FinTech

Review

A Comprehensive Review of Generative AI in Finance

David Kuo Chuen Lee [1], Chong Guan [2], Yinghui Yu [1] and Qinxu Ding [1,*]

1 School of Business, Singapore University of Social Sciences, 463 Clementi Road, Singapore 599494, Singapore
2 SUSS Academy, Singapore University of Social Sciences, Singapore 408601, Singapore
* Correspondence: qinxuding@suss.edu.sg

Abstract: The integration of generative AI (GAI) into the financial sector has brought about significant advancements, offering new solutions for various financial tasks. This review paper provides a comprehensive examination of recent trends and developments at the intersection of GAI and finance. By utilizing an advanced topic modeling method, BERTopic, we systematically categorize and analyze existing research to uncover predominant themes and emerging areas of interest. Our findings reveal the transformative impact of finance-specific large language models (LLMs), the innovative use of generative adversarial networks (GANs) in synthetic financial data generation, and the pressing necessity of a new regulatory framework to govern the use of GAI in the finance sector. This paper aims to provide researchers and practitioners with a structured overview of the current landscape of GAI in finance, offering insights into both the opportunities and challenges presented by these advanced technologies.

Keywords: generative AI; large language models; finance; topic modeling; BERTopic

JEL Classification: G20; O33; C63

1. Introduction

The intersection of generative AI (GAI) and finance has emerged as a rapidly developing area of research and application, revolutionizing various facets of the financial industry. GAI encompasses a broad range of models, such as variational autoencoders (VAEs), generative adversarial networks (GANs), large language models (LLMs), and diffusion models. It has demonstrated significant potential in enhancing financial analytics, improving decision-making processes, and generating synthetic financial data for various applications. This review paper aims to provide a comprehensive overview of recent trends and advancements in GAI's application within the financial sector.

Before the advent of GAI models, traditional methods for financial text mining were primarily used to analyze financial reports and forecast stock prices. However, these methods often relied on the bag-of-words approach to generate word embeddings, which failed to capture the contextual information within sentences. In contrast, pre-trained LLMs, which leverage transformer architecture, are able to capture the complex dependency relationships between words, resulting in more nuanced and contextually aware embeddings. Consequently, GAI models, such as LLMs, have shown significant potential to enhance various financial tasks [1]. Over the past few years, there has been a growing body of research investigating how GAI can address these tasks, evaluating both the results and the associated risks.

In this paper, we aim to evaluate the current state of research on GAI in finance and provide guidance for newcomers in this field. Notably, Ding et al. [2] conducted an extensive analysis of LLMs, while Li et al. [3] provided a broader perspective by examining the current advancements in LLM techniques and their applications. However, neither study explores the intricate relationship between GAI and finance in detail. Lee et al. [4] offered surveys that focus specifically on the impact of LLMs in finance, though they do

check for updates

Citation: Lee, D.K.C.; Guan, C.; Yu, Y.; Ding, Q. A Comprehensive Review of Generative AI in Finance. *FinTech* **2024**, *3*, 460–478. https://doi.org/10.3390/fintech3030025

Academic Editors: Nikiforos T. Laopodis and Eleftheria Kostika

Received: 24 July 2024
Revised: 13 September 2024
Accepted: 15 September 2024
Published: 20 September 2024

not address other GAI models, and their topics were predefined manually. Additionally, Barde and Kulkarni [5], Krause [6], and Mbanyele [7] concentrated on general-purpose LLMs, such as ChatGPT, Bard, and Bing AI.

However, our paper expands the scope by exploring the intersection of GAI and finance, going beyond the focus on LLMs. Additionally, we utilize the advanced topic modeling technique, BERTopic [8], to systematically cluster and analyze the existing research. By leveraging the BERTopic model, we introduce a novel framework for categorizing the current body of work on GAI in finance. In light of the identified gaps and the evolving application of GAI models in finance, this paper aims to address the following research questions:

- RQ1. What are the current trends and advancements in the application of GAI within the financial sector?
- RQ2. How does GAI, beyond LLMs, contribute to solving financial tasks and challenges?
- RQ3. How can BERTopic be used to systematically classify and analyze research on GAI in finance?
- RQ4. What are the risks and challenges associated with the use of GAI in finance, and how have these been addressed in the literature?

The structure of our paper is as follows. In Section 2, we examine the papers relevant to this topic. In Section 3, we introduce the dataset and the topic modeling method used in this study. Section 4 presents the results of our analysis. In Section 5, we provide an in-depth discussion based on the new framework obtained through the topic modeling method. In Section 6, we discuss our contribution, future directions, and potential areas for further research. Finally, we give a conclusion in Section 7.

2. Literature Review

The field of GAI has experienced rapid advancement in recent years. One of the famous GAI methods is the VAE algorithm, developed by Kingma and Welling in 2013 [9]. VAEs applied probabilistic frameworks to generate new data points based on latent representations learned during training. Therefore, it enhances the ability to model complex distributions in data. Subsequently, in 2014, Goodfellow et al. [10] introduced the GAN model, which represented a paradigm shift in generative modeling. It is an adversarial process involving two neural networks: a generative model to capture the data distribution and a discriminative model to distinguish between real data and generated samples. This adversarial framework enabled GANs to produce remarkably realistic outputs across various domains, such as images and texts.

In 2017, Transformers was introduced by Vaswani et al. [11]. It leverages self-attention mechanisms to capture long-range dependencies in sequential data, making them highly effective for various natural language tasks. This innovation inspired the development of LLMs, particularly transformer-based architectures like OpenAI's GPT series [12–15] and Google's BERT [16], which marked a significant advancement in natural language processing. The deployment of LLMs has expanded generative AI into sophisticated applications requiring the comprehension and generation of human-like text. Additionally, Diffusion models have emerged as a novel approach in GAI, focusing on modeling temporal dependencies and irregular patterns in sequential data [17]. These models have shown promise in applications where traditional generative models fall short, particularly in capturing the nuances of dynamic and time-series data.

From the financial researchers' perspective, the impact of LLMs' applications is most significant and far-reaching. In the past two decades, financial text mining has been a popular research area, especially with advancements in computational methods that have made processing large-scale data possible. Beyond conventional financial data sources such as public companies' annual reports and earning announcements, financial researchers have turned to financial news press, regulatory filings, and social media to uncover hidden information and sentimental cues. These insights can be used to predict investment behaviors and trends in stock returns. For instance, Bollen et al. [18] investigated whether measurements of collective mood states derived from large-scale Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. Wisniewski and Yekini [19] analyzed the qualitative part of annual reports of UK-listed companies and used the frequency of words associated with different language indicators to forecast future stock returns. McGurk et al. [20] examined the relationship between investor sentiment and stock returns by employing textual analysis on Twitter posts and found that their investor sentiment measure has a positive and significant effect on abnormal stock returns.

Most of these earlier studies are still exploratory, often reducing text information to "a bag of words" or representing it through dictionary-based sentiment scores. This is primarily because financial texts frequently lack a regular structure, and verbal/textual communication can be subtle and complex. Additionally, financial jargon can vary in meaning depending on the context. To extract more meaningful insights from financial texts, more advanced models are required. LLMs trained on extensive datasets from diverse sources and themes can provide more sophisticated text representations that capture the nuances of financial language. In the literature we surveyed, Gupta [21] simplified the process of assessing annual reports of all the firms by leveraging the capabilities of LLMs, where the insights generated by the LLM are compiled in a Quant-styled dataset and augmented by historical stock price data. Fatemi et al. [1] showcased the remarkable capabilities of LLMs, even smaller models, in both fine-tuning and in-context learning for financial sentiment analysis. Li et al. [22] reported evidence that general-purpose LLMs, especially GPT-4, could outperform domain-specific models in terms of sentiment analysis. Pavlyshenko [23] demonstrated that Llama 2 can be fine-tuned and multitask; when analyzing financial texts, it can return both a structured response and sentiment data in specified JSON format, which can further be loaded directly into predictive models as features. On the other hand, Xing [24] reported that a design framework with heterogeneous LLM agents can be effective in financial sentiment analysis without fine-tuning.

In general, the financial sector, characterized by its vast and complex data, stands to benefit immensely from these advancements in GAI. Traditional data analysis methods often fall short in handling the scale, variability, and intricate patterns inherent in financial data. GAI offers a promising solution by not only managing large datasets effectively but also generating synthetic data close to real-world financial data. This capability is particularly important for applications such as risk management, fraud detection, algorithmic trading, and financial forecasting. Recent years have seen the emergence of specialized GAI models tailored for financial applications. Finance-specific LLMs, such as FinGPT [25,26] and FinPT [27], have been developed to address domain-specific challenges and have shown superior performance compared to general-purpose models in various financial tasks. Despite the promising advancements, the integration of GAI in finance is not without challenges. Issues such as data privacy, model interpretability, regulatory compliance, and the potential for generating biased or misleading data necessitate careful consideration. The ethical and social implications of deploying generative AI in financial decision making further underscore the need for robust frameworks and guidelines to ensure responsible use [28].

Most importantly, there is a limited body of research that surveys the intersection of GAI and finance [3–7], and none of these studies employ advanced topic modeling techniques to mine potential topics from paper abstracts. Using a topic model like BERTopic [8] for paper reviews offers significant advantages over manual review, especially when dealing with large datasets. They usually provide an objective, unbiased grouping of papers, uncovering hidden patterns and emerging trends that may be missed by human reviewers. Additionally, they ensure consistency across the entire dataset and allow researchers to focus on the most relevant areas, making it easier to manage large-scale literature reviews. From the above review, it is evident that while significant progress has been made in applying GAI to finance, many challenges remain underexplored, particularly in terms of integrating various GAI models and uncovering new insights through more sophisticated topic modeling methods.

3. Materials and Methods

This study leverages the Google Scholar database for its extensive coverage, interdisciplinary reach, and up-to-date research indexing. Given that research on GAI, particularly LLMs, is relatively new and rapidly evolving, Google Scholar's comprehensive indexing of reputable authors and institutions is particularly valuable. Many significant papers are submitted to repositories such as arXiv and SSRN, both of which are well indexed by Google Scholar.

We conducted searches using two key combinations: (1) "generative AI and finance" and (2) "large language models and finance." From these searches, we retrieved a total of 90 papers published between 2018 and 2024. These papers were sourced from a diverse range of publishers, including ACM, ACL, arXiv, Curran Associates, Darcy & RoyPress, Elsevier, IEEE, MDPI, Routledge, SSRN, Taylor & Francis Online, and MIT Press. The dataset consists of academic papers from multiple disciplines, with a primary focus on Finance and Computer Science. It also includes papers from related fields such as Economics, Business, Management, and Accounting, ensuring a comprehensive scope. The authors in the dataset are affiliated with a broad spectrum of institutions, ranging from elite universities like Harvard, MIT, and Stanford to tier 1 institutions such as the University of Oxford and the National University of Singapore. Additionally, the dataset includes contributions from leading technology companies like Amazon, Microsoft, and Alibaba, as well as financial institutions such as JP Morgan and Bloomberg. The dataset reflects a geographically diverse set of authors, with substantial representation from the US, Asia, and Europe. This global distribution was intentional to capture a wide range of perspectives on the trends and insights being discussed. As shown in Figure 1, about half of the papers come from Asia, while roughly one-third are from the US. This distribution is consistent with the fact that the US and China are leading countries in research on generative AI models and their applications.

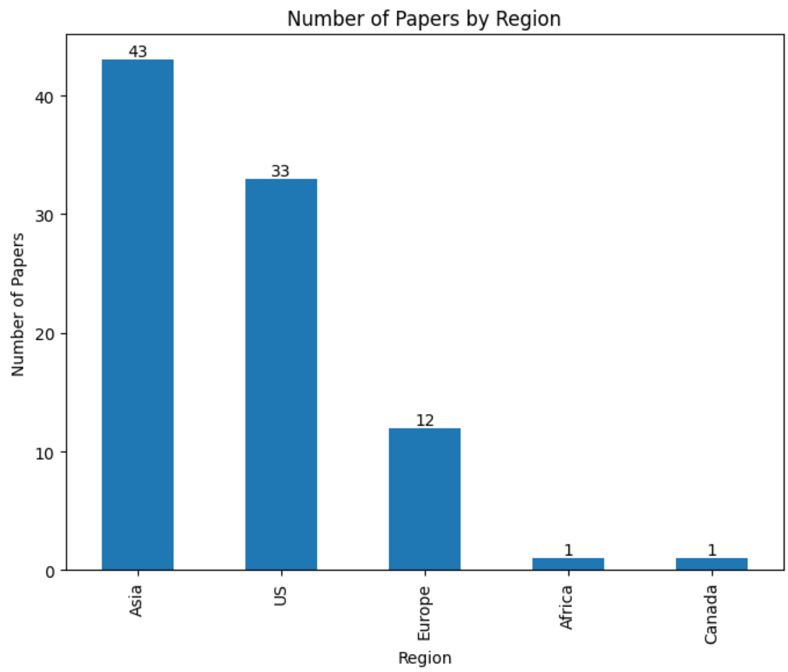

Figure 1. This figure describes data's geographical information. Source: personal processing according to the data.

To effectively cluster and analyze these papers, we applied a robust topic modeling technique known as BERTopic [8]. Topic models are powerful unsupervised tools for uncovering themes and underlying narratives in textual data.

BERTopic Model Specification: To generate coherent topic representations, the BERTopic model will employ three steps. The first step is to generate document embeddings. In this step, the Sentence-BERT (SBERT) framework [29] was employed to convert each paper's abstract into a vector representation. The SBERT uses pre-trained large language models and achieves state-of-the-art performance on various embedding tasks, generating high-quality document vector representations. The second step is to cluster these document embeddings. To achieve a robust cluster result, a dimension reduction technique named UMAP [30] was used, preserving more of the local and global features of high-dimension document embeddings in a lower-dimension space. After that, the HDBSCAN model [31] is used to cluster the reduced embeddings. The benefit of using HDBSCAN is that we can model the outliers as noise, preventing any unrelated documents from being assigned to any cluster. The third step is to model the topic representation with documents in each cluster. The class-based TF-IDF (c-TF-IDF) method [8] is employed to model the importance of a word to a cluster. BERTopic has shown effectiveness in various applications, including systematic reviews [2,21].

BERTopic vs. LDA: While traditional topic modeling methods such as Latent Dirichlet Allocation (LDA) [32] and Non-Negative Matrix Factorization (NMF) [33] represent documents as mixtures of latent topics using a bag-of-words approach, BERTopic enhances this process with advanced techniques. The use of bag-of-words would disregard the semantic relationship among words or lose the context information in a paragraph, resulting in a failure of document representation. In contrast, the BERTopic model takes advantage of pre-trained large language models to encode the meaning of texts into the document embedding. Moreover, the BERTopic model performs better than the LDA and NMF methods with the topic coherence and topic diversity metrics [8].

Our application of BERTopic in this study can be broken down into the following steps:

- Data Preprocessing: First, we convert all text to lowercase to ensure uniformity and reduce redundancy; second, we use nltk.word_tokenize() to split the text into individual tokens and WordNetLemmatizer() to reduce words to their base or root form; finally, we use stopwords.words('english') to eliminate the common stopwords, as they usually do not contribute significantly to the meaning of the text.
- Fit the Model and Transform Documents: We use BERTopic and ClassIFidTransformer to fit the model to our data and transform to discover topics.
- Topics Exploration: After fitting the model, we explore the topics generated by various tools of BERTopic.

4. Results

In this section, we discuss the results obtained from the BERTopic model. After fitting the model, we identified the most frequent topics in our dataset, as shown in Figure 2. Figure 2 reveals four distinct clusters:

- "-1_chatbots_credit_reliable_chatgpt";
- "0_llm_financial_model_task";
- "1_ai_generative_risk_challenge";
- and "2_data_stock_synthetic_market".

Cluster -1 represents all outliers and should be disregarded. Consequently, our focus will be directed towards the examination of the remaining three clusters. Cluster 0 pertains to discussions surrounding the application of LLMs in financial tasks. This cluster highlights the innovative use of LLMs to address various financial modeling and task automation challenges. Cluster 1 delves into the challenges and risks associated with the implementation of GAI within the realm of finance. This cluster underscores the potential risks and regulatory considerations that accompany the deployment of GAI technologies in finance. Cluster 2 centers on the generation of synthetic financial data facilitated by GAI. This cluster emphasizes the role of GAI in creating synthetic datasets, which are essential for tasks such as market simulation and risk assessment.

Additionally, our analysis reveals a significant discrepancy in the distribution of research focus. Specifically, there are approximately 47 papers discussing LLMs in finance, a considerably higher count compared to those addressing the risks and data generation aspects of GAI. This observation underscores that LLMs currently represent the foremost research focus of GAI within the financial domain.

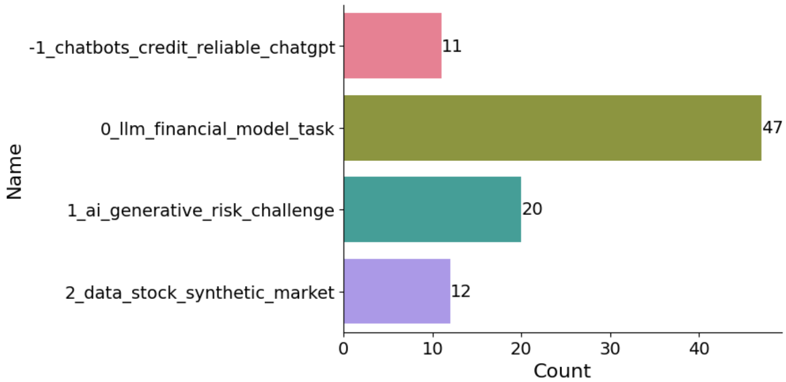

Figure 2. This figure describes the frequent topics obtained from the BERTopic model. Source: personal processing according to the cluster results from BERTopic model.

Observing Table 1, we identify the representative words for each topic generated by the BERTopic model. These words help elucidate the main themes and concepts of each topic, as they are extracted based on their relevance and frequency within the topic. Topic -1 is ignored, as it is an outlier, so our analysis begins with Topic 0: "LLMs for Financial Tasks."

Topic 0: "LLMs for Financial Tasks": In addition to the key terms, such as 'LLM', 'financial', and 'task', that define this topic, other prominent words, like 'model', 'language', and 'large', emphasize the core characteristics of LLMs. These terms highlight the foundational role LLMs play in processing human text and understanding financial language. Financial papers referencing LLMs often focus on the ability of these models to interpret complex financial documents, automate tasks, and assist in decision-making processes. Words like 'benchmark' and 'performance' are frequently used to evaluate how effectively LLMs accomplish financial tasks, whether they are used for explaining phenomena such as market trends, generating reports, or forecasting financial outcomes. The focus on 'benchmark' results indicates the importance of validating LLMs against industry standards, while 'performance' suggests the critical evaluation of their accuracy, speed, and reliability in financial contexts. Thus, these terms collectively indicate that financial research using LLMs aims not only to describe or analyze financial data but also to enhance forecasting and predictive capabilities.

Topic 1: "The Risk and Challenge of Generative AI": Key terms such as 'AI', 'artificial', 'intelligence', 'generative', 'risk', and 'challenge' are central to this topic. These terms reflect the growing discourse around the potential risks and challenges associated with implementing GAI in the financial sector. Additionally, the term 'ethical' frequently appears, highlighting significant concerns about the ethical implications of GAI, particularly in terms of transparency, fairness, and accountability. These ethical considerations are especially critical in the financial industry, where GAI is used to automate decision-making processes, and any misuse or bias in these systems can lead to substantial consequences. Discussions in financial papers often focus on the risks associated with deploying GAI, such as the potential for generating biased or inaccurate financial predictions, or the challenge of ensuring regulatory compliance. Ethical debates are often tied to how GAI is transforming the financial 'industry', affecting everything from risk management to customer service. The common appearance of the word 'paper' in our dataset is largely due to the review-based nature of this article and does not carry significant meaning in this specific context. However, the focus on risks and challenges suggests that researchers are keen on evaluating not only the technical capabilities of GAI but also the broader implications for ethical and responsible use in financial applications.

Topic 2: "Synthetic Financial Data Generation": The words 'synthetic' and 'data' are central to this topic, indicating the focus on the creation of artificial financial datasets. Terms such as 'network', 'GANs', 'learning', and 'adversarial' suggest that many existing studies utilize GANs for generating synthetic financial data, particularly for stock market simulations. In practice, GANs are widely used to produce synthetic 'stock' and 'market' data, such as 'price' 'series', which are essential for financial modeling and analysis. These synthetic datasets play a critical role in financial research, as they allow researchers and practitioners to simulate various market scenarios, test trading strategies, and analyze risk without relying on real-world data, which may be limited or sensitive. In financial papers, the use of synthetic data is typically aimed at overcoming challenges related to data availability, privacy concerns, and the need for large, high-quality datasets for machine learning models. By generating realistic price series and market behaviors, researchers can conduct more robust financial simulations, improving the accuracy and reliability of predictive models. The prominence of terms like 'network' and 'adversarial' reflects the technical focus on GAN architectures, which are key to producing high-fidelity synthetic data that closely mirror actual financial markets. Thus, this topic underscores the practical importance of synthetic data generation for enhancing financial forecasting, risk assessment, and model training.

Table 1. This table includes the representation words of each topic generated by BERTopic model. Source: personal processing according to the cluster results from BERTopic model.

Count	Name	Representation
11	-1_chatbots_credit_reliable_chatgpt	['chatbots', 'credit', 'reliable', 'chatgpt', 'lgp', 'payment', 'user', 'transaction', 'individual', 'process']
47	0_llm_financial_model_task	['llm', 'financial', 'model', 'task', 'language', 'large', 'benchmark', 'performance', 'text', 'instruction']
20	1_ai_generative_risk_challenge	['ai', 'generative', 'risk', 'challenge', 'ethical', 'industry', 'paper', 'intelligence', 'artificial', 'potential']
12	2_data_stock_synthetic_market	['data', 'stock', 'synthetic', 'market', 'network', 'gans', 'learning', 'adversarial', 'series', 'price']

As discussed by Grootendorst [8] in the BERTopic model, the c-TF-IDF method is used to calculate the word score for each cluster or topic. The definition of c-TF-IDF is as follows:

$$W_{x,c} = \|tf_{x,c}\| \times \log(1 + \frac{A}{f_x}), \tag{1}$$

where x is the word, c is the cluster, $tf_{x,c}$ is the frequency of word x in cluster c, f_x refers to word x's frequency across all clusters, and A represents the word's average number per cluster. Figure 3 tells us the 10 most important words for each cluster and their importance. These word scores represent how strongly a word is associated with a particular topic within the model. The higher the score, the more central the word is for defining the topic's theme. Practically speaking, these scores help us identify the most influential terms within each topic, giving us insights into the key concepts driving the discussion in the data. For example, in the context of financial papers, a high score for a word like 'performance' in Topic 0 means that researchers are likely to focus on improving or evaluating financial LLMs' performance in those papers. By identifying the most relevant words, we can better understand the central themes of each topic and how they relate to real-world financial issues like forecasting, strategy optimization, or the evaluation of financial instruments.

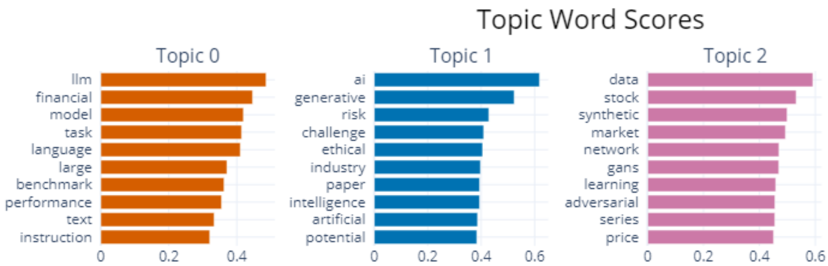

Figure 3. This figure shows the word scores for each topic. Source: personal processing according to the cluster results from BERTopic model.

Figure 4 presents the dendrogram of the hierarchical clustering of the three clusters. The x-axis represents the distance or dissimilarity between clusters, while the vertical lines indicate the points at which clusters are merged. For instance, clusters 0 and 1 merge at a dissimilarity level of approximately 0.6, indicating their relative similarity compared to cluster 2. This visualization helps us understand the relationships and similarities among the identified topics.

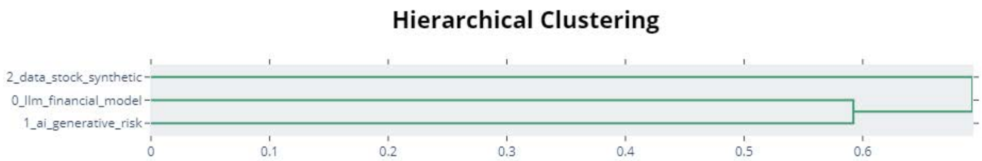

Figure 4. This figure describes the hierarchical clustering result obtained from the BERTopic model. Source: personal processing according to the cluster results from BERTopic model.

5. Discussion

In this section, we delve into three key topics derived from the BERTopic analysis. The first topic explores the application of LLMs in finance across various tasks. This includes discussions on the capability of general-purpose LLMs (e.g., GPTs, Gemini) to address financial problems, the effectiveness of finance-specific LLMs (e.g., FinGPT, FinPT) compared to general-purpose LLMs, and the identification of benchmarks and financial datasets that can be used to fairly evaluate the performance of LLMs in finance. The second topic addresses the potential risks and challenges associated with GAI models for financial applications. This includes an examination of issues such as hallucinations, ethical and social impacts, and financial regulation. Finally, the third topic focuses on the use of GAI for synthetic financial data generation. We will discuss the challenges and areas of focus in this domain, as well as existing work utilizing VAEs, GANs, and diffusion models.

5.1. LLMs for Financial Tasks

5.1.1. General-Purpose LLMs

The rapid advancements in LLMs have ushered in a new era of innovation across various sectors, with finance being a significant beneficiary. Over the past few years, general-purpose LLMs such as GPT-4 have been extensively studied by researchers. Teixeira et al. [34] and Krause [35] presented comprehensive guides to prompt usage in LLMs for financial analysis. Rane et al. [36] explored a comparative analysis of Gemini and ChatGPT, focusing on the discussion of these models' effectiveness and performance in finance and accounting tasks. LLMs have also been tested on a wide range of financial text analytics tasks, demonstrating their versatility and effectiveness [1,21–24,37,38].

General-purpose LLMs can also be used as investment advisors. The potential of LLMs as financial robo-advisors has been rigorously assessed, generally showing good performance [39–42]. Furthermore, Lu et al. [43] demonstrated that ChatGPT could potentially generate portfolios that outperform the markets in out-of-sample tests. Additionally, an innovative LLM multi-agent framework endowed with layered memories has been proposed for stock and fund trading [44]. Their capabilities extend to financial decision making [45,46], financial auditing [47], financial regulatory interpretation [48], financial budgeting [49], financial risk management [50,51], and analyzing climate change issues related to finance [52].

The reasoning capabilities of LLMs have also been studied extensively. Yu et al. [53,54] examined LLMs' ability for explainable financial time series forecasting, demonstrating that these models can generate well-reasoned decisions. Srivastava et al. [55] explored the mathematical reasoning abilities of LLMs on tabular question-answering datasets. Additionally, a comparative study of LLMs for personal financial decision making in a low-resource language, the Yoruba language, was conducted by Sikiru et al. [56]. The results indicate that the performance of LLMs is poor compared to their performance with English financial data, highlighting the need for improvements in low-resource languages. To address the high GPU memory consumption associated with LLMs, Liu et al. [57] presented high-performance GPU-based methods for pretraining and fine-tuning LLMs for financial applications.

5.1.2. Finance-Specific LLMs

General-purpose LLMs offer versatility and adaptability for a wide range of financial tasks but may lack the specialized domain knowledge required for complex financial analyses. In contrast, finance-specific LLMs are exclusively trained on financial data. For instance, BloombergGPT, trained on a diverse range of financial data, showcased superior performance in financial tasks compared to existing general-purpose LLMs [58]. Similarly, FinMA was introduced by fine-tuning LLaMA with a tailored dataset, enabling it to executive various financial tasks [59]. Furthermore, FinGPT emerged as an open-source LLM tailored for the finance sector, providing accessible and transparent resources for researchers and practitioners to develop their FinLLMs [25,26]. Additionally, Li et al. [60] proposed a financial LLM (CFLLM) specially designed to adeptly manage financial texts.

On the other hand, LLMs fine-tuned by financial datasets leverage general-purpose models and enhance their performance in finance-rated tasks through specialized training. For example, Zhang et al. [61] proposed a simple instruction tuning approach to fine-tune general-purpose LLMs, achieving remarkable achievements in financial sentiment analysis. Similarly, Yin et al. [27] introduced FinPT, fine-tuned on LLMs with natural language customer profile text for predictive purposes and pre-trained on a dataset containing Chinese financial data, and a general-purpose dataset. Additionally, Yang et al. [62] presented InvestLM, a financial LLM for investment, tuned on LLaMA with a financial investment instruction dataset. Chen et al. [63] proposed a multiple-expert fine-tuning framework for DISC-FinLLM, a large Chinese financial LLM. Finally, Chu et al. [64] created FLLM, a financial LLM employing multitask prompt-based fine-tuning for data pre-processing and pre-understanding, employing abductive augmentation reasoning (AAR) to overcome manual annotation costs.

Multimodal financial LLMs combine the power of language understanding with the rich information contained in financial data across multiple modalities. By integrating textual, numerical, and visual data, multimodal financial LLMs offer a holistic understanding of financial information, enabling more accurate analyses, predictions, and decision making in the financial domain. For instance, Wang et al. [65] introduced FinVis-GPT, a pioneering multimodal LLM designed to interpret financial charts, marking a significant advancement in the application of multimodal LLMs in finance. Similarly, Bhatia et al. [66] proposed a multimodal financial LLM that integrates textual, numerical, tabular, and image financial data, surpassing the performance of ChatGPT-3.5 in financial tasks.

By leveraging advanced language understanding capabilities and domain-specific knowledge, non-English financial LLMs enable more accurate and nuanced analyses of financial information in languages such as Japanese, Spanish, and beyond. For instance, Hirano [67] developed a Japanese financial-specific LLM through continual pre-training, while Zhang et al. [68] introduced FinMA-ES, an LLM tailored for bilingual financial applications aimed at bridging the gap between Spanish and English financial natural language processing (NLP) capabilities.

5.1.3. Benchmarks of LLMs in Finance

Benchmarks are crucial in evaluating the performance of LLMs in the financial domain. These benchmarks serve as standardized tests or datasets against which various LLMs are assessed, allowing researchers, practitioners, and developers to compare and contrast the effectiveness of different models in handling financial tasks. For instance, Xie et al. [59] proposed a standardized benchmark covering a range of financial tasks, while Zhang et al. [69] introduced FinEval, a benchmark specially tailored for the financial domain knowledge in LLMs, including multiple-choice questions covering various topics in business. Guo et al. [70] presented FinLMEval, offering a comprehensive evaluation of LLMs in financial NLP, and Xie et al. [71] proposed an open-source evaluation benchmark encompassing 35 datasets across 23 financial tasks.

Additionally, Yin et al. [27] provided high-quality datasets on financial risks, Lei et al. [72] introduced CFBenchmark for evaluating LLMs for financial assistance, and Islam et al. [73] proposed FinanceBench for assessing LLMs' performance on open book financial question answering (QA). Zhang et al. [74] curated a practical Text-to-SQL benchmark dataset, beneficial for financial professionals less-skilled in SQL programming. Li et al. [75] introduced the AlphaFin dataset for pretraining or fine-tuning financial analysis LLMs, combining traditional research datasets, real-time financial datasets, and handwritten chain-of-thought (COT) data. Furthermore, Hirano [76] constructed a benchmark specific to the Japanese financial domain, while Xu et al. [77] proposed a benchmark for evaluating Chinese-native financial LLMs.

5.2. The Risk and Challenge of Generative AI

5.2.1. Hallucination

In the realm of GAI, the phenomenon of hallucination manifests when a model generates data or information that deviates from accurately representing the underlying patterns or realities. This occurrence can engender misleading predictions, instill false impressions, or prompt erroneous conclusions based on the generated data. The ramifications of such hallucination are particularly acute within the financial domain, where precision in data and prognostications holds paramount importance for informed decision-making processes. An illustrative case in point is the investigation into GPT-3's efficacy in analyzing climate change related to its financial implications, as undertaken by Leippold [52]. Nevertheless, the inquiry unveiled an issue of hallucination during an interview with the GPT-3 model. Further empirical examination of the financial tasks' hallucination behaviors was conducted by Kang and Liu [78], who evaluated the efficacy of various methods, such as few-shot learning and the retrieval augmentation generation method, in mitigating hallucination in LLMs. Their findings underscored the substantial presence of hallucination behaviors in off-the-shelf LLMs when applied to financial tasks.

Consequently, it is imperative that forthcoming research endeavors prioritize strategies aimed at circumventing hallucination by GAI models. Roychowdhury [79] delineated three major stages to design hallucination-minimized LLM-based solutions tailored for the financial domain's decision-makers: prototyping, scaling, and LLM evolution using feedback. These measures ensure that GAI chatbots, autonomous reports, and alerts are reliable and high-quality, thereby facilitating key decision-making processes.

5.2.2. Ethical and Social Impact

GAI represents a remarkable tool for discerning users, yet it necessitates critical reflection on the ethical implications and societal ramifications of its integration into the financial industry. Rane [28] extensively explored the multifaceted role and challenges encountered by GAI tools within the intricate realms of finance and accounting, elucidating both their transformative potential and the hurdles that must be overcome to genuinely revolutionize the financial landscape. Similarly, Kalia [80] scrutinized the impact of GAI on the financial sector, advocating for the establishment of robust privacy frameworks, stringent enforcement of data protection regulations, and the promotion of responsible and ethical deployment of these technologies by organizations and policymakers. Sarker [81] delved into the myriad perspectives on LLMs, highlighting both their potentiality and associated risk factors, underscored by heightened awareness. Additionally, Krause [82] deliberated on the potential risks posed by GAI tools in finance and proposed comprehensive mitigation strategies for businesses, emphasizing the importance of employing GAI tools within closed networks, utilizing secure training data, the implementation of robust security measures, providing employee training, and monitoring outputs.

Moreover, Remolina [83] emphasized the necessity of context- and sector-specific debates to effectively address the risks and challenges inherent in GAI deployment. For instance, generating financial advice content entails different societal implications than generating imagery of a turtle. Lo and Ross [84] focused on three primary challenges confronting most LLM applications: domain-specific expertise tailored to users' unique circumstances, adherence to moral and ethical standards, and compliance with regulatory guidelines and oversight. Lastly, Yusof and Roslan [85] underscored the imperative of continuous evaluation and adaptation of AI technologies within banking to simultaneously maximize benefits and mitigate associated risks. Collectively, these perspectives underscore the intricate interplay between technological advancement, ethical considerations, and regulatory imperatives within the financial domain, urging stakeholders to navigate these complexities with foresight and diligence.

5.2.3. Financial Regulation

The emergence of robo-advisors and advanced GAI models such as GPT-4 and ChatGPT heralds efficiency gains but also presents distinctive regulatory hurdles. Caspi et al. [86] delved into the regulatory landscape surrounding financial advice in an era increasingly shaped by GAI. Their study scrutinized the extant legal framework governing investment advisors and broker-dealers in the United States, while also examining the ascendancy and impact of robo-advisors and GAI. The pivotal role assumed by AI in the provision of financial advice has necessitated a judicious approach to regulatory strategies. Each regulatory strategy, whether predicated on disclosure mandates or outright prohibitions, carries its own array of benefits and potential challenges. A robust regulatory framework demands more than a cursory understanding; it mandates a comprehensive grasp of these AI technologies and their ramifications.

5.3. Synthetic Financial Data Generation

5.3.1. Challenges of Generating Synthetic Data

The financial services sector produces an enormous amount of highly intricate and diverse data. These data are frequently compartmentalized within organizations for several reasons, such as regulatory compliance and operational requirements. Consequently, the sharing of data both across various business units and with external entities like the research community is greatly restricted. Therefore, exploring techniques for creating synthetic financial datasets that maintain the characteristics of real data while ensuring the privacy of the involved parties is crucial. Assefa et al. [87] emphasized the growing need for the financial domain's effective synthetic data generation and highlighted the following three areas of focus for the academic community:

- Realistic synthetic dataset generation;
- Similarity calculation between real and generated datasets;
- Ensuring privacy constraints with the generative process.

While these challenges are also present in other domains, the financial sector's additional regulatory and privacy requirements add a layer of complexity. This presents a unique opportunity to study synthetic data generation within the context of financial services.

5.3.2. Existing Works by VAE, GAN, and Diffusion Models

From the literature, it is evident that most existing works on financial synthetic data generation employ GANs. For instance, Zhang et al. [88] introduced a novel GAN architecture to forecast stock closing prices, utilizing a Long Short-Term Memory (LSTM) network as the generator. The LSTM generator captures the data distributions of stocks from the given market data, generating data with similar distributions. Takahashi et al. [89] developed the FIN-GAN model for financial time-series modeling, which learns the properties of data and generates realistic time-series data.

Koshiyama et al. [90] proposed using conditional GANs (cGANs) for trading strategy calibration and aggregation, utilizing the generated samples for ensemble modeling. Bezzina [91] examined the correlation characteristics of synthetic financial time series data generated by TimeGAN, demonstrating that TimeGAN preserves the correlation structure in multi-stock datasets. Ramzan et al. [92] explored GANs for generating synthetic data, emphasizing the generation of datasets that mimic the statistical properties of input data without revealing sensitive information. Vuletić et al. [93] investigated GANs for probability forecasting of financial time series, using a novel economics-driven loss function in the generator. Ljung [94] assessed CTGAN's ability to generate synthetic data, while He and Kita [95] employed a hybrid sequential GAN model with three training strategies using S&P500 data.

In addition to GANs, VAEs and diffusion models have also been employed for financial synthetic data generation. Dalmasso et al. [96] introduced PayVAE, a generative model designed to learn the temporal and relational structure of financial transactions directly from data. Applied to a real peer-to-peer payment dataset, PayVAE demonstrated its capability to generate realistic transactions. Huang et al. [97] developed a novel generative framework called FTS-Diffusion, which consists of three modules designed to model irregular and scale-invariant patterns in financial time series.

6. Contribution and Future Research Agenda

This research unveils significant theoretical and managerial implications, crucial for comprehending and harnessing the potential of advanced GAI technologies within the financial domain. By summarizing key past research themes, the paper elucidates the evolving landscape of AI technologies and their applications in finance, providing a comprehensive synthesis that advances understanding and informs future research directions.

6.1. Theoretical Contribution

Theoretically, this review highlights the transformative potential of LLMs within the financial domain. Building upon previous research that summarizes LLMs in finance [3], it emphasizes the need to differentiate between general-purpose and finance-specific LLMs by categorizing research based on training data and application areas. The review then synthesizes findings on the performance and capabilities of prominent LLMs, such as GPT-4 and BloombergGPT, in diverse financial tasks encompassing text analysis, investment advisory, and decision support. This synthesis establishes a foundational framework for future research endeavors to explore critical aspects like performance benchmarks, evaluation criteria, and optimization strategies tailored for LLMs operating in financial contexts. Additionally, the discussion on LLMs' reasoning abilities and their application in financial forecasting and decision making underscores crucial areas for theoretical exploration. These areas include the development of models capable of generating well-reasoned financial decisions and the enhancement of LLMs to operate effectively within low-resource language environments.

Furthermore, the review delves into the significant ethical and risk considerations surrounding GAI models within the financial sector. Previous research has highlighted the need to address ethical, risk, and synthetic data considerations of GAI [28]. By critically examining the phenomenon of hallucination [79], the paper contributes to the theoretical comprehension of the risks associated with GAI. Additionally, the exploration of ethical concerns, encompassing data privacy and responsible AI utilization, lays the groundwork for the development of robust ethical guidelines and frameworks [2,98]. These frameworks can serve as a foundation for guiding the deployment of GAI technologies in a manner that adheres to ethical principles and aligns with societal expectations. This theoretical foundation is paramount in ensuring that future research and applications of AI in finance are demonstrably ethical and in accordance with societal norms.

The review further underscores the critical role of synthetic financial data generation through the utilization of advanced models like GANs, VAEs, and diffusion models. By synthesizing past research on methodologies for constructing realistic synthetic datasets while adhering to privacy concerns and regulatory compliance, the paper contributes to the theoretical underpinnings of synthetic data generation in a financial context. This synthesis serves as a roadmap for future research endeavors to refine the accuracy, utility, and ethical considerations associated with synthetic financial data.

6.2. Managerial Implications

This review offers valuable managerial insights into the practical applications of GAI technologies within the financial sector. By comprehensively summarizing the capabilities of LLMs across various financial tasks, the paper serves as a practical guide for managers seeking to leverage these models to optimize operational efficiency and enhance decision-making processes. The analysis of performance metrics and evaluation criteria for LLMs equips managers with the tools necessary to develop effective implementation strategies for integrating these technologies into financial operations [3].

Furthermore, the review emphasizes the critical need to address ethical considerations and potential risks associated with the deployment of GAI technologies. Financial managers can benefit from the comprehensive frameworks and risk management strategies outlined within the review, which provide guidance for the responsible use of AI and the mitigation of potential legal and reputational risks. This proactive approach ensures the ethical and responsible implementation of AI technologies, fostering trust and confidence amongst stakeholders [48].

Finally, the review highlights the practical applications of synthetic financial data generation, offering managers valuable insights into how these technologies can be leveraged to enhance decision making and bolster the resilience of financial institutions. By investing in the generation and utilization of synthetic financial data, managers gain access to robust datasets that can be employed for financial modeling, stress testing, and scenario analysis, ultimately leading to improved strategic planning and risk management capabilities [87].

6.3. Future Research Agenda

Our research advocates for a future research agenda that prioritizes the exploration of a synergistic relationship between three key areas: differentiation and performance, ethical and risk considerations, and synthetic financial data generation. As depicted in Figure 5, these domains interact dynamically, where advancements in one area can amplify progress in the others.

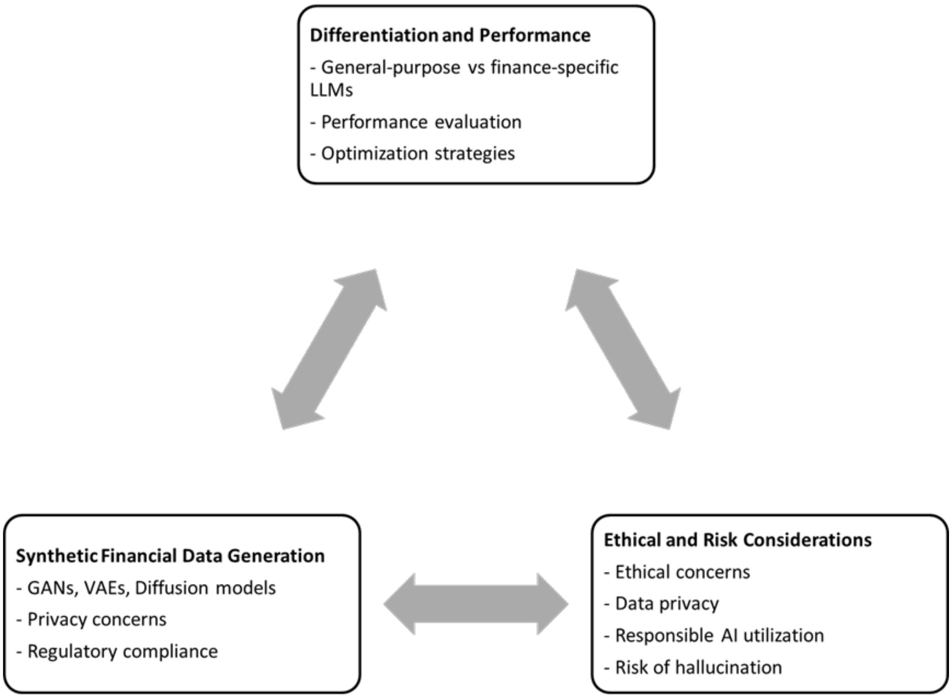

Figure 5. Future research agenda. Source: own processing.

6.3.1. Intertwined Ethics and Performance Optimization

Central to this approach is the integration of ethical considerations within the performance optimization framework for LLMs employed in finance. This entails ensuring that while LLMs are optimized for financial tasks, the resultant outcomes remain fair, transparent, and accountable. Ethical considerations encompassing data privacy, responsible utilization of AI, and the potential for model hallucination (generating irrelevant or misleading outputs) must be embedded within performance evaluation and optimization strategies. This ensures that advancements in model performance do not come at the expense of ethical standards, fostering the development of trustworthy and socially responsible AI applications in the financial domain.

6.3.2. Synthetic Data: A Boon for Performance Benchmarking

Synthetic financial data have the potential to serve as a valuable tool for evaluating and benchmarking the performance of LLMs specifically designed for finance, while simultaneously safeguarding privacy concerns. By leveraging advanced techniques like GANs, VAEs, and diffusion models, researchers can create realistic yet privacy-preserving datasets. These datasets can facilitate extensive testing and validation of models, guaranteeing that performance metrics are reliable and reflect real-world scenarios. It is imperative, however, to address privacy concerns and uphold regulatory compliance during synthetic data generation to maintain public trust and adhere to legal requirements.

6.3.3. Ethical Considerations in Synthetic Data Generation

A crucial aspect of this research agenda involves addressing the ethical risks associated with synthetic data generation. Researchers must meticulously consider the ethical implications and potential risks to ensure responsible use of this data. This entails guaranteeing that the synthetic data are generated in a manner that does not inadvertently introduce biases or other ethical concerns. Specific considerations include potential biases in data generation, the possibility of data misuse, and the need for robust regulatory frameworks to guide ethical practices. By adhering to these principles, researchers can uphold the integrity and reliability of the models trained on such data.

By pursuing research that investigates these interconnected areas, significant contributions can be made towards the development of robust, ethical, and high-performing LLMs within the financial sector. This holistic approach fosters a collaborative environment where advancements in one area bolster developments in the others, ultimately leading to a more integrated and responsible application of LLMs in the realm of finance. Figure 5 serves as a visual representation of the proposed future research agenda, highlighting the critical areas of focus and their interconnections, thereby providing a roadmap for researchers to create comprehensive and ethically sound financial AI systems.

The future research agenda outlined in Figure 5 emphasizes three pivotal areas of exploration to advance the relationship between GAI and finance. First, it calls for a deeper differentiation and performance assessment between general-purpose LLMs and finance-specific models, stressing the need for rigorous performance evaluation and optimization strategies to better align with financial domain requirements. Second, this paper underscores the significance of synthetic financial data generation through sophisticated generative models such as GANs, VAEs, and diffusion models. This approach is paramount for mitigating privacy risks and ensuring adherence to regulatory frameworks, thereby bolstering the trustworthiness and practical utility of GAI in finance. For instance, a worthy avenue of exploration is the potential impact of generative art Non-Fungible Tokens (NFTs) on NFT price volatility [99]. Third, it underscores the necessity of ethical and risk considerations in deploying GAI, particularly regarding data privacy, responsible AI utilization, and minimizing risks such as hallucination. Together, these focal points provide a comprehensive roadmap for future research, advocating for a balanced approach that integrates technical advancement, ethical governance, and domain-specific customization at the intersection of GAI and finance [98].

7. Conclusions

The integration of GAI into the financial sector is revolutionizing various facets of financial technology, particularly in areas such as data analysis, predictive modeling, and synthetic data generation [100]. By harnessing advanced AI techniques, financial institutions can process voluminous datasets with unprecedented efficiency, uncover patterns and trends that were previously hidden, and generate synthetic datasets that are crucial for developing robust models without compromising sensitive information. This review highlights the significant advancements achieved with key generative AI models like VAEs, GANs, LLMs, and diffusion models [101]. These models have opened new avenues for predictive analytics, risk assessment, fraud detection, and personalized financial services. However, while these cutting-edge technologies offer substantial benefits, there are still considerable challenges that need to be addressed, such as ensuring data privacy, improving model interpretability, and maintaining regulatory compliance in an evolving legal landscape. Future research should prioritize enhancing model robustness to better withstand adversarial attacks and market volatility, increasing transparency to build trust among stakeholders, and establishing ethical standards to guide the responsible use of generative AI in finance. This includes developing frameworks for responsible AI governance, ensuring accountability, and minimizing biases that could lead to unfair outcomes [102]. By tackling these challenges comprehensively, the financial industry can fully realize the transformative potential of generative AI, driving innovation while safeguarding the integrity and fairness of financial systems.

Author Contributions: The conceptualization, methodology, validation, formal analysis, investigation, and writing of the paper were completed by all authors equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by MOE Start-up Research Funding (RFE23003).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fatemi, S.; Hu, Y. A Comparative Analysis of Fine-Tuned LLMs and Few-Shot Learning of LLMs for Financial Sentiment Analysis. *arXiv* **2023**, arXiv:2312.08725.
2. Ding, Q.; Ding, D.; Wang, Y.; Guan, C.; Ding, B. Unraveling the landscape of large language models: A systematic review and future perspectives. *J. Electron. Bus. Digit. Econ.* **2023**, *3, 3–19* . [CrossRef]
3. Li, Y.; Wang, S.; Ding, H.; Chen, H. Large language models in finance: A survey. In Proceedings of the Fourth ACM International Conference on AI in Finance, Brooklyn, NY, USA, 27–29 November 2023; pp. 374–382.
4. Lee, J.; Stevens, N.; Han, S.C.; Song, M. A Survey of Large Language Models in Finance (FinLLMs). *arXiv* **2024**, arXiv:2402.02315.
5. Barde, K.; Kulkarni, P.A. Applications of Generative AI in Fintech. In Proceedings of the Third International Conference on AI-ML Systems, Bangalore, India, 25–28 October 2023; pp. 1–5.
6. Krause, D. Large Language Models and Generative AI in Finance: An Analysis of ChatGPT, Bard, and Bing AI. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4511540 (accessed on 7 June 2024).
7. Mbanyele, W. Generative AI and ChatGPT in Financial Markets and Corporate Policy: A Comprehensive Review. 2024. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4745990 (accessed on 7 June 2024).
8. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
9. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
10. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*; The MIT Press: Montreal, QC, Canada, 2014.
11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30*; NeurIPS Foundation: Long Beach, CA, USA, 2017.
12. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: https://hayate-lab.com/wp-content/uploads/2023/05/43372bfa750340059ad87ac8e538c53b.pdf (accessed on 7 June 2024).
13. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
14. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
15. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27730–27744.
16. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
17. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
18. Bollen, J.; Mao, H.; Zeng, X. Twitter mood predicts the stock market. *J. Comput. Sci.* **2011**, *2*, 1–8. [CrossRef]
19. Wisniewski, T.P.; Yekini, L.S. Stock market returns and the content of annual report narratives. In *Proceedings of the Accounting Forum*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 39, pp. 281–294.
20. McGurk, Z.; Nowak, A.; Hall, J.C. Stock returns and investor sentiment: Textual analysis and social media. *J. Econ. Financ.* **2020**, *44*, 458–485. [CrossRef]
21. Gupta, U. Gpt-investar: Enhancing stock investment strategies through annual report analysis with large language models. *arXiv* **2023**, arXiv:2309.03079. [CrossRef]
22. Li, X.; Chan, S.; Zhu, X.; Pei, Y.; Ma, Z.; Liu, X.; Shah, S. Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? A Study on Several Typical Tasks. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, Singapore, 6–10 December 2023; pp. 408–422.
23. Pavlyshenko, B.M. Financial News Analytics Using Fine-Tuned Llama 2 GPT Model. *arXiv* **2023**, arXiv:2308.13032.
24. Xing, F. Designing Heterogeneous LLM Agents for Financial Sentiment Analysis. *arXiv* **2024**, arXiv:2401.05799. [CrossRef]
25. Liu, X.Y.; Wang, G.; Zha, D. Fingpt: Democratizing internet-scale data for financial large language models. *arXiv* **2023**, arXiv:2307.10485.
26. Yang, H.; Liu, X.Y.; Wang, C.D. Fingpt: Open-source financial large language models. *arXiv* **2023**, arXiv:2306.06031. [CrossRef]
27. Yin, Y.; Yang, Y.; Yang, J.; Liu, Q. FinPT: Financial Risk Prediction with Profile Tuning on Pretrained Foundation Models. *arXiv* **2023**, arXiv:2308.00065.
28. Rane, N. Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Finance and Accounting. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4603206 (accessed on 7 June 2024).
29. Reimers, N. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv* **2019**, arXiv:1908.10084.
30. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
31. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [CrossRef]
32. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
33. Févotte, C.; Idier, J. Algorithms for nonnegative matrix factorization with the β-divergence. *Neural Comput.* **2011**, *23*, 2421–2456. [CrossRef]

34. Teixeira, A.C.; Marar, V.; Yazdanpanah, H.; Pezente, A.; Ghassemi, M. 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, Brooklyn, NY, USA, 27–29 November 2023; pp. 340–348.
35. Krause, D. Proper Generative AI Prompting for Financial Analysis. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453664 (accessed on 7 June 2024).
36. Rane, N.; Choudhary, S.; Rane, J. Gemini or ChatGPT? Efficiency, Performance, and Adaptability of Cutting-Edge Generative Artificial Intelligence (AI) in Finance and Accounting. 2024. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4731283 (accessed on 7 June 2024).
37. Callanan, E.; Mbakwe, A.; Papadimitriou, A.; Pei, Y.; Sibue, M.; Zhu, X.; Ma, Z.; Liu, X.; Shah, S. Can gpt models be financial analysts? an evaluation of chatgpt and gpt-4 on mock cfa exams. *arXiv* **2023**, arXiv:2310.08678.
38. Zhao, H.; Liu, Z.; Wu, Z.; Li, Y.; Yang, T.; Shu, P.; Xu, S.; Dai, H.; Zhao, L.; Mai, G.; et al. Revolutionizing finance with llms: An overview of applications and insights. *arXiv* **2024**, arXiv:2401.11641.
39. Niszczota, P.; Abbas, S. GPT has become financially literate: Insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice. *Financ. Res. Lett.* **2023**, *58*, 104333. [CrossRef]
40. Lakkaraju, K.; Vuruma, S.K.R.; Pallagani, V.; Muppasani, B.; Srivastava, B. Can llms be good financial advisors?: An initial study in personal decision making for optimized outcomes. *arXiv* **2023**, arXiv:2307.07422.
41. Huang, Z.; Che, C.; Zheng, H.; Li, C. Research on Generative Artificial Intelligence for Virtual Financial Robo-Advisor. *Acad. J. Sci. Technol.* **2024**, *10*, 74–80. [CrossRef]
42. Liu, Y.; Wang, J. Analysis of Financial Market using Generative Artificial Intelligence. *Acad. J. Sci. Technol.* **2024**, *11*, 21–25. [CrossRef]
43. Lu, F.; Huang, L.; Li, S. ChatGPT, Generative AI, and Investment Advisory. 2023. Available online: https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=4519182 (accessed on 7 June 2024).
44. Li, Y.; Yu, Y.; Li, H.; Chen, Z.; Khashanah, K. Tradinggpt: Multi-agent system with layered memory and distinct characters for enhanced financial trading performance. *arXiv* **2023**, arXiv:2309.03736.
45. Yu, Y.; Li, H.; Chen, Z.; Jiang, Y.; Li, Y.; Zhang, D.; Liu, R.; Suchow, J.W.; Khashanah, K. FinMem: A performance-enhanced LLM trading agent with layered memory and character design. *Proc. AAAI Symp. Ser.* **2024**, *3*, 595–597. [CrossRef]
46. Lakkaraju, K.; Jones, S.E.; Vuruma, S.K.R.; Pallagani, V.; Muppasani, B.C.; Srivastava, B. LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making. In Proceedings of the Fourth ACM International Conference on AI in Finance, Brooklyn, NY, USA, 27–29 November 2023; pp. 100–107.
47. Hillebrand, L.; Berger, A.; Deußer, T.; Dilmaghani, T.; Khaled, M.; Kliem, B.; Loitz, R.; Pielka, M.; Leonhard, D.; Bauckhage, C.; et al. Improving zero-shot text matching for financial auditing with large language models. In Proceedings of the ACM Symposium on Document Engineering 2023, Limerick, Ireland, 22–25 August 2023; pp. 1–4.
48. Cao, Z.; Feinstein, Z. Large Language Model in Financial Regulatory Interpretation. *arXiv* **2024**, arXiv:2405.06808.
49. de Zarzà, I.; de Curtò, J.; Roig, G.; Calafate, C.T. Optimized financial planning: Integrating individual and cooperative budgeting models with llm recommendations. *AI* **2023**, *5*, 91–114. [CrossRef]
50. Chen, B.; Wu, Z.; Zhao, R. From fiction to fact: The growing role of generative AI in business and finance. *J. Chin. Econ. Bus. Stud.* **2023**, *21*, 471–496. [CrossRef]
51. Wang, Y. Generative AI in Operational Risk Management: Harnessing the Future of Finance. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4452504 (accessed on 7 June 2024).
52. Leippold, M. Thus spoke GPT-3: Interviewing a large-language model on climate finance. *Financ. Res. Lett.* **2023**, *53*, 103617. [CrossRef]
53. Yu, X.; Chen, Z.; Lu, Y. Harnessing LLMs for temporal data-a study on explainable financial time series forecasting. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, Singapore, 6–10 December 2023; pp. 739–753.
54. Yu, X.; Chen, Z.; Ling, Y.; Dong, S.; Liu, Z.; Lu, Y. Temporal Data Meets LLM–Explainable Financial Time Series Forecasting. *arXiv* **2023**, arXiv:2306.11025.
55. Srivastava, P.; Malik, M.; Ganu, T. Assessing LLMs' Mathematical Reasoning in Financial Document Question Answering. *arXiv* **2024**, arXiv:2402.11194.
56. Sikiru, R.D.; Adekanmbi, O.; Soronnadi, A. Comparative Study of LLMs for Personal Financial Decision in Low Resource Language. In Proceedings of the 5th Workshop on African Natural Language Processing, Vienna, Austria, 11 May 2024.
57. Liu, X.Y.; Zhang, J.; Wang, G.; Tong, W.; Walid, A. FinGPT-HPC: Efficient Pretraining and Finetuning Large Language Models for Financial Applications with High-Performance Computing. *arXiv* **2024**, arXiv:2402.13533.
58. Wu, S.; Irsoy, O.; Lu, S.; Dabravolski, V.; Dredze, M.; Gehrmann, S.; Kambadur, P.; Rosenberg, D.; Mann, G. BloombergGPT: A Large Language Model for Finance. *arXiv* **2024**. Available online: https://arxiv.org/pdf/2303.17564.pdf (accessed on 7 June 2024).
59. Xie, Q.; Han, W.; Zhang, X.; Lai, Y.; Peng, M.; Lopez-Lira, A.; Huang, J. Pixiu: A large language model, instruction data and evaluation benchmark for finance. *arXiv* **2023**, arXiv:2306.05443.
60. Li, J.; Bian, Y.; Wang, G.; Lei, Y.; Cheng, D.; Ding, Z.; Jiang, C. Cfgpt: Chinese financial assistant with large language model. *arXiv* **2023**, arXiv:2309.10654.

61. Zhang, B.; Yang, H.; Liu, X.Y. Instruct-fingpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. *arXiv* **2023**, arXiv:2306.12659. [CrossRef]
62. Yang, Y.; Tang, Y.; Tam, K.Y. Investlm: A large language model for investment using financial domain instruction tuning. *arXiv* **2023**, arXiv:2309.13064.
63. Chen, W.; Wang, Q.; Long, Z.; Zhang, X.; Lu, Z.; Li, B.; Wang, S.; Xu, J.; Bai, X.; Huang, X.; et al. Disc-finllm: A chinese financial large language model based on multiple experts fine-tuning. *arXiv* **2023**, arXiv:2310.15205.
64. Chu, Z.; Guo, H.; Zhou, X.; Wang, Y.; Yu, F.; Chen, H.; Xu, W.; Lu, X.; Cui, Q.; Li, L.; et al. Data-centric financial large language models. *arXiv* **2023**, arXiv:2310.17784.
65. Wang, Z.; Li, Y.; Wu, J.; Soon, J.; Zhang, X. Finvis-gpt: A multimodal large language model for financial chart analysis. *arXiv* **2023**, arXiv:2308.01430.
66. Bhatia, G.; Nagoudi, E.M.B.; Cavusoglu, H.; Abdul-Mageed, M. FinTral: A Family of GPT-4 Level Multimodal Financial Large Language Models. *arXiv* **2024**, arXiv:2402.10986.
67. Hirano, M.; Imajo, K. Construction of Domain-specified Japanese Large Language Model for Finance through Continual Pre-training. *arXiv* **2024**, arXiv:2404.10555. [CrossRef]
68. Zhang, X.; Xiang, R.; Yuan, C.; Feng, D.; Han, W.; Lopez-Lira, A.; Liu, X.Y.; Ananiadou, S.; Peng, M.; Huang, J.; et al. Dólares or Dollars? Unraveling the Bilingual Prowess of Financial LLMs Between Spanish and English. *arXiv* **2024**, arXiv:2402.07405.
69. Zhang, L.; Cai, W.; Liu, Z.; Yang, Z.; Dai, W.; Liao, Y.; Qin, Q.; Li, Y.; Liu, X.; Liu, Z.; et al. Fineval: A chinese financial domain knowledge evaluation benchmark for large language models. *arXiv* **2023**, arXiv:2308.09975.
70. Guo, Y.; Xu, Z.; Yang, Y. Is chatgpt a financial expert? evaluating language models on financial natural language processing. *arXiv* **2023**, arXiv:2310.12664.
71. Xie, Q.; Han, W.; Chen, Z.; Xiang, R.; Zhang, X.; He, Y.; Xiao, M.; Li, D.; Dai, Y.; Feng, D.; et al. The FinBen: An Holistic Financial Benchmark for Large Language Models. *arXiv* **2024**, arXiv:2402.12659.
72. Lei, Y.; Li, J.; Jiang, M.; Hu, J.; Cheng, D.; Ding, Z.; Jiang, C. Cfbenchmark: Chinese financial assistant benchmark for large language model. *arXiv* **2023**, arXiv:2311.05812.
73. Islam, P.; Kannappan, A.; Kiela, D.; Qian, R.; Scherrer, N.; Vidgen, B. Financebench: A new benchmark for financial question answering. *arXiv* **2023**, arXiv:2311.11944.
74. Zhang, C.; Mao, Y.; Fan, Y.; Mi, Y.; Gao, Y.; Chen, L.; Lou, D.; Lin, J. FinSQL: Model-Agnostic LLMs-based Text-to-SQL Framework for Financial Analysis. *arXiv* **2024**, arXiv:2401.10506.
75. Li, X.; Li, Z.; Shi, C.; Xu, Y.; Du, Q.; Tan, M.; Huang, J.; Lin, W. AlphaFin: Benchmarking Financial Analysis with Retrieval-Augmented Stock-Chain Framework. *arXiv* **2024**, arXiv:2403.12582.
76. Hirano, M. Construction of a japanese financial benchmark for large language models. *arXiv* **2024**, arXiv:2403.15062. [CrossRef]
77. Xu, L.; Zhu, L.; Wu, Y.; Xue, H. SuperCLUE-Fin: Graded Fine-Grained Analysis of Chinese LLMs on Diverse Financial Tasks and Applications. *arXiv* **2024**, arXiv:2404.19063.
78. Kang, H.; Liu, X.Y. Deficiency of Large Language Models in Finance: An Empirical Examination of Hallucination. *arXiv* **2023**, arXiv:2311.15548.
79. Roychowdhury, S. Journey of hallucination-minimized generative ai solutions for financial decision makers. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining, Mérida, Mexico, 4–8 March 2024; pp. 1180–1181.
80. Kalia, S. Potential Impact of Generative Artificial Intelligence (AI) on the Financial Industry. *Int. J. Cybern. Inform.* **2023**, *12*, 37. [CrossRef]
81. Sarker, I.H. LLM potentiality and awareness: A position paper from the perspective of trustworthy and responsible AI modeling. *Discov. Artif. Intell.* **2024**, *4*, 40. [CrossRef]
82. Krause, D. Mitigating Risks for Financial Firms Using Generative AI Tools. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4452600 (accessed on 7 June 2024).
83. Remolina, N. *Generative AI in Finance: Risks and Potential Solutions*; Singapore Management University School of Law Research Paper Forthcoming, SMU Centre for AI & Data Governance Research Paper Forthcoming: Singapore, 2023.
84. Lo, A.W.; Ross, J. Can ChatGPT Plan Your Retirement?: Generative AI and Financial Advice. 2024. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4722780 (accessed on 7 June 2024).
85. Yusof, S.A.B.M.; Roslan, F.A.B.M. The Impact of Generative AI in Enhancing Credit Risk Modeling and Decision-Making in Banking Institutions. *Emerg. Trends Mach. Intell. Big Data* **2023**, *15*, 40–49.
86. Caspi, I.; Felber, S.S.; Gillis, T.B. Generative AI and the Future of Financial Advice Regulation. In Proceedings of the Generative AI and Law Workshop at ICML 2023, Hawaii Convention Center, Honolulu, HI, USA, 28–29 July 2023.
87. Assefa, S.A.; Dervovic, D.; Mahfouz, M.; Tillman, R.E.; Reddy, P.; Veloso, M. Generating synthetic data in finance: Opportunities, challenges and pitfalls. In Proceedings of the First ACM International Conference on AI in Finance, New York, NY, USA, 15–16 October 2020; pp. 1–8.
88. Zhang, K.; Zhong, G.; Dong, J.; Wang, S.; Wang, Y. Stock market prediction based on generative adversarial network. *Procedia Comput. Sci.* **2019**, *147*, 400–406. [CrossRef]
89. Takahashi, S.; Chen, Y.; Tanaka-Ishii, K. Modeling financial time-series with generative adversarial networks. *Phys. A Stat. Mech. Appl.* **2019**, *527*, 121261. [CrossRef]

90. Koshiyama, A.; Firoozye, N.; Treleaven, P. Generative adversarial networks for financial trading strategies fine-tuning and combination. *Quant. Financ.* **2021**, *21*, 797–813. [CrossRef]
91. Bezzina, P. Improving Portfolio Construction Using Deep Generative Machine Learning Models Applying Generative Models on Financial Market Data. Master's Thesis, University of Malta, Msida, Malta, 2023.
92. Ramzan, F.; Sartori, C.; Consoli, S.; Reforgiato Recupero, D. Generative Adversarial Networks for Synthetic Data Generation in Finance: Evaluating Statistical Similarities and Quality Assessment. *AI* **2024**, *5*, 667–685. [CrossRef]
93. Vuletić, M.; Prenzel, F.; Cucuringu, M. Fin-GAN: Forecasting and classifying financial time series via generative adversarial networks. *Quant. Financ.* **2024**, *24*, 175–199. [CrossRef]
94. Ljung, M. Synthetic Data Generation for the Financial Industry Using Generative Adversarial Networks. 2021. Available online: https://www.diva-portal.org/smash/get/diva2:1591892/FULLTEXT01.pdf (accessed on 7 June 2024).
95. He, B.; Kita, E. Stock price prediction by using hybrid sequential generative adversarial networks. In Proceedings of the 2020 International Conference on Data Mining Workshops (ICDMW), Sorrento, Italy, 17–20 November 2020; pp. 341–347.
96. Dalmasso, N.; Tillman, R.E.; Reddy, P.; Veloso, M. Payvae: A generative model for financial transactions. In Proceedings of the AAAI 2021 Workshop on Knowledge Discovery from Unstructured Data in Financial Services Workshop, Virtual, 9 February 2021.
97. Huang, H.; Chen, M.; Qiao, X. Generative Learning for Financial Time Series with Irregular and Scale-Invariant Patterns. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7 May 2024.
98. Guan, C.; Ding, D.; Gupta, P.; Hung, Y.C.; Jiang, Z. A Systematic Review of Research on ChatGPT: The User Perspective. In *Exploring Cyber Criminals and Data Privacy Measures*; IGI Global: Hershey, PA, USA, 2023; pp. 124–150.
99. Guan, C.; Ding, D.; Ren, J.; Guo, J. Unveiling the aesthetic "wow factor": The role of aesthetic incongruity and image quality in NFT art valuation with computer vision. *Electron. Mark.* **2024**, *34*, 38. [CrossRef]
100. Rajath, K. Fintech's Generative AI Revolution How AI is shaping the Future of Banking and Financial Services. *Int. Res. J. Mod. Eng. Technol. Sci.* **2023**, *5*, 1812–1814.
101. Cao, X.; Li, S.; Katsikis, V.; Khan, A.T.; He, H.; Liu, Z.; Zhang, L.; Peng, C. Empowering financial futures: Large language models in the modern financial landscape. *EAI Endorsed Trans. AI Robot.* **2024**, *3*. [CrossRef]
102. Ranković, M.; Gurgu, E.; Martins, O.; Vukasović, M. Artificial intelligence and the evolution of finance: Opportunities, challenges and ethical considerations. *EdTech J.* **2023**, *3*, 20–23. [CrossRef]