finance Research Letters, 48, 103046. https://doi.org/10.1016/j.frl.2022.103046.

[19] Au, C.D., Klingenberger, L., Svoboda, M. and Frere, E. (2021). Business model of sustainable robo-advisors: empirical insights for practical implementation. Sustainability,13(23),13009. https://doi.org/10.3390/su132313009.

[20] Oh, S., Park, M.J., Kim, T.Y. and Shin, J. (2022). Marketing strategies for fintech companies: text data analysis of social media posts. Management Decision, 61(1), 243-268. https://doi.org/10.1108/md-09-2021-1183.

[21] Rodrigues, L.F., Oliveira, A. and Rodrigues, H. (2023). Technology management has a significant impact on digital transformation in the banking sector. International Review of Economics and Finance, 88, 1375-1388. https://doi.org/10.1016/j.iref.2023.07.040.

[22] McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D. (2012). Big data: the management revolution. Harvard Business Review, 90(10), 60-68.

[23] Huang, E.Y. and Lin, C.Y. (2005). Customer-oriented financial service personalization", Industrial Management and Data Systems, 105(1), 26-44. https://doi.org/10.1108/02635570510575171.

[24] Gao, Y. and Liu, H. (2022). Artificial intelligence-enabled personalization in interactive marketing: a customer journey perspective. *Journal of Research in Interactive Marketing*, *17(5)*, 1-18, https://doi.org/10.1108/jrim-01-2022-0023.

[25] Güneş, H. (2022). akademik ve idari personelin bireysel emeklilik okuryazarliğinin tespiti: makü örneği. *Muhasebe Ve Finans İncelemeleri Dergisi*, 5(1), 39-51. https://doi.org/10.32951/mufider.1000689.

[26] Onat, O. K., & Yöntem, H. (2022). Finansal okuryazarliğin bireysel emeklilik sistemi tercihlerine etkisi: burdur mehmet akif ersoy üniversitesinde bir araştirma. *Finansal Araştırmalar Ve Çalışmalar Dergisi*, 14(26), 165-192. https://doi.org/10.14784/marufacd.1055196.

[27] Akgün, M. K., & Bozkurt, Ö. (2023). Bireysel emeklilik sistemindeki katılımcıların memnuniyet ve güven algılarının incelenmesi. *Çalışma İlişkileri Dergisi*, 14(1), 1-19.