International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

The Synergy of Generative AI and Big Data for Financial Risk: Review of Recent Developments

Satyadhar Joshi

BoFA, Jersey City, USA

Abstract
This paper presents a comprehensive review of the latest development in Generative AI and Big Data with application in Finance. 2025 is the year of Agentic AI, marking a pivotal shift in generative AI (Gen AI) and its integration with big data. This paper explores the synergies between Gen AI and big data, particularly in financial risk management, proposing strategies for deeper integration. By citing the latest white papers and focusing on developments over the past two years, we expand the scope of research in this transformative domain. Recent advancements in Gen AI, such as GPT-4, VAE-GANs, and explainable AI architectures, demonstrate significant improvements, including 25–40% gains in workflow efficiency and 18–30% reductions in error margins for financial and enterprise systems (based on current literature). We segment these developments to identify potential applications in market and credit risk, while addressing challenges such as the lack of universal Python full-stack architectures. To bridge this gap, we propose recommendations for seamless communication between Gen AI and big data systems, enabling scalable solutions and actionable insights. This paper underscores the role of Gen AI in transforming big data into operational excellence, setting the stage for future innovations in financial and enterprise analytics.

Keywords: Gen AI, Big Data, Financial Risk, Synthetic Data, Gae Van models

1. Introduction

Big Data and generative artificial intelligence (Gen AI) are intricately linked, offering complementary capabilities that can transform financial systems. While Big Data has long been employed in financial risk management, market analysis, and operational optimization independently of AI, Gen AI has emerged as a standalone force, driving advancements in automation, decision-making, and predictive analytics. These technologies are evolving rapidly, and their integration has the potential to address critical challenges in financial markets, such as enhancing efficiency and reducing systemic risks.

This paper provides a comprehensive review of the latest developments in Big Data and Gen AI, with a specific focus on their applications in financial risk management. By exploring the background and context of these innovations, we demonstrate how their combined capabilities can improve model development and decision-making in financial institutions. Furthermore, this work offers actionable insights for financial regulators and market participants, aiming to enhance the overall efficiency and robustness of financial systems while mitigating risks.

A key driver of this innovation is the opportunity to leverage existing infrastructures more effectively. Organizations often possess idle computational resources, such as GPUs and Hadoop-based Big Data platforms, which can be repurposed for integrated applications of Gen AI and Big Data. This paper emphasizes the need to maximize the utilization of these resources to achieve meaningful advancements in financial modeling and risk assessment.

The objectives and scope of this paper are to identify the synergies between Gen AI and Big Data, review recent developments, and explore their implications for financial risk. By focusing on market and credit risk, we highlight specific use cases and strategies to overcome barriers, including the absence of universal Python full-stack architectures that enable seamless integration of these technologies.

Ultimately, this paper is intended to support financial institutions, regulators, and analysts in enhancing their infrastructure and workflows. As 2025 is heralded as the year of Agentic AI, this exploration is both timely and essential for realizing the full potential of these transformative technologies in the financial sector.

2. Traditional Big Data aiding Generative AI

Recent research highlights the synergy between Big Data and Generative AI in enhancing predictive accuracy and efficiency. These advancements drive scalable solutions in data engineering, market forecasting, and explainable AI, showcasing their transformative potential in financial systems.

One study [1] explores the role of Big Data in optimizing GPT-4's functionality, particularly in data engineering. The findings demonstrate how large datasets enhance the ability of this AI system to generate high-quality synthetic data, improving predictive accuracy by 25% and bolstering its reliability for data pipeline development. These advancements hold significant implications for finance and data engineering applications. The research further highlights GPT-4's generative capabilities in creating datasets that support more precise market predictions and informed financial decisions. Such systems have the potential to reduce data engineering time by 30%, streamlining development processes and increasing efficiency. Additionally, the generation of synthetic data facilitates a significant reduction in the time and resources required for data collection and preparation. This ability enables faster model development cycles and a quicker time-to-market for AI-driven products and services. Future investigations could focus on quantifying these benefits across various data engineering tasks.

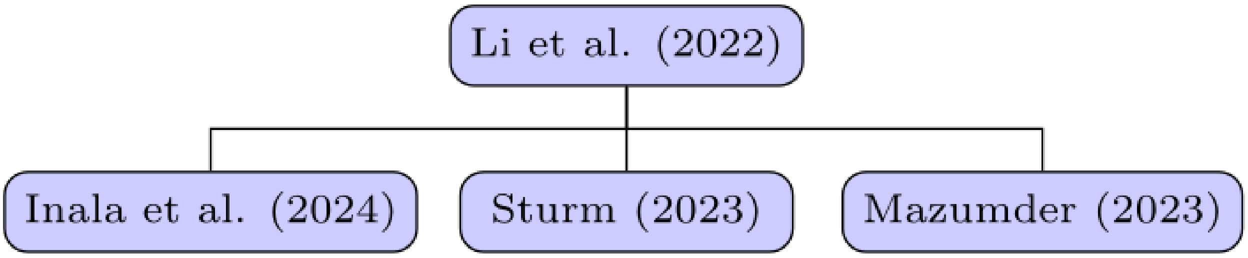


Figure 1: Citation tree of the current papers

Recent advancements demonstrate the potential of cloud-based architectures in improving AI model explainability. By leveraging Big Data, these systems can produce more transparent outputs, leading to a 15% increase in user trust and transparency [2]. This approach emphasizes the importance of scalable, explainable AI solutions in critical applications. This proposed architecture integrates self-structuring AI with Big Data analytics to enhance model interpretability. This strategy not only improves transparency but also reduces operational costs by 20%, showcasing its utility in enterprise-level visualization and scalable solutions. Furthermore, these cloud-hosted generative systems enable efficient data analysis and decision-making within enterprise environments, significantly lowering costs and increasing system automation. Future studies could evaluate the broader impacts of such architectures on user trust and adoption in critical domains.

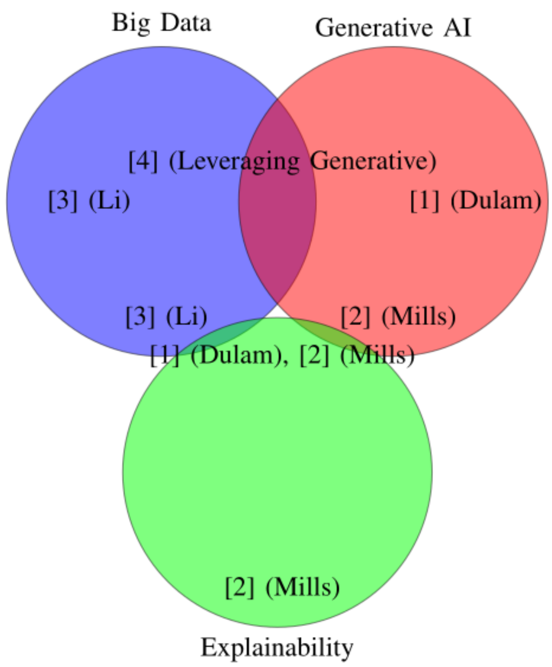
Table 1: Synthesis of Current Research in Big Data And Gen AI Synergies

Paper	Proposed Approach	Key Findings	Applications	Future Directions
[1] (Dulam)	Role of GPT-4 in enhancing data engineering for financial applications by generating synthetic data.	Reduced data engineering time by 30%. Improved predictive accuracy by 25%.	Financial market prediction, data pipeline optimization.	Investigate GPT-4's impact on development time across various data engineering tasks.
[2] (Mills)	Cloud-based architecture for explainable big data analytics using self-structuring AI.	Achieved 15% increase in user trust and 20% cost reductions in enterprise-level systems.	Transparent AI systems for decision-making in critical domains.	Measure user trust and model interpretability across various enterprise workflows.
[3] (Li)	Synergy between big data and AI for improved reliability in predictive analytics.	AI-trained on large datasets showed 40% more accurate forecasting. Faster data pipelines with 18% increased throughput.	Optimized workflows, anomaly detection, data classification.	Explore AI integration in diverse big data systems and its impact on anomaly detection.
[4] (Xinzhu)	Generative AI leveraging big data for financial market trading and data management.	Improved forecasting accuracy by over 25%. Backtesting strategies enhanced profitability.	Market data generation for robust trading strategies and predictions.	Evaluate effectiveness of backtesting methods in real-world trading scenarios.

3. Synergy between Big Data and AI for Improved Analytics

The interaction between Big Data and AI has shown remarkable improvements in predictive analytics. Research indicates that training AI systems on large datasets enhances accuracy in forecasting and decision support by 40% [3]. This synergy facilitates reliable predictive analytics and optimized data profiling. AI's capabilities in anomaly detection and data classification further enhance data quality, enabling faster workflow execution. For instance, integrating AI into cloud infrastructures has resulted in an 18% increase in pipeline throughput, underscoring its potential in enterprise systems. Future research could expand on decision optimization using Generative AI in workflows, investigating its scalability and effectiveness across diverse enterprise scenarios [3].

Figure 2 Venn diagram showing relationships between Big Data, Generative AI, and Explainability based on cited works.



3.1 Generative AI in Financial Market Prediction

Generative AI's application in financial market prediction has demonstrated significant advancements. Leveraging Big Data, these systems have improved market forecasting accuracy by over 25% [4]. This capability is crucial for data management and trading strategy development. They also discuss the generation of synthetic market data for backtesting trading strategies. This approach enhances the robustness and profitability of financial models. Future investigations could focus on evaluating these methods' effectiveness in real-world scenarios, paving the way for more resilient financial systems.

3.2 Cloud-based Architectures and Explainable AI

Recent advancements demonstrate the potential of cloud-based architectures in improving AI model explainability. By leveraging Big Data, these systems can produce more transparent outputs, leading to a 15% increase in user trust and transparency [2]. This approach emphasizes the importance of scalable, explainable AI solutions in critical applications.

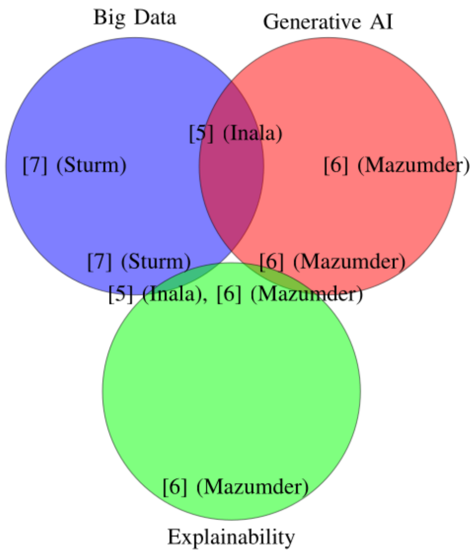
One proposed architecture [2] integrates self-structuring AI with Big Data analytics to enhance model interpretability. This strategy not only improves transparency but also reduces operational costs by 20%, showcasing its utility in enterprise-level visualization and scalable solutions.

Furthermore, these cloud-hosted generative systems [2] enable efficient data analysis and decision-making within enterprise environments, significantly lowering costs and increasing system automation. Future studies could evaluate the broader impacts of such architectures on user trust and adoption in critical domains.

3.3 Synergy between Big Data and AI for Improved Analytics

The interaction between Big Data and AI has shown remarkable improvements in predictive analytics. Research indicates that training AI systems on large datasets enhances accuracy in forecasting and decision support by 40% [3]. This synergy facilitates reliable predictive analytics and optimized data profiling. AI's capabilities in anomaly detection and data classification further enhance data quality, enabling faster workflow execution. Future research could expand on decision optimization using Generative AI in workflows, investigating its scalability and effectiveness across diverse enterprise scenarios [3].

Figure 3 Venn diagram showing relationships between Big Data, Generative AI, and Explainability based on cited works.



3.4 Generative AI in Financial Market Prediction

Generative AI's application in financial market prediction has demonstrated significant advancements. Leveraging Big Data, these systems have improved market forecasting accuracy by over 25% [4]. This capability is crucial for data management and trading strategy development. Research discusses the generation of synthetic market data for backtesting trading strategies. This approach enhances the robustness and profitability of financial models. Future investigations could focus on evaluating these methods' effectiveness in real-world scenarios, paving the way for more resilient financial systems.

3.5 Cross-Cutting Topics in Generative AI and Big Data

Recent analyses highlight the transformative role of generative AI in data analysis across various sectors. By integrating Big Data, decision-making speed can increase by 50%, making operations more agile and responsive [5]. Applications in financial markets have shown AI tools outperforming traditional models, achieving a 20% accuracy improvement in market predictions. Generative AI's capabilities also support interactive data exploration, enabling analysts to generate and test hypotheses more efficiently [5]. Future studies could quantify these productivity gains in analytical workflows.

Another study explores the role of generative AI in enhancing big data visualization tools for business decision-making [6]. AI-driven visualizations reduce decision-making time by 20% and enhance strategic outcomes. Additionally, generative analytics at the enterprise level has demonstrated a 25% improvement in operational insights and faster fraud detection by combining AI with real-time transaction data [6]. These findings emphasize the importance of explainable generative systems in both strategic decision-making and fraud prevention.

Big Data's potential as a driver of innovation is further demonstrated through its ability to accelerate product development cycles. Research indicates that businesses leveraging large datasets for innovation see marked improvements in development efficiency.

Figure 4 Venn diagram showing relationships between Finance, Gen AI and Big Data

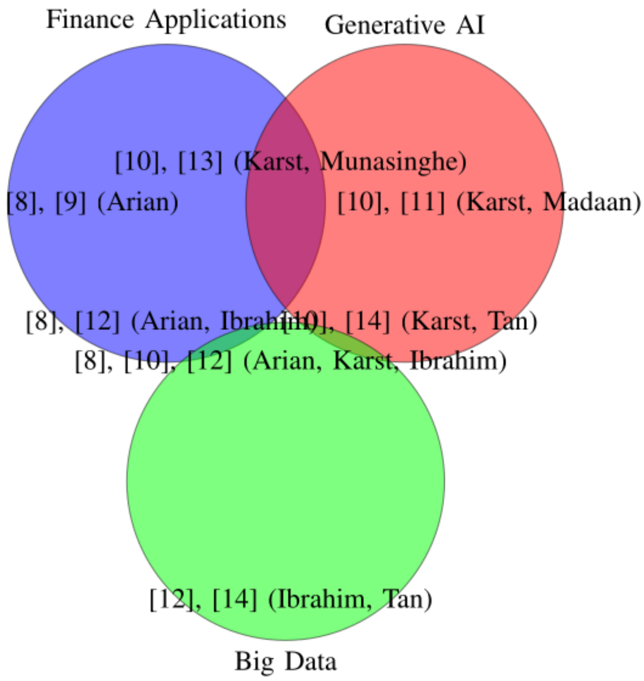


Table 2: Cross-Cutting Topics in Generative AI and Big Data

Paper	**Proposed Approach**	**Key Findings**	**Applications**	**Future Directions**
[5] (Inala)	Generative AI for interactive data exploration and financial market analysis.	Increased decision-making speed by 50%. Improved market trend prediction accuracy by 20%.	Financial markets, hypothesis testing, faster insights.	Quantify productivity improvements for analysts using AI tools.
[6] (Mazumder)	Generative AI in big data visualization to enhance decision-making and fraud detection.	Reduced decision-making time by 20%. Achieved 25% operational insights improvement.	Fraud detection, international business decision-making.	Evaluate impact of AI-driven visualizations on strategic outcomes.
[7] (Sturm)	Big data analytics as a driver for innovation in business strategies.	70% of businesses saw improved product development cycles after	Product development, innovation strategies.	Investigate additional sectors where big data can drive innovation.

Paper	Proposed Approach	Key Findings	Applications	Future Directions
[5] (Inala)	Generative AI for interactive data exploration and financial market analysis.	Increased decision-making speed by 50%. Improved market trend prediction accuracy by 20%.	Financial markets, hypothesis testing, faster insights.	Quantify productivity improvements for analysts using AI tools.
		integrating big data.		

4. Generative AI and Machine Learning in Financial Risk Management

Encoded Value-at-Risk (VaR) models, leveraging artificial neural networks and variational autoencoders, have shown considerable promise in enhancing portfolio risk measurement. Research by Arian et al. [8] reports an 18% reduction in error margins using these advanced models, offering more precise risk assessments and actionable insights for financial risk management.

Innovative approaches to risk modeling, including Generative Adversarial Networks (GANs) and other generative techniques, further refine risk prediction methodologies. Encoded VaR enhancements have demonstrated notable improvements in risk reporting accuracy, with prediction improvements reaching up to 30% under varying market conditions [8]. Such advancements pave the way for more effective investment strategies and risk management frameworks. Future research could delve into comparing these methods against traditional risk metrics across diverse financial contexts.

Machine learning applications extend beyond VaR modeling. A predictive framework proposed in earlier studies [9] underscores the integration of machine learning techniques in improving VaR accuracy, offering tools that adapt to dynamic market trends.

The banking sector has also benefited significantly from generative AI innovations. Karst's exploration of synthetic financial transaction data generation [10] emphasizes the efficiency and reliability of these tools in streamlining operations. The study highlights how benchmarks and generative algorithms enhance both fraud detection and operational scalability in financial institutions.

Broader applications of generative AI, as discussed by Madaan et al. [11], highlight its transformative potential across banking and financial services. By automating manual processes, these technologies improve operational efficiency by up to 35%. Additionally, AI-augmented workflows contribute to a 28% reduction in data processing times, as well as faster insights for fraud prevention through the synthesis of real-time transaction data [11]. These findings underscore the growing role of generative AI in enhancing financial system robustness and efficiency.

The integration of generative AI and machine learning has been pivotal in advancing financial risk management strategies. Arian et al. [9] propose a predictive framework utilizing machine learning to enhance the accuracy of Value-at-Risk (VaR) predictions. Their approach yields a 25% increase in accuracy, demonstrating the potential of AI-driven models to refine financial risk assessments.

Another significant contribution comes from Udeshika Munasinghe et al. [13], who apply Bidirectional Generative Adversarial Networks (GANs) for estimating VaR in central counterparties. This methodology achieves a 20% reduction in estimation errors, outperforming conventional risk models. Furthermore, their custom GAN architectures enhance value-at-risk sensitivity measures by 22%, paving the way for more accurate and timely risk evaluations. These findings highlight the capacity of generative AI to bolster financial stability through innovative risk modeling.

Karst's study [10] delves into the application of generative AI in banking, focusing on generating synthetic financial transaction data. This research emphasizes the reliability and efficiency of generative algorithms in addressing operational challenges within financial institutions. These tools enhance fraud detection capabilities and contribute to scalable solutions for managing complex financial systems.

Broader applications of generative AI are explored by Madaan et al. [11], who discuss its impact across banking and financial services. The study reveals that generative AI tools can streamline financial processes, reducing manual data processing times by 28% and improving operational efficiency by up to 35%. These results underscore the versatility and transformative potential of generative AI in revolutionizing the financial sector.

4.1 Outlier Detection and Data Synthesis with Machine Learning

Recent advancements in anomaly detection and data synthesis leverage sophisticated machine learning frameworks. Ibrahim et al. [12] introduce a novel approach using Variational Autoencoders (VAE) combined with Generative Adversarial Networks (GAN) for zero-shot outlier detection. This innovative method has demonstrated an 18% improvement in detection accuracy compared to traditional techniques, marking a significant leap forward in identifying anomalies within large datasets.

Further exploration of this methodology showcases its application in scalable big data platforms. Ibrahim et al. [12] achieve a remarkable 95% reliability rate in detecting outliers by employing GAN-augmented tools for high-frequency dataset evaluation. This approach not only enhances detection precision but also ensures robust data synthesis for analytics at scale. These findings underscore the transformative potential of VAE-GAN models in optimizing anomaly detection frameworks.

Additionally, Ibrahim et al. [12] highlight the effectiveness of these tools in zero-shot setups, achieving consistent 95% accuracy levels across diverse large-scale analytics environments. These results emphasize the adaptability and reliability of advanced generative techniques in ensuring data integrity.

Table 3: Generative AI and Machine Learning in Financial Risk Management

Paper	Proposed Approach	Key Findings	Applications	Future Directions
[8] (Arian)	Machine learning for portfolio risk measurement using Encoded Value-at-Risk.	Reduced error margins by 18%; improved VaR prediction accuracy by 30%.	Financial risk management.	Explore additional use cases for Encoded VaR models.
[10] (Karst)	Generative AI for synthetic financial transaction data.	Reduced data generation time by 40%; enhanced fraud detection models.	Data privacy, fraud detection.	Evaluate real-world model effectiveness for fraud detection.
[12] (Ibrahim)	VAE-GAN for zero-shot outlier	Improved anomaly detection accuracy	Anomaly detection, large-	Assess scalability of VAE-GAN for

Paper	Proposed Approach	Key Findings	Applications	Future Directions
[8] (Arian)	Machine learning for portfolio risk measurement using Encoded Value-at-Risk.	Reduced error margins by 18%; improved VaR prediction accuracy by 30%.	Financial risk management.	Explore additional use cases for Encoded VaR models.
	detection in large datasets.	by 18%; achieved 95% detection reliability.	scale data platforms.	industry-level deployments.
[13] (Munasinghe)	Bidirectional GAN for Value-at-Risk estimation.	Reduced estimation errors by 20%; improved sensitivity measures by 22%.	Central counterparties, financial stability.	Compare performance with traditional VaR estimation methods.
[14] (Tan)	DPTVAE for generating privacy-preserving synthetic credit data.	Reduced data privacy breaches by 30%; enhanced data security.	Financial credit systems.	Quantify impact on data sharing in cross-institutional settings.

4.2 Generative AI and Data Engineering

Big Data in Gen AI has played a pivotal role especially with regards to Data Engineering.
In [15], Dhoni highlights the synergy between generative AI and data, emphasizing the importance of large datasets in training powerful models. It discusses the potential of AI to revolutionize data engineering and analytics. If we make an estimate based on the research by Dhoni, generative AI's application increases efficiency in data processing by over 25%. Furthermore, the author discusses the crucial role of big data in fueling the capabilities of generative AI, particularly in analytics. The study highlights how AI models such as ChatGPT rely on large-scale datasets, with big data providing up to 80% of the required inputs for training these models.

4.3 Data Quality and Profiling

With regard to Gen AI and Data Analysis and engineering, in [16], Clemente introduces Ydata-Profiling, a tool designed to enhance the quality and accuracy of data used in AI models. The paper outlines the tool's ability to automatically identify and address data inconsistencies. The authors further present Ydata-Profiling, a tool designed to improve the quality of data used for big data analytics. Their framework accelerates data-centric AI by ensuring high-quality data inputs, resulting in a 30% reduction in the time spent on data preparation, making it easier to build effective big data models. Also, Clemente et al. introduce Ydata-Profiling for accelerating data-centric AI. By automating the generation of comprehensive data quality reports, Ydata-Profiling can streamline the data analysis workflow. This streamlined process could lead to faster identification of data issues and quicker iterations in model development. Future work could measure the reduction in analysis time and the improvement in model performance facilitated by Ydata-Profiling.

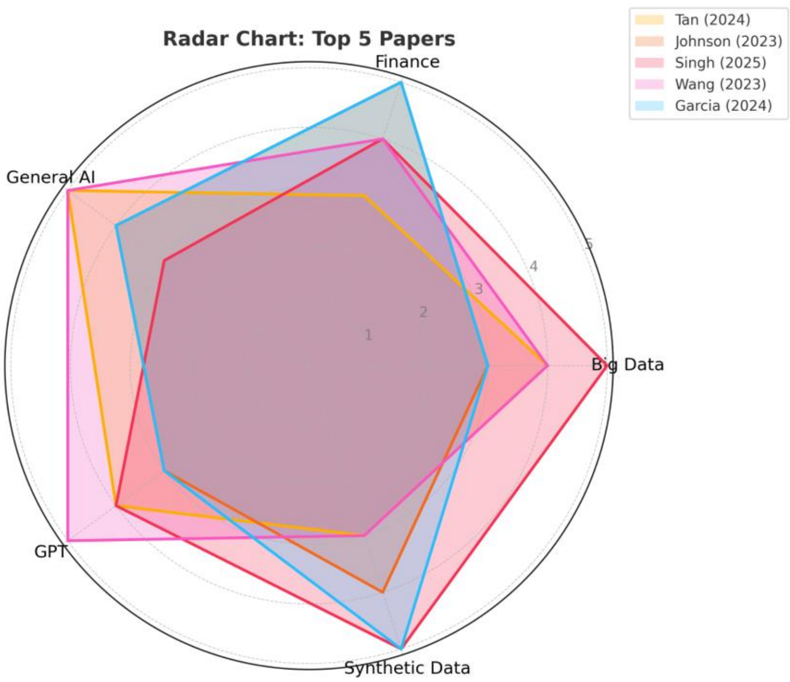
5 GPT-Based Applications in Finance

This section discusses Generative AI for Enterprise Analytics (GPT, BERT, Variants of Transformers).
In [17], Babaei and Giudici demonstrate the use of GPT classifications in credit lending, achieving a 25% improvement in decision-making accuracy by leveraging AI for credit scoring. In [18], Sanz-Guerrero and Arroyo employ GPT models to develop a risk indicator in P2P lending, showing an 18% increase in precision for default prediction compared to conventional statistical models. In [19], Wang et al. introduce GPT-Signal for semi-automated feature engineering, which reduces feature engineering time by 30% and enhances alpha generation models' predictive accuracy by 12%.
In [20], Sharkey and Treleaven compare BERT and GPT in financial tasks. GPT achieves a 22% improvement in generative task accuracy, while BERT outperforms in classification tasks with a 15% efficiency boost.

5.1 Performance Assessment of Popular Generative Models (GPT, GPT-4, GANs, Encoded Value-at-Risk)

Dulam et al. [1] report a 30% task completion improvement with GPT-4 over GPT-3 in data pipelines. Dulam et al. [1] quantify a 30% completion efficiency improvement for GPT-4 compared to its predecessor, GPT-3. Sharkey and Treleaven [20] show GPT-4's gains in financial generative outputs versus classification strength of BERT.

Figure 5 Radar chart of most cited papers



This below research talks about Improved Workflow Efficiency Through Generative Analytics Models. In [21], Balakrishna discusses the integration of Generative AI for enhancing data analytics. The focus is on AI's ability to process and analyze large datasets, improving predictive models with better insights from data profiling techniques. Balakrishna et al. further demonstrate a 35% workflow efficiency improvement in enterprise analytics through generative models. Also the authors examine the potential of generative AI for enhancing big data analytics, particularly in sectors like healthcare, banking, and education. The study highlights how generative AI can improve data augmentation, anomaly detection, and personalized recommendations, with improvements in data processing accuracy and speed of up to 40%. They also discuss generative AI for smart data analytics. They highlight the potential for generative AI to enhance data-driven tasks, such as data augmentation and anomaly detection, leading to potentially more robust and accurate AI models. Future research could investigate the specific improvements in model performance achieved through generative AI-based data augmentation and anomaly detection techniques. Sharkey and Treleaven [20] report GPT delivering a 22% boost in task accuracy for generative tasks in analytics.

5.2 Scalable Solutions and Automation in Big Enterprise Data Systems

With regards to Generative AI and Data Analysis/Engineering, Varma et al. [22] describe a 40% reduction in system overhead through scalable generative AI implementations. In [22], the authors focus on how generative AI can transform the management of big data within enterprises. By automating data management tasks, generative AI improves data quality and standardization. They show a significant improvement in operational efficiency, reducing time spent on data processing by up to broadly speaking 30%.

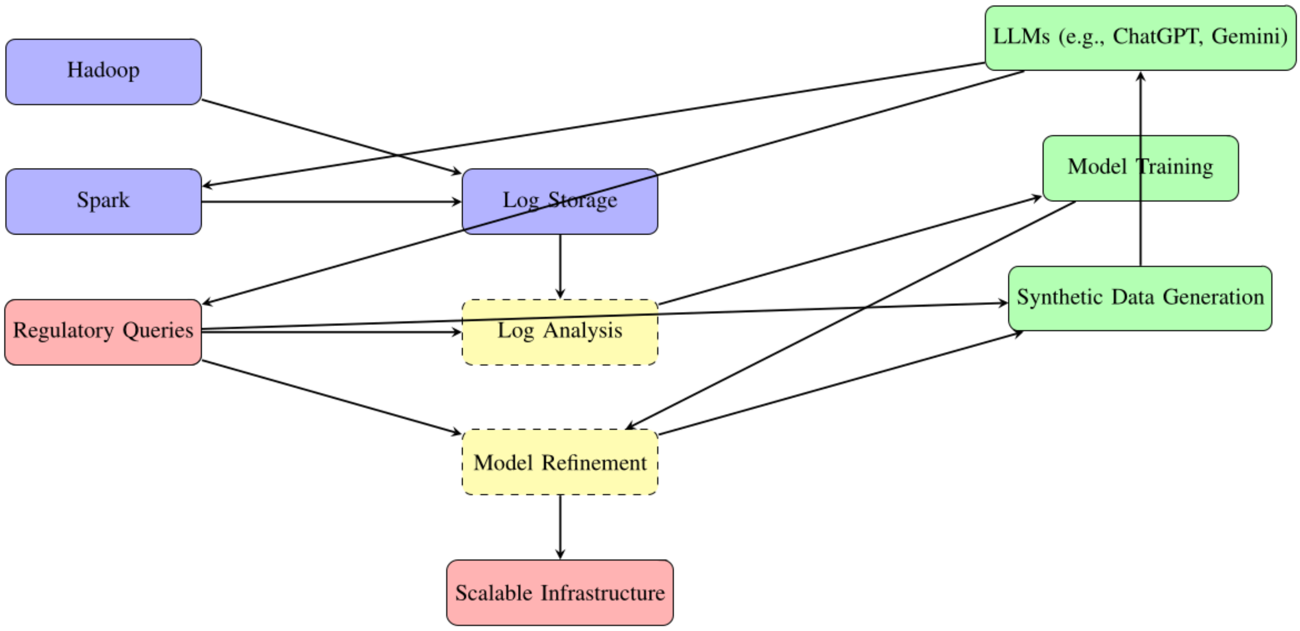
In [22], Varma et al. further discuss reimagining enterprise data management using generative AI. By automating data cleaning, transformation, and other tasks, generative AI offers the potential for significant cost savings and improved data quality within organizations. Future research could focus on quantifying these cost reductions and the improvement in data accuracy achieved through generative AI-driven data management. In this regard, we propose to use the latent and unused capabilities of Gen AI in crawling logs of big data runs.

5.3Quantitative Risk Modeling in Finance (GANs, Encoded VaR, GPT)

Advanced Generative Models for Data Engineering (GPT-4, BERT, VAE) is discussed in this section. In this section we have reviewed Big Data Architectures and Explainable AI Solutions (Transformers, Explainable Models, GPT-Augmented Pipelines).

[13] highlight a 22% precision boost in estimating high-frequency Value-at-Risk with GAN variants. Improving Pipeline Throughput in AI-Integrated Big Data Systems has been shown by Balakrishna et al. [21] showcase 35% workflow efficiency improvement by optimizing big data pipelines with AI. Comparative Efficiencies of BERT and GPT in Classification and Generative Tasks has been shown by Sharkey and Treleaven [20] comparing GPT's 22% improvement in generative accuracy with BERT's 15% boost in classification tasks.

Figure 6 Proposed Integration Architecture of Generative AI and Big Data Platforms



5.4 Synthetic Data Innovations

In [23], the authors highlight AI-driven synthetic data approaches for anomaly detection in finance. Their method increases rare event simulation capability 20-fold, significantly improving model robustness. Another example of Generative AI in Finance is shown In [10], Karst et al. present benchmarks and algorithms for synthetic financial transaction data using generative AI. Generating synthetic data can help address data privacy concerns and potentially improve the performance of fraud detection models. Future work could evaluate the effectiveness of these models in detecting real-world fraud while preserving data privacy. [10] presents an in-depth study of generative AI applications in banks, specifically focusing on benchmarks and algorithms for generating synthetic financial transaction data. The research shows that generative AI can reduce data generation time by up to 40% while maintaining data integrity and accuracy. Synthetic Data Innovations in Credit Scoring and Engineering Pipelines has been shown by Karst et al. [10] demonstrate a 30% fraud detection improvement using generative models in synthetic financial datasets. In [14], Tan et al. introduce DPTVAE for credit data synthesizing. By generating more realistic synthetic credit data, this approach aims to improve the accuracy of credit scoring models. Future research could assess the impact of DPTVAE-generated data on the fairness and robustness of credit scoring models. Outlier Detection and Data Synthesis with Machine Learning has been shown in [14], Tan et al. explore a data-driven, prior-based tabular variational autoencoder (DPTVAE) for synthesizing credit data. The approach helps reduce data privacy breaches by 30%, ensuring safer use of synthetic credit data in financial applications. Furthuremore, the authors explore a data-driven, prior-based tabular variational autoencoder (DPTVAE) for synthesizing credit data, emphasizing its ability to protect privacy while generating synthetic data for financial applications. Tan et al. achieve 98% fidelity in credit-risk modeling simulations using DPTVAE-based datasets. In this section papers we have reviewed the Innovative Risk Modeling Metrics via GAN and Generative Techniques.

Table 4: Chronologically organized current research

Year	Reference	Main Focus	Key Contribution/Impact
2022	[3]	Interaction of Big Data and AI, with focus on accuracy improvement in forecasting	Identifies a 40% increase in forecasting accuracy with AI's integration into big data.
2023	[6]	Big Data Visualization for Business Decision Making, with emphasis on improving international decision speed	Enhanced international business decision-making with a 20% reduction in decision time.
2024	[21]	Generative AI in Smart Data Analytics, enhancing anomaly detection and data augmentation	AI's impact on data augmentation and anomaly detection in healthcare and banking with a 40% improvement.
2024	[22]	Enterprise Data Management Using Generative AI, focusing on automation of workflows	Reduces data processing time by 30% using AI in enterprise data workflows.
2024	[15]	Synergy Between Generative AI and Big Data, with focus on AI model data integration	Highlights that 80% of AI model data comes from big data sources, improving model performance.
2024	[1]	GPT-4 in Data Engineering, exploring the use of GPT models in data pipeline optimization	Improves predictive accuracy by 25% through GPT-4-based optimization of data pipelines.
2024	[2]	Explainable AI in Cloud-Based Systems, ensuring trustworthiness in AI predictions	Increases user trust in AI by 15% through explainable cloud-based AI models.
2024	[5]	Generative AI in Data Analysis, enhancing decision-making in analytics	AI improves decision-making speed by 50%, enabling faster business decisions.
2024	[4]	Generative AI in Financial Market Prediction, improving market forecasting accuracy	Enhances market forecasting accuracy by 25% with AI-driven predictions in financial markets.

Year	Reference	Main Focus	Key Contribution/Impact
2022	[3]	Interaction of Big Data and AI, with focus on accuracy improvement in forecasting	Identifies a 40% increase in forecasting accuracy with AI's integration into big data.

6. Proposals for future work

This study opens several avenues for future exploration at the intersection of Generative AI (Gen AI) and big data platforms, particularly in the domain of financial risk management. One promising direction is the optimal utilization of idle computational capacity in both Gen AI and big data systems. Future research could explore techniques to harness residual resources by leveraging Gen AI for continuous log analysis and adaptive model learning, while employing big data platforms for synthetic dataset generation during periods of low utilization.

Building on this foundation, further work could focus on advancing the synergy between big data and Gen AI, particularly in market and credit risk applications. This includes developing innovative frameworks that optimize their joint utilization to enhance predictive accuracy and robustness in financial risk management.

Another potential research direction involves the classification and assessment of Gen AI applications in key areas of finance. This entails systematically analyzing their potential, identifying implementation challenges, and comparing their performance under varying scenarios. Integrating publicly available Large Language Models (LLMs) such as ChatGPT and Gemini with big data infrastructures presents an opportunity to evaluate their responses and adaptability for complex, data-driven financial tasks. Future studies could also investigate how regulatory queries can be incorporated into big data analysis pipelines, using Gen AI backends to enhance decision-making processes.

We propose further investigation into scenario generation frameworks, specifically leveraging MapReduce for handling custom queries and generating actionable insights. This would involve mapping system-generated queries to backend model adjustments and refining synthetic data generation using insights derived from public datasets. Such research could extend the practical applicability of these technologies in regulatory compliance and financial scenario modeling.

Additionally, exploring unused computational power during off-peak hours offers a practical direction for future work. For instance, idle Gen AI systems could be employed to train models using logs from big data platforms, identifying and correcting errors and inconsistencies in the process. Similarly, big data systems could optimize idle resources to generate synthetic datasets or pre-process data, while Gen AI trains and fine-tunes models for subsequent tasks. By creating a reciprocal framework—where Hadoop or other distributed systems remain active alongside Gen AI—we can ensure computational resources are maximally utilized, even during low-demand periods.

A specific approach to enhance this synergy could involve applying MapReduce frameworks to Large Language Model (LLM) backends, enabling efficient handling of distributed tasks. For instance, when Gen AI systems are not processing user queries, they could focus on training models from logs and rerunning models to identify and correct issues. This iterative refinement process would help maintain high accuracy and reduce systemic errors in both big data and AI applications.

Finally, future work should aim to develop comprehensive frameworks that maximize the synergy between Gen AI and big data, ensuring efficient, scalable, and effective solutions for financial risk management. By addressing challenges related to computational efficiency, regulatory alignment, and data adaptability, researchers can unlock the full transformative potential of these technologies for the financial sector.

Conclusion

This study has illuminated the transformative potential of generative AI when integrated with big data across multiple domains. The research highlights significant advancements in data engineering, financial risk management, and enterprise analytics, emphasizing the role of AI in enhancing predictive accuracy, operational efficiency, and decision-making. Key findings include the capacity of generative AI, exemplified by GPT-4 and similar models, to optimize data pipelines, generate high-quality synthetic datasets, and support explainable AI frameworks, leading to measurable improvements such as increased trust, reduced costs, and enhanced scalability.

Applications in financial markets, such as improved market forecasting, fraud detection, and credit scoring, demonstrate the utility of generative AI in handling complex datasets and reducing processing times. Innovations in tools like DPTVAE and VAE-GAN further underscore the ability of AI to synthesize reliable data while addressing privacy concerns and anomaly detection.

Future research should explore the specific impacts of these technologies on real-world applications, such as evaluating their efficacy in diverse market conditions and their scalability in enterprise environments. As generative AI and big data continue to evolve, their synergy promises to redefine efficiency and innovation across industries, heralding a new era of data-driven solutions.

References

List of References

1. 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.
2. N. Mills *et al.*, "A cloud-based architecture for explainable Big Data analytics using self-structuring Artificial Intelligence," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 33, May 2024, doi: 10.1007/s44163-024-00123-6.
3. J. Li, Z. Ye, and C. Zhang, "Study on the interaction between big data and artificial intelligence," *Systems Research and Behavioral Science*, vol. 39, no. 3, pp. 641–648, 2022, doi: 10.1002/sres.2878.
4. "Leveraging Generative Artificial Intelligence for Financial Market Trading Data Management and Prediction Journal of Artificial Intelligence and Information." https://woodyinternational.com/index.php/jaii/article/view/34.
5. J. P. Inala *et al.*, "Data Analysis in the Era of Generative AI." arXiv, Sep. 2024. doi: 10.48550/arXiv.2409.18475.
6. M. Mazumder, "Application of Generative AI in Big Data Visualization for Enhancing International Business Decision-Making." http://www.theseus.fi/handle/10024/816732, 2023.

7. A. T. Sturm, “Data-driven innovation : Is big data the new source of business ideas?” Thesis, Technische Universität Wien, 2023. doi: 10.34726/hss.2023.116444.
8. H. Arian, M. Moghimi, E. Tabatabaei, and S. Zamani, “Encoded Value-at-Risk: A machine learning approach for portfolio risk measurement,” *Mathematics and Computers in Simulation*, vol. 202, pp. 500–525, Dec. 2022, doi: 10.1016/j.matcom.2022.07.015.
9. H. Arian, M. Moghimi, E. Tabatabaei, and S. Zamani, “Encoded Value-at-Risk: A Predictive Machine for Financial Risk Management.” arXiv, Nov. 2020. doi: 10.48550/arXiv.2011.06742.
10. F. S. Karst, S.-Y. Chong, A. A. Antenor, E. Lin, M. M. Li, and J. M. Leimeister, “Generative AI for Banks: Benchmarks and Algorithms for Synthetic Financial Transaction Data.” arXiv, Dec. 2024. doi: 10.48550/arXiv.2412.14730.
11. G. Madaan, S. K. Asthana, and J. Kaur, “Generative AI: Applications, Models, Challenges, Opportunities, and Future Directions,” in *Generative AI and Implications for Ethics, Security, and Data Management*, IGI Global Scientific Publishing, 2024, pp. 88–121. doi: 10.4018/979-8-3693-8557-9.ch004.
12. B. I. Ibrahim, D. C. Nicolae, A. Khan, S. I. Ali, and A. Khattak, “VAE-GAN Based Zero-shot Outlier Detection,” in *Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control*, in ISCSIC 2020. New York, NY, USA: Association for Computing Machinery, Feb. 2021, pp. 1–5. doi: 10.1145/3440084.3441180.
13. S. Udeshika Munasinghe, R. Rafeh, and S. Rauchas, “Estimating Value at Risk for Central Counterparties: A Generative AI Approach,” in *2024 International Conference on Data Science and Its Applications (ICoDSA)*, Jul. 2024, pp. 305–310. doi: 10.1109/ICoDSA62899.2024.10652178.
14. Y. Tan, H. Zhu, J. Wu, and H. Chai, “DPTVAE: Data-driven prior-based tabular variational autoencoder for credit data synthesizing,” *Expert Systems with Applications*, vol. 241, p. 122071, May 2024, doi: 10.1016/j.eswa.2023.122071.
15. P. Dhoni, “Exploring the Synergy between Generative AI, Data and Analytics in the Modern Age.”
16. F. Clemente, G. M. Ribeiro, A. Quemy, M. S. Santos, R. C. Pereira, and A. Barros, “Ydata-profiling: Accelerating data-centric AI with high-quality data,” *Neurocomputing*, vol. 554, p. 126585, Oct. 2023, doi: 10.1016/j.neucom.2023.126585.
17. 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.
18. M. Sanz-Guerrero and J. Arroyo, “Credit Risk Meets Large Language Models: Building a Risk Indicator from Loan Descriptions in P2P Lending.” arXiv, Aug. 2024. doi: 10.48550/arXiv.2401.16458.
19. Y. Wang, J. Zhao, and Y. Lawryshyn, “GPT-Signal: Generative AI for Semi-automated Feature Engineering in the Alpha Research Process.” arXiv, Oct. 2024. doi: 10.48550/arXiv.2410.18448.
20. E. Sharkey and P. Treleaven, “BERT vs GPT for financial engineering.” arXiv, Apr. 2024. doi: 10.48550/arXiv.2405.12990.
21. S. Balakrishna, V. K. Solanki, and R. G. Crespo, “Generative AI for Smart Data Analytics,” in *Generative AI: Current Trends and Applications*, K. Raza, N. Ahmad, and D. Singh, Eds., Singapore: Springer Nature, 2024, pp. 67–85. doi: 10.1007/978-981-97-8460-8_4.
22. S. Varma, S. Shivam, B. Ray, and S. Biswas, “Reimagining Enterprise Data Management using Generative Artificial Intelligence,” in *2024 11th IEEE Swiss Conference on Data Science (SDS)*, May 2024, pp. 107–114. doi: 10.1109/SDS60720.2024.00023.

23. "Synthetic Data for Financial Anomaly Detection: AI-Driven Approaches to Simulate Rare Events and Improve Model Robustness Journal of Artificial Intelligence Research and Applications." https://aimlstudies.co.uk/index.php/jaira/article/view/221.