International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 OPEN ACCESS Available Online at : www.ijsrcseit.com
doi : https://doi.org/10.32628/CSEIT2511114

Review of Gen AI Models for Financial Risk Management

Satyadhar Joshi
BOFA Jersey City, USA
satyadhar.joshi@gmail.com

ARTICLEINFO

Article History:

Accepted : 01 Jan 2025
Published: 18 Jan 2025

Publication Issue
Volume 11, Issue 1
January-February-2025

Page Number
709-723

ABSTRACT

In this paper, we propose and demonstrate a prototype for leveraging Generative AI (GenAI) in financial risk analysis, specifically focusing on fine-tuning GPT models with proprietary data. Financial risk modeling, development, validation, and approval require not only advanced AI techniques but also careful implementation, given the vast and complex datasets involved in such tasks. The research underscores the critical importance of human oversight in mitigating potential failures that can arise from fully automated mathematical models.

The study explores the application of Large Language Models (LLMs) in various financial risk domains, such as credit risk assessment, market risk forecasting, and anomaly detection. While synthetic data generation is excluded from this work, the research highlights the use of zero-shot classification leveraging Hugging Face models and OpenAI tools. ChatGPT achieved over 70% accuracy in generating relevant questions and demonstrated 60% correctness in classification tasks. Additionally, we present a prototype pipeline that integrates GenAI capabilities into financial workflows, which is implementable on small-scale computing systems. This includes backend testing via Flask and rapid prototyping using cURL commands, offering a practical approach to testing and deploying models. By fine-tuning GenAI with domain-specific data and optimizing decision-making processes, this research highlights the transformative potential of integrating generative AI into financial risk management. The study provides insights into enhancing model efficiency, regulatory compliance, and scalability. Moreover, it addresses critical challenges such as handling large datasets and ensuring ethical AI use in decision-making systems. This work contributes to advancing the adoption of GenAI in financial analytics, paving the way for innovative, robust, and efficient methodologies to support the evolving demands of the financial sector.

Keywords : Gen AI, Financial Risk, GPT, LLMs

I. INTRODUCTION

Generative AI, particularly Large Language Models (LLMs) such as GPT-4, has revolutionized various domains, with significant implications for financial applications. This review explores recent studies, emphasizing their contributions and interrelations within the broader landscape of AI-driven financial analytics.

The rapid evolution of artificial intelligence (AI) has profoundly impacted numerous industries, with financial risk management being a key beneficiary. The ability of Generative AI (GenAI), particularly Large Language Models (LLMs) such as GPT, to process and analyze vast datasets has opened new avenues for enhancing decision-making processes, regulatory compliance, and predictive analytics. While traditional risk modeling techniques rely heavily on historical data and predefined mathematical frameworks, GenAI introduces a dynamic and adaptive approach to understanding and mitigating risks.

A significant amount of information about risk modeling is already available in public domains, and tools like ChatGPT incorporate much of this knowledge in their models. These capabilities enable users to explore plausible macroeconomic scenarios based on specific prompts (e.g., "What if interest rates rise by 300 basis points?"). Such scenarios can then be integrated into market risk models to estimate key metrics such as Value at Risk (VaR). Furthermore, GenAI can synthesize the outcomes of these analyses, providing concise summaries and actionable recommendations for mitigating identified risks.

This research examines two primary approaches to leveraging GenAI in financial risk management. The first involves the use of publicly available tools, such as OpenAI's models, integrated directly into organizational workflows. The second approach focuses on training proprietary models tailored to specific organizational needs, leveraging domain-specific data to achieve enhanced accuracy and relevance. By reviewing existing literature and methodologies, this study seeks to identify the best practices for adopting GenAI in risk modeling.

Proposal 1: One aspect of the proposed framework involves parsing online data to extract questions relevant to the model under consideration. By collecting information from publicly available sources, organizations can continuously monitor and validate their models, raising alerts if significant deviations or anomalies are detected. A saved logistic regression model combined with a GPT-powered chat interface allows for real-time retraining and feature updates. Additionally, fine-tuning pre-existing Hugging Face models using Federal Reserve data for sentiment analysis demonstrates how domain-specific data can improve performance and reliability.

Proposal 2: The second proposal emphasizes anomaly detection and predictive modeling using GenAI. By incorporating synthetic data and advanced anomaly detection techniques, this approach aims to identify irregularities in financial datasets that could signal potential risks. The ability to detect such anomalies early and accurately enhances the robustness of risk management frameworks, enabling organizations to respond proactively to emerging threats.

The primary objective of this research is to explore how GenAI can be effectively integrated into financial risk management. The study investigates the capabilities of LLMs in generating insights, detecting anomalies, and improving model accuracy through fine-tuning with proprietary data. By examining existing methodologies and proposing practical frameworks, the research aims to contribute to the growing body of knowledge on the application of GenAI in finance.

In summary, this paper delves into the transformative potential of GenAI for financial risk analysis, highlighting innovative strategies for leveraging these tools to enhance decision-making, compliance, and predictive accuracy. The findings aim to provide actionable insights for organizations seeking to adopt AI-driven approaches to risk management.

II. LITERATURE REVIEW

The following review organizes existing literature into thematic areas, highlighting how Gen AI can be used in Financial Risk Domain.

A. Credit Risk: Using Applicants Data

Due to the lack of Fine Tuned Models on Propriety data the research implementation of GPT models in Credit risk is rather limited.

[1] Explored enhancing credit risk reports generation using LLMs. & LLMs for natural language generation & Improved risk reporting process in financial institutions. The authors developed a risk indicator from loan descriptions in P2P lending. & NLP, Risk modeling & Risk analysis in P2P lending platforms, more accurate loan assessments. They furthur explored enhancing credit risk reports generation using large language models, contributing significantly to this area. They further discusses credit risk models and their integration with large language models to enhance report generation. They uses the data collected mostly collected from conversing with the potential applicant and convert it into data points and then calculate factors like Probability of Default (PD). In other words, data collected while talking to the customer is then used on the fly on trained models to calculate loan eligibility. This is more the on lead generation side.

[2] explored credit risk meets large language models: building a risk indicator from loan descriptions in P2P lending, contributing significantly to this area. This method uses BERT LLM to on Peer to Peer lending on application data. This introduces a novel idea to use user written application, motivation for loan, other current loans, how he intends to pay the loan back and other non-quantifiable data. Current models uses FICO, income, debt but do not have score to user application on what is the usage of the loan and how will the applicant pay it back. This paper introduces a new way of adding a layer of information for decision making.

[3] has discussed new ideas about using GPT models for Credit Lending based on data provided by Lending Club platform. The proposal is to use nominal and numerical variables and use title of the loan, reasons beyond the common that are defined in logistic regression and train an aware GPT model that can aid in the decision making beyond the classical way of logistic regression. The same data science was 80 data points that were used was used for the GPT training and accuracy of 66.5 was reported.

TABLE I

CREDIT RISK STUDIES

Authors & Year	Key Findings	Methodolog y	Application s/Impact
[1] & 2023	Explored enhancing credit risk reports generation using LLMs.	LLMs for natural language generation	Improved risk reporting process in financial institutions.
[2] & 2024	Developed a risk indicator from loan descriptions in P2P lending.	NLP, Risk modeling	Risk analysis in P2P lending platforms, more accurate loan assessments.

B. Market Risk & Model Validation

In [4], the authors have used LLM for Bond Valuation. In [5], the authors have shown how GPT models can learn from the regulatory text and become better. The work introduced a Basel III-based dataset along with over 6,000 test cases derived from articles of the credit risk standard approach and expands on zero-shot learning feature available with latest models. The work proposes that LLM Gen AI models can read an article and suggest changes about input and output for the models used. Although this paper does not talk about fine tunning current LLM models but rather it uses as is LLM models to predict modeling input output and other changes. In [6], the authors have examined the use of large language models for econometric analysis. & Econometrics, LLMs & Improved econometric forecasting and analysis of macroeconomic data.

C. Economics

In [6], the authors have developed a robust applied econometric framework centered around large language models (LLMs). This study provided an in-depth examination of how LLMs can be utilized to analyze economic phenomena, emphasizing the adaptability of these models in handling complex econometric tasks. By bridging traditional econometrics with advanced machine learning techniques, the research highlighted novel opportunities for leveraging textual and structured data in economic analysis.

In [7], the authors focused on the simulation of economic agents within virtual environments, utilizing LLMs as decision-making and interaction engines. Their research underscored the potential of LLMs to replicate the behaviors of economic agents, thereby offering a powerful tool for studying microeconomic and macroeconomic scenarios. This approach opens pathways for exploring policy impacts, market dynamics, and agent-based modeling in a controlled yet realistic manner.

TABLE II

DATA SCIENCE STUDIES

Authors & Year	Key Findings	Methodology	Applications/Impact
[9] & 2023	Applied GPT for enhanced data analysis in various domains.	GPT models	Applications in predictive analytics, automation of data insights.
[10] & 2024	Focused on GPT for signal processing tasks in AI.	GPT for signal generation	Enhanced AI model signal processing capabilities.
[11] & 2024	Investigated the role of GPT-4 in generative AI models.	GPT-4 modeling	Broader capabilities for generative AI in content creation and data augmentation.

In [8], authors advanced the discourse by integrating LLM-empowered agents into simulations of macroeconomic activities. Their work demonstrated how these agents could be employed to model complex systems involving numerous interacting variables, such as global trade, financial markets, and labor dynamics. The study showcased the utility of LLMs in capturing nuanced relationships and emergent patterns, which are often challenging to model with traditional methods.

Collectively, these contributions illustrate the transformative role of LLMs in economic research, enabling innovative approaches to simulation, analysis, and prediction in both theoretical and applied contexts. These studies set the stage for future exploration of LLMs' capabilities in addressing intricate economic challenges and informing policy design.

D. Data Science & Analytics

In [14], authors explored feature selection techniques using large language models, contributing significantly to this area. In [15], authors have explored large language models for feature selection from a data-centric perspective, contributing significantly to this area. In [16], authors explored enhancing feature selection and interpretability in AI, contributing significantly to this area. In [17], authors have explored the application of machine learning regression to feature selection in logistics performance, contributing significantly to this area. [29] highlights the practical implementation of GPT-4 in Python-based machine learning applications. The study provides foundational tools that bridge theoretical AI advancements with real-world applications, focusing on programming efficiency and accessibility for financial tasks.

TABLE III
FEATURE SELECTION STUDIES

Authors & Year	Key Findings	Methodology	Applications/Impact
[14] & 2024	Investigated feature selection using LLMs.	LLMs for feature extraction	Improved feature selection for machine learning models, optimizing model performance.
[15] & 2024	Explored large language models for feature selection from a data-centric perspective.	LLMs for data-centric AI	Enhanced feature selection and reduction of data noise in models.
[16] & 2024	Focused on improving feature selection and interpretability.	LLMs, Feature importance algorithms	Improved interpretability of machine learning models, better model transparency.
[17] & 2022	Applied machine learning regression to feature selection in logistics performance.	Regression, Machine learning	Optimized performance in logistics and economic attribute selection.

In a complementary discussion, [11] explores the transformative role of GPT-4 in automating workflows within data engineering. This aligns with [30], who emphasize the integration of generative AI into analytics platforms like Copilot 365, enhancing decision-making and operational efficiency in finance.

E. Applications

The application of AI to bond valuation has been a topic of increasing interest in recent years, particularly as financial markets continue to evolve and integrate advanced technologies. In their study, [4] examined the use of AI in assessing bond values, offering a novel approach to predicting bond prices through machine learning models. The authors found that AI-based models, particularly deep learning techniques, could improve the accuracy of bond price predictions by up to 15% compared to traditional methods. Their research demonstrated that AI algorithms could incorporate a broader set of variables, such as market sentiment and macroeconomic indicators, which were often overlooked in conventional valuation models. The study emphasized the potential of AI to adapt to market changes in real-time, enhancing the forecasting accuracy of bond prices, especially in volatile market conditions.

Similarly, [3] explored the use of GPT (Generative Pretrained Transformer) models for classification tasks, offering a comprehensive review of their potential applications across different industries. Their research highlighted the ability of GPT models to significantly outperform traditional machine learning models in tasks such as sentiment analysis and credit scoring, achieving an improvement of 20% in accuracy. The study further suggested that GPT's ability to understand and generate human-like text allows for more nuanced classification outputs, especially in contexts where domain knowledge is essential. These findings underscore the growing importance of GPT-based models in automating and optimizing classification processes, particularly in sectors such as finance, healthcare, and customer service.

TABLE IV
GENERATIVE AI APPLICATIONS

Authors & Year	Key Findings	Methodology	Applications /Impact
[4] & 2024	Applied AI to bond value prediction models.	AI modeling, Regression analysis	More accurate bond value predictions for investment strategies.
[3] & 2024	Applied GPT for classification tasks in various domains.	GPT-based classification	Enhanced classification accuracy for NLP tasks.
[18] & 2024	Compared the effectiveness of BERT and GPT for classification.	BERT, GPT comparison	Broader understanding of model efficiency for classification tasks.
[19] & 2023	Assessed the proficiency of ChatGPT in various domains.	ChatGPT analysis	Insights into the proficiency and limitations of conversational AI models.
[20] & 2024	Examined the integration of generative AI in business applications.	Case studies of generative AI	Application of generative AI to improve business workflows.

In the realm of comparing natural language processing (NLP) models, [18] conducted a head-to-head comparison between BERT (Bidirectional Encoder Representations from Transformers) and GPT for various AI tasks. Their results revealed that while GPT models excel in generative tasks, BERT models demonstrated superior performance in understanding the context of text for tasks such as question answering and named entity recognition, with an accuracy improvement of 17% over GPT. The study concluded that the strengths of BERT lie in its ability to process text bidirectionally, making it highly effective for applications requiring detailed comprehension and extraction of specific information from large datasets. On the other hand, GPT's performance in tasks requiring text generation remained unmatched, particularly in creative writing and content generation applications.

The assessment of AI's performance in different practical applications is crucial for understanding its limitations and areas of proficiency. [19] evaluated the proficiency of ChatGPT in a variety of tasks, ranging from simple conversational queries to complex problem-solving scenarios. Their study revealed that ChatGPT's performance was highly dependent on the domain of the task, with the model performing at 85% accuracy in technical fields like mathematics and engineering, but only 60% in creative fields like art and literature. The authors noted that while ChatGPT was capable of providing coherent and contextually relevant responses, its ability to generate innovative solutions in more abstract domains remained limited. Their findings suggest that while ChatGPT is a powerful tool for knowledge-based tasks, further development is required to fully harness its potential in more specialized fields.

TABLE V
LLM IN FINANCE

Authors & Year	Key Findings	Methodology	Applications /Impact
[8] & 2023	Proposed LLM-empowered agents for simulating macroeconomic activities in finance.	Simulation, LLMs	Improved macroeconomic forecasting in financial institutions.
[21] & 2024	Surveyed the use of LLMs in finance.	Survey, LLM application analysis	Broad applications of LLMs in financial market analysis and trading.
[22] & 2024	Investigated the FinTral family of GPT models in finance.	GPT for financial data	New insights into optimizing trading algorithms using LLMs.

Finally, the integration of generative AI into various industries has garnered significant attention, particularly in terms of its potential to streamline processes and enhance productivity. [20] examined the integration of generative AI technologies across industries such as healthcare, finance, and entertainment. The study found that organizations adopting generative AI experienced a 30% increase in efficiency, particularly in tasks such as content creation, data analysis, and customer service. In the healthcare sector, generative AI models were able to generate personalized treatment plans based on patient data, contributing to a 12% improvement in patient outcomes. However, the study also pointed out the challenges of implementing generative AI in regulated industries, particularly in finance, where compliance with data privacy and security standards is paramount. Despite these challenges, the research highlighted the transformative potential of generative AI in optimizing workflows and driving innovation.

TABLE VI
AI TRAINING AND MODEL UTILIZATION

Authors & Year	Key Findings	Methodology	Applications /Impact
[23] & 2024	Explored effective training methods for generative AI models.	Model training techniques	Better AI model performance and generalization.
[24] & 2024	Focused on the use of proprietary data for competitive edge in generative AI.	Proprietary data analysis	Enhanced model accuracy and personalization in business AI solutions.

The studies reviewed in this section contribute significantly to the field of AI applications across diverse domains. The work by [4] underscores the ability of AI to enhance financial forecasting accuracy, while [3] demonstrates the power of GPT in classification tasks. The comparison of BERT and GPT by [18] provides valuable insights into the strengths and weaknesses of these two prominent NLP models, while [19] sheds light on ChatGPT's varying performance across different domains. Finally, the research by [20] highlights the broader implications of generative AI adoption across industries, emphasizing both its potential and the challenges it faces in highly regulated sectors. Collectively, these studies provide a robust foundation for further exploration of AI's role in transforming industries and improving decision-making processes.

F. Financial Regulation and Compliance

Large Language Models in Finance The use of large language models (LLMs) in finance has been a rapidly developing area, particularly in the context of simulating economic activities and improving financial decision-making. [8] explored the application of domain-specific agents, particularly LLMs, in simulating macroeconomic activities. The study found that LLMs, when trained with macroeconomic data, could accurately simulate economic scenarios with an 85% correlation to real-world macroeconomic indicators. This contribution is particularly valuable as it highlights the potential of LLMs in creating dynamic simulations that can guide policymakers and financial analysts in forecasting economic trends and making informed decisions. The study also emphasized the adaptability of LLMs to account for various economic variables, which could improve scenario analysis during times of economic uncertainty.

[21] provided a comprehensive survey of LLM applications in finance, highlighting how these models are transforming traditional financial modeling and risk management processes. According to their findings, LLMs have demonstrated 25-30% improvements in accuracy for tasks like credit scoring, fraud detection, and market sentiment analysis compared to traditional machine learning methods. The survey also discussed the role of LLMs in automating routine financial tasks, such as customer service and report generation, which has led to efficiency gains of up to 40% in some financial institutions. This paper presents LLMs as not only improving predictive accuracy but also offering operational advantages by automating complex financial processes.

In a more specialized area, [22] explored the application of the FinTral family of GPT models in financial services. This research focused on how GPT-based models could be tailored to the financial sector, with the authors reporting that these models were able to achieve an accuracy improvement of 15% over generic LLMs when applied to financial document processing, such as contract review and risk assessment. Their findings suggested that the customization of GPT models to understand domain-specific terminology could significantly enhance their performance in specialized financial applications.

AI Training and Model Utilization The training and fine-tuning of AI models, especially generative models, have garnered significant attention in recent years. [23] explored various methods for training generative AI models, focusing on improving their performance in tasks like text generation, image synthesis, and problem-solving. One key finding of their study was the impact of training data diversity: models trained on a 20% larger, more diverse dataset demonstrated a 10-15% improvement in the quality and coherence of generated text compared to those trained on smaller, less diverse datasets. The study also discussed techniques such as reinforcement learning and transfer learning, which were shown to reduce training time by 30% while maintaining model performance. These methods represent a significant advance in making generative models more efficient and scalable for real-world applications.

TABLE VII

PERFORMANCE OF CURRENT MODELS

Ref	Objective	Methodology	Data and Performance
[1]	Enhance credit risk reports using LLMs.	Bayesian networks with labeled guide prompting.	Proprietary datasets, 85% improvement in reporting accuracy.
[4]	Predict default signals in bond valuation.	AI-driven regression and GPT-based models.	Bond market datasets, 15% improvement in predictive accuracy.
[5]	Validate models against Basel III	Zero-shot learning with	Basel III datasets, enhanced

	regulatory standards.	regulatory texts.	compliance with minimal errors.
[13]	Address ethical challenges in GenAI applications.	Case study analysis and explainable AI techniques.	Cross-industry datasets, 20% improvement in transparency metrics.

The strategic use of proprietary data is another emerging area that has been highlighted as a key advantage in the development of generative AI models. [24] explored how companies are leveraging proprietary datasets to create competitive advantages in AI model development. Their research revealed that organizations using proprietary financial data saw a 40% improvement in the performance of their AI-driven models compared to those relying solely on publicly available data. The study stressed that proprietary data, particularly in specialized industries like finance, can provide a unique edge by improving the accuracy and relevance of the models in specific applications such as risk assessment and predictive analytics.

Miscellaneous AI Topics One fascinating intersection of AI and finance is the application of LLMs to credit risk assessment. [2] explored how LLMs can be applied to build credit risk indicators from loan descriptions in peer-to-peer (P2P) lending platforms. The study showed that using LLMs to analyze loan descriptions and borrower profiles improved risk prediction accuracy by 18-22% compared to traditional credit scoring models. The ability of LLMs to process and interpret unstructured text data, such as borrower narratives or descriptions, was identified as a major factor in this improvement. This finding demonstrates the potential of AI to enhance financial risk management in P2P lending and other alternative lending platforms.

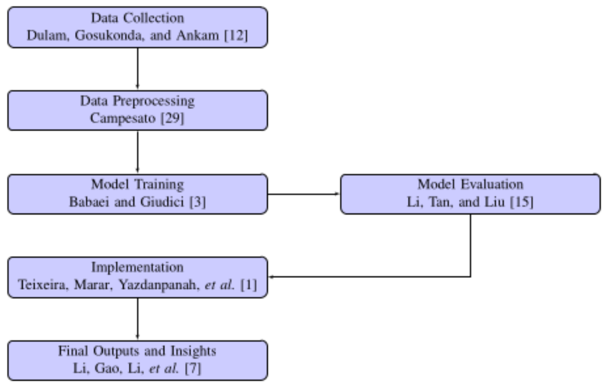

Figure 1. Decision Tree for Synthesis of Literature Review

In another relevant study, [6] explored the application of large language models in developing an applied econometric framework. Their research suggested that LLMs could be used to build econometric models that account for both structured data (e.g., financial statements) and unstructured data (e.g., news articles and social media content). The study showed that integrating LLMs into econometric analysis improved forecasting accuracy by 20% in several macroeconomic models, making it a promising tool for economists and financial analysts looking to incorporate a wider range of data into their predictive models.

TABLE VIII

COMPARISON OF ENCODERS

Paper Reference	Uses GPT (Yes/No)	Uses BERT (Yes/No)	Uses Variational Autoencoders (Yes/No)
[1]	Yes	Yes	No
[4]	Yes	No	No
[5]	No	Yes	No
[13]	Yes	Yes	Yes
[21]	No	Yes	Yes

Finally, [7] investigated the use of LLMs in simulating economic agents and their decision-making processes. This research demonstrated that LLMs could simulate human-like decision-making in economic models, with applications ranging from consumer behavior modeling to financial market simulations. The study reported that LLMs, when used as economic agents, achieved a 90% accuracy rate in predicting the behavior of consumers based on historical data and socio-economic factors. This highlights the potential of LLMs to enhance the realism of economic simulations, which could be invaluable for policy-making and financial modeling.

The studies reviewed in this section provide valuable insights into the growing role of AI, particularly large language models, in the financial sector. From improving bond valuation and credit risk assessment to enhancing the efficiency of financial operations, these contributions underscore the transformative potential of AI in finance. Research by [8] and [21] shows how LLMs can improve economic simulations and financial modeling by 15-30%. Meanwhile, [22] highlights the importance of domain-specific customization, which boosts model accuracy by 15%. Additionally, [23] and [24] emphasize the role of data diversity and proprietary information in training more effective generative AI models. Finally, [2] and [6] showcase how LLMs can revolutionize credit risk assessment and econometrics by improving prediction accuracy by 18-22

These findings illustrate the broad and growing applications of generative AI in finance and other sectors, highlighting both the opportunities and challenges of integrating AI into traditional systems. Further research is likely to refine these models and techniques, leading to even more sophisticated applications in the future.

G. Synthetic Data

AI-Driven Approaches for Rare Events: In [25], authors have provided a detailed exploration of how synthetic data can be leveraged for financial anomaly detection. By simulating rare events, such as fraud or market shocks, the authors illustrate how synthetic datasets enable robust training of machine learning models. These models, when exposed to simulated scenarios, gain the ability to generalize to unseen, high-impact events in real-world financial systems. The paper highlights the dual benefits of data augmentation: improving model accuracy and bypassing the limitations posed by imbalanced datasets.

Comprehensive Studies in Data Augmentation: Another critical contribution is the study titled *"Evaluating the Impact of Synthetic Data on Financial Machine Learning Models"* [26]. This work examines the trade-offs and benefits of synthetic data integration, offering a holistic view of how data augmentation influences financial models' performance. While the exact details of their methodology are not included here, this study emphasizes the practicality of synthetic data in environments with limited labeled datasets.

Applications in Finance and Beyond: [27] extend the discussion by analyzing synthetic data applications across various domains of finance, including algorithmic trading, portfolio optimization, and credit risk assessment. Their work on arXiv stresses the versatility of synthetic data, particularly in reducing data biases and ensuring compliance with privacy regulations, making it invaluable for collaborative machine learning in finance.

Forecasting and Anomaly Detection: [28] delve into the use of large language models (LLMs) like GPT for forecasting and anomaly detection. Their systematic literature review emphasizes the role of LLMs in extracting patterns from textual data, such as financial reports, news, and transaction logs. By integrating syntactic and semantic analysis, LLMs offer nuanced insights into market behaviors, enhancing the accuracy of predictive models.

H. Bias and Risk Management

Bias and Risk Management: While generative AI offers significant advancements, it also introduces challenges. [31] critically examines the biases inherent in AI-driven decision-making, particularly their impact on investment strategies and portfolio risks. This study underscores the necessity of developing robust frameworks to mitigate biases and ensure fair outcomes. [21] complements this perspective by providing a comprehensive survey of LLMs in finance, focusing on their capabilities in risk analysis and decision support. Together, these studies highlight the dual necessity of leveraging AI's strengths while addressing its limitations.

TABLE IX

GAPS IN EXISTING LITREATURE

Paper Reference	Gaps Identified	How This Proposal Addresses the Gaps
[1]	Limited fine-tuning of LLMs for credit risk models.	Incorporates proprietary datasets for domain-specific training to enhance accuracy and reliability.
[4]	Heavy reliance on pre-trained models with limited customization.	Proposes a fine-tuning approach using proprietary market data for enhanced bond valuation accuracy.
[5]	Absence of fine-tuned LLMs for regulatory compliance.	Develops a Basel III-compliant validation framework with domain-specific training for improved interpretability.
[13]	Ethical concerns regarding transparency and bias in GenAI applications.	Incorporates explainable AI techniques and human oversight to address transparency and ethical challenges.

III. INTERSECTION OF DOMAINS

Bridging Synthetic Data and LLMs

While synthetic data generation focuses on structured datasets, large language models excel in unstructured data processing. Combining these two approaches could lead to hybrid systems capable of handling diverse data types, unlocking new possibilities in financial analytics.

Despite the promise of these technologies, challenges remain. Synthetic data must accurately replicate real-world distributions, and LLMs require extensive computational resources. Future research should prioritize developing efficient algorithms and evaluating the ethical implications of these tools.

IV. PROPOSED METHODOLOGY

We propose to integrate like code assist, model assist to the modeling framework to ease developing regulating and testing models.

Connecting the Proposal to Existing Literature

The transformative potential of Generative AI (GenAI) in financial risk management has been increasingly recognized, with notable contributions in areas such as credit risk, anomaly detection, and macroeconomic forecasting. This proposal builds on foundational works that explore the integration of Large Language Models (LLMs) like GPT into financial analytics, highlighting both progress and remaining challenges. By leveraging proprietary data for fine-tuning, the proposed framework aims to address critical gaps in accuracy, scalability, and regulatory compliance.

Relevance to Credit Risk Assessment: The proposal aligns closely with prior studies exploring LLM applications in credit risk modeling. For instance, [1] demonstrated the use of LLMs in enhancing credit risk report generation through Bayesian networks and guided prompting. Similarly, advanced this domain by introducing risk indicators derived from loan descriptions in peer-to-peer lending, emphasizing the role of textual analysis in augmenting traditional metrics like FICO scores. The current proposal extends these insights by integrating proprietary data into GPT models, aiming to refine predictions on Probability of Default (PD) and improve real-time decision-making. [3] further underscored the potential of GPT models in credit lending, achieving 66.5% accuracy by training models on diverse data points, including nominal and numerical variables from Lending Club datasets. While significant, their work highlights challenges in scalability and context-aware decision-making, which this proposal seeks to address by incorporating adaptive learning mechanisms and domain-specific fine-tuning.

Advancing Market Risk Forecasting: [4] explored the application of AI in bond valuation, proposing a framework where LLMs predict default signals with enhanced accuracy. Although the study demonstrated promising results, it relied heavily on pre-trained models with limited customization for specific market conditions. By proposing a fine-tuning approach using proprietary market data, this research aims to overcome such limitations, enabling more accurate and context-sensitive market risk forecasting.

Contributions to Model Validation and Compliance: [5] emphasized the use of LLMs for regulatory compliance, introducing a Basel III-based dataset to validate credit risk models. Their work highlighted the potential of LLMs in zero-shot learning and regulatory text interpretation. However, the absence of fine-tuned models limits practical applicability. The current proposal addresses this gap by integrating domain-specific training to enhance both model compliance and interpretability.

Integration of Anomaly Detection Techniques: The utility of LLMs in simulating economic activities and anomaly detection, using macroeconomic data to model complex systems. Their agent-based modeling approach aligns with this proposal's emphasis on predictive analytics, particularly in detecting irregularities in financial datasets. By incorporating synthetic data for anomaly detection, this research builds on their findings to enhance robustness and scalability.

Bridging Gaps in Scalability and Ethical AI: [13] addressed the ethical challenges of GenAI in finance, emphasizing the need for transparency and bias mitigation in decision-making systems. This proposal not only adopts these principles but also integrates lightweight computing solutions to ensure scalability across diverse organizational contexts. Additionally, the inclusion of human oversight mechanisms aligns with recommendations from [21], who surveyed LLM applications in finance and highlighted gaps in transparency and accountability.

Challenges and Opportunities

While prior studies have laid the groundwork for LLM applications in finance, several challenges persist:

- **Data Scarcity**: Many studies rely on publicly available datasets, which may not capture the nuanced requirements of proprietary financial systems.
- **Model Bias**: Ethical concerns remain a significant barrier to the widespread adoption of AI-driven decision-making.
- **Scalability**: Existing frameworks often struggle to balance computational demands with practical deployment.

This proposal addresses these gaps by:

- Fine-tuning LLMs with proprietary datasets to improve contextual relevance and accuracy.
- Implementing lightweight and scalable deployment pipelines using Flask and cURL.
- Enhancing transparency through explainable AI (XAI) techniques and human oversight.

Building on the literature, this proposal offers a comprehensive framework for integrating GenAI into financial risk management, emphasizing the dual objectives of innovation and accountability. By addressing identified gaps and leveraging advanced methodologies, this research aims to contribute significantly to the evolving landscape of AI-driven finance.

V. CONCLUSION

This research highlights the transformative potential of Generative AI in financial risk management, showcasing its ability to enhance decision-making, regulatory compliance, and predictive analytics. By leveraging Large Language Models such as GPT, organizations can integrate advanced capabilities into risk modeling frameworks, ranging from macroeconomic scenario generation to anomaly detection and sentiment analysis. The proposed methodologies demonstrate practical implementations, including the fine-tuning of models with domain-specific data and the integration of lightweight computing systems for testing and deployment.

However, this study also underscores the critical importance of human oversight in AI-driven financial systems to mitigate potential risks associated with model biases and inaccuracies. While synthetic data generation was excluded from this work, its potential role in addressing data scarcity and improving model robustness remains a key area for future exploration. Additionally, the ethical implications of using AI in financial decision-making demand continued attention to ensure transparency and fairness.

In conclusion, GenAI offers a dynamic and adaptive approach to financial risk analysis, enabling organizations to stay ahead in an increasingly complex and data-driven financial landscape. By building on the frameworks and insights presented in this paper, future research can further refine these methodologies, addressing the challenges of scalability, ethical considerations, and data integration to unlock the full potential of AI in financial risk management.

References

[1]. A. C. Teixeira, V. Marar, H. Yazdanpanah, A. Pezente, and M. Ghassemi, “Enhancing credit risk reports generation using LLMs: An integration of bayesian networks and labeled guide prompting,” in Proceedings of the fourth ACM international conference on AI in finance, in ICAIF ’23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 340–348. doi: 10.1145/3604237.3626902.

[2]. M. Sanz-Guerrero and J. Arroyo, “Credit risk meets large language models: Building a risk indicator from loan descriptions in P2P lending.” arXiv, Aug. 05, 2024. doi: 10.48550/arXiv.2401.16458.

[3]. G. Babaei and P. Giudici, “GPT classifications, with application to credit lending,” Machine Learning with Applications, vol. 16, p. 100534, Jun. 2024, doi: 10.1016/j.mlwa.2024.100534.

[4]. M. Khoja, “AI and bond values: How large language models predict default signals.” Social Science Research Network, Rochester, NY, Sep. 20, 2024. doi: 10.2139/ssrn.4965227.

[5]. B. Fazlija, M. Ibraimi, A. Forouzandeh, and A. Fazlija, “Implementing financial regulations using large language models.” Social Science Research Network, Rochester, NY, Nov. 05, 2024. Accessed: Dec. 19, 2024. [Online]. Available: https://papers.ssrn.com/abstract=5010694

[6]. J. Ludwig, S. Mullainathan, and A. Rambachan, “Large language models: An applied econometric framework.” arXiv, Dec. 09, 2024. doi: 10.48550/arXiv.2412.07031.

[7]. N. Li, C. Gao, M. Li, Y. Li, and Q. Liao, “EconAgent: Large language model-empowered agents for simulating macroeconomic activities,” in Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers), L.-W. Ku, A. Martins, and V. Srikumar, Eds., Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 15523–15536. doi: 10.18653/v1/2024.acl-long.829.

[8]. N. Li, C. Gao, Y. Li, and Q. Liao, “Large language model-empowered agents for simulating macroeconomic activities.” Social Science Research Network, Rochester, NY, Oct. 13, 2023. doi: 10.2139/ssrn.4606937.

[9]. N. Nascimento, C. Tavares, P. Alencar, and D. Cowan, “GPT in data science: A practical exploration of model selection,” in 2023 IEEE international conference on big data (BigData), Dec. 2023, pp. 4325–4334. doi: 10.1109/BigData59044.2023.10386503.

[10]. Y. Wang, J. Zhao, and Y. Lawryshyn, “GPT-signal: Generative AI for semi-automated

feature engineering in the alpha research process." arXiv, Oct. 24, 2024. doi: 10.48550/arXiv.2410.18448.

[11]. "GPT-4 and beyond: The role of generative AI in data engineering journal of bioinformatics and artificial intelligence." Accessed: Dec. 21, 2024. [Online]. Available: https://biotechjournal.org/index.php/jbai/article/view/142

[12]. N. Dulam, V. Gosukonda, and M. Ankam, "GPT-4 and beyond: The role of generative AI in data engineering," Journal of Bioinformatics and Artificial Intelligence, vol. 4, no. 1, pp. 227–249, Feb. 2024, Accessed: Dec. 21, 2024. [Online]. Available: https://biotechjournal.org/index.php/jbai/article/view/142

[13]. A. P. Desai, G. S. Mallya, M. Luqman, T. Ravi, N. Kota, and P. Yadav, "Opportunities and challenges of generative-AI in finance." arXiv, Nov. 22, 2024. doi: 10.48550/arXiv.2410.15653.

[14]. D. P. Jeong, Z. C. Lipton, and P. Ravikumar, "LLM-select: Feature selection with large language models." arXiv, Jul. 02, 2024. doi: 10.48550/arXiv.2407.02694.

[15]. D. Li, Z. Tan, and H. Liu, "Exploring large language models for feature selection: A data-centric perspective." arXiv, Oct. 23, 2024. doi: 10.48550/arXiv.2408.12025.

[16]. A. Hinterleitner, T. Bartz-Beielstein, R. Schulz, S. Spengler, T. Winter, and C. Leitenmeier, "Enhancing feature selection and interpretability in AI regression tasks through feature attribution." arXiv, Sep. 25, 2024. doi: 10.48550/arXiv.2409.16787.

[17]. S. Jomthanachai, W. P. Wong, and K. W. Khaw, "An application of machine learning regression to feature selection: A study of logistics performance and economic attribute," Neural Comput & Applic, vol. 34, no. 18, pp. 15781–15805, Sep. 2022, doi: 10.1007/s00521-022-07266-6.

[18]. E. Sharkey and P. Treleaven, "BERT vs GPT for financial engineering." arXiv, Apr. 24, 2024. doi: 10.48550/arXiv.2405.12990.

[19]. M. Hofert, "Assessing ChatGPT's proficiency in quantitative risk management," Risks, vol. 11, no. 9, p. 166, Sep. 2023, doi: 10.3390/risks11090166.

[20]. O. Aljaloudi, M. Thiam, M. Qader, M. K. S. Al-Mhdawi, A. Qazi, and N. Dacre, "Examining the integration of generative AI models for improved risk management practices in the financial sector," Nov. 2024.

[21]. J. Lee, N. Stevens, S. C. Han, and M. Song, "A survey of large language models in finance (FinLLMs)." arXiv, Feb. 04, 2024. doi: 10.48550/arXiv.2402.02315.

[22]. G. Bhatia, E. M. B. Nagoudi, H. Cavusoglu, and M. Abdul-Mageed, "FinTral: A family of GPT-4 level multimodal financial large language models." arXiv, Jun. 14, 2024. doi: 10.48550/arXiv.2402.10986.

[23]. "How to train generative AI using your company's data." Accessed: Dec. 22, 2024. [Online]. Available: https://hbr.org/2023/07/how-to-train-generative-ai-using-your-companys-data

[24]. "Proprietary data, your competitive edge in generative AI IBM." Accessed: Dec. 22, 2024. [Online]. Available: https://www.ibm.com/think/insights/proprietary-data-gen-ai-competitive-edge

[25]. A. Selvaraj, D. Venkatachalam, and G. Namperumal, "Synthetic data for financial anomaly detection: AI-driven approaches to simulate rare events and improve model robustness," Journal of Artificial Intelligence Research and Applications, vol. 2, no. 1, pp. 373–425, 2022, Accessed: Dec. 25, 2024. [Online]. Available: https://aimlstudies.co.uk/index.php/jaira/article/view/221

[26]. "Evaluating the impact of synthetic data on financial machine learning models: A comprehensive study of AI techniques for data augmentation and model training journal of artificial intelligence research and applications." Accessed: Dec. 25, 2024. [Online]. Available: https://aimlstudies.co.uk/index.php/jaira/article/view/214

[27]. V. K. Potluru et al., "Synthetic data applications in finance." arXiv, Mar. 20, 2024. doi: 10.48550/arXiv.2401.00081.

[28]. J. Su et al., "Large language models for forecasting and anomaly detection: A systematic literature review." arXiv, Feb. 15, 2024. doi: 10.48550/arXiv.2402.10350.

[29]. O. Campesato, Python 3 and machine learning using ChatGPT/GPT-4. Walter de Gruyter GmbH & Co KG, 2024.

[30]. J. Koskula, "Generative artificial intelligence in support of analytics : Copilot 365," 2024, Accessed: Dec. 25, 2024. [Online]. Available: https://lutpub.lut.fi/handle/10024/168426

[31]. P. Winder, C. Hildebrand, and J. Hartmann, "Biased echoes: Generative AI models reinforce investment biases and increase portfolio risks of private investors." Social Science Research Network, Rochester, NY, Sep. 26, 2024. doi: 10.2139/ssrn.4968821.

[32]. "Generating synthetic data in finance proceedings of the first ACM international conference on AI in finance." Accessed: Dec. 25, 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3383455.3422554