Politécnico
de Viseu

Escola Superior
de Tecnologia
e Gestão de Viseu

Enhancing Financial Product Recommendations through Chatbots and Large Language Models: A User-Centric Approach

Pedro Manuel Noutel Pereira

Trabalho de Projeto
Mestrado em Engenharia Informática - Sistemas de Informação

Trabalho efetuado sob a orientação de
Professor Doutor Carlos Augusto da Silva Cunha

Fevereiro de 2024

Politécnico
de Viseu

Escola Superior
de Tecnologia
e Gestão de Viseu

Enhancing Financial Product Recommendations through Chatbots and Large Language Models: A User-Centric Approach

Pedro Manuel Noutel Pereira

Trabalho de Projeto
Mestrado em Engenharia Informática - Sistemas de Informação

Trabalho efetuado sob a orientação de

Professor Doutor Carlos Augusto da Silva Cunha

Fevereiro de 2024

. . .

Acknowledgements

This dissertation is the culmination of a challenging and enriching journey, which would not have been possible without the support, encouragement and guidance of various people and institutions.

First of all, I would like to express my deepest gratitude to my advisor, Prof. Dr. Carlos Cunha, from the School of Technology and Management of the Polytechnic Institute of Viseu, for his support, patience and valuable contributions throughout this work. His guidance and knowledge were fundamental to the completion of this study. Special thanks to my colleagues and friends who, directly or indirectly, contributed with suggestions, support and encouragement during this academic journey.

I would like to express my gratitude to my family for their support, which has been essential in overcoming the challenges of this journey.

A very special thank you to my partner, Andreia Almeida Lopes, for her unconditional support, patience and encouragement throughout this journey, because without her it wouldn't have been difficult to continue. In the most challenging moments, you were my safe haven, giving me the strength to carry on and believe in my work, always helping me to find the solution when it didn't seem to exist. Thank you for being part of my life and may we continue to hold on when things seem lost.

Abstract

In an era where the intersection of artificial intelligence and digital finance is becoming increasingly relevant, the application of Large Language Models (LLM) to investor risk profiling is a revolutionary possibility. Traditional risk profiling methods rely too heavily on structured questionnaires, which fail to capture the nuances of investor behaviour, resulting in stereotyped categorization and more often than not, inaccurate representations of risk tolerance. This thesis explores how LLMs, specifically using OpenAI GPT series, can enhance the accuracy and personalisation of risk assessment models by leveraging natural language understanding and adaptive learning capabilities.

This study explains the limitations of conventional profiling techniques and proposes an LLM-based model that is dynamic in nature and adjusts according to investors' responses, making the classification more precise with interactive discussion. To identify the effectiveness of the model proposed, the Morningstar risk profile is taken as a training and testing dataset for the LLM, while the traditional approach is analyzed, based on Google Forms, for comparing dynamic and static models.

The results indicate that LLMs can significantly contribute to investor profiling by providing a conversational, real-time assessment process that complements the accuracy and user engagement, rendering the monitoring process more personalized. They also show that LLM achieves a significant improvement in investor profiling through a more flexible and interactive classification procedure. The model outperforms conventional methods in cases where investor responses are ambiguous or subjective, ensuring better alignment with risk tolerance. In addition, it was observed that the GPT-based classification tends to be conservative in the case of uncertain responses, which can serve to avoid excessive risk-taking. These results validate the potential of leveraging LLMs to augment financial advisory services by providing a more accurate and responsive risk assessment to individual investor profiles.

By demonstrating the advantages and limits of AI-based investor profiling, the thesis contributes to the growing body of literature on LLM application for making financial decisions. The findings pinpoint the transition from rigid, questionnaire-based questionnaires to adaptive, context-aware systems, with new potentials for augmenting financial advisory services in an increasingly digitalized economy.

Keywords: Large Language Models, Risk Profiling, Morningstar, Financial Advisory, Artificial Intelligence, Fintech, Investor Behaviour

Resumo

Numa era onde a interseção entre inteligência artificial e finanças digitais se torna cada vez mais relevante, a aplicação de Large Language Models (LLM) à perfilagem de risco de investidores representa uma possibilidade revolucionária. Os métodos tradicionais de avaliação de risco dependem excessivamente de questionários estruturados, que não conseguem captar as nuances do comportamento dos investidores, resultando em categorização estereotipada e, frequentemente, representações imprecisas da tolerância ao risco. Esta tese explora como os LLMs, especificamente a série GPT da OpenAI, podem melhorar a precisão e a personalização dos modelos de avaliação de risco, aproveitando as capacidades de compreensão de linguagem natural e aprendizagem adaptativa.

Este estudo explica as limitações das técnicas convencionais de perfilagem e propõe um modelo baseado em LLMs que é dinâmico por natureza e ajusta-se às respostas dos investidores, tornando a classificação mais precisa através de uma discussão interativa. Para identificar a eficácia do modelo proposto, o perfil de risco da Morningstar é utilizado como *dataset* de treino e teste para os LLMs, enquanto a abordagem tradicional é analisada com base em Google Forms, permitindo a comparação entre modelos dinâmicos e estáticos.

Os resultados indicam que os LLMs podem contribuir significativamente para a perfilagem de investidores ao fornecer um processo de avaliação conversacional e em tempo real, que complementa a precisão e o envolvimento do utilizador, tornando o acompanhamento mais personalizado. Indicam também que os LLMs podem contribuir significativamente para a perfilagem de investidores, proporcionando um processo de avaliação conversacional e em tempo real que complementa a precisão e o envolvimento do utilizador, tornando o acompanhamento mais personalizado. Mostram também que os LLMs permitem uma melhoria significativa na perfilagem de investidores através de um procedimento de classificação mais flexível e interativo. O modelo supera os métodos convencionais em situações onde as respostas dos investidores são ambíguas ou subjetivas, garantindo um melhor alinhamento com a tolerância ao risco. Além disso, observou-se que a classificação baseada em GPT tende a ser conservadora em casos de respostas incertas, o que pode servir para evitar a assunção excessiva de riscos. Estes resultados validam o potencial da utilização de LLMs para melhorar os serviços de consultoria financeira, proporcionando uma avaliação de risco mais precisa e adaptável aos perfis individuais dos investidores.

Ao demonstrar as vantagens e limitações da perfilagem de investidores baseada em IA, esta tese contribui para o crescente corpo de literatura sobre a aplicação de LLMs na tomada de decisões financeiras. Os resultados destacam a transição de questionários rígidos e padronizados para sistemas adaptativos e sensíveis ao contexto, abrindo novas possibilidades para o aprimoramento dos serviços de consultoria financeira numa economia cada vez mais digitalizada.

Palavras-chave: Large Language Models, Perfilagem de Risco, Morningstar, Consultoria Financeira, Inteligência Artificial, Fintech, Comportamento do Investidor

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Risk assessment is an important aspect of financial decision-making, determining investment choices, credit rating, and overall financial health [Feng et al., 2023].

Traditional approaches to creating investor risk profiles rely on formal questionnaires with predetermined response formats, such as in Morningstar analysis [Rabbani and Nobre, 2022]. While effective in assigning standardized risk categories, these approaches fail when they need to be adaptable and representative of the investor's true risk tolerance. Investors may struggle to discern the most appropriate responses, which will result in an excessive simplification that do not necessarily reflect their financial decisions or choice processes. Most investors also experience issues with knowing what they are being asked because (1) they do not understand the words employed, (2) they do not understand the subject matter itself, or (3) their answer is not represented by the alternative answers [Mazzoli and Palmucci, 2023]. This can result in incorrect responses and, consequently, less precise risk assessments, making it harder to analyze their profiles.

Advances in recent artificial intelligence, particularly Large Language Models such as GPT-4 and LLaMA 2 [Hens and Nordlie, 2024], provide a groundbreaking opportunity to enhance risk assessment techniques. Unlike static questionnaires, LLMs can comprehend inputs in natural language, adapt dynamically to user feedback, and improve assessments through relevant follow-up questions.

This study analyzes the development of an LLM-based risk assessment model that goes beyond traditional methods, offering a more personalized, interactive and detailed investor profiling process. As part of offering an enhanced degree of accuracy in classification, the technique pursues an user-friendly and interactive assessment solution that will offer better aid for investors at any level of financial literacy during the process.

Using instruction-based learning [Nie et al., 2024] and multitask assessment [Feng et al., 2023], the model integrates the framework and logic of traditional questionnaires, adding an adaptative mechanism for better investor classification. This approach enhances user experience, breaks the limitations of strict classification, and ensures risk assessments that align with individual financial inclinations. Furthermore, by identifying inconsistencies or ambiguities in responses, the model aims to provide a more comprehensive view of an investor's risk profile, with an enhanced degree of classification accuracy.

This thesis contributes to the nascent field of AI-driven financial analysis by demonstrating the practical application of LLMs in calculating investor risk profiles. Through the integration of financial domain expertise and advanced natural language processing capabilities, this study seeks to fill the gap between conventional structured analyses and adaptive user-friendly AI options. The findings indicate the potential of LLMs not only to improve risk profiling but also to redefine how financial institutions interact with investors in the digital economy.

1.1 Motivation

The revolutionary potential of Large Language Models in financial risk analysis is the main motivator for this thesis. Traditional investor profiling methods follow prescriptive and formulaic approaches that have a tendency to disregard distinct levels of financial knowledge and different investment practices. Institutional models are largely based on static questionnaires with pre-assigned responses, making the process impersonal, advanced, and, in certain instances, inaccessible, frequently using very technical or incomprehensible language [Hens and Nordlie, 2024]. Therefore, investors, and especially less financially sophisticated ones, can't connect with risk estimates that do not reflect their real decision-making behavior.

LLMs offer the potential for evading such limitations through the development of a dynamic, user-oriented risk assessment model. LLMs differ from static questionnaires in permitting interactive and conversational assessment of users, with greater user understanding and engagement [Wang et al., 2023]. LLMs have the capability to simplify advanced financial terminology, clarify ambiguous questions, and adapt the assessment using contextually suitable follow-up questions. This interactivity enables more accurate and inclusive depiction of an investor's actual risk tolerance.

Apart from refining personal risk profiling, AI-powered financial advisory systems have the ability to democratize access to personalized advice. With real-time AI-driven recommendations, LLMs can make financial literacy possible and enable more individualized risk assessment accessible to more people, enabling informed decisions. The models also have the capability to empower financial institutions to better align their services with investor interest and user intent through dynamic and adaptative user interaction learning.

1.2 Definition of the Problem

Investor profiling remains a rigid process, often unable to capture the evolving nature of financial decision-making [Hens and Nordlie, 2024]. Traditional approaches rely on static questionnaires and predefined risk categories, lacking the flexibility needed to reflect changes in investors' financial situations, goals, or market perceptions. This rigidity limits the accuracy and relevance of investment recommendations, reducing trust in financial advisory systems.

A primary limitation is the reliance on static, pre-defined classifications that are not capable of adaptive and personalized assessments. Current methods are not able to interact with investors dynamically and are therefore unable to refine their responses to be more accurate based on contextualized feedback. This leads to financial institutions being unable to offer truly personalized, concise and dynamic investment recommendations that reflect individual risk appetites.

This thesis explores the transition to adaptive LLM-based models from static investor profiling with the goal of enhancing personalization, flexibility, and trust in financial advisory services. By interpreting investor responses in a conversational way and adapting dynamically to contextual nuances, LLMs offer the potential to analyze more accurate and personalized investor profiles, reducing the misclassification problems associated with rigid questionnaire-based methods. Further, their ability to integrate context and facilitate explainability improves transparency and user engagement.

The goal is to demonstrate how LLMs can overcome those limitations by rendering financial decision-making more lively and data driven. The contribution should be in the form of a more dynamic and investor-oriented financial advisory system, in harmony with the sophistication of contemporary financial conduct.

1.3 Objectives

The primary objectives of this thesis are as follows:

1. **Develop an Adaptive Investor Profiling Framework with LLMs:**

   Design and implement a risk assessment framework based on the use of LLMs that will dynamically compute and classify investor profiles. This involves:

- Departing from static questionnaire assessments through enabling context-based and dialogue-type interactions.
- Developing methods for real-time adjustment to inputs from investors in a way that reduces misclassifications due to vagueness or misreading.
- Assessing the extent to which LLMs can balance personalization against standardization in investor profiling.

2. **Compare LLM-Based Profiling vs. Traditional Approaches:**

   Compare the accuracy and effectiveness of LLM-driven investor profiling with static questionnaire-based methods by:

   - Investigating the extent to which GPT-based categorizations correlate with conventional financial risk estimates (e.g., Google Form results).
   - Identifying key areas of variation and their implications for investor categorization and financial decision-making.
   - Assessing whether GPT-based assessments improve precision, or if they introduce new bias or incongruities.

3. **Identify Challenges and Constraints of LLM-Based Profiling:**

   Critically examine the risks and constraints of financial profiling through LLMs, such as:

   - The implications of biases in LLMs and how they could be involved in investor classifications.
   - The difficulty in converting qualitative, subjective financial preferences into Artificial Intelligence (AI) models.
   - The importance of human participation in providing reliable and effective financial advice.

This research aims to demonstrate the potential of LLMs in improving investor profiling while highlighting the practical considerations for their implementation. By bridging the gap between static risk assessment models and AI-driven dynamic profiling, this study seeks to contribute to the ongoing development of intelligent financial advisory systems and how LLMs can complement or enhance traditional profiling.

1.4 Expected Results

This thesis aims to demonstrate the effectiveness of LLMs in enhancing investor profiling through adaptable and personalized risk assessment methods. The expected results focus on analysing the potential benefits and limitations of LLM-based methods in comparison to traditional methods.

1. **Improved Accuracy and Flexibility in Investor Profiling:**

    - LLM-based models are expected to provide a more dynamic and personalized classification of investor risk profiles by facilitating interactive, conversational assessments instead of relying on rigid questionnaires.
    - With the use of natural language understanding, the model is expected to reduce the number of misclassifications due to ambiguous or misunderstood questions in traditional methods.

2. **Identification of Strengths and Weaknesses in LLM-Based Profiling:**

    - A cross-comparison of LLM-based profiling against traditional approaches is expected to reveal both areas of intersection and divergence.
    - Results should highlight whether LLM-based profiling can function as an independent tool, a supporting method, or if finetuning is required for it to align with financial advisory best practices.

3. **Future of AI in Financial Advisory Services:**

    - The research is expected to provide some insight into how LLMs are capable of enhancing user engagement in financial decision-making by offering a more intuitive and user-friendly risk assessment process.
    - The study anticipates that AI-based profiling models could improve financial inclusion by providing broader access to quality investor assessments for individuals with diversified financial literacy.
    - Results should reflect directions for integrating LLMs into existing financial advisory architectures to enable more flexible and personalized investor support.

1.5 Work Plan

In order for the results of this research to be achieved in the appropriate timeframe (12 months), the following schedule of tasks will be followed, as presented in Figures 1.1 and 1.2:

- **Phase 1: Literature Review** - Reviewing existing academic literature, research papers, case studies, and industry reports on machine learning applications in finance, with a focus on chatbots and LLMs in financial product recommendation systems.

- **Phase 2: Methodology Development** - Developing the methodology for integrating LLMs into financial product recommendation systems. This includes defining the data requirements, the model architecture, and the algorithms for processing and analysis.

- **Phase 3: Prototype Design and Development** - Designing and developing a prototype of the financial product recommendation system. This phase involves coding the model, integrating and setting up the LLM for processing user inquiries and data.
- **Phase 4: Testing and Refinement** - Conducting extensive testing of the prototype to assess performance, accuracy, and user satisfaction. Based on feedback and testing results, the model will be refined and optimized.
- **Phase 5: Evaluation and Analysis** - Systematically evaluating the model using the defined criteria and success metrics. This phase includes collecting and analyzing user feedback, performance data, and comparing the outcomes with existing models.
- **Phase 6: Documentation and Dissemination** - Documenting the research findings, model development process, testing and evaluation results. Preparing the thesis document and planning for dissemination through academic publications or presentations.

	Task Mode	Task Name	Duration	Start	Finish	Predecessors
		Literature Review	30 days	Wed 31/01/24	Wed 28/02/24	
		Methodology Development	31 days	Thu 29/02/24	Fri 29/03/24	1
		Prototype Design and Development	61 days	Sun 31/03/24	Tue 28/05/24	2
		Testing and Refinement	61 days	Wed 29/05/24	Fri 26/07/24	3
		Evaluation and Analysis	61 days	Sat 27/07/24	Mon 23/09/24	4
		Documentation and Dissemination	25 days	Tue 24/09/24	Fri 18/10/24	5

Figure 1.1: Work plan

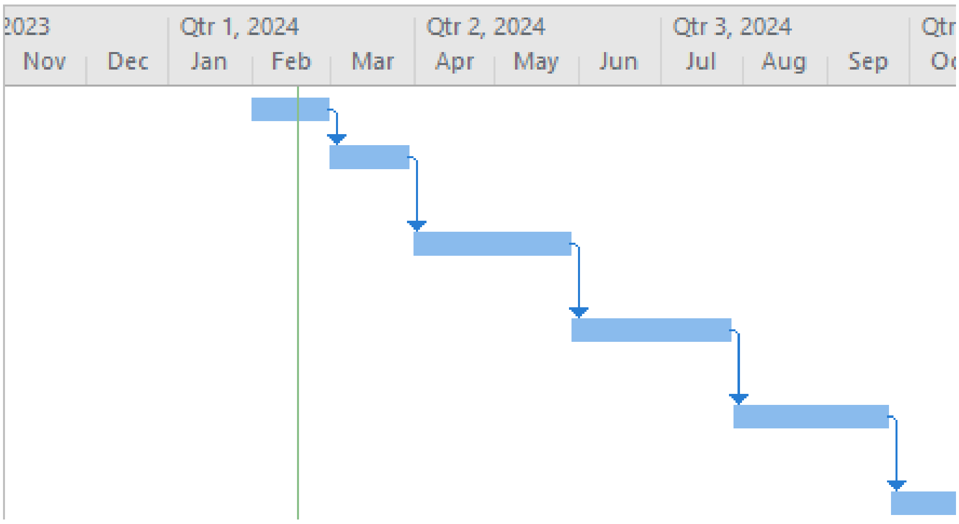

Figure 1.2: Work plan-Gantt Graphic

To structure this research, the following chapters are organized as follows:

- **Chapter 2** contains the Literature Review, describing the theoretical foundations of financial risk profiling and LLM utilization in financial advisory. It also reviews the pertinent literature, revealing gaps in current practice.
- **Chapter 3** revisits the Contextualization and Problem Specification, providing an in-depth analysis of the key challenges in investor profiling and justifying the need for a dynamic, LLM-based assessment model.
- **Chapter 4** outlines the Methodology and Requirements Analysis, detailing the research design, data collection methods, and the development of an AI-driven investor profiling system.
- **Chapter 5** presents the Results and Discussion, evaluating the performance of the proposed model compared to traditional methods, analyzing its accuracy, adaptability, and user engagement.
- **Chapter 6** concludes the study by summarizing the key findings, discussing limitations, and proposing directions for future research in AI-driven financial advisory systems.

Chapter 2

Literature Review

This chapter addresses how emerging technologies like machine learning, chatbots, and large language models are transforming investor profile and financial product recommendation systems to satisfy demand for personalization. The chapter begins by exploring the theoretical foundations that underpin this thesis, addressing key concepts, approaches, techniques, and methods that support the application of LLMs and other AI-driven tools in financial services, providing a comprehensive understanding of the technological advancements and fundamental principles necessary to contextualize the research.

It also examines related work, presenting a critical review of previous research relevant to the topic. This includes an analysis of the main contributions, methodologies, and results in the field, highlighting the shortcomings of current approaches. By identifying gaps in the existing literature, this section delineates the research space that this thesis aims to address, demonstrating how the proposed work builds upon and extends prior studies.

By providing both a theoretical and contextual framework, the chapter lays the ground work for the central exploration of how LLMs can enhance investor profiling and financial advisory systems. This focus on theoretical foundations and related work ensures a clear understanding of the research problem while situating the contribution of this thesis within the broader academic and technological landscape.

2.1 Theoretical Basis

Artificial intelligence, in its evolution and influence, has changed the world of finance over time. It had major breakthroughs in machine learning and Large Language Models. Previously, Machine Learning (ML) arrived as a very revolutionary tool in which systems could automatically learn from their data, enabling them to distinguish patterns and further predict things on their own instead of explicit programming. This basically revolutionized the analysis of complex financial data and provided landmark insights into investor behaviors and decision-making procedures. With the advancement of technology, LLMs furthered these by adding capabilities to process and interpret unstructured textual data such as market commentary, financial news, and investor communications. They added a whole new dimension to financial analysis, enabling deeper contextual understanding and more personalized ways of offering financial services.

The integration of LLMs has made chatbots much more sophisticated tools, allowing them to engage with investors in a much more personalized and conversational way. They can answer queries, provide tailored advice, and even collect nuanced information about investor preferences and behaviors. Investor surveys have also been improved by adding AI-driven features to collect much more detailed and accurate data.

Combining technical innovation with practical application can help redefine how financial institutions are supposed to approach the area of investor profiling, risk assessment, and decision making. This chapter talks about the revolutionary ability of such technologies to reshape financial systems toward increasing personalization and accuracy for investor-focused solutions.

2.1.1 Machine Learning in Finance

Machine Learning has revolutionized the financial sector by enabling systems to analyze large volumes of data more efficiently and autonomously, identifying complex patterns that would be difficult or even impossible to detect manually [Alpaydin, 2020]. In the financial context, ML applications encompass a wide range of tasks, including risk analysis, market forecasting, fraud detection, portfolio management, and asset allocation optimization.

Its ability to process both structured and unstructured data allows financial institutions to leverage information from media sources and transactional histories to improve their forecasts and strategies. The reinforcement learning have been integrated into financial systems to optimize operations and provide personalized solutions, such as financial product recommendations or the categorization of investor risk profiles.

2.1.2 Risk Assessment and Investor Profiling

Risk assessment in the financial sector is a critical process for identifying, analyzing, and controlling potential risks that could impact the achievement of financial objectives, whether at an individual or organizational level. This process goes beyond individual investors, encompassing applications in credit analysis, transportation management, and regulatory compliance, providing actionable insights that aid decision-making, minimize losses, and maximize returns [Zhao et al., 2024].

Traditional risk assessment techniques, such as Value at Risk (VaR) [Nagpal, 2024], estimate the maximum potential loss a portfolio might face within a specific confidence interval, while stress tests evaluate portfolio performance under adverse economic scenarios [Canabarro, 2024].

In traditional risk assessments, questionnaires have played a fundamental role in evaluating investors' risk tolerance and financial behavior. Tools such as the Morningstar Risk Questionnaire assess key factors, including financial goals, time horizon, income, and reactions to hypothetical market scenarios. The Finametrica Risk Tolerance Assessment focuses on psychological attitudes toward risk, while Riskalyze assigns a numerical risk score to balance emotional and financial capacity for risk-taking. However, it is important to note that despite their widespread use, these tools have limitations due to their static nature, which makes them unable to adapt easily to changing market conditions or investors' preferences [Mazzoli and Palmucci, 2023]..

In addition to questionnaires, alternative risk assessment methods have emerged that do not rely on direct responses from investors. For example, portfolio analysis tools, such as Monte Carlo simulations, assess the probability of achieving financial goals under various market conditions.

Advances in machine learning have enabled the analysis of transactional data, macroeconomic trends, and behavioral patterns to predict potential risks with high precision. Stress testing and scenario analysis further enhance risk assessment by simulating adverse economic conditions to test portfolio resilience, which is particularly important for regulatory compliance [Strong et al., 2009].

With advancements in Artificial Intelligence and Machine Learning, modern risk assessment methods have become significantly more dynamic and precise. They utilize supervised, unsupervised, and reinforcement learning techniques to identify complex patterns in financial data, improving the accuracy of risk predictions [Enkhsaikhan and Jo, 2024]. These innovations not only refine risk forecasting but also strengthen risk mitigation strategies, such as portfolio diversification and the use of hedging instruments, which aim to manage and reduce financial losses [Baz et al., 2020].

Investor profiling complements risk assessment by categorizing individuals based on their financial capacity, risk tolerance, investment horizon, and financial objectives. It is a fundamental process in the financial sector, enabling the personalization of investment strategies and aligning recommendations with each client's unique interests. However, traditional approaches to investor profiling have limitations, as they fail to adapt to changing client preferences and market conditions, resulting in static and potentially outdated profiles. Machine learning algorithms now analyze transactional data, investment histories, and even social media activity to generate adaptive profiles that evolve with client behavior and market dynamics. Additionally, Natural Language Processing (NLP) techniques further enrich this process by extracting qualitative insights from client-advisor interactions, capturing preferences and priorities that conventional questionnaires might overlook [Lhoest et al., 2021].

It is important to emphasize that the intersection between risk assessment and investor profiling is essential for creating robust and personalized financial strategies tailored to individual needs. Furthermore, compliance with international regulations, such as Basel III and MiFID II, highlights the importance of integrating these processes to ensure ethical, transparent, and adaptive financial practices [Jasrotia et al., 2020].

As the financial sector continues to evolve, the combined application of modern risk assessment and investor profiling techniques represents a transformative opportunity. By leveraging the capabilities of AI and ML, financial institutions can create systems that are more precise, adaptive, and aligned with their clients' needs, simultaneously enhancing performance and satisfaction.

2.1.3 LLMs in Chatbots

Natural Language Processing is a critical field of artificial intelligence that empowers machines to understand and generate human language by leveraging syntactic, semantic, and pragmatic analyses. This capability has paved the way for the development of chatbots and conversational agents that interpret complex queries, provide detailed information, and engage in human-like interactions. Over time, Large Language Models have revolutionized NLP by employing transformer-based architectures and attention mechanisms, enabling the processing and generation of text with unprecedented precision. Trained on vast datasets, these models excel at recognizing complex linguistic patterns and have been applied to tasks such as text classification, content generation, and conversational interfaces [Roberts, 2024].

In the financial sector, the capabilities of LLMs have been transformative. Their ability to process and analyze unstructured data—including financial news, market commentary, and investor communications—makes them invaluable tools for market analysis, personalized financial advisory, and advanced recommendation systems. By providing real-time insights into market trends, sentiments, and emerging risks, LLMs empower financial institutions, traders, and investors to make swift and informed decisions. Beyond advisory applications, LLMs are increasingly used in automating compliance monitoring and fraud detection. By analyzing large datasets, such as transaction histories and regulatory documents, these models can detect suspicious patterns and anomalies indicative of misconduct. This not only ensures adherence to complex regulations, such as Basel III and MiFID II, but also helps mitigate financial penalties and reputational risks.

A key application of LLMs within the financial domain is the development of advanced chatbots. Modern chatbots harness the power of LLMs to deliver context-aware, adaptive, and multi-turn conversations, far surpassing the limitations of earlier rule-based systems such as ELIZA (Weizenbaum, 1966). These conversational agents rely on the theoretical foundations of language modeling and computational linguistics, utilizing deep learning techniques such as Sequence-to-Sequence learning and transformer architectures like GPT and BERT [Ghojogh and Ghodsi, 2020]. This enables chatbots to provide responses that are contextually relevant and dynamically adapt to user interactions, making them highly effective in roles ranging from customer support to personalized financial guidance.

The interdisciplinary design of chatbots integrates knowledge from psychology, sociology, and human-computer interaction to improve user experience and usability. For instance, understanding user behavior and preferences enhances the ability of chatbots to provide tailored financial recommendations and support. In the financial context, these conversational agents play a pivotal role in streamlining client interactions by offering immediate assistance, answering investment-related questions, and even simulating complex financial scenarios to aid decision-making.

By combining the advanced capabilities of LLMs with the interactive nature of chatbots, financial institutions can offer highly personalized and efficient services to their clients. This integration not only improves user engagement and satisfaction but also enhances operational efficiency by automating repetitive tasks and ensuring compliance with regulatory requirements. As the financial industry continues to evolve, the synergy between LLMs and chatbots represents a transformative opportunity to redefine how institutions interact with their clients and address their needs.

2.1.4 Surveys for Investor Profiling

Surveys have traditionally been one of the primary tools for assessing risk profiles in the financial sector. They consist essentially of structured forms designed to assess an individual's financial goals, risk tolerance, and investment horizon. Typically, can include scenarios such as: "How would you react if your portfolio lost 20% of its value?" with predefined response options that categorize investors into risk profiles, such as conservative, moderate, or aggressive. Although traditional, straight forward, and widely used, they present significant limitations, including the inability to adapt to changing investor circumstances, the oversimplification of complex preferences, and the potential biases introduced by the formulation or interpretation of the questions.

With artificial intelligence and machine learning advancements, risk profile surveys have evolved into more dynamic and adaptive tools. The most modern AI-based surveys can personalize questions based on real-time responses, integrate data from external sources such as transactional histories or market behaviors, and adapt over time to reflect changes in an investor's financial situation [van de Poll et al.,]. These approaches address the rigidity and lack of depth of traditional methods, enabling a more comprehensive and precise understanding of an investor's risk tolerance and financial goals.

Technologies like Natural Language Processing have been employed to analyze open-ended responses, extracting deeper insights into investor preferences. As a result, these advanced surveys provide more accurate risk assessments and personalized financial recommendations, improving the user experience and the reliability of financial advisory systems [Kurochkin et al., 2024].

2.1.5 Search Strategy

The selection of scientific articles for this study followed a systematic and rigorous approach to ensure a solid theoretical foundation in analyzing the application of Large Language Models in the creation of investor risk profiles. The main objective of this strategy is to identify, evaluate, and synthesize relevant literature, ensuring that the study is based on the most recent and high-quality research available. To achieve a comprehensive coverage of the existing literature, searches were conducted in the following academic databases: Google Scholar, IEEE Xplore, SpringerLink, ScienceDirect, and ACM Digital Library.

To ensure that the search results were aligned with the study objectives, a set of search terms and Boolean operators was defined. The queries included the following keywords and logical connectors: "Large Language Models" AND "Investor Risk Profile," "Financial Risk Assessment" AND "LLMs," "Artificial Intelligence in Finance" AND "Investor Profiling," and "Machine Learning" OR "LLMs" AND "Investment Risk." The use of Boolean operators (AND, OR, NOT) allowed the search to retrieve relevant studies while avoiding irrelevant or overly broad results.

To ensure the relevance and quality of the selected articles, the following inclusion and exclusion criteria were applied. The inclusion criteria involved studies published in the last five years, peer-reviewed journal articles, conference papers, and technical reports, research focused on LLMs, AI, or ML applied to finance, and studies discussing methodologies for creating investor risk profiles. The exclusion criteria eliminated articles that do not focus on LLMs or the creation of investor risk profiles, studies without peer review or with low credibility, articles that deal exclusively with traditional statistical models without any AI component, and research that does not present a clear methodological framework.

Although the search strategy was designed to be comprehensive, some challenges and limitations were encountered, including restricted access to certain articles due to paywalls, scarcity of specific literature on LLMs in the creation of financial risk profiles, requiring references to adjacent areas, potential publication bias, as AI-based financial models tend to highlight successes rather than failures, and methodological differences in existing studies, making direct comparisons challenging. To mitigate these issues, the study incorporated a broader range of sources and included older but highly relevant studies when necessary.

This search strategy ensured a rigorous and systematic identification of relevant literature, providing a solid foundation for analyzing LLMs in the creation of investor risk profiles. The use of multiple databases, refined keyword queries, structured filtering, and quality assessment contributed to a comprehensive and reliable review process. These findings will guide the subsequent stages of this study, ensuring that the theoretical framework is well-grounded in existing research.

2.1.6 Inclusion Criteria

The establishment of specific inclusion criteria was put in place for the selection of literature on relevant material and quality to ensure depth in the review, filtering those studies that will contribute directly to research in Large Language Models applied to the creation of investor risk profiles, with the maintenance of methodological rigor and practical applicability.

1. **Publication Date:** only the researches which were published within the last five-year period were included so that only current and relevant studies are represented.

2. **Peer-Reviewed Sources:** the findings are based only on data from peer-reviewed journals, conference proceedings, and technical reports to assure credibility and reliability.

3. **Relevance to LLMs in Finance:** only those studies were included where the use of LLM, AI, or ML for financial applications was discussed clearly, specifically in developing an investor's risk profile.

4. **Methodological Rigor:** only the research contributions that presented a clear methodological framework, including data sources, model architecture, and evaluation criteria, were taken into consideration.

5. **Empirical or Theoretical Contributions:** it also invited both empirical studies with real-world financial data and theoretical analyses that could provide insight into how LLMs can be applied to investor profiling.

6. **Availability in English:** only studies published in English were included to ensure the accessibility and feasibility of a uniform analysis.

2.2 Related Work

The evolution of investor profiles and financial advisory systems has been shaped by a wide range of research efforts, from traditional static methods to the emergence of advanced AI-based approaches. Understanding the current state of the field is essential to contextualize the contributions of this thesis and highlight the challenges and opportunities in applying LLMs to investment profile analysis. This section aims to review related work on the topic, exploring key developments in analyzing investor profiles through the application of ML and artificial intelligence in financial systems and the specific advancements enabled by LLMs.

This section is structured into the following categories:

- Risk and Credit Assessment Using LLMs;
- Decision Support and Financial Advisory using LLMs;
- Challenges and Transformations in Financial Models using LLMs.

2.2.1 Risk and Credit Assessment Using LLMs

The researchers behind [Cao et al., 2024] proposed RiskLabs, a framework that uses LLMs for financial risk prediction, addressing gaps in prior studies focused on summarization and stock prediction. Their approach integrates multimodal data to improve forecast accuracy and decision-making in the financial field, including transcripts and audio of earnings conference calls (ECC), market time series, sentiment extraction, tone, contextual financial news, and other characteristics. Time-series data are processed with BiLSTM networks and autoregressive methods to capture historical trends. Experimental results demonstrated that RiskLabs outperformed traditional methods (e.g., GARCH, LSTM) and state-of-the-art models (e.g., HTML (Hierarchical transformer-based multi-task learning)) in risk forecasting, particularly for short- and medium-term horizons.

The study conducted by [Hens and Nordlie, 2024] examine the effectiveness of natural language models, such as ChatGPT and Bard, in categorizing investor risk profiles and comparing their performance to banking experts. Accurate risk profile categorization is essential for strategic asset allocation, as mandated by regulations like MiFID and FIDLEG in Europe and Switzerland. The methodology involved analyzing 10 predefined client profiles using a standardized 1-to-5 scale, with both LLMs and bankers responding to identical questions. The results, evaluated using Welch's t-tests, showed no significant evidence for half of the clients with negligible economic differences except in specific situations. However, LLMs failed to support evaluations, as they rely on general principles rather than clients' specific details, and demonstrated "extrinsic hallucinations" by returning information that cannot be verified. This study concludes that LLMs are helpful as a complementary tool to validate risk profiles but cannot replace human experts, given their limited capability for reasoning and contextual analysis.

The authors of [Dai et al., 2021] proposed a machine learning framework to predict bank credit ratings and assess financial risk. Their approach used a data set of 302 companies and combines LASSO regression with recursive feature elimination to identify critical features, such as days with zero or negative income. These selected features were used to train machine learning models. The evaluation results showed that SVM achieved the highest accuracy (86%) and sensitivity (84%), while Random Forest and Gradient-Boosted Trees demonstrated superior specificity (95%). Additionally, survival analysis using Cox and Kaplan-Meier methods highlighted that negative and zero-income days strongly affect credit ratings.

The article of [Mhlanga, 2021] proposed the use of artificial intelligence and machine learning to improve credit risk assessment and financial inclusion in emerging economies, addressing the exclusion of under-banked groups, such as women, youth, and small businesses, due to a lack of collateral or credit histories. The study emphasized the potential of alternative data sources—such as public records, social media interactions, and satellite imagery—combined with supervised techniques like decision trees and logistic regression and unsupervised methods, such as clustering, to predict borrower behavior and assess default risk. A literature review and conceptual analysis of secondary sources, including peer-reviewed articles and institutional reports, formed the methodological basis of the research. The findings revealed that algorithms that used mobile usage data and social media interactions significantly reduced default rates, improved borrower classification accuracy, and enabled financial institutions to extend credit access to previously excluded groups.

The contributors of [Feng et al., 2023] proposed the Credit and Risk Assessment Language Model (CALM), leveraging Large Language Models for multitasking financial risk and credit assessment. Traditional credit scoring systems often lack generalization across financial activities, which CALM addresses by fine-tuning Llama2-chat with the LoRA strategy to manage computational costs. The model was evaluated using a curated benchmark of nine datasets, including 45,000 instruction-tuning samples, covering tasks such as credit scoring, fraud detection, financial distress identification, and claims analysis. The evaluation used metrics such as accuracy (Acc), Matthews Correlation Coefficient (MCC), and F1-score, alongside fairness measures like Equal Opportunity Difference (EOD) and Average Odds Difference (AOD). CALM outperformed traditional expert systems and other LLMs, demonstrating superior accuracy, adaptability on imbalanced datasets, and knowledge transfer across tasks.

2.2.2 Decision Support and Financial Advisory using LLMs

The researchers behind [Lakkaraju et al., 2023] investigate the role of LLMs in financial advisory services, focusing on efficacy and fairness for tasks such as product discovery and multi-product interaction. Comparing ChatGPT and Bard against SafeFinance, a rule-based system, this research emphasizes key challenges that LLMs face, including inconsistencies, biases, and limitations in numerical reasoning, despite fluent responses. Quantification metrics such as ISIP and ISA were thus developed to measure bias and consistency. The responses for SafeFinance were traceable and consistent, but not adaptable. The authors thus feel that a hybrid model that scales the LLM's scalability with the precision of rule-based systems could handle these challenges much more effectively.

The authors of [de Zarzà et al., 2023] investigate the use of Large Language Models in financial planning, addressing the limitations of traditional methods in meeting dynamic and personalized financial needs. The authors propose an optimization framework that integrates econometric models with LLM recommendations to optimize income allocation for both individual and cooperative budgeting scenarios. It will use mathematical optimization techniques and metrics such as maximization of savings or keeping within financial constraints. The paper proposed a retrieval-augmented generation approach for LLM recommendations in order to make the recommendations verifiable. Results are shown to prove that LLMs allow for improvements in personalization and efficiency regarding financial planning, LLMs provide economically sound advice, and enhance cooperative decision-making within households. Even so, this research makes a case for keeping humans in the loop so that recommendations can be reliable, ethical, and unbiased.

The article of [Wang et al., 2023] proposed an autonomous LLM-powered agent that combines planning, personalized memory, and external tools to enhance recommendation capabilities named Recmind. Its Self-Inspiring (SI) algorithm improves reasoning by leveraging multiple historical decision paths, surpassing existing methods like Chain-of-Thought and Tree-of-Thoughts. Additionally, the system uses tools such as SQL queries and web searches to retrieve real-time data and enrich its recommendations. The authors demonstrated the effectiveness of RecMind through experiments on Amazon and Yelp datasets. Results showed that RecMind outperforms baseline methods in rating prediction and sequential recommendation. These findings highlight the potential of combining LLMs with external tools to overcome the limitations of current recommendation systems.

The authors of [Li et al., 2024] present InvestorBench, a benchmark to study how well LLM-based agents can carry out financial decision-making tasks by filling in the void of a lack of all-encompassing frameworks and standardized benchmarks in the financial sector. The approach focuses on developing an agent with modules for stratified memory, data processing, and action mechanisms, while InvestorBench includes tasks like stock trading, cryptocurrency trading, and ETFs. These experiments were performed on realistic market conditions simulated over the multimodal data from Yahoo Finance and SEC EDGAR with 13 different LLMs tested in total, covering open-source, proprietary, and domain-specific settings. The results indicated that proprietary LLMs outperformed their open-source counterparts in more complex financial tasks, particularly in highly fluctuating markets. Although larger and more specialized models showed their results with higher accuracy and robustness, general-purpose LLM-based agents are more variable in performance.

2.2.3 Challenges and Transformations in Financial Models using LLMs

The article of [Zhao et al., 2024] has deeply analyzed the effect of Large Language Models in changing the world of finance by underlining a series of use cases, from generating financial reports, predicting market trends, and analyzing investor sentiment to personalized advisory. The large language models are built on a transformer-based architecture that is immense in terms of handling large volumes of data in order to come up with actionable insights that can boost efficiency and satisfaction. It combines a systematic review with practical testing of the capabilities of GPT-4 in tasks such as sentiment analysis, entity recognition, and prediction of stock movements using multimodal data sets from financial news, earnings reports, and social media. The performance metrics included precision, contextual understanding, and mathematical reasoning. The results show that, although LLMs do well in sentiment analysis and zero-shot learning, they can underperform on more quantitative tasks, such as time series forecasting and risk modeling. Thus, there is a need to integrate them with specific models. It finally concludes that LLMs have revolutionary powers but also have to be integrated with traditional systems and guided by human oversight regarding their limitation in data interpretation and alignment to ethics.

The authors of [Zhou et al., 2024] introduce the Financial Bias Indicators (FBI) framework to evaluate and mitigate biases in Large Language Models used in the financial sector, drawing on principles of behavioral finance. The study examines how intrinsic biases, such as risk preferences and framing effects, can undermine financial decision-making. Using the FBI framework, 23 LLMs, including general-purpose and finance-specific models (FinLLMs), were assessed through simulated scenarios with financial data and tailored prompts to analyze biases like loss aversion and situational dependency. These are mitigated by using techniques for bias mitigation such as Chain-of-Thought (CoT) reasoning with the support of the FinCausal dataset. Results indicate that FinLLMs show more variability and irrationality onrisk-related tasks, while general-purpose models tend to have fewer biases but often suffer from interpretability challenges. The CoT methods decreased the bias for some cases, but for less-biased models, this technique sometimes increased bias. The authors, therefore, conclude that the FBI framework provides great potential to enhance rationality in LLMs but also puts forth that continuous refinement will be necessary so that financial decision-making becomes ethical and stable.

2.3 Research Gap

Although Large Language Models have demonstrated great potential in analyzing investor profiles and financial applications, several significant research gaps persist. Existing models struggle to understand the specific behavior of investors due to their reliance on generalized principles and the lack of integration of personalized and contextual information, particularly for user monitoring. This limitation compromises the accuracy of profile analysis and the identification of risk tolerances.

In spite of breakthroughs in AI-driven financial evaluation, investor profiling is still a static and inflexible process and does not reflect the complexity of individual financial conduct and perception of risk. Traditional questionnaires are based on pre-determined categories that do not allow investors to accurately convey their decision-making behaviors due to many reasons like inadequate knowledge of finance, if they are inexperienced, or inadequate time to be able to answer all the questions, leading to incorrect classifications and unjustified risk estimations.

1. **Lack of flexibility in investor responses:** the majority of investors cannot select appropriate responses in structured questionnaires as their interests do not align with the provided options. This leads to constrained decisions that do not correctly reflect their true financial preference.

2. **Challenges in understanding financial terminology:** a significant portion of investors, especially those with poor financial literacy, do not comprehend financial jargon. This impedes adequate self-testing and increases the potential for erroneous responses.

3. **Lack of interactive and adaptable assessment tools:** traditional methods do not allow investors to belong to an adaptive conversation where they can update responses in light of clarifications or follow-up queries. This is a rigidity that finds itself converted into frozen and at times obsolete risk profiles.

Although LLMs have been explored in numerous financial applications, their interactive risk profiling capability has yet to be extensively researched. Existing studies have yet to explore systematically how LLMs can improve upon the limitations of structured questionnaires by providing more personalized, simpler-to-understand, and more adaptive risk categorization.

This thesis aims to fill this gap by exploring how LLMs can make investor profiling more user-friendly, conversational and adaptive. By bridging the gap between mathematical risk assessment models and an interactive AI-based profiling process, this study aims to assist in the creation of more user-friendly and effective financial advisory systems.

Chapter 3

Contextualization and Problem Specification

3.1 Contextualization

The financial technology sector is currently experiencing a significant shift towards more interactive and user-centric approaches, particularly in understanding individual investor profiles.

Traditional methods have so far focused on static questionnaires and rigid categorizations, which are now facing significant limitations in capturing the dynamic nature of human behaviors and investor preferences. These methods fail to reflect certain fluctuations, such as changes in financial goals, risk tolerance, or reactions to constantly evolving market conditions. As one can imagine, all these variables can compromise the effectiveness of financial decisions, impacting both the institutions' strategies and the clients' attempts to make the best possible choices for themselves.

The application of LLMs in modeling investor profiles offers a novel approach, as their core strength lies in their ability to process and integrate large volumes of diverse data. This enables the creation and analysis of dynamic profiles that adapt to contextual and behavioral changes in real-time.

This thesis positions itself at the intersection of technological innovation and personalization in financial services, presenting a new concept focused on investor profile analysis. However, its primary focus is on exploring the potential of adaptive and continuously evolving LLMs, providing a detailed analysis of the methodologies, challenges, and opportunities associated with applying this technology in dynamic financial environments.

3.2 Problem Specification

Financial advisory services have evolved over the years, adapting to new technological challenges and market trends. However, significant challenges persist, particularly in aligning recommendations with the unique needs and preferences of each investor, helping them make the best financial decisions. This aspect becomes even more relevant in a context where financial decisions are influenced by multiple factors, including personal goals, economic conditions, and market dynamics.

Traditionally, investor profiling relies on static questionnaires that classify investors into generic categories based on their risk tolerance, financial goals, and investment horizons. While this approach has been effective in more predictable contexts, namely in stable environments and within a static framework, it exposes several critical limitations, such as the inability to reflect the continuous and dynamic changes in investors' lives and behaviors, especially in relation to their individual economic context. Additionally, there are intrinsic issues within the questionnaire process itself, which often compromise the quality of responses and, consequently, the accuracy of the profiles created.

One of the most common problems is the lack of time and motivation among investors to carefully analyze and respond to the questions. Many end up providing random or standardized answers that fail to accurately reflect their true preferences and financial goals. Another problem is the technical language and complex formulation of certain questions, which can hinder understanding, particularly for investors with lower financial literacy. This not only increases the risk of incorrect responses but also generates frustration and demotivation during the process.

Investors often also struggle to fit their opinions and preferences into the available categories, being forced to choose options that do not accurately represent their financial reality. For instance, an investor who considers themselves to have a moderate risk tolerance may feel uncomfortable choosing between broad categories such as "conservative" or "aggressive." Another factor is that static questionnaires provide no clarification or context about certain questions, leading investors to answer randomly.

This misalignment between the available options and the investor's true preferences creates imprecise profiles, limiting the system's ability to offer relevant and reliable recommendations. For example, an investor who begins their journey with a high risk tolerance, possibly supported by adequate financial resources, and who, over time, with changes in income, new family responsibilities, or even a greater aversion to market volatility may completely alter their perspective, introducing new variables that need to be analyzed. Traditional methods, which capture preferences at a single point in time, are unable to keep up with these transformations. As a result, financial advisory systems become ineffective at offering relevant recommendations, undermining the user experience and eroding trust in the process, thereby complicating access to new financial products or services.

In this context, LLMs emerge as a transformative new opportunity for financial advisory services, challenging conventional approaches. They also possess the capability to process vast amounts of data and identify complex patterns in real time, integrating information from conversations, transaction patterns, and even external data, such as market trends, to create investor profiles that are continuously updated and refined. More importantly, they can detect subtle changes in investors' preferences and behaviors, allowing financial recommendations to be dynamically adapted, making them more realistic and reducing the need for extensive human analysis.

Based on this potential, this thesis proposes to explore how LLMs can overcome the limitations of traditional methods for evaluating investor profiles, supporting their needs and clarifying questions about what they truly seek in their investments.

Chapter 4

Methodology and Requirements Analysis

This chapter presents the methodology used in this study to explore the potential of Large Language Models in the analysis of investor risk profiles. The outlined approach integrates a research design that includes quantitative surveys and model training for this purpose, ensuring an analysis of the LLMs' capabilities to provide personalized financial information.

The chapter begins by describing the research design, highlighting the integration of various data collection methods adapted to understanding user preferences and experiences, followed by the presentation of data analysis techniques used to interpret both quantitative and qualitative information.

Additionally, the strategies implemented to mitigate limitations are discussed, ensuring transparency and strengthening the reliability of the conclusions. This chapter concludes implementation processes, as well as an overview of the requirements that guided the system's development.

Through this structured approach, the study aims to provide valuable insights into how LLMs can revolutionize financial services by offering an innovative, user-centric solution for risk profile analysis.

4.1 Methodology of Study

The methodology of this study follows a structured and iterative process, aiming at the implementation and evaluation of the use of Large Language Models for the analysis of investor risk profiles. This process is divided into several essential phases, ensuring a systematic and well-founded approach.

As illustrated in Figure 4.3, the approach progresses from the review of traditional risk assessment methods to the deployment of the LLM-based model and its subsequent validation with real users. The workflow begins with the selection of the appropriate methodology, followed by the realization of the traditional questionnaire, mapping responses in the LLM model, and conducting iterative testing to refine and enhance the prototype.

4.1.1 Analysis of Existing Risk Assessment Methods

Initially, an analysis of existing risk assessment methods is conducted to understand traditional approaches and identify gaps or opportunities for the application of LLMs. Risk assessment in the financial sector is an essential process for making informed investment decisions. Traditionally, this process relies on quantitative and qualitative methods to classify investors based on their risk tolerance, financial goals, and investment horizon.

The most common methods include:

- **Risk Profile Questionnaires:** these structured surveys, such as those used by Morningstar, are widely applied to categorize investors as conservative, moderate, or aggressive. They are based on questions concerning financial situation, investment experience, and reactions to market scenarios [van de Poll et al.,].

- **Statistical and Mathematical Models:** techniques such as Value at Risk, standard deviation, and risk-adjusted return analysis (Sharpe Ratio) are used to quantify the volatility and risk exposure of an investor or portfolio [Nagpal, 2024].

- **Behavioral Analysis:** some methods incorporate psychological and emotional factors into the risk assessment process by analyzing investors' behavioral patterns and decision-making tendencies [Blanchard et al., 2011].

Based on this analysis, the most appropriate method is selected, serving as the basis for the development of the proposed system, in this specific case, the Risk Profile Questionnaires were used with a focus on Morningstar.

The Risk Profile Questionnaires method were chosen due to their widespread acceptance in financial risk assessment, their structured and validated approach, and their ease of implementation within an automated system. The methodology adopted by Morningstar provides a robust framework that enables a clear classification of investor profiles. Furthermore, this method is well aligned with the capabilities of Large Language Models, allowing for structured question mapping and greater adaptability.

To mirror the traditional approach, a questionnaire was created based on Morningstar's but using the Google Forms tool, replicating the structure and methodology of the selected risk assessment model. The form functions as a direct implementation of the conventional method, facilitating the collection and classification of data, without any personalized follow-up.

Morningstar's Risk Tolerance Questionnaire is a scientific tool for identifying an investor's risk profile based on assessment of his or her investment time horizon, financial goals and ability to withstand market fluctuation. It is constructed with a series of multiple-choice questions carefully designed to classify investors into different levels of risk like conservative, moderate or aggressive [Rabbani and Nobre, 2022].

The evaluation process is based on fundamental financial principles:

- **Time Horizon** - Determines how many years an investor intends to continue holding their investments before they start making withdrawals.
- **Risk Appetite** - Measures an investor's willingness to stomach short-term market volatility in the expectation of long-term return.
- **Portfolio Preferences** - Establishes the investor's preference between stability and return potential in the composition of their portfolio.
- **Response to Market Downturns** - Evaluates how an investor would react to downturns in the market, helping to determine their tolerance during volatile markets.

Each question is constructed to provide codified responses that correspond to pre-conceived risk profiles, which are then translated into pre-defined categories to achieve consistency and objectivity of classification. The methodical approach earns Morningstar's questionnaire widespread acceptance in financial risk measurement in terms of the standardized investor profiling method.

At the end of the questionnaire, a total score is calculated, based on the score obtained in each question, taking into account the following logic:

- The first two questions ascertain the time horizon of the investor and assign a relative score.
- The remaining questions are used to calculate the risk tolerance score, which measures risk tolerance in terms of the preferences and behaviors stated by the investor as illustrated on Figure 4.1.

The crossing of the Time Horizon Score and the Risk Tolerance Score provides a final ranking that helps to establish the investor's risk profile and recommend an appropriate investment strategy.

4.1.2 Development of a Risk Assessment Model

After implementing the traditional questionnaire, a GPT-based model was developed to generate a more personalized risk assessment, allowing for a dynamic and adaptive evaluation of investors' profiles, surpassing the rigid classifications of conventional methods.

A crucial issue with traditional questionnaires is the difficulty investors face in selecting the correct response and the demotivation associated with the process. Static forms often fail to capture the nuances of an individual's financial behavior, resulting in generalized classifications that may not accurately reflect the investor's true risk tolerance [Mazzoli and Palmucci, 2023]. In this case, instead of forcing investors into predefined categories, the model interprets responses in a more flexible manner, dynamically adapting to user input and refining the risk classification in real time. This approach not only improves the accuracy of the assessment, but also enhances user engagement and motivation, making the evaluation experience more intuitive and tailored to their needs.

The GPT-based model is designed to interpret and assimilate Morningstar's traditional questionnaire, comprising not only the structure of the questions but also the possible response options, in order to establish a framework by mapping them through the insertion of the full questionnaire. This allows the model to simulate the traditional risk assessment process, while enabling dynamic adjustments based on investors' responses, ensuring that they are not only classified, but also understood within a broader financial context. That is, by generating additional sub-questions when answers are unclear or ambiguous, guiding the user towards a more accurate rating by refining their input through contextualized follow-up queries.

In addition, it allows for a contextualized assessment, ensuring that investor profiles are classified with a deeper understanding by analyzing the coherence of the answers in the different questions, identifying contradictions or uncertainties that may require further clarification to refine the investor's risk profile.

4.1.3 Question Design and Mapping Strategy

For the development of each question, the structure of the questionnaire was designed to maintain both its original intent and the logic behind the answer choices. In the first phase, the questions were categorized according to their function, such as risk tolerance, investment experience, financial goals, and reactions to market scenarios. In the second phase, the investment horizon, the aim is to ensure that long-term and short-term investment objectives are properly identified and integrated into the overall assessment.

The mapping ensured that each answer corresponded precisely to one of the predetermined options, allowing the model to correctly read and classify investor responses. The structured format allowed for the correct allocation of risk profiles and minimized misinterpretations that could have arisen from ambiguous or conflicting answers. Based on the Morningstar questionnaire that was used, an investor's answer to the initial questions about their investment time horizon has a direct impact on their subsequent assessment of risk tolerance. Consider the following:

1. *When do you expect to begin withdrawing money from your investment account?*
   a. *Less than 1 year*
   b. *1 to 2 years*
   c. *3 to 4 years*
   d. *5 to 7 years*
   e. *8 to 10 years*
   f. *11 years or more*
2. *Once you begin withdrawing money from your investment account, how long do you expect the withdrawals to last?*
   a. *As one lump sum distribution or in less than 1 year*
   b. *Over a period of 1 to 4 years*
   c. *Over a period of 5 to 7 years*
   d. *Over a period of 8 to 10 years*
   e. *Over a period of 11 years or more*
3. *Some investors are more willing than others to accept declines in the value of their portfolio as a trade-off for potentially higher long-term returns. Which response best represents your attitude toward the following statement?*

   "*I am willing to experience large and frequent declines in the value of my investment if it will increase the likelihood of achieving high long-term returns.*"
   a. *Strongly Agree*
   b. *Agree*
   c. *Disagree*
   d. *Strongly Disagree*

4. *Suppose you had a choice between three hypothetical investments. The table below lists the return possibilities associated with each investment. In any given one-year period, the portfolio may be losing money, earning a return between 0 and 7 percent, or earning more than 7 percent. Of the 3 portfolios shown below, which best matches the risk/return trade-off you would be most comfortable with when investing?*

Portfolio	Chance of Losing Money	Chance of Earning 0-7%	Chance of Earning More than 7%
A	5%	85%	10%
B	15%	65%	20%
C	25%	45%	30%

Table 4.1: Portfolio risk and return probabilities

   a. *Portfolio A*
   b. *Portfolio B*
   c. *Portfolio C*

5. *Investments with the best chance of high returns typically have the most uncertainty. For example, the portfolio with the potential for the biggest gain in the best case scenario also has the potential for the biggest loss in the worst case scenario.*

   "*How much return are you willing to give up in order to feel comfortable?*"

   a. *I will accept the lowest return in exchange for the portfolio with the smallest and least frequent changes in value.*
   b. *I will accept a moderate return in exchange for less frequent and smaller changes in value.*
   c. *I will only accept a high return. I understand that the portfolio will frequently have large changes in value.*

6. *Inflation can greatly erode the return on your investments, especially over time. For example, in a year with a 4% inflation rate, an investment with a 7% return before inflation would only produce an increase in purchasing power of 3%.*

   "*Please specify one of the following statements that best summarizes your attitude toward investing and inflation.*"

   a. *I aim to limit the volatility of my portfolio while keeping pace with inflation.*
   b. *I aim to moderately exceed inflation while accepting a moderate level of risk.*
   c. *I aim to significantly exceed inflation while accepting a significant level of risk.*

d. *I aim to maximize my return while accepting a large and dramatic level of risk.*

7. *The following graph shows the probable range of returns and losses of five hypothetical portfolios over a one-year period. Notice that portfolios with high returns also have the probability of experiencing large losses. In which of these portfolios would you prefer to invest?*

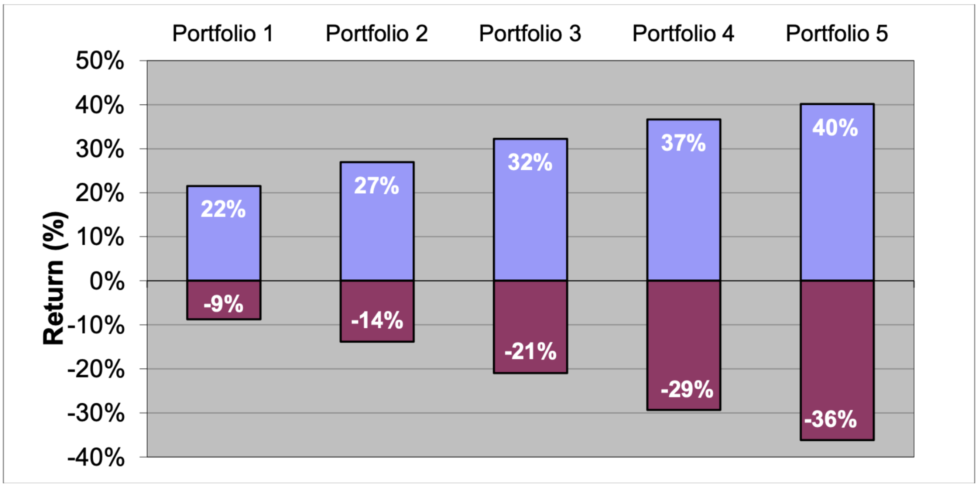

Figure 4.2: Five (5) hypothetical portfolios

a. *Portfolio 1*

b. *Portfolio 2*

c. *Portfolio 3*

d. *Portfolio 4*

e. *Portfolio 5*

8. *The following table presents a supposed potential worst-case loss, probable gain, and potential best-case gain of \$100,000 invested in five hypothetical portfolios over a one-year period. Which one of the following would you prefer to hold in your account?*

Portfolio	Potential Best Case (\$)	Probable Gain (\$)	Potential Worst Case (\$)
1	121,500	105,100	91,300
2	126,900	105,900	86,100
3	132,200	106,700	79,000
4	136,600	107,500	70,700
5	140,100	108,100	63,800

Table 4.2: Potential outcomes for different portfolios

a. *Portfolio 1*

b. *Portfolio 2*

c. *Portfolio 3*
d. *Portfolio 4*
e. *Portfolio 5*

9. *The following table shows the average return and probability of experiencing a loss in five different hypothetical investments over a twenty-year period. Which of the following investments would you prefer?*

Investment	**Likely Annual Return**	**Number of Negative Years**
A	3%	2
B	4%	4
C	5%	5
D	6%	6
E	7%	7

Table 4.3: Comparison of investments: annual return and negative years

a. *Investment A*
b. *Investment B*
c. *Investment C*
d. *Investment D*
e. *Investment E*

A structured approach was taken to context awareness, ensuring that certain answers dynamically influence subsequent questions, similar to the personalized approach of a financial advisor.

In cases where the model could not accurately map a response, it was designed to generate additional questions within the context of each category to refine the assessment. To achieve this objective, prompt engineering was applied, providing example responses and clarifications to guide the model in simulating a personalized investor assessment. The follow prompt was used for training was as follows:

1. Using the Morningstar uploaded file, create a survey equal to the file uploaded. However, when asking the questions, the questions should roughly be the same, but should not present the options available to answer. These should be rewritten into the question or you should try and taunt the answer out of the user.
2. If the answer given does not match the options of that question from the Morningstar file, you should try to clarify and continue until you get an answer that matches the answers provided in the Morningstar file. Remember to not explicitly present the options

from that file at all costs. You should also not indicate that there are options available; you should try and get an answer from the user without influencing their decision with defined options. The user's answer must be mapped to one of the hidden options in each question. For example, if on question 2, the user answers "2 years", the final answer should map to "b. Over a period of 1 to 4 years" as 2 years is included in that time range.

3. Once a question is answered, you should proceed to the next question without diverging from the flow.
4. For questions that depend on an image, table, or graph, do not refer to the visual content at all; ask the user to refer to their "Images" document provided to them. That includes question 10, which includes a graph.
5. Do not engage in discussions outside of the provided questionnaire, even if asked.
6. Exclude Question 10 from the survey.
7. Once the user has finished the questionnaire, it should map the answers given to one of the options listed in the Morningstar file for each question and then export the generated map and its question. The format must be: Question Number, followed by the question from the Morningstar file; under the Question, the Answer given by the user, and under the Answer, the Mapped Option, in .pdf format. The file should be named "Investor Profile Results".

4.1.4 Limitations and Challenges

However, throughout the model creation process, certain limitations did emerge, affecting its performance and accuracy. One of the main challenges is its inability to directly read images. That is, it could only retain images for a short period, meaning that any visual elements, such as charts or diagrams, had to be hosted externally via URLs and referenced in the text. This dependency on external hosting introduced potential issues, such as accessibility difficulties or delays in loading the referenced content.

There was also difficulty in converting multiple-choice options into coherent textual responses. While structured options work well in traditional questionnaires, transforming them into a continuous and context-aware text format proved to be a challenge. Additionally, maintaining the correct correspondence between the options and responses was problematic, often resulting in inconsistent or misaligned interpretations.

Another limitation that existed throughout the process was the fact that the GPT model with a free license could only issue a maximum of 10 prompts per day. In other words, in this case the user would not be able to answer all 10 questions, since they had already spent a prompt when they started, which posed some challenges when collecting samples.

Furthermore, the model occasionally altered the context of certain questions, either by rewording them in a way that changed their meaning or by making incorrect assumptions about investor preferences. This issue was particularly evident when questions contained more complex financial terminology, where small variations in wording could lead to different risk classifications. To mitigate this problem, manual adjustments and prompt fine-tuning were necessary to ensure the model preserved the original intent of each question.

4.1.5 Optimizations and Adjustments

To achieve this, several manual refinements were applied to improve the model's accuracy:

- **Rewriting ambiguous questions:** a few questions in the Morningstar file included implicit assumptions that were not properly understood by the model. These were rephrased so that they prompted answers that could be mapped to predefined options without actually listing them.

- **Follow-up question improvement:** when a response from a respondent investor was provided outside the predetermined options, initially the model did not direct the user to the right response. Manual tuning included adding structured explanations and reworded follow-up questions to redirect the investor towards a matching response without biasing their choice.

- **Prevention of leading questions:** in order to maintain neutrality, adjustments were made in questions that tended to elicit leading user responses by default. This was particularly critical where risk appetite questions were concerned, and word tweaks were necessary to prevent suggestive wording.

Regarding its inability to directly read images, external URLs were incorporated and tested to ensure stability and proper visualization for the user, minimizing accessibility issues and delays in content loading.

For the conversion of multiple-choice options into coherent textual responses, custom formatting rules were introduced to maintain logical flow and contextual integrity. This included the explicit mapping of answer choices to predefined response models, reducing inconsistencies. The model was adjusted to preserve the original question structure, ensuring that the options were correctly aligned with their intended meanings.

To prevent misinterpretations and context alterations, reinforcement techniques in prompt engineering were applied. This involved creating clear and unambiguous prompts, including examples of correct responses, to guide the model in maintaining the original intent of the questions. Additionally, dynamic question adjustments were incorporated, allowing the model to generate additional clarifications if it detected uncertainty in its own interpretation.

With these improvements, the model's ability to replicate the structured logic of the traditional questionnaire was significantly enhanced, while also increasing its adaptability and personalization for the user.

4.1.6 User-Based Model Validation

The samples for 5 users were collected after the model implementation and training had been completed, which included subjects with various experiences in the risk investment market. It started with novice investors who were just taking their first steps into financial markets, moved into investors with some experience but looking for more knowledge, and finally included well-established experienced investors with elaborate investment strategies. This sampling was done to see the model's adaptability - whether it could correctly estimate and classify investors independent of their level of experience. Eventually, this feedback from them was quite helpful in locating further improvements that the model could use to give personalized and contextually relevant risk assessments.

This methodological process ensures that the application of LLMs in risk profile analysis is well-grounded, tested, and optimized, contributing to an innovative and efficient approach in the financial sector.

4.2 Requirements Analysis

Shifting from traditional, static methods of financial advisory to dynamic, LLM-driven solutions requires an appropriate set of criteria for developing such a model. In ensuring that the system may give personal risk profiling while adapting to evolving investor preferences, it became very important to define requirements. These requirements were designed to support the accuracy, reliability, and usability of the model, including regulatory and ethical considerations. It is based on these principles that the system ensures smooth user interactions, accurate financial recommendations, and adherence to industrial standards.

4.2.1 Model Requirements

- **Investor Risk Profiling:** the model, based on input provided by the user, should generate an investor risk profile using Artificial Intelligence and set criteria and assessment techniques.

- **Natural Language Understanding:** it needs for the LLM to understand what a user wants through direct or indirect financial intent.
- **Recommendation:** the system should be able to make recommendations of financial products on the basis of individual investor profiles and specific financial goals.
- **Reliability:** the model should assure high availability and fault tolerance, hence giving consistent and reliable risk assessments.
- **Usability:** the system shall be intuitive for all types of users, unsophisticated and sophisticated, in financial matters so that it guarantees an exciting interaction with the user.
- **Adaptability:** the model will be dynamic in the ever-changing market conditions and investor objectives, with real-time and context-aware insight.
- **Regulatory Compliance:** the system shall be based on financial regulations and data protection laws, confirming security and legality to industry standards.
- **Data Processing and Integration:** it also needs to provide for efficient volume processing of financial data and must be perfectly integrated with prevailing financial advisory frameworks so as to increase decision-making.

By meeting these requirements, the model ensures a sophisticated, dynamic, and trustworthy financial advisory system that enhances user engagement and decision-making efficiency.

The following table shows the point value for the time horizon questions.

Time Horizon

Question 1	Question 2
A. 0	A. 0
B. 1	B. 2
C. 3	C. 4
D. 6	D. 5
E. 9	E. 6
F. 11	

Responses to questions 1 and 2 are added together to arrive at the time horizon level. (_____) [1]

Risk Tolerance Score

The risk tolerance portion of the scoring is taken from questions 3 through 10. Morningstar Investment Management assigns the point value for each response according to its ability to quantify the risks involved in investing and its ability to effectively convey these risks to investors. The highest points are awarded to the most aggressive answer choice. The risk tolerance score ranges from zero (most conservative) to 100 (most aggressive).

Risk Tolerance

Question 3	Question 4	Question 5	Question 6
A. 13	A. 4	A. 4	A. 3
B. 8	B. 7	B. 7	B. 5
C. 5	C. 11	C. 11	C. 8
D. 3			D. 13

Question 7	Question 8	Question 9	Question 10
A. 0	A. 0	A. 0	A. 13
B. 4	B. 4	B. 4	B. 8
C. 7	C. 7	C. 7	C. 5
D. 11	D. 11	D. 11	D. 3
E. 13	E. 13	E. 13	

Responses to questions 3 – 10 are added together to arrive at a risk tolerance score. (______) [2]

© 2024 Morningstar Investment ManagementQuestionnaire Scoring System ALPS Sector SPDR Trust
Page 9

Figure 4.1: Morningstar questionnaire-Score Calculation

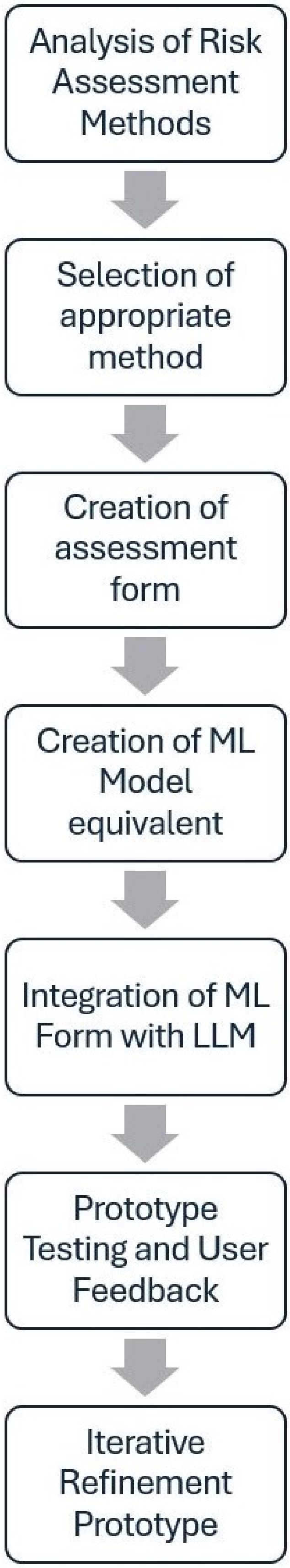

Figure 4.3: Development Process

Chapter 5

Results and Discussion

This chapter presents the results, quantifying the validity of the GPT model in forecasting the investor profile and comparing GPT responses to Google Forms. Possible differences in investment profile classification are analyzed and graphed, considering time horizon and risk tolerance. Also, quantitative and qualitative differences between classifications achieved by the model and original responses of the respondents are analyzed to better understand the alignment as well as potential deviations in profiling.

The major convergences and divergences between the classifications proposed by GPT and existing methods are discussed, taking into consideration variables such as investment horizon, risk tolerance, and expected returns.

5.1 Results

The analysis of the results is organized into several sections, each focused on a key aspect of the comparison between GPT and traditional methods, such as:

- **Alignment of GPT and Google Forms Responses:** evaluating the degree of alignment between user-provided answers and the GPT classifications.
- **Analysis of Divergences:** identifying the main areas where GPT deviates from traditional assessments and their impact on investor profiling.
- **Time Horizon Distribution:** assessing how investors classify their investments over different time spans and how well GPT replicates this behavior.

- **Risk Tolerance and Portfolio Choice:** investigating whether GPT accurately maps investors' risk preferences to appropriate portfolio selections.
- **Investment Adjustment Strategies:** understanding how the model handles subjective and less defined responses in investment decision-making.

5.1.1 Google Forms and GPT alignment

Table 5.1: Alignment between user responses and GPT (Questions 1 to 5)

Question	**Parameters**	**Users**				
		User 1	**User 2**	**User 3**	**User 4**	**User 5**
	Form	3 to 4 years	5 to 7 years	11 years or more	11 years or more	5 to 7 years
Question 1	GPT	3 to 4 years	10 years	11 years or more	11 years or more	10 years
	Divergence	No	Yes	No	No	Yes
	Form	Lump sum $<$ 1 year	Lump sum $<$ 1 year	8 to 10 years	11+ years	1 to 4 years
Question 2	GPT	Lump sum $<$ 1 year	$<$ 1 year	8 to 10 years	11+ years	5 to 7 years
	Divergence	No	No	No	No	Yes
	Form	Disagree	Agree	Strongly Agree	Strongly Agree	Strongly Agree
Question 3	GPT	Stability	Balance	Strongly Agree	Strongly Agree	Disagree
	Divergence	Yes	No	No	No	Yes
	Form	Port. A	Port. B	Port. B	Port. B	Port. A
Question 4	GPT	Stability, low returns	Port. B	Port. C	Port. B	Port. A
	Divergence	No	No	Yes	No	No
	Form	Minimal return	Moderate return	High return	Moderate return	Moderate return
Question 5	GPT	Stability, short-term	Balanced approach	High return	0.1	Moderate return
	Divergence	Yes	No	Yes	No	No

Table 5.2: Alignment between user responses and GPT (Questions 6 to 9)

Question	**Parameters**	**Users**				
		User 1	**User 2**	**User 3**	**User 4**	**User 5**
Question 6	Form	Beat inflation	Beat inflation	Significantly beat	Beat inflation	Beat inflation
	GPT	Moderate risk	Moderate risk	Beat moderate	Moderate gains	Moderate risk
	Desvio	No	No	Yes	No	No
Question 7	Form	Port. 1	Port. 3	Port. 4	Port. 1	Port. 1
	GPT	Port. 1	Port. 2	Port. 4	Port. 1	Port. 1
	Desvio	No	Yes	No	No	No
Question 8	Form	Port. 1	Port. 3	Port. 4	Port. 1	Port. 1
	GPT	Port. 1	Port. 2	Port. 4	Port. 1	Port. 1
	Desvio	No	Yes	No	No	No
Question 9	Form	Inv. D	Inv. A	Inv. C	Inv. E	Inv. A
	GPT	Wait 1 year	Risk balance	Inv. C	Inv. E	Inv. C
	Desvio	No	Yes	No	No	Yes

The tables 5.1 and 5.2 showed that there was a high level of correspondence between participants' initial responses and the alternatives to which they were mapped by the GPT, particularly in the case of long-term investors. This suggests that GPT is effective in interpreting and mapping user inputs to structured questionnaire options. However, discrepancies exist, notably in subjective or less well-defined responses.

To better understand GPT's classification performance and its potential to complement or replace traditional approaches, the discussion will focus on:

- Interpretation of the investment time horizon;
- Impact of risk tolerance on profile classification;
- Choice of portfolio decisions compared to GPT classifications;
- Investment adjustment strategies.

These variables allow for an assessment of the model's accuracy, interpretative capabilities for subjective responses, and overall utility for financial decision-making.

5.1.1.1 Investment Time Horizon Analysis (Q1 - Q2)

When comparing investors with an investment horizon of 5 to 7 years, there were some slight differences between the answers to the forms and the GPT categorizations. As shown in 5.1, there were two divergent answers in Question 1 and 1 divergent answer in Question 2.

The model, in trying to classify these answers, associated them with selections in the questionnaire that represented somewhat different time horizons, but within an acceptable understanding. This shows that when the answers are not explicit or not correctly formulated, the GPT interprets the nuances of the investor's intention and adapts them to the closest available option.

5.1.1.2 Impact of Risk Tolerance on Classification (Q3 - Q6)

The results revealed that, in most cases, GPT accurately classified investors in terms of the degree of risk specified by users. High-risk investors were accurately placed in more aggressive categories and risk-averse investors in more conservative ones. However, when the answers were poorly defined or unclear, the model was biased, placing potentially riskier investors in lower risk brackets.

This shows that in situations where the answer was ambiguous or neutral, GPT chose a more conservative classification, avoiding a high risk profile when it had no explicit evidence. This is, for example, evidenced in the divergence of Question 6, where the user's answer changed from one questionnaire to the next, having selected 'I intend to significantly exceed inflation, accepting a significant level of risk' in the form, and yet the GPT classified the answer as 'I intend to moderately exceed inflation, accepting a moderate level of risk'.

5.1.1.3 Correspondence of Portfolios (Q7 - Q8)

When comparing the GPT category classifications with responses obtained from participants, there was high correspondence among investors who made a selection of *Portfolio 1*, being a conservative plan. The model precisely mapped the responses with that selection in the questionnaire. However, in Question 7, where investors had to choose between four portfolios with varying risk levels, a deviation was identified. The same also occurred in Question 8, where it was assessed the investors' preference for portfolios based on loss and gain scenarios. The user had selected "Portfolio 3" in both cases, yet the GPT put these responses into slightly less risky categories, classifying them as "Portfolio 2" instead.

Since these portfolios had similar characteristics, these discrepancies may have resulted from the wording used by the participants, taken by the model who favoured a less risky profile than actually occurred.

5.1.1.4 Investment Adjustment Strategy (Q9)

The results are that, for answerers who provided direct and objective answers, the GPT was capable of maintaining high correspondence with their choices. In cases where the answers were less subjective or open, however, the model tended to classify the answers in more conservative options. This pattern confirms the hypothesis that, where the answer isn't clearly stated, the GPT will adopt a less risky approach, grouping investors more conservatively than they might have presented in their own responses.

5.1.2 Proportion Of Identical Vs Divergent Responses

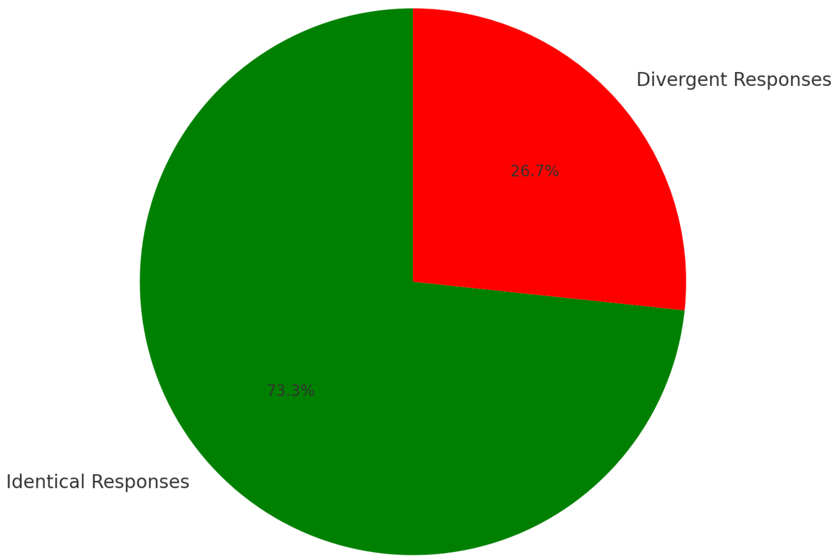

Figure 5.1: Proportion Of Identical Vs Divergent Responses

The circle graph on Figure 5.1 illustrates the number of similar and dissimilar responses between the GPT model language and the human responses collected on the form. The main point of this examination is to consider the level of agreement of the model with participants' choices under the context of their investor type and risk measure.

The green zone in the Figure shows the instances where the response was provided by the GPT in the form of a recommendation that was identical to the one pointed out by the user, and the red zone shows the instances where there was a mismatch between the responses. The ratio of this distribution allows us to understand the extent to which the model fits the human decision-making patterns, and a high percentage of comparable answers can indicate an effective performance in replicating investors' reasoning. A proportion of 26.7% of divergent answers suggests that the model may be interpreting the decision criteria in a different way from the participants, which may indicate that there is more scope to adjust the model training or the definition of its decision rules.

5.1.3 Distribution Of Same Vs Different Responses - GPT Vs Forms (Q1 - Q2)

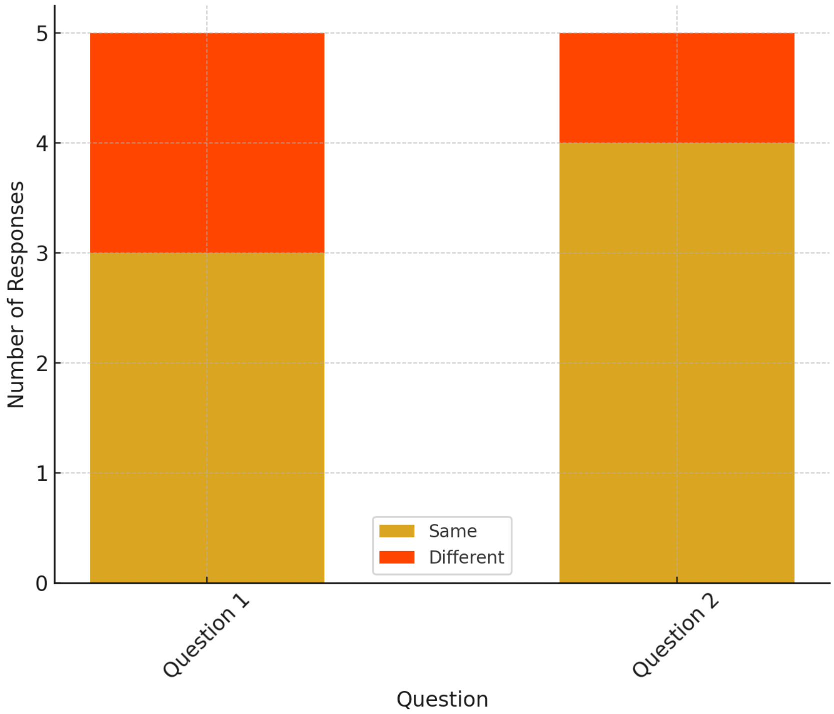

Figure 5.2: Distribution graphic over Time Horizon

The Figure 5.2 analyzes the distribution of responses regarding the investment time horizon, representing the period investors plan to hold their assets before making adjustments or withdrawals. Understanding the time horizon is crucial for classifying investors, as it significantly impacts their investment strategy and the selection of appropriate financial instruments.

The X-axis lists the two questions concerning investment time horizons, while the Y-axis indicates the number of responses for each category. By aggregating responses from both the original Google Forms and the GPT mapping, the graph reveals a slight divergence between the two responses. While the majority of classifications align, a portion of responses falls into the "Different" category, indicating cases where the GPT-based classification does not match the investor's self-reported time horizon.

Key insights from this analysis include:

- **High Agreement Between GPT and Human Responses**

  The majority of the answers fall under "Same", which suggests that GPT's classification is very much in line with the human responses collected through

Google Forms. This means that GPT can be a useful tool in investor profiling, particularly in ascertaining investment time horizons.

- **Presence of Some Discrepancies**

  While GPT generally aligns with investor responses, a non-negligible number of classifications fall into the "Different" category. These discrepancies indicate cases where the AI model interpreted responses differently from human participants, which may highlight edge cases or differences in how investors express their time horizon preferences.

- **Consistent Response Patterns Across Questions**

  The distribution of "Same" and "Different" responses remains stable across the two questions, suggesting that investment time horizon responses are generally consistent across different formulations. This reinforces the robustness of time horizon as a stable characteristic in investor classification.

This analysis focuses positively on GPT's classification in aligning responses with Morningstar's predefined investor profiles, while the observed discrepancies suggest areas for refinement.

5.1.4 Percentage of divergences between GPT and Forms

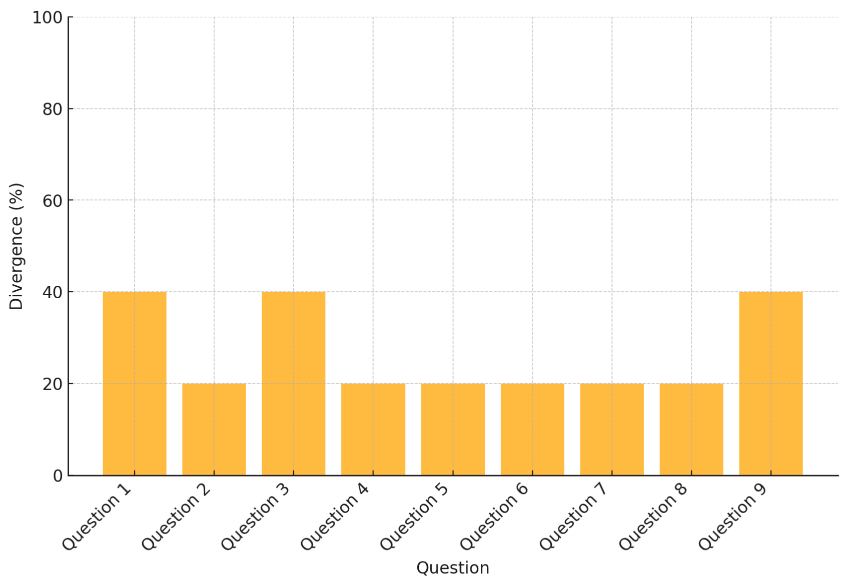

Figure 5.3: Divergences between GPT and Forms

Figure 5.3 presents the alignment between the user responses of Google Forms and the responses from the GPT language model. The rate of divergence between the two sources was analyzed in order to determine in which questions the model comes closest to or farthest from the views of investors.

It can be seen from the Figure that the divergence percentages vary across the different questions. Surprisingly, Questions 1, 3 and 9 have the highest divergence, which indicates that the answers of these questions by the GPT model are significantly different from the answers of human subjects. These types of questions can be linked to more qualitative aspects of investor profiling, i.e., risk perception, behavioral characteristics, or personalized investment strategies, where human experience and intuition play a significant role that the model cannot necessarily replicate.

In contrast, questions related to investment experience evaluation or portfolio selection produce lower divergence percentages, suggesting that the model has learned to recognize some patterns of investor decision-making. This itself suggests that, in some areas, human financial reasoning can be replicated by the GPT model with relative precision.

Finally, measuring the divergence between GPT and Google Forms responses is a key measure of the model's ability to identify investor profiles. The areas of greatest divergence are where adjustments can be made, such as by enhancing the training data with more accurate investor data or developing calibration tools to make the model's responses more closely align with the reality of how financial markets function. All of these advancements would allow for broader practical application of LLMs to investor profiling and credit risk assessment.

5.2 Discussion

This section discusses the results presented in this chapter, addressing the divergences between the GPT model in identifying investor profiles by comparing its classifications with responses obtained through Google Forms.

The analysis explores whether AI can replace, complement, or improve traditional profiling methods, focusing on key aspects such as investment horizon and risk tolerance.

5.2.1 Investment Time Horizon

The results indicate that GPT successfully classified most users into appropriate investment time horizons (Q1, Q2). Discrepancies arose in cases where investment intervals were contiguous (e.g., 5-7 years vs. 10 years), making it difficult to determine the most accurate classification.

These slight differences suggest that minor deviations may be due to interpretation gaps rather than errors. Additionally, non-professional investors might not have a rigid distinction between close time intervals, leading to natural ambiguities in responses.

This means that GPT has a general idea of the investment patterns, but the detail in the time frames can lead to minute discrepancies. The difference may also be due to the subjectivity of the user's decision-making, as investment withdrawal periods can vary based on individual preferences and financial circumstances.

5.2.2 Risk Tolerance and Portfolio Classification

GPT generally aligned well with user-indicated risk tolerance levels (Q3 - Q6), particularly in classifying highly risk-averse and aggressive investors. However, when responses were less explicit, GPT exhibited a conservative bias, often categorizing users into safer risk brackets.

This conservatism in classification appears to be a protective measure when uncertainty is present, rather than an outright misinterpretation. For instance, during the assessments, GPT sometimes placed users in lower-risk categories even when their responses indicated a higher tolerance for volatility. This highlights a tendency for the model to prioritize stability over aggressive investment behaviors unless explicitly stated otherwise.

In portfolio selection (Q4 - Q8), GPT demonstrated strong alignment for conservative portfolios, while slight discrepancies appeared in riskier portfolios, especially when portfolios present similar characteristics. It can also be argued that if the users chose more moderate answers in other questions, the GPT may have assumed that certain responses would be more in line with their investment strategy, justifying the adjustment of the answer. This suggests that the model's interpretation of risk-return trade-offs could be improved to better reflect investors' actual preferences.

5.2.3 Investment Adjustment Strategies

In Q9, GPT showed high accuracy when user responses were explicit and well-defined. However, when responses were more subjective, GPT tended to recommend more balanced or conservative investment strategies. This aligns with the previous observations that the model defaults to a lower-risk assessment when faced with uncertainty. A notable example is the case where a user selected 'Investment A' but GPT classified their response as 'Investment C,' which represents a riskier alternative.

This shift may have been caused by contextual interpretation across multiple responses rather than an isolated misclassification. Understanding these nuances can help refine how GPT processes individual risk preferences.

5.2.4 Understanding GPT Divergences

- **High Accuracy in Risk Assessment and Time Horizon**: the model is highly successful in categorizing users into appropriate investment horizons and risk profiles in most cases, indicating the potential of the model as a decision-support tool.

- **Conservatism of Uncertain Answers**: GPT always defaults to conservative classifications when there is uncertainty about user answers, which may be good in some financial contexts but may also misrepresent aggressive investors.
- **Portfolio Choice Nuances**: even though GPT is in line with conservative portfolios, occasionally it misclassifies riskier choices as well, suggesting that portfolio selection requires additional contextual information.
- **Subjective Responses Create Challenges**: GPT has difficulties when answers are not clearly detailed, affirming the necessity for organized, clearly defined input forms to maximize model precision.

The research suggests that GPT has significant potential as an investor profiling tool, particularly in highly structured decision-making situations. The most challenging issue appears to be the handling of subjective or imprecise responses, with the model choosing to arrive at conservative approximations. The accuracy could be improved by developments in training and response mapping, particularly where investor tastes remain undefined. Such results are contributing to the wider understanding of the role of AI in financial advisory services and in investor profiling activities.

Chapter 6

Conclusion

The aim of this research was to develop and validate an adaptive investor profile model based on LLMs to assess its viability for application in dynamic risk profile classification and establish its superiority over or complementarity with traditional static questionnaire methods. For this purpose, a solution was developed where more contextual and interactive responses were allowed, reducing the likelihood of classification errors due to uncertainties or incorrect answers. The results showed that the model was able to classify most of the investors accurately with high similarity between the GPT classifications and the profiles provided by conventional methods. In particular, there was high agreement in the investment horizon analysis and conservative portfolio classification.

Compared to traditional approaches, the results indicate that AI models have the potential to be more adaptable and contextual in the analysis of investor profiles, and can thus overcome some of the limitations of static questionnaires. Divergences existed in risk tolerance classification, especially between the model's classifications and user responses in open or subjective questions, where the model exhibited a tendency to make more conservative classifications. Although LLMs hold good promise for investor profiling, certain scenarios revealed divergences between the model's classification and user feedback. In particular, inconsistencies seemed to occur in risk tolerance classification since the model tended to categorize more conservative profiles for open-ended or subjective responses. Also, in portfolio selection, certain investors preferring higher-risk investments were classified with more moderate profiles, alluding to wariness in uncertain cases. These inconsistencies point out areas that have to be improved before LLMs can effectively be integrated into the risk-assessment systems of finance.

The constraints and challenges of using LLMs in investor profile categorization were also explored, namely the impact of bias in response interpretation and the inability to convert qualitative and subjective preferences into objective classification. These findings highlight the importance of refining prompt engineering techniques to improve the model's accuracy and contextual understanding, reducing the reliance on extensive human intervention for monitoring and calibration. By optimizing prompts, AI could serve as an effective complement to, or in some cases a replacement for, conventional financial analysis.

As future work, we plan to refine interpretation of subjective responses, reduce risk classification biases in order to improve the accuracy of the model. Exploring reinforcement learning with expert feedback would also help further calibrate the model to actual investor behavior. Further advances in this direction could further enable LLMs to adapt better to risk profiling, and thus become even more dynamic, flexible, and accessible, and result in more efficient and customized financial advice.

References

[Alpaydin, 2020] Alpaydin, E. (2020). *Introduction to machine learning*. MIT press. [Cited on page 10]

[Baz et al., 2020] Baz, J., Davis, J., Sapra, S., Gillmann, N., and Tsai, J. (2020). A framework for constructing equity-risk-mitigation portfolios. *Financial Analysts Journal*, 76(3):81–98. [Cited on page 11]

[Blanchard et al., 2011] Blanchard, D. C., Griebel, G., Pobbe, R., and Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience & Biobehavioral Reviews*, 35(4):991–998. [Cited on page 28]

[Canabarro, 2024] Canabarro, E. (2024). Model risk management in stress testing: The road up to here. *Journal of Risk Management in Financial Institutions*, 17(4):370–382. [Cited on page 11]

[Cao et al., 2024] Cao, Y., Chen, Z., Pei, Q., Dimino, F., Ausiello, L., Kumar, P., Subbalakshmi, K., and Ndiaye, P. M. (2024). Risklabs: Predicting financial risk using large language model based on multi-sources data. *arXiv preprint arXiv:2404.07452*. [Cited on page 16]

[Dai et al., 2021] Dai, Z., Yuchen, Z., Li, A., and Qian, G. (2021). The application of machine learning in bank credit rating prediction and risk assessment. In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pages 986–989. IEEE. [Cited on page 17]

[de Zarzà et al., 2023] de Zarzà, I., de Curtò, J., Roig, G., and Calafate, C. T. (2023). Optimized financial planning: integrating individual and cooperative budgeting models with llm recommendations. *AI*, 5(1):91–114. [Cited on page 18]

[Enkhsaikhan and Jo, 2024] Enkhsaikhan, B. and Jo, O. (2024). Risk-averse reinforcement learning for portfolio optimization. *ICT Express*. [Cited on page 11]

[Feng et al., 2023] Feng, D., Dai, Y., Huang, J., Zhang, Y., Xie, Q., Han, W., Chen, Z., Lopez-Lira, A., and Wang, H. (2023). Empowering many, biasing a few: Generalist credit scoring through large language models. *arXiv preprint arXiv:2310.00566*. [Cited on pages 1, 2, and 17]

[Ghojogh and Ghodsi, 2020] Ghojogh, B. and Ghodsi, A. (2020). Attention mechanism, transformers, bert, and gpt: tutorial and survey. [Cited on page 13]

[Hens and Nordlie, 2024] Hens, T. and Nordlie, T. (2024). How good are llms in risk profiling? *Swiss Finance Institute Research Paper*, (24-30). [Cited on pages 1, 2, 3, and 16]

[Jasrotia et al., 2020] Jasrotia, S. S., Mishra, H. G., and Sharma, R. L. (2020). Capital adequacy norms: banks compliance with basel-iii norms. *International Journal of Electronic Banking*, 2(1):16–37. [Cited on page 12]

[Kurochkin et al., 2024] Kurochkin, S. V., Pavlov, N., Tkachenko, M., and Yaremich, E. A. (2024). Machine learning methods in investor risk profiling. *AlterEconomics*, 21(3):527–552. [Cited on page 14]

[Lakkaraju et al., 2023] Lakkaraju, K., Jones, S. E., Vuruma, S. K. R., Pallagani, V., Muppasani, B. C., and Srivastava, B. (2023). Llms for financial advisement: A fairness and efficacy study in personal decision making. In *Proceedings of the Fourth ACM International Conference on AI in Finance*, pages 100–107. [Cited on page 18]

[Lhoest et al., 2021] Lhoest, Q., Del Moral, A. V., Jernite, Y., Thakur, A., Von Platen, P., Patil, S., Chaumond, J., Drame, M., Plu, J., Tunstall, L., et al. (2021). Datasets: A community library for natural language processing. *arXiv preprint arXiv:2109.02846*. [Cited on page 12]

[Li et al., 2024] Li, H., Cao, Y., Yu, Y., Javaji, S. R., Deng, Z., He, Y., Jiang, Y., Zhu, Z., Subbalakshmi, K., Xiong, G., et al. (2024). Investorbench: A benchmark for financial decision-making tasks with llm-based agent. *arXiv preprint arXiv:2412.18174*. [Cited on page 19]

[Mazzoli and Palmucci, 2023] Mazzoli, C. and Palmucci, F. (2023). Reconciling self-assessed with psychometric risk tolerance: A new framework for profiling risk among investors. *Journal of Behavioral Finance*, pages 1–14. [Cited on pages 1, 11, and 30]

[Mhlanga, 2021] Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International journal of financial studies*, 9(3):39. [Cited on page 17]

[Nagpal, 2024] Nagpal, K. M. (2024). Portfolio stress testing and value at risk (var) incorporating current market conditions. *arXiv preprint arXiv:2409.18970*. [Cited on pages 11 and 28]

[Nie et al., 2024] Nie, Y., Kong, Y., Dong, X., Mulvey, J. M., Poor, H. V., Wen, Q., and Zohren, S. (2024). A survey of large language models for financial applications: Progress, prospects and challenges. *arXiv preprint arXiv:2406.11903*. [Cited on page 2]

[Rabbani and Nobre, 2022] Rabbani, A. G. and Nobre, L. H. N. (2022). Financial risk tolerance. [Cited on pages 1 and 29]

[Roberts, 2024] Roberts, J. (2024). Do large language models learn to human-like learn? In *Proceedings of the AAAI Symposium Series*, volume 3, pages 588–591. [Cited on page 12]

[Strong et al., 2009] Strong, R. A., Steiger, N. M., and Wilson, J. R. (2009). Introduction to financial risk assessment using monte carlo simulation. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 99–118. IEEE. [Cited on page 11]

[van de Poll et al.,] van de Poll, J., Yong, Y., and van de Weijer, N. Validity of ai-generated one-off questionnaires for organizational transformation. [Cited on pages 14 and 28]

[Wang et al., 2023] Wang, Y., Jiang, Z., Chen, Z., Yang, F., Zhou, Y., Cho, E., Fan, X., Huang, X., Lu, Y., and Yang, Y. (2023). Recmind: Large language model powered agent for recommendation. *arXiv preprint arXiv:2308.14296*. [Cited on pages 2 and 18]

[Zhao et al., 2024] Zhao, H., Liu, Z., Wu, Z., Li, Y., Yang, T., Shu, P., Xu, S., Dai, H., Zhao, L., Mai, G., et al. (2024). Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*. [Cited on pages 11 and 19]

[Zhou et al., 2024] Zhou, Y., Ni, Y., Liu, X., Zhang, J., Liu, S., Ye, G., and Chai, H. (2024). Are large language models rational investors? *arXiv preprint arXiv:2402.12713*. [Cited on page 19]