Natural language processing in finance: A survey

Kelvin Du [a], Yazhi Zhao [b], Rui Mao [a], Frank Xing [c], Erik Cambria [a],*

[a] *College of Computing and Data Science, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore*
[b] *Visa Inc, 71 Robinson Road, Singapore, 068895, Singapore*
[c] *School of Computing, National University of Singapore, 15 Computing Drive, Singapore, 117418, Singapore*

ARTICLE INFO

ABSTRACT

This survey presents an in-depth review of the transformative role of Natural Language Processing (NLP) in finance, highlighting its impact on ten major financial applications: (1) financial sentiment analysis, (2) financial narrative processing, (3) financial forecasting, (4) portfolio management, (5) question answering, virtual assistant and chatbot, (6) risk management, (7) regulatory compliance monitoring, (8) Environmental, Social, Governance (ESG) and sustainable finance, (9) explainable artificial intelligence (XAI) in finance and (10) NLP for digital assets. With the integration of vast amounts of unstructured financial data and advanced NLP techniques, the study explores how NLP enables data-driven decision-making and innovation in the financial sector, alongside the limitations and challenges. By providing a comprehensive analysis of NLP applications combining both academic and industrial perspectives, this study postulates the future trends and evolution of financial services. It introduces a unique review framework to understand the interaction of financial data and NLP technologies systematically and outlines the key drivers, transformations, and emerging areas in this field. This survey targets researchers, practitioners, and professionals, aiming to close their knowledge gap by highlighting the significance and future direction of NLP in enhancing financial services.

1. Introduction

In the rapidly evolving landscape of finance, the integration of Natural Language Processing (NLP) technologies has emerged as a transformative force, unlocking new dimensions of data-driven decision-making and innovation. This survey aims to explore the multifaceted applications of NLP techniques within the financial sector, highlighting its profound impact on areas such as financial sentiment analysis, natural language-based financial forecasting, portfolio management, financial narrative processing, question answering, virtual assistant and chatbot, risk management, regulatory compliance, ESG and sustainable finance, explainable AI in finance and NLP for digital assets from both academic and industrial research perspective. As financial institutions increasingly rely on vast amounts of unstructured data, from corporate releases to news articles and reports to social media and beyond, the capability to efficiently process, understand, and act upon this information has become a competitive necessity. In the field of NLP in finance, Fisher et al. [1] executed a thorough exploration, mainly concentrating on tasks related to classification and prediction. Chen et al. [2] presented a broad review of financial application areas, including three categories: Know Your Customer (KYC), Know Your Product (KYP), and Satisfy Your Customer (SYC).

While prior reviews have often leaned towards NLP or finance, our survey seeks to provide an all-encompassing review that connects recent research across both dimensions. Crucially, our study involves a detailed investigation of the latest advancements in NLP within the finance sector, providing insights from both the perspectives of NLP methodologies and financial applications. Moreover, we systematically analyze NLP in finance by considering financial textual data, NLP techniques, and financial applications, along with emerging trends in both NLP techniques, such as large language models (LLMs), as well as financial applications, including NLP for ESG and sustainable finance.

This survey begins with an overview of NLP's foundational concepts and techniques, providing a solid groundwork for understanding its application in finance. It then delves into specific applications where NLP has been successfully implemented, together with the financial data adopted, illustrating the technology's versatility and power. Further, it addresses the challenges and limitations of applying NLP in the finance sector, including issues related to data quality, privacy concerns, and the need for highly specialized models. It also explores the future trajectory of NLP technologies in finance, considering advances in AI and computational linguistics that promise to further enhance the sophistication and effectiveness of financial services.

---

* Corresponding author.

E-mail addresses: zidong001@e.ntu.edu.sg (K. Du), yazzhao@visa.com (Y. Zhao), rui.mao@ntu.edu.sg (R. Mao), xing@nus.edu.sg (F. Xing), cambria@ntu.edu.sg (E. Cambria).
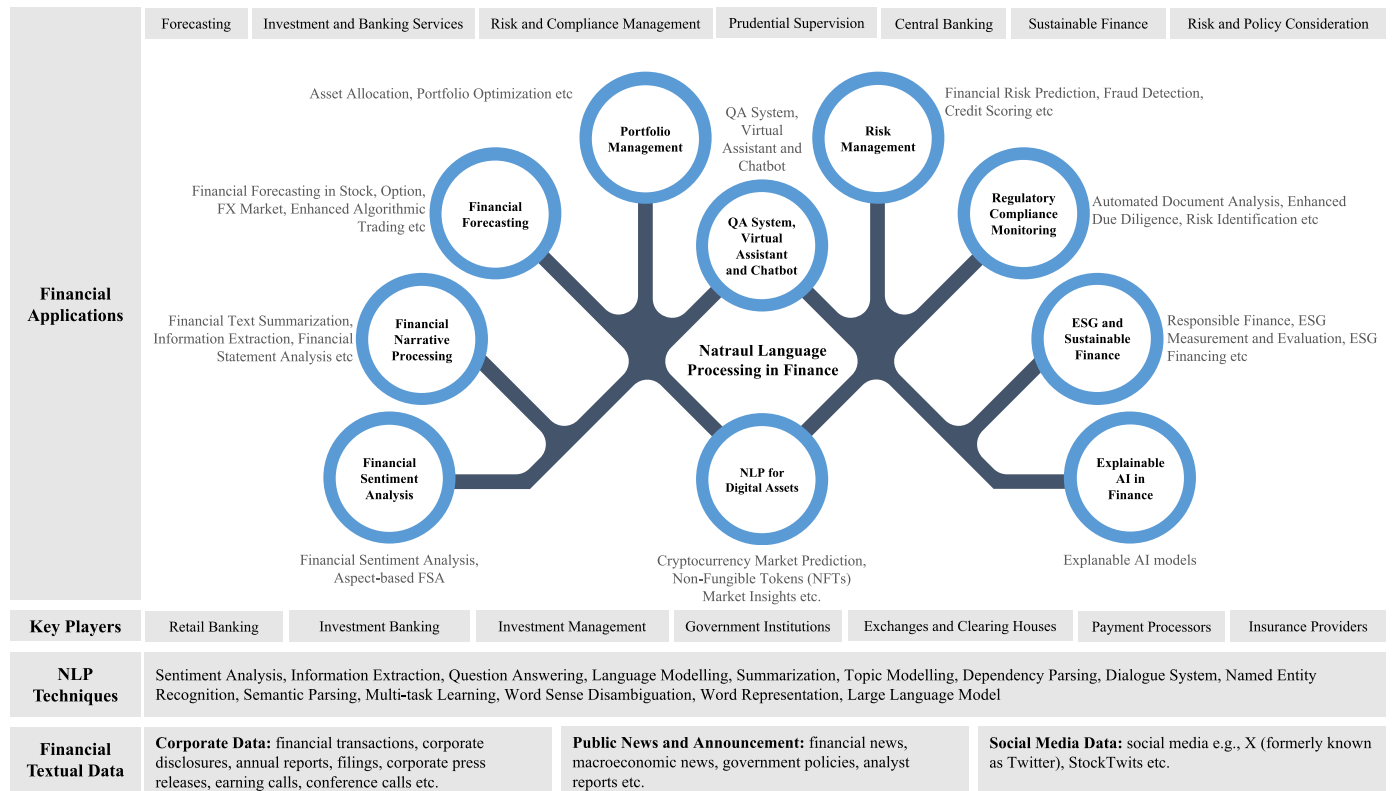
Fig. 1. NLP in finance: Financial textual data, NLP techniques, and financial applications.

Through this comprehensive exploration, we aim to provide academics, industry professionals, and technology enthusiasts with a thorough understanding of the current state and future potential of NLP in the financial domain. By showcasing the technology's significant contributions and pondering its future directions, the survey aims to foster a deeper appreciation for the role of NLP in shaping the next generation of the financial sector. Specifically, this study aims to answer the following three groups of research questions:

1. What are the key drivers behind the increasing importance of NLP in the finance sector, and how do these drivers address the unique challenges and opportunities within this industry?
2. What emerging areas of application for NLP in finance are being explored, and what potential do these areas hold for the future of financial services?
3. What are the prevailing trends in NLP research within the finance sector, and how do these trends reflect the evolving challenges and opportunities in the field?

Our contributions can be outlined in the following three aspects:

1. We presented a detailed literature review on NLP techniques in finance, bridging a significant gap by providing a definitive reference for researchers and practitioners. It encapsulates the evolution and diverse applications of NLP across both academic research and industrial implementations.
2. We developed a comprehensive review framework for NLP in finance, which maps the relationship between data, NLP techniques, and financial applications. We identified ten key financial application areas that are potentially revolutionized by NLP, and our framework provides a systematic overview to comprehending the intricacies of financial data, NLP techniques, and their real-world applications in the finance industry.
3. We analyzed key research contributions that have significantly influenced NLP's application in the finance sector, highlighting

transformative effects on services, products, and operations. This includes insights into practical challenges and the transformative potential of these technologies.

The remainder of this survey is structured as follows: Section 2 describes the literature review framework; Section 3 conducts an in-depth review of studies on NLP in finance by financial application areas; Section 4 demonstrates the research findings of this review including trends, challenges, opportunities, and future directions; Section 5 provides concluding remarks.

2. Literature review framework

Fig. 1 presents our literature review framework, focusing specifically on the financial textual data, NLP techniques, key players and financial applications. We started with a review of a study on AI in finance conducted by the International Monetary Fund (IMF) [3]. We then zoomed into each potential application area, examining the relevant literature. We also summarized the data sources available in finance. The financial data can be broadly categorized into three types: corporate data, public news and announcements, and social media data. Corporate data includes data produced by corporations such as financial transactions, financial statements, customer feedback, disclosures, annual reports, filings, press releases, earnings calls, and conference calls etc. The public news and announcements category encompasses financial and economic news, government policies, and analyst reports, among others. Social media data consists of information from social media platforms, e.g., X, StockTwits, and various online forums. Subsequently, we defined the scope of our study to include retail banking, investment banking, investment management, government institutions, exchanges and clearing houses, payment processors, and insurance providers. We proposed key financial application areas significantly enhanced by adopting various NLP techniques in the finance sector. This includes financial sentiment analysis, financial forecasting, portfolio management, financial narrative processing, question

Table 1
Key words for literature search.

Research topic	Key words
Financial sentiment analysis	financial sentiment analysis
Financial narrative processing	finance AND (summarization OR information extraction OR information retrieval OR financial narrative processing)
Financial forecasting	(nlp OR text mining OR news OR tweet) AND (market prediction OR financial forecasting)
Portfolio management	(nlp OR text mining OR news OR tweet) AND portfolio AND (selection OR optimization OR management)
Question answering, virtual assistant and chatbot	(question answering OR virtual assistant OR chatbot) AND (finance OR financial services)
Risk management	(nlp OR text mining OR news OR tweet) AND (risk management OR risk prediction OR credit risk OR fraud detection)
Regulatory compliance monitoring	(nlp OR text mining) AND (regulatory compliance OR anti-money laundering)
ESG and sustainable finance	(nlp OR text mining) AND (esg OR (sustainable finance) OR (environmental AND social AND governance))
Explainable artificial intelligence (XAI) in finance	(nlp OR text mining) AND (explainable artificial intelligence OR explainable AI OR explainability) AND finance
NLP for digital assets	(nlp OR text mining OR news OR tweet) AND (digital asset OR cryptocurrency OR non-fungible token OR blockchain)

answering, virtual assistant and chatbot, risk management, regulatory compliance monitoring, explainable AI in finance, ESG and sustainable finance and NLP for digital assets. Additionally, we examine how these applications are adopted across businesses in the finance sector in subsequent sections.

Following PRISMA guidelines, we designed a three-step process, namely identification, screening and inclusion, to select the literature for review. Firstly, leveraging industry experience, we defined keywords, as illustrated in Table 1, which closely align with the topics from the International Monetary Fund's (IMF) study on AI in finance. Subsequently, we used these keywords to perform targeted searches on Web of Science and Google Scholar, prioritizing relevance as defined by the database. Our aim is to ensure the coverage of the range of NLP tasks applicable to this field. We assessed each article by screening through the abstracts to ascertain their relevance and contribution, which determined their inclusion in the review. We balanced the breadth of coverage to prevent overwhelming the review given the broad topic. Our focus extended to the coverage of tasks, the recency of the research, and its impact within the field. This methodical approach not only ensured the thoroughness of our review but also presented the status and achievements of AI applications in specific fields, laying a foundation for future research directions.

The number of publications across the ten research topics are presented in Fig. 2. The fields are ranked by the volume of publications, in descending order, as follows: risk management, question answering, virtual assistant and chatbot, financial sentiment analysis, financial narrative processing, financial forecasting, NLP for digital assets, portfolio management, ESG and sustainable finance, regulatory compliance monitoring and explainable AI in finance. All fields have demonstrated an increasing trend in publication volume over the past decade, highlighting a growing interest in NLP applications within the finance sector. Notably, emerging fields such as ESG and sustainable finance, regulatory compliance monitoring and explainable AI in finance are also drawing significant scholarly attention.

3. An in-depth review of NLP in finance

NLP in finance refers to the broader application of NLP techniques to the financial sector, which generally includes corporate data, public financial news and announcements, and social media data such as StockTwits. It encompasses a wide array of tasks aimed at understanding, interpreting, and extracting information from textual data in the financial domain. This includes processing corporate data, public news and announcement and social media data, and any other form of textual data that can impact financial markets and decision-making. NLP in finance leverages techniques such as sentiment analysis, information extraction, named entity recognition, language modeling, LLMs and more, to analyze financial texts for various purposes like financial sentiment analysis, financial forecasting, portfolio management, financial narrative processing, question answering, virtual assistant and chatbot, risk management, financial regulatory compliance monitoring, explainable AI in finance, ESG and sustainable finance and NLP for digital assets.

3.1. Financial sentiment analysis

Financial Sentiment Analysis (FSA) is one of the widely adopted NLP techniques in the finance domain due to the advancements of sentiment analysis in the general domain [4–7]. Du et al. [8] has conducted a comprehensive survey on FSA which categorizes FSA into two principal research tracks. The FSA techniques concentrate on identifying specific tasks such as document-level [9], paragraph-level [10], sentence-level [11] and aspect-level [12] sentiment analysis and aim to propose methodologies to improve the performance of different FSA tasks by using human-annotated datasets. The FSA applications emphasize the application of financial sentiments, whether direct or indirect, in subsequent applications within financial markets, attracting greater research attention.

Presented in Table 2, we explore the advancement of research in FSA techniques, tracing a path from the lexicon-based method, through traditional machine learning models, to the deep learning and pre-trained language models. Notably, a majority of FSA research has converged on the established benchmarks of PhraseBank [13], SemEval 2017 Task 5 [14], and FiQA Task 1 [15] for evaluation purposes. The construction of financial lexicons remains a focal point for researchers, and lexicons can be employed either independently or in conjunction with learning-based methods for FSA [16]. A notable shift in lexicon development moves from simplistic single-word expression to complex, multi-word, and context-aware phrases, a critical adaptation for the financial sector where the implication of terms can dramatically change with context. At the same time, there is a trend towards automated approaches, highlighted by studies such as Oliveira et al. [17] and Du et al. [18], which contrasts with the traditional manual compilation of lexicons, demanding significant expert input. In the context of machine learning models, the process of feature engineering stands out as a pivotal stage, with the categorization into lexicon-based, linguistic, domain-specific, and word embedding features, etc. Among these, domain-specific attributes, especially numerical data or combinations of keywords with numerical values, though not extensively examined, have shown their potential in FSA. An example is the term "revenue" coupled with a positive sign and a numerical value, indicating a positive financial forecast. Furthermore, the forefront of FSA includes deep learning approaches, notably through the use of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, as well as pre-trained language models, which have significantly improved FSA performance. Specifically, a BERT variant tailored for the financial sector, known as FinBERT [19,20], has been developed through the incorporation of a variety of data sources including Reuters Corpora, Yahoo Finance, and financial reports, marking a significant leap in FSA research methodologies. The latest study by Du et al. [16] achieved unparalleled results on FiQA Task 1 and SemEval 2017 Task 5 by integrating the knowledge into the process of fine-tuning language models such as RoBERTa, thus pushing the performance boundaries of FSA technique research further.

The development of domain-specific transformer models, with notable examples, including FinBERT [19,20], has significantly enhanced FSA. However, the potential and applicability of autoregressive decoder
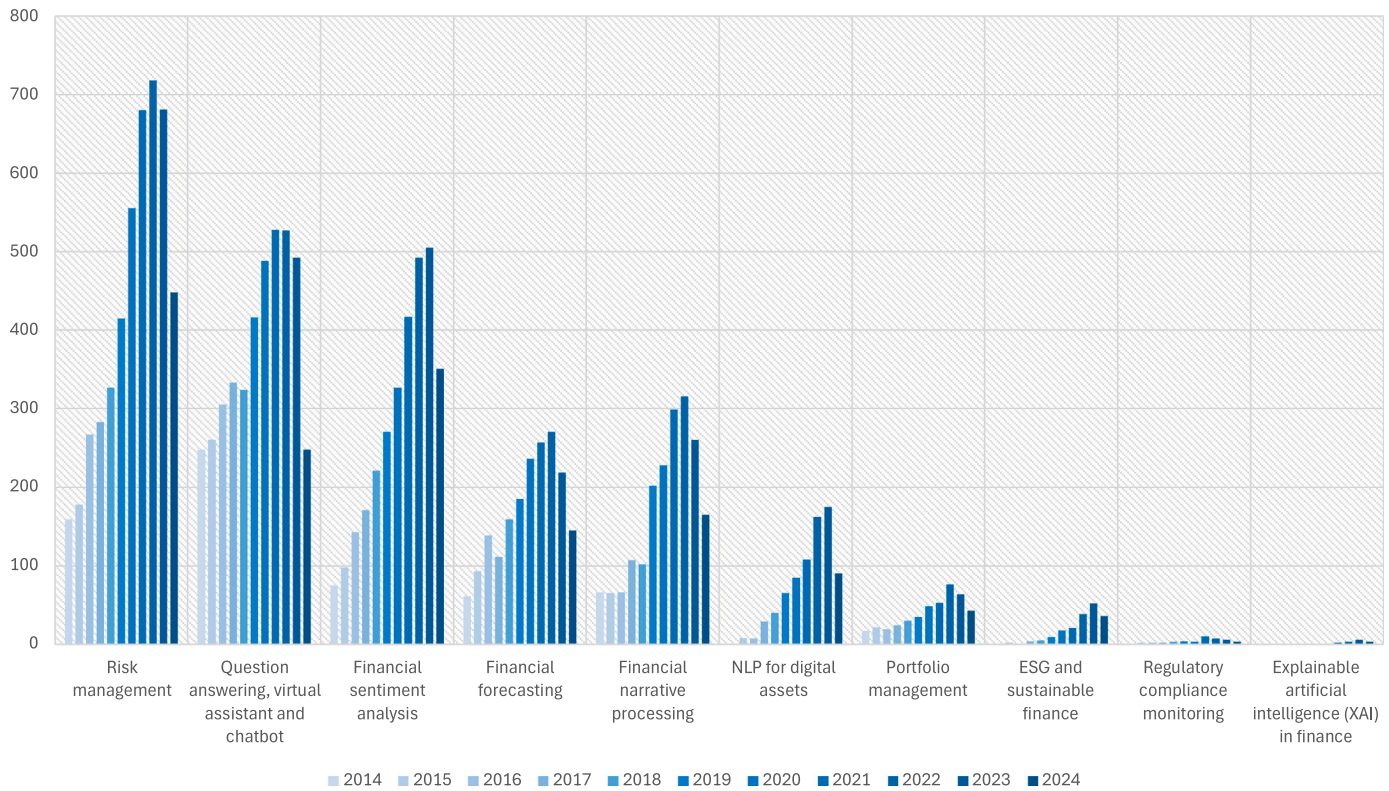
Fig. 2. Number of publications in the field of NLP in finance.

architectures such as Generative Pre-trained Transformer (GPT) [21] in the context of FSA have not been thoroughly investigated yet [22]. In recent studies, Fatouros et al. [23] adopted zero-shot prompting to examine multiple ChatGPT prompts on a curated dataset comprised of forex-related news headlines, measuring performance using a range of metrics such as precision, recall, F1-score, which outperforms Fin-BERT. Zhang et al. [24] presented a framework that integrates a retrieval-augmented mechanism with LLMs specifically for FSA. The framework consists of two key modules which are instruction-finetuned LLMs and a retrieval-augmented module. The performance metrics, specifically accuracy and F1 score, are significantly improved, under-scoring the efficacy of the framework in FSA. Fatemi and Hu [22] conducted a thorough comparative analysis to examine the effective-ness of fine-tuning LLMs versus employing few-shot learning techniques in the context of FSA. In particular, the in-context learning is adopted on GPT-3.5-turbo and the fine-tuning is performed on Flan-T5. The study highlights the remarkable capabilities of LLMs, even smaller mod-els, in both in-context learning and fine-tuning for the FSA task. Zhang et al. [25] introduced a novel approach named Instruct-FinGPT, which transforms a limited subset of supervised FSA data into instruction data to fine-tune a general-purpose LLM. The Instruct-FinGPT approach has demonstrated significant advancements in its ability to perform zero-shot generalization across other financial datasets. Hu et al. [26] conducted a study to evaluate the sentiment of management's dis-cussion and analysis disclosures using three distinct methods: GPT-3, FinBERT, and the dictionary-based method. The findings reveal that GPT-3 surpasses the dictionary-based method in its effectiveness for financial sentiment classification tasks. However, it was observed that GPT-3 slightly underperforms when compared to FinBERT. Deng et al. [27] implemented a semi-supervised learning approach using LLMs to generate weak sentiment labels for Reddit posts and then used that data to train a small model which is to be served in production. Additionally, Bloomberg has introduced BloombergGPT [28], an LLM tailored for financial contexts, which has demonstrated superior performance in financial NLP tasks, including FSA, named entity recognition, news classification, question answering, among others. Du et al. [29] pro-posed a prompting framework to evaluate the reasoning capabilities of LLMs for FSA by identifying six specific financial attributes that potentially influence financial sentiment. Xing [30] designed special-ized LLM agents for FSA, based on guiding knowledge. Mao et al. [31] analyzed the market sentiment perception from the perspective of cognition, and they employed MetaPro [32,33] to parse concept mappings from metaphorical expressions, and then detected the distinct cognitive patterns during different market conditions. Manro et al. [34] examined the influence of CEOs' cognitive states on stock price trends by analyzing the opinions expressed in their letters to shareholders using MetaPro.

3.2. Financial forecasting

The primary application of FSA is to investigate the causality and correlation between financial sentiment and financial markets, and to perform financial forecasting [8]. Natural language-based financial forecasting was brought up by Xing et al. [50] with a focus on the stock market. Subsequently, Du et al. [8] broadened the scope of review to include the FOREX and cryptocurrency markets. Additionally, Du et al. [8] brought attention to the interconnected relationship between financial sentiment, investor sentiment, and market sentiment, arguing that financial forecasting is an essential application of FSA. This is because financial sentiment, whether used implicitly or explicitly, is crucial in predicting financial market movements. Natural language-based financial forecasting serves to augment algorithmic trading by leveraging NLP to analyze news articles and social media, thereby informing automated trading decisions, enabling algorithms to parse and interpret vast amounts of textual data in real time and enhancing decision-making processes and market responsiveness.

Predicting market movements in financial markets is notably chal-lenging, attributed to their volatile and unpredictable nature. The foundation of earlier financial market prediction research rests on his-torical trading data, technical indicators, and macroeconomic variables,

Table 2
FSA techniques.

Literature	Dataset	Task	Method	Performance
Park et al. [35]	PhraseBank	Sentence-level FSA	Lexicon	Precision: 0.8238, Recall: 0.8128 F1-Score: 0.8105
Du et al. [18]	PhraseBank, SEntFiN, etc.	Sentence-level FSA	Lexicon	Accuracy: 0.7619, F1-Score: 0.7216
Krishnamoorthy [36]	PhraseBank	Sentence-level FSA	Machine learning	Precision: Pos: 0.83, Neg: 0.93, Neut: 0.86 Recall: Pos: 0.82, Neg: 0.93, Neut: 0.81 F1-Score: Pos: 0.83, Neg: 0.93, Neut: 0.83
Araci [37]	PhraseBank, FiQA Task 1	Sentence-level FSA	Pre-trained language model	PhraseBank (Accuracy: 0.86, F1-Score: 0.84) FiQA Task 1 (MSE: 0.07)
Liu et al. [20]	PhraseBank	Sentence-level FSA	Pre-trained language model	Accuracy: 0.94, F1-Score: 0.93
Zhao et al. [38]	CCF BDCI, CCKS	Sentence-level FSA	Pre-trained language model	Accuracy: 96.774%
Wu et al. [28]	PhraseBank, FiQA Task 1	Sentence-level FSA	LLMs	PhraseBank: 0.5107 FiQA Task 1: 0.7507
Du et al. [29]	PhraseBank, Fin. News	Sentence-level FSA	LLMs	PhraseBank: (Accuracy: 0.9639) Fin. News: (Accuracy: 0.8234)
Xing [30]	PhraseBank, Fin. News, etc.	Sentence-level FSA	LLMs	PhraseBank: Accuracy: 0.8048, F1-Score: 0.8141 FiQA: Accuracy: 0.9391, F1-Score: 0.9522 SEntFiN: Accuracy: 0.7745, F1-Score: 0.7693
Jiang et al. [39]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Machine learning	WCS (News: 0.7779, Microblogs: 0.7107)
Schouten et al. [40]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Machine learning	WCS (News and Microblogs: 0.7050)
Dridi et al. [41]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Machine learning	WCS (News: 0.655, Microblogs: 0.726)
Akhtar et al. [42]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Hybrid approach	WCS (News: 0.786, Microblogs: 0.797)
Ghosal et al. [43]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Hybrid approach	WCS (News: 0.697, Microblogs: 0.751)
Mao et al. [44]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Metaphor Paraphrasing	Accuracy: 0.696, F1-Score: 0.696
Xiang et al. [45]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Pre-trained language model	WCS (News: 0.8441, Microblogs: 0.8333)
Sinha et al. [46]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Pre-trained language model	Accuracy: 0.9429, F1-Score: 0.9327
Du et al. [16]	SemEval 2017 Task 5	Targeted, Sentence-level FSA	Pre-trained language model	WCS (News: 0.8483, Microblogs: 0.8122)
de França Costa and da Silva [47]	FiQA Task 1	Targeted, Aspect-level FSA	Machine learning	Aspect Extraction: F1-Score (News: 0.4240, Post: 0.5775) Sentiment Analysis: MSE (News: 0.1052, Microblogs: 0.1281)
Piao and Breslin [48]	FiQA Task 1	Targeted, Aspect-level FSA	Deep learning	Aspect Extraction: F1-Score (0.6530) Sentiment Analysis: MSE (0.0926)
Xiang et al. [45]	FiQA Task 1	Targeted, Aspect-level FSA	Pre-trained language model	MSE (0.0717), R^2 (0.4878)
Du et al. [16]	FiQA Task 1	Targeted, Aspect-level FSA	Pre-trained language model	MSE: 0.0490
Luo et al. [49]	Financial documents	Document-level FSA	Hierarchical Query-driven Attention	Accuracy: 0.9446, F1-Score: 0.9449
Mao et al. [31]	Financial news	Perception-level FSA	MetaPro and Statistical Analysis	
Manro et al. [34]	Letters to shareholders	Perception-level FSA	MetaPro and Statistical Analysis	

with recent advancements incorporating financial textual data using NLP techniques for enhanced financial forecasting [50–53]. Established financial news providers, including Bloomberg and Thomson Reuters, have been at the forefront of offering commercial FSA service [52, 54]. Mudinas et al. [52] highlighted that financial institutions e.g., D. E. Shaw and Two Sigma have integrated sentiment signals from financial texts with traditional structured transaction data, to enhance their machine learning models for algorithmic trading. Conventionally, fundamental analysis and technical analysis are the two principal approaches in stock market analysis [55]. Financial forecasting using natural language can be classified as part of technical analysis, since it does not alter the intrinsic value of assets. Tables 3 and 4 showcase studies in financial forecasting in stock and FOREX markets, respectively, demonstrating the effectiveness of incorporating financial textual information. The financial forecasting for the cryptocurrency market is reviewed in Section 3.10 NLP for Digital Assets.

Investor sentiment, or the deviation of investors' beliefs about future firm valuations from fundamental information, significantly impacts stock prices and market activities [56]. Research in stock market prediction spans various methodologies, including time series models, machine learning, deep learning, and reinforcement learning, focusing on aspects like stock indexes, prices, movements, return rates, and volatility. Notable approaches include the use of neural tensor networks for event-driven predictions [57], combining news and event

tuple features with trading data [58], and integrating sentiment indicators with technical analysis for enhanced price prediction [59]. Novel strategies also involve parsing earnings calls, which are shown to influence short-term investor sentiment and predict stock returns [60–63]. Further advancements include the application of Graph Convolutional Neural Networks to incorporate company relationships [64], utilizing social media texts [65], and employing transformers and self-supervised learning for dynamic market prediction and improved stock forecasting accuracy [53,66,67]. Additionally, combining various sentiment sources through Principal Component Analysis and Kalman Filter has proven effective for generating robust daily sentiment indicators, crucial for accurate market predictions [54]. Event-driven trading strategies have also been developed to capitalize on corporate events detected in news articles, offering substantial returns [68]. Investor sentiment not only influences time-series returns but also affects the cross-section of stock returns, with certain stocks being more susceptible to sentiment-driven fluctuations [69]. Recently, Du et al. [70] proposed a dual-graph neural network, which dynamically generates and integrates price relationships and semantic relationships between companies for stock price movement prediction.

The focus on predicting stock market trends has been significant, with the FOREX market receiving less attention. Early research explored the link between macroeconomic fundamentals and short-term

Table 3

Financial forecasting in stock market.

Literature	Data source	Method	Task	Performance
Ding et al. [57]	Thomson Reuters, Bloomberg	CNN	Price prediction S&P500 prediction	Accuracy: 0.6548, MCC: 0.41 Accuracy: 0.6421, MCC: 0.40
Nguyen and Shirai [71]	Yahoo Finance Message Board	SVM	Price movement prediction	
Oliveira et al. [54]	Twitter	MR, NN, SVM, RF, Ensemble	Daily return Daily trading volume Daily volatility	NMAE: NDQ, 7.58 NMAE: DJIA, 5.84 NMAE: DJIA, 2.79
Liu et al. [58]	Thomson Reuters	SVM, LSTM	Price movement prediction	Accuracy: 0.5544, F1-Score: 0.7133
Xu and Cohen [72]	Twitter	GRU, VAEs	Price movement prediction	Accuracy: 0.5823, MCC: 0.080796
Wu et al. [73]	Twitter	Cross-modal attention-based Hybrid RNN	Price movement prediction	Accuracy: 0.5915
Guo and Li [74]	Twitter	Linear Regression	Price prediction	Accuracy: 0.6722
Keith and Stent [63]	Earning calls	Ridge/Logistic Regression LSTM, Ensemble	Change in target price	Regression (MSE: 0.00137, R^2: 0.1718) Classification (Accuracy: 0.482, F1-Score: 0.475)
Jin et al. [75]	Stocktwits	Empirical Mode Decomposition-based LSTM	Price prediction	Accuracy: 0.7056, MAPE: 1.65%, MAE: 2.39
Sawhney et al. [65]	Twitter	Graph Attention Network	Price movement prediction	F1-Score: 0.605, Accuracy: 0.608, MCC: 0.195
Zhou et al. [68]	PRNewswire Businesswire	Multi-class Classification with MLM Loss	Price movement prediction	Trade at End Strategy: Average Return: 1.74% Trade at Best Strategy: Average Return: 9.11%
Wu et al. [76]	EastMoney.com	CNN, LSTM	Price prediction	MAE: 2.38, MSE: 7.27, RMSE: 2.69
Soun et al. [67]	Twitter	Attentive LSTM	Price movement prediction	Accuracy (BIGDATA22: 54.81%, ACL18: 58.72%, CIKM18: 55.86%) MCC (BIGDATA22: 0.0952, ACL18: 0.2065, CIKM18: 0.0899)
Ma et al. [53]	Hundsun Electronics	Pre-training, BiLSTM and GCN	Price movement prediction	Accuracy: 0.6726, MCC: 0.3452
Du et al. [70]	Twitter	Attentive LSTM, GATs	Price movement prediction	Accuracy (BIGDATA22: 0.5833, CIKM18: 0.5730) MCC (BIGDATA22: 0.1556, CIKM18: 0.0389)

Table 4

Financial forecasting in FOREX market.

Literature	Data source	Method	Task	Performance
Jin et al. [77]	Bloomberg	Linear Regression	Change in currency value and generate warning messages	Precision (Argentina: 0.18, Brazil: 0.28, Chile: 0.33, Colombia: 0.25) Recall (Argentina: 0.60, Brazil: 0.63, Chile: 1, Colombia: 1)
Nassirtoussi et al. [55]	MarketWatch.com Google RSS	SVM	FOREX price movement prediction	Accuracy: 0.8333, Precision (Pos: 0.6667, Neg: 0.8889) Recall (Pos: 0.6667, Neg: 0.8889)
Seifollahi and Shajari [78]	MarketWatch.com	SVM	FOREX price movement prediction	Accuracy: 0.5926, Precision: 0.5710, Recall: 0.5735
Semiromi et al. [79]	Forexfactory.com,	SVM, RF, XGB	FOREX price movement prediction	Accuracy (EUR/USD: 0.638, GBP/USD: 0.663)
	First Word FX News			Accuracy (USD/CHF: 0.631, USD/JPY: 0.641)
Xing et al. [80]	Dow Jones Newswire	SVM	FOREX price movement prediction	Accuracy: 0.503, F1-Score: 0.538

exchange rate volatility using the Flexible Fourier Form regression, focusing on the impact of macroeconomic announcements (e.g., interest rates, GDP, consumer confidence indexes) on USD/EUR volatility [81]. Findings from data segmented into 5-min intervals over a few months indicated that such news significantly affects exchange rate volatility immediately post-announcement, with the impact varying by news type and country. Subsequent studies confirmed that macroeconomic news influences currency markets over time, with immediate absorption of the direct price impact but delayed total reflection [82]. The Forex-foreteller (FF) model was introduced to forecast currency market movements using news articles, language models, topic clustering, and sentiment analysis in conjunction with historical data, proving effective for generating predictive signals [77]. Other approaches have leveraged text mining of news headlines, such as TF–IDF weighted by sentiment scores using SentiWordNet, to predict FOREX market movement [55]. Further advancements include sentiment analysis of news headlines for predicting EUR/USD directional movement with enhanced accuracy [78], incorporating news story events from the economic calendar to forecast currency pair movements with SVM, RF, and XGBoost algorithms [79], and evaluating high-frequency news sentiment without other semantic features, where a FinBERT-based model showed

that news sentiment alone offers predictive value, albeit limited, for FOREX price movements [80]. Lastly, researchers and practitioners are expanding their focus to include other financial forecasting areas, like option pricing and IPO pricing [83], and firm value [84] etc.

3.3. Portfolio management

Portfolio optimization is an important element of portfolio management, which involves strategies to identify the optimal investment mix to maximize expected returns for a given risk level, or to minimize risk for a targeted return. The use of NLP in this field is innovative, focusing on enhancing decision-making by analyzing qualitative data (see Table 5). In portfolio optimization, NLP helps by capturing semantic relations of investment targets [85] or incorporating opinions from investor posts for better stock market investments [86]. This shift towards utilizing financial sentiment in portfolio management has led to various approaches, including Follow-the-Loser strategies, based on stock microblogs [87], and models predicting the quality of investment opinions for portfolio creation [88].

Moreover, techniques like Gaussian inverse reinforcement learning, model market dynamics and investor sentiment for trading systems that

Table 5
Portfolio management.

Literature	Data source	Method	Performance
Tu et al. [88]	StockTwits	Linear regression for opinion quality score	Cumulative return: 30%
Koyano and Ikeda [87]	Yahoo! JAPAN textream	MLP, Follow-the-Winner/Loser strategy	Cumulative return: 1.151
Sun et al. [98]	Sina Guba, Eastmoney Guba, RESSET	Probabilistic Neural Network	Accuracy: 86.3%
Yang et al. [89]	Thomson Reuters	Gaussian inverse reinforcement learning	Return: 17.39%, Sharpe ratio: 0.85 Sterling Ratio: 0.76
Malandri et al. [86]	Financial and sentiment data	EW, LSTM, MLP, RF	Wealth of the portfolio (LSTM + Sentiment) Five portfolios: 2.23, 2.72, 2.30, 2.81 and 1.65
Xing et al. [85]	Company profiles	Doc2vec, covariance matrix correction	CAGR: 0.2368, Sharpe ratio: 1.01
Ye et al. [91]	Financial news	State augmented reinforcement learning	Portfolio value improved (Bitcoin: 140.9%, HighTech: 15.7%) Sharpe ratio (Bitcoin: 14.78, HighTech: 7.73)
Du and Tanaka-Ishii [92]	Wall Street Journal (WSJ), Reuters, Bloomberg	Deep learning	Average realized annual gain (WSJ: 17.2%, R&B: 35.5%)
Chen et al. [94]	RESSET database	DT, LR, SVM, RF	Accuracy: 79.6%, Holding Period yield: 5.41
Sawhney et al. [95]	Twitter, Wind.com.cn	Time-aware LSTM (t-LSTM)	US S&P 500: Return ratio: 1.34, Sharpe ratio: 0.96 China and Hong Kong: Return ratio: 1.44, Sharpe ratio:1.19
Koratamaddi et al. [93]	Wharton Research Data Services	Deep reinforcement learning	Sharpe ratio: 2.07, Annualized return: 22%
Ghosh et al. [99]	Bloomberg	RF, LSTM	Daily return (LSTM: 0.64%, RF: 0.54%)
Hung et al. [96]	Reuters, CNBC	GRU, Black–Litterman	Annualized return: 46.6262%, Sharpe ratio: 13.02% Sortino ratio: 17.96%, MDD: 4.45%, Turnover: 68.35%
Ma et al. [97]	RESSET, Qichacha, Hundsun Electronics	Multitask learning, pre-training	IRR: 0.3923, Sharpe ratio: 2.0334, MDD: 0.0754

filter out noisy signals [89]. Strategies e.g., PROFIT (Policy for Return Optimization using FInancial news and online Text) leverages financial news and social media for trading decision optimization, showing significant performance improvements over benchmarks [90]. Similarly, State Augmented Reinforcement Learning (SARL) integrates diverse data sources for learning portfolio management strategies, achieving superior results [91]. Stock embedding, representing stocks as neural network-generated vectors from news and price history, enhances portfolio optimization beyond price prediction [92]. Including online public mood in models has proven to consistently outperform equal-weighted strategies, highlighting the performance benefits of incorporating financial sentiment [86]. Sentiment-aware models for daily portfolio allocation demonstrate the robustness and higher returns compared to benchmarks [93]. Additionally, sentiment-based models adapted to market volatility, like those for Chinese stocks, yield superior returns [94]. Time-aware LSTM models, which capture market trends and rank stocks, based on predicted returns, outperform state-of-the-art methods in both intra-day situations and risk-adjusted returns [95]. Hung et al. [96] constructed a portfolio using news sentiment measured by BERT and demonstrated that Black–Litterman portfolio model with GRU model to predict stock price yields the highest performance. Ma et al. [97] developed a multi-task learning framework, learning risk and return factors, simultaneously, for portfolio optimization. The framework also leveraged data from multiple sources, e.g., news, numerical features, and company relationships.

3.4. Financial narrative processing

Financial Narrative Processing (FNP) [100], is a subset of NLP in finance with a specific focus on analyzing narrative content within financial documents. While NLP in finance has a broader application range, encompassing any textual analysis that can inform financial decisions and strategies, FNP is more specifically focused on the narrative and qualitative aspects of financial texts. It involves understanding the stories and qualitative disclosures that companies and financial analysts provide, such as management's discussion and analysis section in annual reports, earnings call transcripts, analyst reports, and financial news. The key objective is to extract insights from narrative texts that are not readily quantifiable but have significant implications for financial decision-making. El-Haj et al. [100] has organized the

FNP workshop since 2018.[1] Moreover, the automation of financial reporting [101] and financial statement analysis [102] using NLP, particularly with LLMs, has also emerged as a promising area of research. In our review, we focus on the financial text summarization and information extraction task.

3.4.1. Financial text summarization

Companies generate a multitude of reports incorporating both textual narratives and numerical data throughout their financial year, such as financial annual reports, earnings announcements, and conference calls. The information encapsulated within these financial reports holds significant importance for both companies and investors. It serves to showcase company accomplishments, garner support from shareholders in the stock market, as well as unveil risks and opportunities crucial for investor decision-making. Moreover, it plays a pivotal role in the due diligence procedures during mergers, acquisitions, and audit processes. However, the vast volume of financial data poses challenges in terms of navigation, management, and monitoring. Consequently, there arises a critical need for automated financial text summarization.

Financial text summarization endeavors to generate succinct, informative, and non-repetitive summaries from either single or multiple input texts. This objective can be pursued through two distinct methodologies: the extractive approach and the abstractive approach. The extractive method involves the identification and ranking of pertinent subsets within the input text [103]. Conversely, the abstractive approach generates summaries from scratch, based on the contents of the input document [104].

Table 6 showcases a variety of techniques and methods ranging from traditional machine learning techniques to high-performing deep learning models and word embeddings. The majority of techniques have converged on the established benchmark dataset from FNS Shared Tasks since 2020 [105–107], with a focus on producing concise summaries for annual financial reports of firms listed on the stock exchanges of the UK, Greece, and Spain. Extractive summarizer continues to be dominant on either sentence-level or section-level text summarization within the FNS shared Task dataset. El-Haj and Ogden [108] proposed to use the TF–IDF representations of the sentence and K-means clustering model to rank the importance of sentences, and combine them into a final summary. As deep learning models have demonstrated effectiveness in generating pertinent text features, the approach

---

[1] http://wp.lancs.ac.uk/cfie.

Table 6
Financial text summarization.

Literature	Dataset	Type of summarization	Method	Performance
El-Haj and Ogden [108]	FNS Shared Task	Extractive at sentence level	TF–IDF + Clustering	Rouge-1F (English: 0.317, Spanish: 0.448, Greek: 0.334)
Gokhan et al. [109]	FNS Shared Task	Extractive at sentence level	SentenceBERT-based Clustering	Rouge-1F: 0.48
La Quatra and Cagliero [110]	FNS Shared Task	Extractive at sentence level	BERT	Rouge-1F: 0.424
Abdaljalil and Bouamor [111]	FNS Shared Task	Extractive at sentence level	BERT	Rouge-1F: 0.38
Orzhenovskii [112]	FNS Shared Task	Extractive at sentence level	T5	Rouge-1F: 0.54
Shukla et al. [114]	FNS Shared Task	Extractive at section level	LR, k-maximal word allocation, top-k summarizer	Rouge-1F: 0.508
Shukla et al. [115]	FNS Shared Task	Extractive at section level	GPT3.5	Rouge-1F (English: 0.45, Spanish: 0.41, Greek: 0.12)
Li et al. [116]	Financial News	Abstractive	Transformer-BiLSTM encoder, Graph attention decoder	Rouge-1F (LCFNS: 0.3814, Fin-LCSTS, 0.3669, CNN/Daily Mail: 0.4164)
Liu et al. [117]	FINDSum	Abstractive	BART-large	Rouge-1F (FINDSum-ROO: 0.5390, FINDSum-Liquidity: 0.5412)
Zmandar et al. [118]	FNS Shared Task	Hybrid	Reinforcement learning	Rouge-1F: 0.52
Singh [119]	FNS Shared Task	Hybrid	Pointer Network + T5	Rouge-1F: 0.466
Mukherjee et al. [120]	ECTSum	Hybrid	FinBERT + T5	Rouge-1F: 0.467

of fine-tuning language models has been employed to produce sentence-level embeddings customized for specific financial contexts. Gokhan et al. [109] introduced a sentence-BERT methodology, leveraging BERT to generate sentence-level embeddings and subsequently clustering these embeddings using the K-means model. Additionally, both BERT and T5 have been utilized through fine-tuning processes to enhance performance in ranking the importance of sentences [110–112].

Financial reports, comprising financial statements and extensive information, often suffer from repetitiveness, posing a challenge when summarizing them at the sentence level. Rather than summarizing the entire report, a strategy has been adopted to identify key narrative sections within the report and generate summaries, based on these identified sections. As outlined by Litvak et al. [113], financial documents typically feature 13 predefined narrative section titles. Abdaljalil and Bouamor [111] employed a similar approach for section extraction. It compares BERT embeddings of reference sections to extracted ones, naming the extracted sections after the closest reference sections by cosine similarity. A frequency counter is then utilized to determine the weight of each section in reference summaries. Following a clustering process to identify the top three sections for each document, the section with the highest frequency weight is retained, and the top 1000 words of that section are included in the final summary. Shukla et al. [114] built an annotated dataset to define sections as "narrative" or "non-narrative". Tables of content within the report are parsed to extract section features: section name, page number, and length. Utilizing these features, a classification model is employed to categorize each section as narrative ("true") or non-narrative ("false"), based on the annotated dataset. The weight of a section is defined as the probability of it being narrative, as assigned by the model. Subsequently, it adopts the "K-Maximal Word Allocation" approach, which optimally distributes the required number of words among sections, based on their weights and the number of available words in each section. The final summary of the report is then generated, based on a set of pairs consisting of narrative sections and the number of words to be generated. Guided by DiMSum framework [114], Shukla et al. [115] showcased the significant enhancement in financial report summarization quality through the utilization of LLMs. It employs Retrieval Augmented Generation (RAG) to generate few-shot examples from the DiMSum annotated dataset, facilitating LLM-based few-shot classification to identify narrative sections and determine their weights. Subsequently, the LLM summarizer generates summaries, based on section name, content, and word count.

In scenarios where gold summaries have been removed from original annual reports, abstractive models offer the potential for greater conciseness by generating summaries from scratch. Sequence-to-Sequence model [121] and RNN-based encoder–decoder technique [122] have demonstrated effective performance in generating succinct summaries for short input and output sequences in the abstractive summarization task. Building upon these techniques, Li et al. [116] introduced an enhanced Seq2Seq model named TLGA. TLGA utilizes a hierarchical Transformer-BiLSTM encoder to capture long-range interactions and sequential semantics, while employing a Graph Attention-based decoder to leverage historical information of decoded tokens and capture key causal relations. This approach captures both qualitative and quantitative information as well as latent causal relationships from financial news.

Several other studies [123,124] have suggested integrating extractive and abstractive methodologies to enhance performance. Zmandar et al. [118] proposed a reinforcement learning model utilizing the standard policy gradient method to construct an end-to-end trainable computation graph that combines both extractor and abstractor agents. Singh [119] implemented pointer networks to extract important narrative sentences, subsequently using T5 to paraphrase extracted sentences into a concise yet informative sentence. Shukla et al. [115] introduced a novel approach to enhance the summarization of financial reports, the first instance of applying LLMs to the financial narrative summarization task. The emergence of various formats of financial text data has ignited research interest in techniques for financial text summarization. Mukherjee et al. [120] presented ECTSum,[2] a novel benchmark dataset for telegram-style bullet-point summarization of long transcripts of earning calls (ECT). Additionally, ECT-BPS is proposed, a hybrid approach that fine-tunes FinBERT to identify the most relevant sentences from the input ECT documents, followed by fine-tuning T5 [125] to paraphrase these sentences to the telegram-style format of target summary sentence. Liu et al. [117] proposed FINDSum, a large-scale dataset encompassing both long text and multiple tables, enabling extractive summaries to incorporate quantitative descriptions of critical metrics in tables. The model processes long text by adopting the Maximum Marginal Recall Gain (MMRG) method to select salient text segments as inputs, while transforming each table cell into a tuple and treating tuple selection as a binary classification task for tabular content. Three types of methods are evaluated on FINDSum dataset.[3] 1. combine-and-generate (CG) concatenates selected text segments and tuples as two types of input to a sequence-to-sequence summarizer 2. generate-and-combine (GC) parallelly generates summary on selected text segments and tuples, and concatenate two summaries as one. 3. generate-combine-and-generate (GCG) adopts a tuple-to-text generator to produce text descriptions of input tuples, concatenate tuples' text descriptions with selected text segments, and generate a final summary. Overall, CG and GCG methods demonstrate superior performance over text-only baselines on FINDSum's two subsets.

---

[2] https://github.com/rajdeep345/ECTSum.
[3] https://github.com/StevenLau6/FINDSum.

Table 7
Information extraction.

Literature	Type of extraction	Method	Performance
Xiao et al. [126]	Corporate event	FLAN-T5, FLAN-UL2, FLAN-Alpaca	F1-Score (Acquisition: 0.70, SEO: 0.79, CBI: 0.61)
Jacobs and Hoste [127]	Corporate event	BERT, RoBERTa, XLM, and XLNet	Precision: 0.737, Recall: 0.564, F1-Score: 0.591
Jacobs et al. [128]	Corporate event	SVM, RNN-LSTM	Precision: 0.80, Recall: 0.71, F1-Score: 0.73
Ein-Dor et al. [129]	Corporate event	BERT	Precision: 0.93, Recall: 0.95, F1-Score: 0.94
Carta et al. [130]	Corporate event	Hierarchical clustering	Silhouette Coefficient: 0.3, Dunn Index: 0.46
Guo et al. [131]	Corporate event	GNN + Transformer	Precision: 0.844, Recall: 0.791, F1-Score: 0.817
Cheng et al. [132]	Corporate event and relation	OpenIE, Bidirectional LSTM	Macro-F1: 0.7686, Micro-F1: 0.7515, Weighted-F1: 0.7527
Zhou et al. [68]	Corporate event	Bi-level event detector	Win Rate: 0.545 Big Win Rate: 0.342
Balashankar et al. [133]	Causal relationship	Predictive causal graph	RMSE: 0.020
Izumi and Sakaji [134]	Causal relationship	Causal-chain search algorithm	F1-Score (Financial Summaries: 0.71, Newspaper: 0.83)
Takayanagi et al. [135]	Causal relationship	ChatGPT	FinCausal (Precision: 0.646, Recall: 0.795, F1-Score: 0.713)
Rajpoot and Parikh [136]	Relation extraction	LLMs	F1-Score: 0.718
Hillebrand et al. [137]	Relation extraction	KPI-BERT	F1-Score: 0.7032
Zhang and Zhang [138]	Entity extraction	FinBERT-MRC	F1-Score: (ChFinAnn: 0.9278, AdminPunish: 0.9680)
Oral et al. [139]	Entity extraction	BiLSTM-CRF	Macro-F1: 0.8925, Micro-F1: 0.9223

3.4.2. Information extraction

Information Extraction (IE) focuses on identifying specific types of entities, relationships, events, and other factual information from unstructured textual sources [140]. IE is extensively used in the finance sector, particularly for the identification and detection of entities, corporate events, relations, causal relationships, financial document structure, etc (see Table 7). Event extraction is a fundamentally challenging task in NLP, extensively applied to manage vast and rapidly expanding collections of financial corpus which often contain multiple events with their elements scattered and mixed across the documents, complicating the extraction process [131]. In terms of event extraction from financial texts, Jacobs et al. [128] demonstrated that a feature-engineered linear kernel SVM approach obtains better performance than RNN-LSTM word-vector in the identification of corporate events from news articles. Jacobs and Hoste [127] proposed SEN-TiVENT, an annotated English economic news corpus of fine-grained company-specific events at the sentence level, and performed text classification using pre-trained BERT, RoBERTa, XLM, and XLNet models. Xiao et al. [126] fine-tuned the FLAN series of models to evaluate the effectiveness of LLMs in the context of corporate event prediction. A weakly supervised approach has also been adopted for event extraction. For example, Ein-Dor et al. [129] leveraged information contained in Wikipedia to generate a weakly labeled sentence dataset, demonstrating its effectiveness for company-specific events in news articles. In terms of unsupervised approaches, Carta et al. [130] proposed a hierarchical clustering-based approach that integrates news articles and StockTwits messages to detect corporate events.

The construction of knowledge graphs and their integration into event extraction has garnered significant attention from researchers. Guo et al. [131] enhanced the document-level event extraction by integrating a knowledge graph capturing entity relations and their attributes. Specifically, the proposed model employed a graph encoder to embed knowledge graph information for entities, applied named entity recognition to identify potentially relevant entities, and incorporated a transformer and linear layer to integrate the graph embeddings of the extracted entities into the event extraction model. Cheng et al. [132] introduced an event embedding framework, based on a knowledge graph, which not only extracts structured events from unstructured texts, but also builds the knowledge graph using the entities and relations mentioned therein, and then incorporates the knowledge graph information into the objective function of an event embedding learning model using a joint model. The learned representations serve as inputs for subsequent quantitative trading strategies. Meanwhile, Zhou et al. [68] presented a bi-level event detection model to detect corporate events from news articles which are subsequently used to predict stock movement for the event-driven trading strategy. The model includes low-level and high-level event detectors. The low-level detector identifies specific event-describing sub-sequences by classifying each

token, while the high-level takes the predicted results and integrates them with the global contextual information from the input article to predict the probabilities of each event's existence. The multi-task learning approach is also explored in identifying financial events. For instance, Li and Zhang [141] introduced a unified model for financial event classification, detection, and summarization. Specifically, the model leveraged pre-trained BERT to encode financial documents, with the encoded information shared across the event type classification, detection, and summarization components. A transformer structure serves as the decoder for the event summarization task. Alongside the input document encoded by BERT, the decoder incorporates predicted event types and cluster information, enabling it to focus on specific aspects of the event type during the summary generation.

Extraction of relations and causality [142,143] represents a prominent research domain within information extraction. Relation extraction stands as a pivotal task within NLP, focusing on identifying and categorizing connections between entities referenced in text. Rajpoot and Parikh [136] proposed GPT-FinRE, which explored the potential of GPT with two retrieval strategies, KNN and Efficient Prompt Retrieval (EPR), on the financial relation extraction task on REFinD dataset [144]. Hillebrand et al. [137] introduced a novel model that includes three main building blockers, a BERT-based sentence encoder, a named entity recognition decoder, and a relation extraction decoder, to extract and link key performance indicators from financial reports. Causality detection, meanwhile, seeks to unveil the causal relationships between entities, with applications ranging from stock movement prediction [133] and supporting financial information services [134]. For example, Balashankar et al. [133] proposed the Predictive Causal Graph by discovering the influence between words, based on temporal co-variance and demonstrated that PCG outperforms other graph-based methods in stock market prediction. To assess GPT's performance in causal reasoning, Takayanagi et al. [135] conducted a comprehensive evaluation of ChatGPT's causal text mining capabilities and observed that causal hallucination is evident in the low precision on the FinCausal dataset when implementing ChatGPT to detect causality relationship. Sharma et al. [145] released FinRED,[4] a relation extraction dataset curated from financial news and earning call transcripts containing relations from the finance domain. Lastly, Becquin [146] formulates the causality extraction into a span extraction and sequence labeling task. Span extraction aims to extract text spans, such as words or phrases, from plain texts [147]. Specifically, features are generated from an input text using transformers with high-performance tokenizers, and question-answering state-of-the-art architecture is implemented with two spans extracted (cause, effect) pairs. The architecture not only can extract cause and effect but also predict the number of cause/effect relationships in a document.

---

[4] https://github.com/soummyaah/FinRED.

Financial named entity recognition (FinNER), which refers to the process of identifying and categorizing specific entities, such as companies, currencies, dates, and numerical values, within financial texts, is a challenging task in financial text information extraction [138]. Zhang and Zhang [138] proposed a new model termed FinBERT-MRC by formulating FinNER task as a machine reading comprehension (MRC) problem, and achieved competitive performance. Oral et al. [139] extracted named entities with a BiLSTM-CRF architecture from text-intensive and visually rich scanned documents and predicted binary relations for each entity pair using a graph-based MLP architecture. Shah et al. [148] proposed a novel domain-specific model named FLANG (Financial LANGuage) which enhances masking using financial keywords and phrases, incorporating span boundary objective and in-filing objective. Together, an open-source comprehensive suite of benchmarks for the financial domain, named Financial Language Understanding Evaluation (FLUE), is released. It includes new benchmarks across NLP tasks which are sentiment analysis, FinNER, financial structure boundary detection, and financial question answering. The source dataset for FinNER is from [149].

Lastly, researchers also explored financial information extraction tasks such as financial document structure extraction [150], which extracts table-of-contents (TOC) from financial documents by detecting the document titles and organizing them hierarchically into a TOC, and structure boundary detection [151], an extension of boundary detection, which extracts the boundaries of sentences, lists and list items, including structure elements like header, footer and tables.

3.5. Question answering, virtual assistant and chatbot

Question answering systems, virtual assistants, and chatbots leveraging NLP techniques serve distinct yet overlapping roles in the realm of dialogue interactions. Question answering systems are optimized for precision and are adept at fetching exact answers from structured and unstructured data, primarily used in settings requiring high accuracy such as academic research or customer support [152]. Virtual assistants are crafted for a wide range of functionalities, from managing daily activities to handling enterprise processes. These systems prioritize user interaction and strive to simplify and enhance the user experience. Chatbots, particularly those integrated with LLMs, focus on simulating conversational human interactions to enhance customer engagement and service with responses that are context-aware and increasingly personalized. Despite their differences, all these technologies interpret and respond to human language, tailoring their functionalities to specific user needs and contexts.

3.5.1. Question answering

Question Answering (QA) involves designing a system that can answer questions posed by humans in natural language. It requires the integration of several fundamental components, such as tokenization, entity extraction, intent detection, and more. Key challenges include understanding the intent behind the query and retrieving the most relevant information from a set of documents [153]. In addition to general QA which answers natural language questions [15], the finance domain requires complex numerical reasoning and understanding of heterogeneous representations [154]. Recent research in QA has delved into various aspects, including the development of datasets [155] that challenge the reading comprehension abilities of models, like SQuAD [156] and TriviaQA [157] and the enhancement of reasoning capabilities necessary for answering questions more accurately (see Table 8). Numerical reasoning capability is another area of focus for QA in finance. Chen et al. [158] propose ConvFinQA, a new large-scale dataset aiming to study the chain of numerical reasoning in conversational question answering. Li et al. [159] introduced FinMath, which injects a tree-structured neural model to perform multi-step numerical reasoning, aiming to improve the numerical reasoning capacity of the model. Chen et al. [154] released a new large-scale

dataset, FINQA, for numerical reasoning over financial reports. Other areas of focus include open-domain question answering [160], which involves extracting answers from large, unstructured datasets using sophisticated retrieval and reading strategies. The quest for sophisticated QA systems in finance also extends to the ability to handle hybrid data, combining both tabular and textual content. For example, Zhu et al. [161] introduced TAT-QA, a benchmark leveraging both tabular and textual content, demonstrating the importance of context in finance-related queries, and further proposed a novel QA model termed TAGOP, which is capable of reasoning over both tables and text. Deng et al. [162] also presented the PACIFIC dataset to facilitate conversational question answering (CQA) over hybrid contexts in finance. Further, the knowledge-enhanced QA system, which enables information retrieval from an external knowledge base or knowledge graph, has improved the performance on QA tasks [163,164]. Lastly, the evaluation of LLM's capability including mathematical reasoning for QA over financial documents has also attracted researchers' attention significantly [165, 166].

3.5.2. Virtual assistant and chatbot

Unlike research in QA, which primarily leverages golden datasets, virtual assistant and chatbot are more application-oriented, spanning various functions in finance such as customer service [167], personal finance advice [168], cryptocurrency [169], and anti-fraud [170] etc. Virtual assistants are designed to perform tasks, based on user requests. These tasks can range from finding information online, like the best local restaurants, to more complex activities such as scheduling appointments. The effectiveness of a virtual assistant hinges on its ability to accurately parse the user's intent and to retrieve and interact with the relevant information effectively. Research in this domain often focuses on enhancing the assistants' understanding of user intents through improved dialogue management systems [171, 172] and by refining the query processing capabilities that help in better understanding and actioning user requests. The development of LLMs has significantly influenced the evolution of chatbots. These models, trained on vast text collections, are capable of generating responses that mimic human conversation and handle a wide range of linguistic tasks. The quality of these responses is heavily dependent on the context provided. Major tech companies like OpenAI, Meta, and Google primarily oversee these sophisticated models. Factors such as high computational costs, extensive data collection requirements, specialized expertise and labor, infrastructure and maintenance needs, licensing and intellectual property costs, and substantial energy consumption make it prohibitively expensive for corporations to develop and train these models from scratch. Consequently, practical applications of LLMs often involve methods like fine-tuning or advanced context management strategies. Recent innovations in this field include RAG techniques [173], which bolster the domain-specific abilities of chatbots, and developments like MEM-GPT [174], which approach LLMs as operating systems capable of both long and short-term memory management to efficiently update and manage context. Furthermore, the infinite attention mechanism [175] employs transformers to dynamically manage context length, addressing the inherent token limitations of LLMs. Additional research has concentrated on model fine-tuning strategies such as parameter-efficient fine-tuning (PEFT) by Xu et al. [176] and LOw Rank Adoption (LORA) by Hu et al. [177], which train LLMs through low-rank parameterization or in computationally efficient ways. Recent alignment efforts in LLMs include techniques like Reinforcement Learning with Human Feedback (RLHF) by Kaufmann et al. [178] and Supervised Fine-Tuning (SFT) by Mecklenburg et al. [179]. Reinforcement Learning with Generative Adversarial Feedback [180] provides an alternative alignment method that overcomes the challenges posed by RLHF, which is constrained by the expertise and productivity limitations of human evaluators, and SFT, which relies on additional, carefully selected expert demonstrations.

Table 8

Question answering, virtual assistant and chatbot.

Literature	Data sources	Method	Performance
Li et al. [159]	TAT-QA	RoBERTa-large, Auto-regressive sequence-to-tree model	Exact Match: 0.586, numeracy-focused F1: 0.641
Zhu et al. [161]	TAT-QA	RoBERTa, Sequence tagging, aggregation operator, scale prediction	Exact Match: 0.501, numeracy-focused F1: 0.580
Chen et al. [158]	CONVFINQA	RoBERTa-large, FinQANet	Execution accuracy: 0.6890, program accuracy: 0.6824
Chen et al. [154]	FinQA	RoBERTa-large, FinQANet	Execution accuracy: 0.6124, program accuracy: 0.5886

Lastly, the trends and characteristics of chatbot development in finance include personalization [181], decentralization [182], and ontology-based approaches [183], among others. As chatbots become more integral to finance operations, ensuring they operate reliably and ethically is crucial. This involves designing systems that are not only accurate but also trustworthy and free from biases. Current research [184] is exploring how to implement safety features and ethical guidelines to prevent the generation of harmful or misleading responses. These guardrails are essential for maintaining user trust and ensuring that interactions are both helpful and safe.

3.6. Risk management

3.6.1. Financial risk prediction

Research in financial risk prediction leverages a variety of data sources including financial reports [185–188], financial news [187, 188] and earning calls [189] (see Table 9). Wang et al. [186] explored the impact of sentiment words on financial risk, by establishing a regression model to predict future real-value risk, based on sentiment analysis and a ranking model to assess risk levels, using financial reports i.e., management's discussion and analysis of financial conditions and results of operations in the 10-K Form. The findings reveal that models developed using sentiment words outperform those trained on original texts, underscoring the significance of financial sentiment lexicons in predicting financial risk. Furthermore, the results also indicate a strong correlation between financial risk and financial sentiment words. Wang and Hua [189] defined financial risk as stock price volatility in the subsequent week and investigated its relationship with earnings calls. It utilized uni-grams, bi-grams, part-of-speech tags, named entities, and probabilistic frame-semantic features to construct a Gaussian copula model, evaluated by Kendall's tau and Spearman's correlation. The research examined three datasets consisting of transcribed quarterly earnings calls from the U.S. stock market during the period of the Great Recession. The datasets were categorized into three periods: pre-2009 (2006–2008), which marks the onset of the economic downturn; 2009, capturing the global spread of the financial crisis; and post-2009 (2010–2013), during the global economic recovery. It has demonstrated the potential of using quarterly earnings calls to predict short-term stock volatility. Another study by Rekabsaz et al. [190] explored volatility forecasting using SVM with Radial Basis Function (RBF) by employing textual features extracted from financial disclosures, such as Term Count (TC), Term Frequency (TF), TF–IDF, and BM25, combined with a dimension reduction strategy using Principal Component Analysis (PCA). Additionally, the study integrated market features like current volatility and sector specifics from factual market data. This sentiment analysis approach notably surpassed existing leading methods, demonstrating that data from 10-K reports are valuable for predicting volatility. Xing et al. [191] introduced the Sentiment-Aware Volatility Forecasting (SAVING) model, combining symbolic and sub-symbolic AI approaches by infusing grounded knowledge into neural networks. This model uses market sentiment to enhance predictions of stock return fluctuations. It surpasses traditional statistical models like GARCH and its variants, the Gaussian-process volatility model, and the latest neural stochastic volatility model and variational recurrent neural networks. In a separate study, Deveikyte et al. [192] developed a classifier for market volatility using Latent Dirichlet Allocation for topic modeling. The study demonstrates a significant negative correlation between positive social media sentiments, such as tweets, and

the subsequent day's market volatility, by analyzing the relationship between financial news, tweets, and the FTSE100 index. The findings also highlight that the accuracy of the model depends on the selected number of topics.

3.6.2. Fraud detection

Fraud detection in finance is a broad field that encompasses various strategies and methodologies to identify unusual patterns that may indicate fraudulent activities. In the finance sector, fraudulent activities can be categorized into transaction fraud, mortgage fraud, corporate and financial statement fraud, money laundering fraud, securities and commodities fraud, insurance fraud, and cryptocurrency fraud [193]. The advent of NLP techniques marks a significant transformation in detecting and preventing fraudulent activities. By leveraging NLP's power to sift through vast amounts of unstructured data, ranging from customer correspondences and transaction narratives to financial statements and digital social interactions, financial institutions are now better equipped to identify suspicious patterns and inconsistencies that may indicate fraud.

The primary research focus of NLP in fraud detection centers on identifying fraudulent transactions as well as detecting fraud within corporate and financial statements (see Table 10). Rodríguez et al. [194] presented a novel approach for fraud transaction detection, based on the transformer model. Boulieris et al. [195] proposed FraudNLP, the first anonymized dataset accessible to the public for detecting online fraud, and benchmarked both machine learning and deep learning techniques with multiple evaluation measures. Additionally, they argue that online behaviors adhere to patterns similar to natural language, suggesting that natural language processing techniques can effectively address these phenomena. Yang et al. [196] introduced FinChain-BERT, demonstrating a successful application of BERT models in detecting financial fraud and showing effectiveness in handling complex financial text information. In terms of corporate and financial statement fraud, Chen et al. [197] employs a fusion of NLP, Queen Genetic Algorithm (QGA), and SVM techniques to create a novel approach for identifying fraudulent content within the narratives of annual reports. Dong et al. [198] automatically captures various signals including sentiment and emotional attributes, topic characteristics, lexical features, and social network dynamics from financial social media data, which are subsequently integrated into machine learning classifiers to detect corporate fraud. Craja et al. [199] detects fraud from financial statements using a hierarchical attention network (HAN) to extract text features from annual reports including the management's discussion and analysis section. Achakzai and Peng [200] proposed a Dynamic Ensemble Section (DES) algorithm to detect fraudulent firms. A novel approach involves integrating a knowledge graph for fraud detection. Specifically, Mao et al. [201] developed a related-party transactions (RPTs) knowledge graph, by using techniques such as information extraction, entity extraction, relation extraction, and attribute extraction, to detect financial fraud.

3.6.3. Credit scoring

This section explores how NLP can enhance the qualification and quantification of an individual's creditworthiness. Traditional credit scoring methods rely heavily on factors such as credit history, income, and outstanding debts. However, NLP techniques offer an alternative approach by analyzing an individual's digital footprint, including social media messages, website interactions, and online browsing and

Table 9
Financial risk.

Literature	Data source	Method	Performance
Wang et al. [186]	10-K reports	SVR	Regression (MSE: 0.14894), Ranking (Kendall's Tau: 0.60458, Spearman's Rho: 0.63403)
Wang and Hua [189]	Earning calls	Linear Regression, Linear SVM, Gaussian SVM, Gaussian Copula Models	Spearman (Pre-2009: 0.425, 2009: 0.422, Post-2009: 0.375) Kendall: (Pre-2009: 0.315, 2009: 0.310, Post-2009: 0.282)
Oliveira et al. [54]	Twitter	MR, NN, SVM, RF, Ensemble	Daily return (Lowest NMAE: NDQ, 7.58) Daily trading volume (Lowest NMAE: DJIA, 5.84) Daily volatility (Lowest NMAE: DJIA, 2.79)
Rekabsaz et al. [190]	10-K reports	GARCH, SVM	MSE: 0.111, R^2: 0.527
Xing et al. [191]	StockTwits	SAVING, (E)GARCH, TARCH, GJR, GP-vol, VRNN, NSVM, LSTM, s+LSTM	Negative Log-Likelihood (NLL): −3.0642
Deveikyte et al. [192]	RavenPack, Twitter, Thomson Reuters	Logistic Regression	Accuracy (Headlines: 65%, Tweets: 65%, Stories: 67%) F1-Score (Headlines: 64%, Tweets: 64%, Stories: 63%)

Table 10
Fraud detection.

Literature	Task	Method	Performance
Rodríguez et al. [194]	Fraudulent transactions	RF, IF, SVM, KNN, LSTM, transformer	F1-Score: 0.830, MCC: 0.801, AUC: 0.871
Yang et al. [196]	Fraud detection from financial texts	FinChain-BERT	Precision: 0.97, Recall: 0.96, Accuracy: 0.97
Boulieris et al. [195]	Fraudulent transactions	LR, RF, KNN, SVM, LSTM, CNN, TCN	AUPRC: 0.404, F1 Score: 0.467
Craja et al. [199]	Financial statement fraud	LR, RF, SVM, XGB, ANN, HAN, GPT-2 w/ Attn	AUC: 0.9264, F1-score: 0.7500
Chen et al. [197]	Financial statement fraud	QGA-SVM	Accuracy: 0.852482
Achakzai and Peng [200]	Financial statement fraud	Dynamic ensemble selection algorithm	AUC: 0.763, MCC: 0.354, AP: 0.443
Seemakurthi et al. [202]	Financial statement fraud	LR, SVM, NN, LDA, Ensemble	Accuracy: 0.887, Specificity: 0.58, AUC: 0.72
Dong et al. [198]	Corporate fraud detection	SVM, NN, DT, LR	Accuracy: 0.8000, Recall: 0.8304, F1-Score: 0.7908, AUC: 0.8503
Mao et al. [201]	Corporate fraud detection	LR, DT, RF, XGB	Accuracy: 0.8502, AUC: 0.7508

purchasing behaviors, to assess their ability to repay debts. NLP-based credit scoring has the potential to provide financial solutions to unbanked communities that may not have access to traditional credit evaluation methods. By leveraging alternative data sources, NLP techniques can enhance the assessment of an individual's creditworthiness without the need for intrusive data collection methods, thereby offering a more accessible and inclusive approach (see Table 11). In the realm of business credit scoring, the extensive corpus of texts issued by or related to companies, including management reports, news articles, and social media content, offers significant opportunities for gaining deeper insights into a company's financial health [203].

Markov et al. [204] reviewed recent trends and considerations, noting a sustained increase in research focused on credit scoring. The research on NLP application in credit scoring can be broadly categorized into consumer credit scoring [205–207] such as default risk and business credit scoring [208–210] such as default risk, corporate distress and bankruptcy prediction [211]. Within the consumer credit scoring category, the focus is predominantly on peer-to-peer (P2P) lending, Wang et al. [205] proposed a novel consumer credit scoring approach using attention-based LSTM by incorporating the sequence of events in the online user operation behavior data, outperforming models without the operation behavior data, suggesting the value of user actions for improved credit risk assessment. Kriebel and Stitz [206] adopted machine learning and deep learning methods to predict the risk of default using user-generated text such as the brief description of themselves and the reason for their loan request and the results indicate that the text provided by users is valuable for predictions. Netzer et al. [207] implemented text mining and machine learning techniques to automatically process and analyze raw text in loan requests, and the findings indicated that incorporating textual information from the loan into the predictive model significantly enhances the prediction of loan defaults. In terms of business credit scoring, Stevenson et al. [208] proposed a deep learning approach using BERT to demonstrate the value of text for small business default prediction. Matin et al. [209] predicted corporate distress using deep learning of text segments in annual reports and Mai et al. [210] predicted corporate bankruptcy forecasting using textual disclosures in annual reports. The credit scoring task is formulated into a classification problem. The machine learning models such as Multilayer Perceptron (MLP), Linear Discrimination Analysis

(LDA), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and XGBoost (XGB), and deep learning models, such as Average Embedding Neural Networks (AE), CNN, RNN, LSTM, BERT and RoBERTa, are widely adopted.

3.7. Financial regulatory compliance monitoring

The potential of AI in enhancing financial regulatory compliance monitoring is significantly bolstered by NLP technologies [216]. The integration of NLP into financial regulatory compliance monitoring has become increasingly transformative, potentially reshaping the landscape of regulatory compliance management [217]. Despite these advancements, research in specific application areas remains sparse. NLP is applied to analyze communications within financial institutions to ensure compliance with regulatory requirements. It examines emails, chats, and documents for any suspicious content or language that might suggest manipulative actions or insider trading, boosting the surveillance effectiveness and proactively preventing potential breaches of compliance [218]. Additionally, Al-Shabandar et al. [219] emphasized the importance of real-time monitoring of transaction data, the identification of anomalies, and the application of NLP to align regulatory rules with institutional data. Furthermore, Abualhaija et al. [220] introduced an automated approach for monitoring and analyzing changes in financial regulations.

3.8. Explainable AI in finance

The notion of explainability holds tremendous importance in highly regulated domains such as finance where decisions can have significant consequences [221,222]. Yeo et al. [223] has categorized the explanation procedure into feature relevance, simplification, by example, visual, and textual. Most studies in NLP in finance, such as credit scoring, financial forecasting, and FSA, emphasize explainability through visual, feature relevance, and simplification procedures. Specifically, in visual explanation, Kumar et al. [224] introduced a method called Class Enhanced Attentive Response (CLEAR), which utilizes deconvolution on the penultimate layer before the output to create a visual attention map. This method graphically highlights the periods with the highest degree of attention by the stock-picking agent, along with a separate

Table 11
Credit scoring.

Literature	Data source	Method	Performance
Wang et al. [205]	Online operation behavior data of borrowers for peer-to-peer (P2P) lending	Attentive LSTM	AUC: 0.672, KS: 0.28
Kriebel and Stitz [206]	Text such as brief description of themselves and the reason for their loan request provided by users	RF, XGB, CNN, RNN, AE, BERT, Convolutional RNN, and RoBERTa	AUC (Incl. text: 0.712, Excl. text: 0.702)
Yang et al. [212]	Personality mining from social media information of borrowers	LDA, NB, SVM, MLP, and LR	Precision: 0.8325, Recall: 0.9437 F1-Score: 0.8830, Accuracy: 0.8757
Stevenson et al. [208]	Short description of the applicant's business context, the requirement of the loan etc.	LR, RF, BERT	AUC (New: 0.86, Existing: 0.88)
Netzer et al. [207]	Text in loan requests	LR, RF	AUC: 0.7260
Gao et al. [213]	Written description of the borrower's financial situation and purpose of the loan	Hypothesis testing	Marginal effects are reported to show statistical significance
Matin et al. [209]	Corporate distress prediction using annual reports and financial	CNN-LSTM, XGB, LR	AUC: 0.844, Log score: 0.1064
Mai et al. [210]	Corporate bankruptcy prediction using textual disclosures in annual reports	CNN, LR, SVM, RF	Accuracy: 0.712, AUC: 0.856
Ahmadi et al. [203]	Corporate bankruptcy prediction using business management reports	Dependency Sensitive Convolutional Neural Networks	Cohen's Kappa: 0.6544, Accuracy: 0.8414, F1-Score: 0.8760
El-Qadi et al. [214]	Comments including company description, business context, and activities	LR, SVM, RF, XGB, LightGBM, MLP	F1-Score: LR-0.209, Accuracy: MLP-0.923 Precision: RF-0.340
Nguyen and Huynh [215]	Assess manager's reflections towards company performance in financial reports	LR	Pseudo R^2: 0.2008

plot corresponding to the sentiment class of the stock. Deng et al. [225], on the other hand, applied knowledge graphs to visually link event entities that are extracted from stock news, offering a graphical representation of the relationships between features and their corresponding predictions. In terms of feature relevance, Carta et al. [226] explored the explainability of feature relevance by adopting various configurations of a permutation importance technique to prune less significant technical indicators, followed by the implementation of decision tree techniques for predicting stock market trends. The proposed method was found to be more reliable than LIME. Similarly, Ong et al. [227] used aspect-based sentiment analysis to investigate the correlation between stock price movements and key aspects identified in tweets, determining the sentiment of each aspect through a SenticNet-based graph convolutional network (GCN) [228]. This approach mirrors the feature relevance technique, with its focus on identifying key contributing aspects and their associated polarity values, emphasizing the interplay among financial variables rather than direct financial forecasting. For simplification procedure, Bandi et al. [229] combined sentiment analysis and technical analysis to develop a random forest model for stock forecasting, explained through LIME. Moreover, Gite et al. [230] employed LIME together with the LSTM-CNN model to effectively identify keywords that align with the target sentiment. In related work, Yuan and Zhang [231] adopted GPT-2 [232] for generating text explanations, incorporating specific keywords in the generated text. The presented method, named SC-DBA (soft-constrained dynamic beam allocation), uses a distinct network designed for analyzing news titles to extract keywords associated with different tiers of anticipated market volatility. The efficacy of the proposed method is evaluated, based on the fluency and practical relevance of the generated explanation. Specifically for NLP, Danilevsky et al. [233] categorized XAI for NLP into local versus global and self-explaining versus post-hoc, with explainability techniques including feature importance, surrogate model, example-driven and provenance-based, and declarative induction.

3.9. ESG and sustainable finance

Lim [234] conducted an extensive review of ESG and AI in the finance sector. However, our primary interest lies in the application of NLP techniques within this domain. While research like that of Capelli et al. [235], which integrates structured ESG data into financial forecasting, falls within the broader category of ESG and AI in finance, it does not align with our specific focus on NLP methodologies. The NLP research in ESG and sustainable finance encompasses two primary

streams, which have attracted researchers from both academia [236] and industry [237–240]. The first focuses on ESG disclosure, measurement, and governance including the identification of ESG topics and issues [236,241–245], addressing how NLP can enhance transparency, accuracy, and effectiveness in reporting and managing ESG criteria. The second stream explores the integration of ESG factors into broader financial applications, seeking innovative ways to incorporate ESG considerations into sentiment analysis [246], portfolio management [247], and risk management [248–251] etc. This distinction emphasizes the role of NLP in enhancing the ESG reporting framework and its potential to weave ESG considerations into the broader landscape of financial decision-making and product development.

Table 12 presents the research on NLP in ESG and sustainable finance. From the standpoint of NLP techniques, the fine-tuning of transformer-based models for ESG criteria has emerged as a prominent area of research [239,244,251]. Specifically, Raman et al. [240] used pre-trained BERT, XLNet, and RoBERTa to analyze the transcripts of corporate earning calls and detect historical trends of ESG discussions. Mehra et al. [251] proposed ESGBERT for classification tasks related to corporate ESG practices. Chen et al. [236] adopted BERT-like language models with data augmentation for multi-lingual ESG issue identification. The same ESG issue identification task is performed by Glenn et al. [237] using synthetic data and transfer learning and Wang et al. [238] by contrastive learning with BERT.

3.10. NLP for digital assets

Another promising application of NLP in finance is within the realm of digital assets. NLP for digital assets is a distinct area of application, reflecting our belief in NLP's potential to significantly influence the landscape of digital assets. Over the past decade, there has been an explosion of digital assets and digitalization of financial services [258]. Cryptocurrencies and Non-Fungible Tokens (NFTs) represent two of the most significant types of digital assets. Cryptocurrencies, or digital currencies, leverage cryptographic techniques to offer secure and decentralized financial transactions. Bitcoin (BTC) and Ethereum (ETH) are the most well-known cryptocurrencies. On the other hand, NFTs, share the digital token format with cryptocurrencies but differ significantly in their nature and usage. Unlike cryptocurrencies, which are fungible and can be exchanged on a one-to-one basis like traditional money, NFTs are inherently unique and cannot be exchanged on a like-for-like basis. This uniqueness allows NFTs to represent ownership of specific, often one-of-a-kind digital items such as artwork, collectibles, and even real estate in virtual worlds. Both cryptocurrencies and NFTs

Table 12
NLP in ESG and sustainable finance.

Literature	Task	Method	Performance
Guo et al. [248]	Predict stock volatility	BERT, Bayesian Model Ensemble	RMSE (MSCI-US: 0.289, AC-EU: 0.281) MAE (MSCI-US: 0.249, AC-EU: 0.241)
Apel et al. [249]	Approximate changes in transition risk from climate-related news events	BERT and multivariate regression	Significance is observed for pure-play and decarbonized indices
Jan [250]	Detect information asymmetry for sustainable development	RNN, LSTM	Accuracy: 0.9488, F1-Score: 0.9231, AUC: 0.9586
Mehra et al. [251]	Classify environmental score change	BERT	Accuracy for Change/No Change: 0.6709 Accuracy for Positive/Negative: 0.793
Sandwidi and Mukkolakal [246]	ESG aspect-oriented sentiment analysis	RoBERTa, ESG fine-tuning, Attention Augmentation	Accuracy: 0.9130, F1-Score: 0.9020
Koloski et al. [241]	Classify financial texts as sustainable or unsustainable	Knowledge-based latent heterogeneous representation	Precision: 0.90, Recall: 0.89, F1-Score: 0.89
Chen et al. [236]	Identify multi-lingual ESG issue	BERT-like models with data augmentation	F1-Score (English: 0.69, French: 0.78, Chinese: 0.392)
Haase and Sassen [252]	Uncover lobbying strategies in sustainable finance disclosure regulations	Pre-trained hierarchical topic model	Metadata helps to identify stakeholders and lobbying strategies
Kouloukoui et al. [243]	Analyze tweets and identify the network of ESG topics and their trend	Structural topic model	The popularity of Twitter as a platform to discuss CSR and ESG issues, clean and cleaner production has increased significantly
Lee et al. [253]	Classify E, S, or G using ESG documents and calculate ESG rating	BERT, RoBERTa, ALBERT	Accuracy: 0.8030, Recall: 0.79, Precision: 0.79, F1-Score: 0.79
Schimanski et al. [254]	Classify Environment, Social, Governance	FinBERT-ESG	Accuracy: (E: 0.957, S: 0.934, G: 0.897)
Sokolov et al. [255]	ESG categorization	BERT	AUC: 0.88 to 0.97, AUPR: 0.19 to 0.76
Pasch and Ehnes [239]	Classify ESG behavior	Fine-tuning transformer models for ESG	Accuracy: 0.79, F1-Score: 0.78
Raman et al. [240]	Detect historical trends of ESG discussion	BERT, XLNet and RoBERTa	Accuracy: 0.782, F1-Score: 0.784
Nugent et al. [244]	Detect ESG topics	BERT, Data Augmentation	Precision: 0.84, Recall: 0.84, F1-Score: 0.84
Huang et al. [256]	Classify ESG-related discussion	FinBERT	Accuracy: 0.895, Precision: 0.90 Recall: 0.895, F1-Score: 0.896
Fan and Wu [257]	Study effect of environmental regulations on firm valuation and policies	Multivariate regression	Stricter EPA regulations reduce pollution and increase firm innovation, especially green innovation

operate on blockchain technology, ensuring transparency, security, and permanence in transactions and ownership records.

Research in NLP for digital assets primarily focuses on predicting trends in the cryptocurrency market and providing insights into the NFTs market. Cryptocurrencies have seen unprecedented value growth, surpassing significant historical bubbles [259]. Recent studies have deep-dived into cryptocurrency market dynamics, as shown in Table 13, focusing on causality and correlation to understand behavior over time [260–262]. It has been shown that investor sentiment plays a crucial role in predicting returns of major cryptocurrencies, with nonlinear effects [263]. A hidden Markov Model was developed to predict market trends, based on sentiment, trading volume, and price, finding that markets react differently to sentiments depending on the overall market condition; however, this research focused solely on Bitcoin [264]. Another study employed an LSTM-based RNN model to analyze Chinese social media sentiment's impact on cryptocurrency prices, showing improved precision and recall over traditional autoregressive models [265].

In terms of market insights for NFTs, Leitter and Cambria [266] analyzed 200,000 tweets about NFTs using state-of-the-art neurosymbolic AI tools, which aims to decode the drivers of online conversations and sentiments surrounding NFTs, thereby uncovering factors that contribute to their perceived value. Additionally, Meyns and Dalipi [267] examined the human perceptions of, or attitudes towards, NFTs, specifically aiming to identify concerns expressed by social media users engaged with NFTs on Twitter.

3.11. Benchmarks and evaluation metrics

We summarize the common benchmark datasets and evaluation metrics for various applications in Table 14. The first type of metrics measures the closeness between the predicted value and ground truth in the context of a regression task. The popular metrics include

Weighted Cosine Similarity (WCS), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination or R-squared (R^2).

$$WCS = \frac{|P|}{|G|} \times \frac{\sum_{i=1}^{n}(G_i \times P_i)}{\sqrt{\sum_{i=1}^{n}(G_i^2)} \times \sqrt{\sum_{i=1}^{n}(P_i^2)}} \tag{1}$$

where P is the vector of scores predicted by the model and G is the vector of ground truth scores.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}| \tag{3}$$

where y_i is the gold standard score and $\hat{y}_i$ is the score predicted by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y}_i)} \tag{4}$$

where y_i is the gold standard score and $\hat{y}_i$ is the score predicted by the model.

The second type of metrics measures the categorical accuracy between predicted value and ground truth in the context of classification tasks. The popular metrics include Accuracy, Matthews Correlation Coefficient (MCC), Area Under the Curve (AUC), F1-Score, and the more generalized F_β-Score where β can be 1 and 2 for instance.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{6}$$

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

Table 13

NLP for cryptocurrency market.

Literature	Data source	Method	Task	Performance
Shen et al. [260]	Twitter	Granger causality test Vector Autoregression	Volume, return and realized volatility	Significant for Bitcoin RV and volume
Kraaijeveld and De Smedt [261]	Twitter	Granger causality test	Cryptocurrency price return	Significant predictive power (p < 0.05) on Bitcoin, Bitcoin Cash and Litecoin
Gao et al. [268]	Twitter	Regression	Return and volatility	Sentiment from Twitter can predict Bitcoin returns and volatility, based on significance level
Kim et al. [264]	Google Trends Twitter	Hidden Markov Model	Market movement prediction	Accuracy: 54%, AUC: 52%
Huang et al. [265]	Sina-Weibo	Autoregression, LSTM	Market movement prediction	Precision: 87%, Recall: 92.5%
Saha et al. [269]	Twitter	LSTM	Financial sentiment, BTC volume correlation	Sentiment: Accuracy: 0.8352, F1-Score: 0.8515 BTC Volume: Pearson's R: 0.1584
Oikonomopoulos et al. [262]	Twitter	Granger causality test Vector Autoregression	Price	MAPE: 0.0038

Table 14

Benchmarks and evaluation metrics.

Research topic	Benchmark dataset	Common evaluation metrics
Financial sentiment analysis	PhraseBank, SemEval 2017 Task 5, FiQA Task 1, StockSen, SentiEcon GS-1000, SEntFiN, FinLin, Twitter Financial News	Accuracy, MCC, F1-Score, WCS, MSE, R^2
Financial forecasting	BIGDATA22, ACL18, CIKM18, EDT	Accuracy, MCC, F1-Score, WCS, MSE, MAE, R^2
Portfolio management	No public benchmark dataset has been identified	MAE, Accuracy, MCC, Return, Sharpe ratio, Sterling Ratio, Sortino Ratio
Financial narrative processing	FNP, FINDSum, ECTSum, EDT, CGRAPH, REFinD, Chinese FinNER	ROUGE, F1-Score, Precision, Recall, Silhouette Coefficient, Dunn Index
Question answering, virtual assistant and chatbot	ConvFinQA, FinQA, TAT-QA, PACIFIC	Accuracy, Exact Match, Numeracy-focused F1-Score, ROUGE
Risk management	No public benchmark dataset has been identified	Accuracy, F1-Score, Precision, Recall, AUC, AUPRC, Kendall's Tau, Spearman's Rho, R^2, MSE, NMAE, NLL, Cohen's Kappa
Regulatory compliance monitoring	No public benchmark dataset has been identified	No common metrics has been identified
ESG and sustainable finance	FinSim-ESG	Accuracy, AUC, F1-Score, Precision, Recall, RMSE, MAE
Explainable AI in finance	No public benchmark dataset has been identified	No common metrics has been identified
NLP for digital assets	No public benchmark dataset has been identified	Accuracy, AUC, F-Score, MAPE, Pearson's Correlation Coefficient

$$\text{AUC} = \int_0^1 \text{TPR}\, d(\text{FPR}) \tag{9}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

$$F_\beta\text{-Score} = (1 + \beta^2) \times \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}} \tag{13}$$

In addition, Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is widely employed in assessing the quality of automatic summarization, machine translation and question answering. ROUGE-1 evaluates the overlap of individual words between the generated text (i.e., candidate) and a reference. ROUGE-2, on the other hand, measures the overlap of bigrams. ROUGE-N generalizes this approach to N-grams, which can include both unigrams (one word) and bigrams (two words). Lastly, ROUGE-L focuses on the Longest Common Subsequence, assessing the longest sequence of words that appears in both the generated text and the reference in the same order. Given candidate C and reference R, O_cr represents the number of overlapping N-grams, L_cr represents the longest common subsequence, N_c is the number of N-grams in candidate and N_r is the number of N-grams in reference, ROUGE-N and ROUGE-L metrics can be calculated as follows:

$$\text{Recall}_{ROUGE-N} = \frac{O\_cr}{N\_r} \tag{14}$$

$$\text{Precision}_{ROUGE-N} = \frac{O\_cr}{N\_c} \tag{15}$$

$$\text{F1-Score}_{ROUGE-N} = 2 \times \frac{\text{Precision}_{ROUGE-N} \times \text{Recall}_{ROUGE-N}}{\text{Precision}_{ROUGE-N} + Recall_{ROUGE-N}} \tag{16}$$

$$\text{Recall}_{ROUGE-L} = \frac{L\_cr}{N\_r} \tag{17}$$

$$\text{Precision}_{ROUGE-L} = \frac{L\_cr}{N\_c} \tag{18}$$

$$\text{F1-Score}_{ROUGE-L} = 2 \times \frac{\text{Precision}_{ROUGE-L} \times \text{Recall}_{ROUGE-L}}{\text{Precision}_{ROUGE-L} + Recall_{ROUGE-L}} \tag{19}$$

Exact-match accuracy (EM) and numeracy-focused F1 are introduced into the evaluation of question answering. EM is set to 1 when the characters of the prediction exactly match the characters of (one of) the true answer(s), otherwise will be 0. Numeracy-focused F1 defines F1 to be 0 when there is a number mismatch between the gold and predicted answers, regardless of other word overlap. In addition to execution accuracy, Chen et al. [154] also proposed program accuracy, by replacing all the arguments in a program with symbols, and then evaluating if two symbolic programs are mathematically equivalent. For example, the following two programs are equivalent programs:

$$\text{add}(a_1, a_2), \text{add}(a_3, a_4), \text{subtract}(\#0, \#1)$$

$$\text{add}(a_4, a_3), \text{add}(a_1, a_2), \text{subtract}(\#1, \#0)$$

For ranking tasks, Kendall's Tau (τ) and Spearman's Rho (ρ) are commonly used non-parametric metrics to measure the strength and direction of the association between two ranked variables.

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \tag{20}$$

where n represents the sample size, n_c is the number of concordant pairs, and n_d is the number of discordant pairs.

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{21}$$

where d_i is the difference in rankings for each object i, $i \in \{1, 2, \ldots, n\}$.

Portfolio management involves comparing the returns of an investment against its associated risks. Commonly used metrics for this purpose include the Sharpe ratio, Sterling Ratio, and Sortino Ratio,

which are employed to evaluate the effectiveness of the proposed portfolio selection strategy. Given R_p the return of a portfolio, R_f the risk-free rate, σ_p the standard deviation of the portfolio's excess return, and σ_d the standard deviation of the downside.

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \tag{22}$$

The Sterling Ratio evaluates the risk-adjusted returns of an investment by focusing on drawdowns rather than volatility.

$$\text{Sterling Ratio} = \frac{R_p - R_f}{\text{Average Annual Max Drawdown}} \tag{23}$$

The Sortino ratio, a modified version of the Sharpe ratio, distinguishes detrimental volatility from overall volatility by using the standard deviation of negative returns in a portfolio.

$$\text{Sortino Ratio} = \frac{R_p - R_f}{\sigma_d} \tag{24}$$

4. Main findings

The application of NLP in the financial sector has been transformative across various facets, from enhancing operational efficiencies to enabling more personalized customer engagements and improving strategic decision-making processes. We summarize our main findings from the following aspects.

4.1. Growing importance of NLP in finance

To address our first research question: What are the key drivers behind the increasing importance of NLP in the finance sector, and how do these drivers address the unique challenges and opportunities within this industry? The growing importance of NLP in the finance sector is driven by several key factors. First, the exponential growth in data, including unstructured data from financial documents, news, reports, social media, etc, necessitates sophisticated tools like NLP for efficient data processing and insight generation. Second, the advancement of NLP technologies, particularly from pre-trained language models to large language models, plays a crucial role in propelling the use of NLP in finance. NLP enables enhanced financial forecasting and sentiment analysis, allowing firms to make more informed decisions. Additionally, the complexity and volume of financial regulations demand advanced solutions for compliance monitoring, where NLP can automate and streamline the extraction and management of pertinent information. Moreover, as financial services increasingly move towards digital interfaces, NLP-powered virtual assistants and chatbots are essential for improving customer interaction and service efficiency. Lastly, the need for transparency and better risk management prompts the integration of NLP in areas like fraud detection and financial risk prediction, helping firms identify potential threats faster and with greater accuracy. These drivers collectively address the unique challenges of managing vast amounts of diverse data, ensuring compliance, enhancing user engagement, and bolstering security in the finance industry.

4.2. Stages of maturity and adoption of NLP in finance

To address our second research question: How has NLP transformed financial services and products from the past to the present, and what are the anticipated future developments in this area? What emerging areas of application for NLP in finance are being explored, and what potential do these areas hold for the future of financial services? We assess the stages of maturity and adoption of various NLP applications in finance as shown in Fig. 3. We aim to evaluate their potential to address real-world business challenges and to identify new opportunities for innovation and growth within the financial sector. Currently,

NLP in XAI in finance, portfolio management, ESG and sustainable finance, and financial forecasting are at the initial awareness stage, and have garnered significant enthusiasm for their potential to drive technological breakthroughs, although their practical impact may not yet be fully realized. Generative AI, particularly promising in wealth management, has the capability to revolutionize the industry by synthesizing information and creating personalized advisory services and investment strategies. Meanwhile, NLP applications in financial risk prediction, financial regulatory compliance monitoring, fraud detection, and credit scoring are now undergoing a more realistic assessment of their challenges and limitations during operational integration, striving to meet the expectations for accuracy and reliability of early adopters. NLP technologies, such as financial sentiment analysis, information extraction, financial text summarization, QA systems, and virtual assistants, and chatbots, are beginning to demonstrate clear benefits for enterprises and are becoming more widely understood, analyzing financial documents and markets, supporting decision-making processes across the finance sector.

4.3. Trends, challenges, and opportunities

To address our third research question: What are the prevailing trends in NLP research within the finance sector, and how do these trends reflect the evolving challenges and opportunities in the field? We summarize the trends, challenges, and opportunities from the following key perspectives.

4.3.1. Increasing reasoning capabilities

The ability to perform complex numerical reasoning and understand heterogeneous representations is particularly important in the finance domain. Financial data encompasses a broad spectrum of formats and content types, from numerical tables to varied textual documents. As such, NLP models operating within the finance domain must not only parse and interpret the data but also reason across. Applications such as QA demand the ability to infer answers from fragmented or implicit data, while FSA requires models to assess the sentiment and potential market impacts of nuanced textual information. Similarly, tasks like financial textual summarization and information extraction call for the capacity to distill and synthesize essential information from extensive textual and numerical data efficiently. Hence, increasing the reasoning capabilities of NLP models is crucial, as it enhances their ability to make informed decisions and predictions by effectively navigating and integrating the complex, multi-faceted landscape of financial data.

4.3.2. Understanding tabular and textual data

The application of NLP in finance demands adept handling of both tabular and textual data, reflecting the complex nature of financial information which encompasses both structured numerical data and unstructured textual content. Tabular data, which includes financial ratios, stock prices, and transaction histories, provides essential quantitative insights fundamental to financial analysis. On the other hand, textual data from news articles, financial reports, and regulatory filings introduces qualitative information that can profoundly influence market movements and investment strategies. This combination of capabilities enables NLP to not only extract meaningful insights from the vast and varied data streams but also enhance more nuanced and context-aware financial decision-making. Current research in fields such as QA, FSA, and FNP focus on advancing these capabilities, improving how models comprehend and integrate both forms of data in financial contexts.
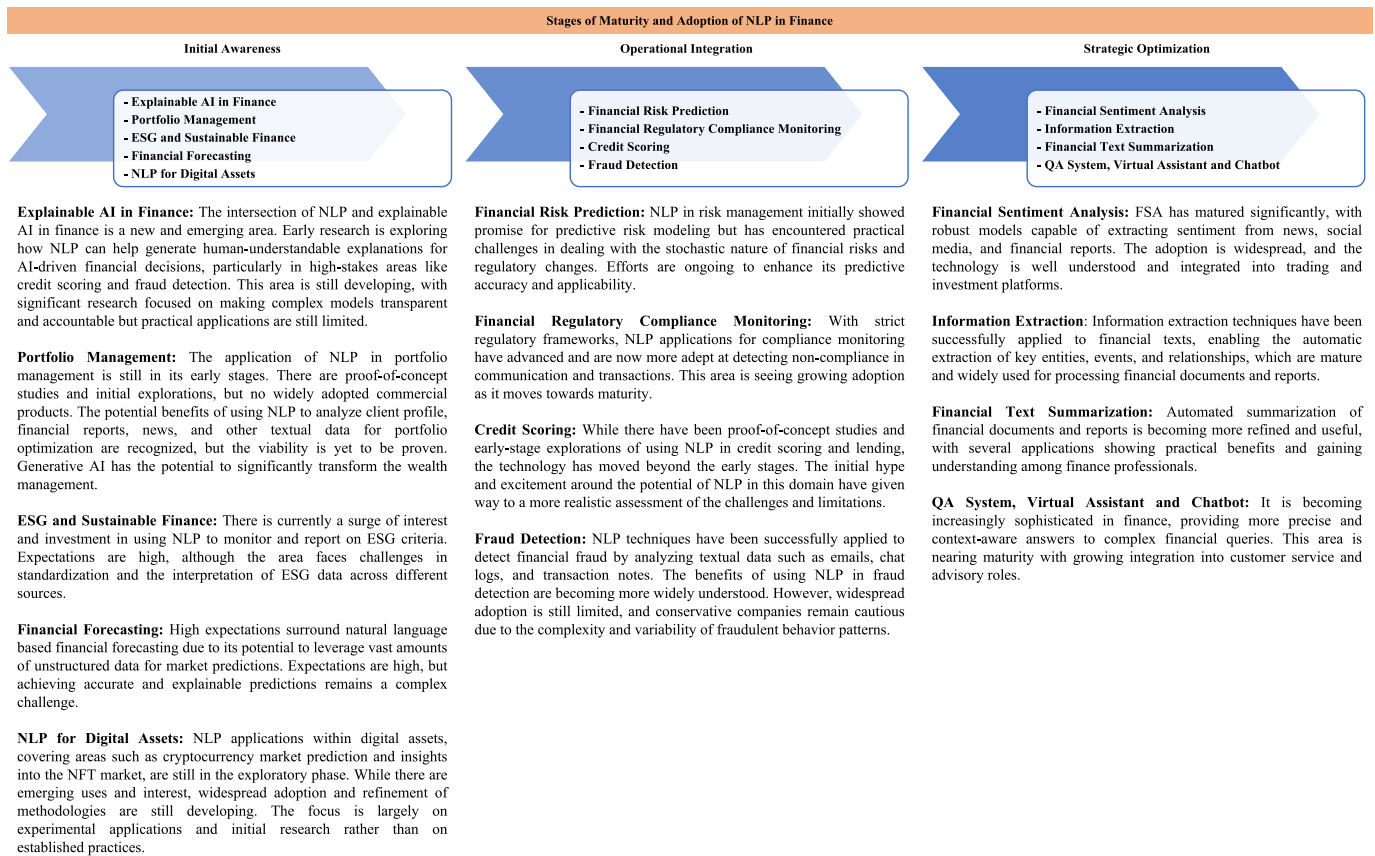
Stages of Maturity and Adoption of NLP in Finance

Initial Awareness	Operational Integration	Strategic Optimization
- Explainable AI in Finance - Portfolio Management - ESG and Sustainable Finance - Financial Forecasting - NLP for Digital Assets	- Financial Risk Prediction - Financial Regulatory Compliance Monitoring - Credit Scoring - Fraud Detection	- Financial Sentiment Analysis - Information Extraction - Financial Text Summarization - QA System, Virtual Assistant and Chatbot

Explainable AI in Finance: The intersection of NLP and explainable AI in finance is a new and emerging area. Early research is exploring how NLP can help generate human-understandable explanations for AI-driven financial decisions, particularly in high-stakes areas like credit scoring and fraud detection. This area is still developing, with significant research focused on making complex models transparent and accountable but practical applications are still limited.

Portfolio Management: The application of NLP in portfolio management is still in its early stages. There are proof-of-concept studies and initial explorations, but no widely adopted commercial products. The potential benefits of using NLP to analyze client profile, financial reports, news, and other textual data for portfolio optimization are recognized, but the viability is yet to be proven. Generative AI has the potential to significantly transform the wealth management.

ESG and Sustainable Finance: There is currently a surge of interest and investment in using NLP to monitor and report on ESG criteria. Expectations are high, although the area faces challenges in standardization and the interpretation of ESG data across different sources.

Financial Forecasting: High expectations surround natural language based financial forecasting due to its potential to leverage vast amounts of unstructured data for market predictions. Expectations are high, but achieving accurate and explainable predictions remains a complex challenge.

NLP for Digital Assets: NLP applications within digital assets, covering areas such as cryptocurrency market prediction and insights into the NFT market, are still in the exploratory phase. While there are emerging uses and interest, widespread adoption and refinement of methodologies are still developing. The focus is largely on experimental applications and initial research rather than on established practices.

Financial Risk Prediction: NLP in risk management initially showed promise for predictive risk modeling but has encountered practical challenges in dealing with the stochastic nature of financial risks and regulatory changes. Efforts are ongoing to enhance its predictive accuracy and applicability.

Financial Regulatory Compliance Monitoring: With strict regulatory frameworks, NLP applications for compliance monitoring have advanced and are now more adept at detecting non-compliance in communication and transactions. This area is seeing growing adoption as it moves towards maturity.

Credit Scoring: While there have been proof-of-concept studies and early-stage explorations of using NLP in credit scoring and lending, the technology has moved beyond the early stages. The initial hype and excitement around the potential of NLP in this domain have given way to a more realistic assessment of the challenges and limitations.

Fraud Detection: NLP techniques have been successfully applied to detect financial fraud by analyzing textual data such as emails, chat logs, and transaction notes. The benefits of using NLP in fraud detection are becoming more widely understood. However, widespread adoption is still limited, and conservative companies remain cautious due to the complexity and variability of fraudulent behavior patterns.

Financial Sentiment Analysis: FSA has matured significantly, with robust models capable of extracting sentiment from news, social media, and financial reports. The adoption is widespread, and the technology is well understood and integrated into trading and investment platforms.

Information Extraction: Information extraction techniques have been successfully applied to financial texts, enabling the automatic extraction of key entities, events, and relationships, which are mature and widely used for processing financial documents and reports.

Financial Text Summarization: Automated summarization of financial documents and reports is becoming more refined and useful, with several applications showing practical benefits and gaining understanding among finance professionals.

QA System, Virtual Assistant and Chatbot: It is becoming increasingly sophisticated in finance, providing more precise and context-aware answers to complex financial queries. This area is nearing maturity with growing integration into customer service and advisory roles.

Fig. 3. Stages of maturity and adoption of NLP in finance.

4.3.3. Leveraging large language models

Generative AI, particularly LLMs, is revolutionizing the financial sector by automating complex decision-making processes, personalizing customer experiences, synthesizing information, generating financial insights, and providing advisory services [270]. By leveraging vast structured and unstructured datasets, they not only enhance operational efficiency but also drive the creation of innovative financial products and services, transforming banking into a more agile and customer-centric industry. LLMs are reshaping the landscape of NLP in finance [271–273], influencing areas from FSA, FNP to QA, Virtual Assistant and Chatbot. Advancements in prompt engineering, retrieval-augmented generation (RAG), and multi-agent systems are demonstrating significant potential for use in industry. Additionally, ongoing research is addressing challenges faced by LLMs, such as hallucination, enhancing the ability of NLP systems to generate contextually relevant and highly precise responses in complex financial scenarios.

4.3.4. Capturing the sustainability wave

Leveraging NLP to sift through the massive volumes of ESG data facilitates informed decision-making for sustainability investments and ensures compliance with evolving regulatory frameworks. This critical application of NLP is pivotal in aligning financial strategies with global sustainability objectives. By focusing primarily on the utilization of NLP methodologies within the finance sector, we identify two main research streams in the domain of ESG and sustainable finance. The first stream emphasizes ESG disclosure, measurement, and governance, spotlighting the role of NLP in enhancing transparency, accuracy, and effectiveness in reporting and managing ESG criteria. The second stream delves into the integration of ESG factors into broader financial applications such as sentiment analysis, portfolio management, and risk management. This highlights the transformative potential of NLP to embed ESG considerations seamlessly into the financial decision-making process

and product development, thereby promoting a more sustainable and transparent financial landscape.

4.3.5. Transforming portfolio management

NLP is potentially revolutionizing the field of portfolio management by enhancing the speed and intelligence of decision-making processes. It enables real-time data processing and analysis, offering predictive insights that refine investment strategies using current market data and trends. An essential application of NLP is in portfolio optimization, where it analyzes quantitative and qualitative data to identify the best investment mix to maximize returns or minimize risk. Techniques such as Gaussian inverse reinforcement learning are employed to model investor sentiments and market dynamics, enabling the development of trading systems that effectively filter out market noise. Additionally, strategies, e.g., PROFIT and SARL utilize financial news and social media to optimize trading decisions, showing marked improvements over traditional benchmarks. Furthermore, NLP models incorporate sentiments from investor posts and public moods online, enhancing stock market investments and portfolio allocations. This sentiment-driven approach not only aligns with dynamic market conditions but also consistently outperforms equal-weighted strategies. Advanced models, e.g., time-aware LSTM and those using BERT for sentiment analysis, have proven their efficacy by significantly surpassing conventional methods in predicting stock returns and managing portfolio risks. The integration of these innovative NLP strategies into portfolio management is setting new standards for achieving superior financial outcomes.

4.3.6. Improving risk management and regulatory compliance monitoring

NLP is transforming the landscape of financial risk management and regulatory compliance monitoring. By leveraging the rich information contained in various financial documents, such as earnings calls, financial reports, and social media data, NLP facilitates more precise and

predictive risk assessment models. In financial risk prediction, studies have leveraged sentiment analysis and textual features from financial reports, news, and earnings calls to forecast stock volatility and financial risk with notable accuracy. For instance, models that utilize sentiment from financial statements have shown strong correlations with financial risk indicators, thereby improving prediction capabilities over traditional models. In fraud detection, NLP techniques analyze vast arrays of unstructured data, such as transaction narratives and corporate financial statements, to identify suspicious patterns indicative of fraudulent activities. Innovative NLP applications like transformer models and knowledge graphs have demonstrated enhanced effectiveness in detecting complex financial fraud scenarios. Furthermore, NLP has introduced new methodologies in credit scoring by assessing individuals' digital footprints, providing inclusive financial solutions especially beneficial to unbanked populations. Additionally, in the realm of regulatory compliance, NLP has been instrumental in monitoring internal communications within financial institutions, ensuring adherence to regulatory norms, and proactively preventing compliance breaches. Through these applications, NLP not only enhances the precision and efficiency of risk and compliance management processes but also supports more informed decision-making in the finance sector.

4.3.7. Augmenting algorithmic trading

Natural language-based financial forecasting augments the landscape of algorithmic trading by enabling the rapid processing and integration of financial documents, news, and social media content, thereby improving model performance. This allows trading systems to quickly capitalize on market movements driven by real-time events, significantly enhancing the speed and effectiveness of trading strategies. Natural language-based financial forecasting in this area includes forecasting in stock, foreign exchange, cryptocurrency, options pricing, firm valuation, etc. A significant trend involves incorporating company relationships into financial forecasting models, refining methods for preprocessing textual data, and enhancing data representation. By analyzing textual data from diverse sources, NLP helps in identifying sentiment and key information that directly impacts financial markets, and predict market signals, enabling algorithmic traders to execute more timely and informed trades, based on the latest news and social media trends, leading to potentially higher returns and more robust trading strategies. From this perspective, NLP serves as a powerful tool in the arsenal of modern algorithmic trading, bridging the gap between unstructured data and actionable trading insights.

4.3.8. Enhancing financial narrative processing

Financial text summarization and information extraction are two primary application areas of FNP, each significantly enhancing operational efficiency. A prominent trend in financial text summarization is the increasing reliance on both extractive and abstractive techniques, with extractive methods particularly dominant in sentence or section-level summarization using pre-trained language models. A key challenge in this field stems from the repetitiveness and density of financial texts, which complicate the summarization process. To address this, researchers have proposed models that focus on extracting and summarizing key narrative sections rather than entire reports, thus improving the relevance and utility of summaries. The integration of extractive and abstractive methods, capitalizing on the strengths of both, has shown considerable potential. Innovative approaches such as reinforcement learning models and pointer networks, which combine the extraction of key sentences with advanced paraphrasing techniques, are showing promise. Additionally, the use of LLMs for narrative section identification and summarization is emerging as a powerful tool, enhancing the quality of financial report summaries. There is also a growing interest in developing specialized datasets and methods for summarizing diverse financial document formats, such as earnings call transcripts, which could further refine the precision and applicability of financial text summarization technologies.

In the field of information extraction, a key trend is leveraging pre-trained language models like BERT, RoBERTa, and GPT for tasks such as event extraction and financial relation identification, which are demonstrating promising results in improving the accuracy and depth of extracted information, with knowledge graphs further enhancing the models' contextual understanding. The challenges include handling the complex, dispersed information typical of financial texts and the resource-intensive nature of creating well-annotated datasets necessary for effective model training. However, there is potential for expanding the use of weakly supervised and unsupervised learning methods to reduce dependency on extensive labeled datasets. Exploring multitask learning frameworks that simultaneously perform classification, detection, and summarization of financial events can lead to more efficient processing of financial documents. Moreover, the development of methods for detecting causality, such as those using predictive causal graphs, offers new avenues for understanding the intricate dynamics of financial markets, providing insights into cause-and-effect relationships crucial for market prediction. Lastly, automated financial reporting and financial statement analysis using NLP, particularly with LLMs, has also emerged as a promising area of research.

4.3.9. Robustness and trustworthiness

NLP technologies have undergone significant advancements, becoming increasingly sophisticated to address the specialized linguistic challenges inherent in the financial sector. This progression enhances precision and resilience, leading to reduced operational risks associated with errors and biases. As NLP systems evolve, their ability to understand and process complex financial jargon and semantics improves, thereby enhancing decision-making processes and compliance monitoring in finance. Trustworthiness forms a crucial pillar in evaluating AI systems within the financial industry. It measures the extent to which AI models and algorithms can be relied upon to function as expected, make decisions that are both accurate and ethical, and mitigate potential adverse impacts. This concept is closely associated with "Intention Awareness", which pertains to the AI's capability to align its operations with intended goals. However, trustworthiness also encompasses explainability and interpretability, key aspects of what is often referred to as XAI. In finance, XAI facilitates transparency by enabling stakeholders to understand and trust the decision-making processes of AI systems, thus bridging the gap between complex AI operations and user-friendly outputs. These advancements signify a pivotal shift towards more reliable and accountable NLP applications in the financial sector, paving the way for broader acceptance and integration of AI technologies in high-stakes environments. Lastly, the presence of algorithmic bias, which inherits biases from the training data, can lead to decisions that disproportionately impact certain individuals and groups such as credit scoring. Consequently, the pursuit of fairness and equity in models is crucial within financial services. Lastly, the robustness and trustworthiness also extend to considerations of data privacy and security, especially when managing sensitive compliance-related data, which requires robust measures to ensure data privacy and prevent unauthorized access, safeguarding integrity and confidentiality.

4.3.10. NLP system implementation

This section zooms into the practical aspects of NLP system implementation, focusing on data and platform integration, process integration, risk management, regulatory compliance, data privacy, scalability and performance, ethics and fairness, and ROI estimation. Firstly, financial institutions are expected to seamlessly integrate both structured and unstructured data sources into their enterprise data platforms. This requires the development of robust data ingestion and preprocessing pipelines that cleanse, normalize, and prepare data for NLP systems. Integrating these platforms necessitates compatibility with existing IT infrastructures, which may require upgrading legacy systems or deploying new middleware solutions. Next, integrating NLP systems into the routine workflows of financial institutions demands forward planning.

It is crucial to synchronize NLP systems with established business processes. Adopting change management strategies is essential to facilitate smooth transitions and foster user acceptance, ensuring the benefits of NLP technologies are fully realized. Additionally, implementing NLP systems introduces new risks associated with data accuracy, algorithmic bias, and system reliability. Comprehensive risk management protocols must be established to mitigate these issues. This includes continuous model monitoring, periodic updates to models to reflect evolving data, and implementing fallback mechanisms to prevent system failures. Another key consideration is regulatory compliance, which requires NLP systems to be developed with an awareness of the regulatory frameworks governing financial institutions. Compliance with financial regulations, data protection laws, and international standards is paramount. In terms of data privacy, financial institutions must prioritize the safeguarding of sensitive information. NLP systems should be designed to anonymize and secure personal data, ensuring compliance with stringent data protection laws. Integrating robust data privacy measures into system design and operations is essential, underpinned by clear policies and procedures to prevent data breaches and unauthorized access. Scalability and performance are also crucial for NLP systems. As financial institutions manage ever-larger datasets and require faster processing for timely decision-making, scalability and performance are paramount. NLP systems must be designed to efficiently handle scalability, ensuring they can accommodate increasing data volumes without compromising performance. This involves leveraging scalable cloud infrastructures, employing efficient data storage solutions, and utilizing advanced algorithms capable of high-speed data processing. Moreover, the deployment of NLP systems must also consider ethical implications and strive for fairness in automated decisions. Financial institutions should develop frameworks that ensure NLP models are devoid of biases that may lead to discriminatory treatment, based on demographic factors. Regular audits and updates of models to identify and rectify biases, coupled with transparency in decision-making processes, are essential for upholding fairness and ethical standards in automated systems. Lastly, the ROI estimation and economic value of NLP systems need to be measured, potentially from perspectives of revenue impact and productivity gains.

4.3.11. Future directions

The landscape of NLP in finance is poised for transformative advancements with several emerging areas ripe for exploration. The adoption of time series transformers (TST) in financial forecasting signifies a significant advancement in the modeling of financial time series data. TST models, known for their ability to capture complex temporal relationships, promise to greatly enhance forecasting accuracy, especially in volatile financial environments. Reinforcement learning is emerging as a potent tool for portfolio management, leveraging dynamic decision-making to optimize investment strategies in real time. This approach could revolutionize portfolio management by continuously learning and adapting to new market conditions. Neuro-symbolic AI presents a compelling frontier, integrating the interpretability of symbolic AI with the learning capabilities of neural networks. This hybrid approach is particularly promising for complex financial tasks that require both deep learning for pattern recognition and symbolic reasoning for rule-based decision-making. In the realm of QA systems, enhancing numerical reasoning capabilities to interpret and answer queries involving textual and tabular data is crucial. This will improve decision-making processes by providing precise and contextually relevant answers to financially oriented questions. FNP is another promising area, where the accuracy of information extraction and summarization from financial texts can be greatly enhanced. The integration of knowledge graphs into NLP models could enable more sophisticated processing of financial narratives, identifying relationships and insights that are not readily apparent.

Fine-grained FSA within financial documents can be targeted more precisely with NLP, focusing on specific aspects such as market trends, corporate performance, or economic indicators, thus providing more nuanced sentiment insights that are crucial for investment decisions. Lastly, in the realm of compliance, there is a growing need to demonstrate the effectiveness of NLP applications. Explainable AI models in particular are crucial in finance, where stakeholders require clear, understandable explanations for automated decisions to ensure compliance with regulatory standards and to build trust in AI-driven processes. Overall, the application of NLP in finance is transitioning from its initial, task-driven early stages to a more in-depth exploration and adoption of innovative algorithms. This evolution enhances the system's ability to effectively handle complex data, thereby improving functionality and precision in financial contexts.

5. Conclusion

This survey conducted a comprehensive review of research in NLP in finance, targeting researchers, practitioners, and industry professionals. We began with the basic structure of an IMF industry report and expanded to include a detailed review of research literature. This review serves as a reference for both researchers and practitioners, highlighting the evolution and diverse applications of NLP across both academic research and industrial implementations. Our review framework systematically maps the relationships between financial data, NLP techniques, key players, and financial applications, identifying ten principal areas where NLP has potentially revolutionized the finance sector. Additionally, we analyzed key research contributions that have significantly influenced the application of NLP in finance, emphasizing its transformative effects on services, products, and operations. This survey not only addresses practical challenges but also highlights the transformative potential of NLP technologies in the finance industry. Lastly, we summarized the current trends, challenges, and opportunities from nine perspectives, and provided eight promising directions for future research.

6. Limitations

This review is designed to explore broad themes within the field of NLP in finance. It does not include detailed data extraction, data quality assessment, and risk of bias evaluation due to the expansive nature of NLP's application in finance. This is intended as part of the review's design which allows for a more extensive discussion across a broader array of topics, which was deemed more beneficial for our intended audience of researchers, practitioners, and professionals in the finance sector. By delineating these boundaries, we aim to clarify this article's intended scope and encourage readers to interpret the findings within the context of an exploratory and thematic analysis.

CRediT authorship contribution statement

Kelvin Du: Writing – original draft, Methodology, Investigation, Conceptualization. **Yazhi Zhao:** Writing – original draft, Investigation, Conceptualization. **Rui Mao:** Writing – review & editing, Methodology, Formal analysis. **Frank Xing:** Writing – review & editing, Formal analysis, Conceptualization. **Erik Cambria:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Data availability

No data was used for the research described in the article.

References

[1] I.E. Fisher, M.R. Garnsey, M.E. Hughes, Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research, Intell. Syst. Account. Finance Manage. 23 (3) (2016) 157–214.

[2] C.-C. Chen, H.-H. Huang, H.-H. Chen, NLP in FinTech applications: past, present and future, 2020, arXiv preprint arXiv:2005.01320.

[3] E.B. Boukherouaa, M.G. Shabsigh, K. AlAjmi, J. Deodoro, A. Farias, E.S. Iskender, M.A.T. Mirestean, R. Ravikumar, Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance, International Monetary Fund, 2021.

[4] R. Mao, X. Li, Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 13534–13542.

[5] A. Valdivia, V. Luzón, E. Cambria, F. Herrera, Consensus vote models for detecting and filtering neutrality in sentiment analysis, Inf. Fusion 44 (2018) 126–135.

[6] E. Cambria, X. Zhang, R. Mao, M. Chen, K. Kwok, SenticNet 8: Fusing emotion AI and commonsense AI for interpretable, trustworthy, and explainable affective computing, in: Proceedings of International Conference on Human-Computer Interaction, HCII, Washington DC, USA, 2024.

[7] L. Zhu, R. Mao, E. Cambria, B.J. Jansen, Neurosymbolic AI for personalized sentiment analysis, in: Proceedings of International Conference on Human-Computer Interaction, HCII, Washington DC, USA, 2024.

[8] K. Du, F. Xing, R. Mao, E. Cambria, Financial sentiment analysis: Techniques and applications, ACM Comput. Surv. 56 (9) (2024) http://dx.doi.org/10.1145/3649451.

[9] R. Moraes, J.F. Valiati, W.P.G. Neto, Document-level sentiment classification: An empirical comparison between SVM and ANN, Expert Syst. Appl. 40 (2) (2013) 621–633.

[10] P. Ferguson, N. O'Hare, M. Davy, A. Bermingham, P. Sheridan, C. Gurrin, A.F. Smeaton, Exploring the Use of Paragraph-Level Annotations for Sentiment Analysis of Financial Blogs, Technical Report, Dublin City University, 2009.

[11] P. Zhang, Z. He, Using data-driven feature enrichment of text representation and ensemble technique for sentence-level polarity classification, J. Inf. Sci. 41 (4) (2015) 531–549.

[12] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S.M. Jiménez-Zafra, G. Eryiğit, SemEval-2016 task 5: Aspect based sentiment analysis, in: S. Bethard, M. Carpuat, D. Cer, D. Jurgens, P. Nakov, T. Zesch (Eds.), Proceedings of the 10th International Workshop on Semantic Evaluation, SemEVal-2016, Association for Computational Linguistics, San Diego, California, 2016, pp. 19–30.

[13] P. Malo, A. Sinha, P. Korhonen, J. Wallenius, P. Takala, Good debt or bad debt: Detecting semantic orientations in economic texts, J. Assoc. Inf. Sci. Technol. 65 (4) (2014) 782–796.

[14] K. Cortis, A. Freitas, T. Daudert, M. Huerlimann, M. Zarrouk, S. Handschuh, B. Davis, Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news, in: Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval-2017, 2017, pp. 519–535.

[15] M. Maia, S. Handschuh, A. Freitas, B. Davis, R. McDermott, M. Zarrouk, A. Balahur, WWW'18 open challenge: financial opinion mining and question answering, in: Companion Proceedings of the the Web Conference 2018, 2018, pp. 1941–1942.

[16] K. Du, F. Xing, E. Cambria, Incorporating multiple knowledge sources for targeted aspect-based financial sentiment analysis, ACM Trans. Manage. Inf. Syst. 14 (3) (2023) 23, http://dx.doi.org/10.1145/3580480.

[17] N. Oliveira, P. Cortez, N. Areal, Stock market sentiment lexicon acquisition using microblogging data and statistical measures, Decis. Support Syst. 85 (2016) 62–73.

[18] K. Du, F. Xing, R. Mao, E. Cambria, FinSenticNet: A concept-level lexicon for financial sentiment analysis, in: 2023 IEEE Symposium Series on Computational Intelligence (SSCI), 2023, pp. 109–114.

[19] D. Araci, FinBERT: Financial sentiment analysis with pre-trained language models, 2019, arXiv:1908.10063.

[20] Z. Liu, D. Huang, K. Huang, Z. Li, J. Zhao, FinBERT: A pre-trained financial language representation model for financial text mining, in: IJCAI, 2020, pp. 4513–4519.

[21] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, et al., Improving Language Understanding by Generative Pre-Training, OpenAI, 2018, pp. 1–12.

[22] S. Fatemi, Y. Hu, A comparative analysis of fine-tuned LLMs and few-shot learning of LLMs for financial sentiment analysis, 2023, arXiv:2312.08725.

[23] G. Fatouros, J. Soldatos, K. Kouroumali, G. Makridis, D. Kyriazis, Transforming sentiment analysis in the financial domain with ChatGPT, 2023, arXiv:2308.07935.

[24] B. Zhang, H. Yang, T. Zhou, M. Ali Babar, X.-Y. Liu, Enhancing financial sentiment analysis via retrieval augmented large language models, in: Proceedings of the Fourth ACM International Conference on AI in Finance, 2023, pp. 349–356.

[25] B. Zhang, H. Yang, X.-Y. Liu, Instruct-FinGPT: Financial sentiment analysis by instruction tuning of general-purpose large language models, 2023, arXiv:2306.12659.

[26] N. Hu, P. Liang, X. Yang, Whetting All Your Appetites for Financial Tasks with One Meal from GPT? A Comparison of GPT, FinBERT, and Dictionaries in Evaluating Sentiment Analysis, Technical Report, SSRN: , 2023, https://ssrn.com/abstract=4426455.

[27] X. Deng, V. Bashlovkina, F. Han, S. Baumgartner, M. Bendersky, What do llms know about financial markets? a case study on reddit market sentiment analysis, in: Companion Proceedings of the ACM Web Conference 2023, 2023, pp. 107–110.

[28] S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, G. Mann, BloombergGPT: A large language model for finance, 2023, arXiv:2303.17564.

[29] K. Du, F. Xing, R. Mao, E. Cambria, An evaluation of reasoning capabilities of large language models in financial sentiment analysis, in: IEEE Conference on Artificial Intelligence, IEEE CAI, Singapore, 2024, pp. 189–194.

[30] F. Xing, Designing heterogeneous LLM agents for financial sentiment analysis, ACM Trans. Manag. Inf. Syst. (2024) http://dx.doi.org/10.1145/3688399.

[31] R. Mao, K. Du, Y. Ma, L. Zhu, E. Cambria, Discovering the cognition behind language: Financial metaphor analysis with MetaPro, in: 2023 IEEE International Conference on Data Mining, ICDM, IEEE, Shanghai, China, 2023, pp. 1211–1216.

[32] R. Mao, X. Li, K. He, M. Ge, E. Cambria, MetaPro Online: A computational metaphor processing online system, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations), Vol. 3, 2023, pp. 127–135.

[33] R. Mao, K. He, C.B. Ong, Q. Liu, E. Cambria, MetaPro 2.0: Computational metaphor processing on the effectiveness of anomalous language modeling, in: Findings of the Association for Computational Linguistics: ACL, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 9891–9908.

[34] R. Manro, R. Mao, L. Dahiya, Y. Ma, E. Cambria, A cognitive analysis of CEO speeches and their effects on stock markets, in: Proceedings of the 5th International Conference on Financial Technology, ICFT, Singapore, 2024.

[35] J. Park, H.J. Lee, S. Cho, Automatic construction of context-aware sentiment lexicon in the financial domain using direction-dependent words, 2021, arXiv preprint arXiv:2106.05723.

[36] S. Krishnamoorthy, Sentiment analysis of financial news articles using performance indicators, Knowl. Inf. Syst. 56 (2) (2018) 373–394.

[37] D. Araci, Finbert: Financial sentiment analysis with pre-trained language models, 2019, arXiv preprint arXiv:1908.10063.

[38] L. Zhao, L. Li, X. Zheng, A BERT based sentiment analysis and key entity detection approach for online financial texts, in: 2021 IEEE CSCWD, 2021, pp. 1233–1238.

[39] M. Jiang, M. Lan, Y. Wu, Ecnu at semeval-2017 task 5: An ensemble of regression algorithms with effective features for fine-grained sentiment analysis in financial domain, in: Proceedings of SemEval-2017, 2017, pp. 888–893.

[40] K. Schouten, F. Frasincar, F. de Jong, Commit at semeval-2017 task 5: Ontology-based method for sentiment analysis of financial headlines, in: Proceedings of SemEval-2017, 2017, pp. 883–887.

[41] A. Dridi, M. Atzeni, D.R. Recupero, FineNews: fine-grained semantic sentiment analysis on financial microblogs and news, Int. J. Mach. Learn. Cybern. (2019) 1–9.

[42] M.S. Akhtar, A. Kumar, D. Ghosal, A. Ekbal, P. Bhattacharyya, A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis, in: Proceedings of EMNLP, 2017, pp. 540–546.

[43] D. Ghosal, S. Bhatnagar, M.S. Akhtar, A. Ekbal, P. Bhattacharyya, IITP at SemEval-2017 task 5: an ensemble of deep learning and feature based models for financial sentiment analysis, in: Proceedings of SemEval-2017, 2017, pp. 899–903.

[44] R. Mao, X. Li, M. Ge, E. Cambria, MetaPro: A computational metaphor processing model for text pre-processing, Inf. Fusion (ISSN: 1566-2535) 86–87 (2022) 30–43.

[45] C. Xiang, J. Zhang, F. Li, H. Fei, D. Ji, A semantic and syntactic enhanced neural model for financial sentiment analysis, Inf. Process. Manage. 59 (4) (2022) 102943.

[46] A. Sinha, S. Kedas, R. Kumar, P. Malo, SEntFiN 1.0: Entity-aware sentiment analysis for financial news, J. Assoc. Inf. Sci. Technol. (2022).

[47] D. de França Costa, N.F.F. da Silva, INF-ufg at FiQA 2018 task 1: predicting sentiments and aspects on financial tweets and news headlines, in: Companion Proceedings of the the Web Conference 2018, 2018, pp. 1967–1971.

[48] G. Piao, J.G. Breslin, Financial aspect and sentiment predictions with deep neural networks: an ensemble approach, in: Companion Proceedings of the the Web Conference 2018, 2018, pp. 1973–1977.

[49] L. Luo, X. Ao, F. Pan, J. Wang, T. Zhao, N. Yu, Q. He, Beyond polarity: Interpretable financial sentiment analysis with hierarchical query-driven attention, in: IJCAI, 2018, pp. 4244–4250.

[50] F.Z. Xing, E. Cambria, R.E. Welsch, Natural language based financial forecasting: a survey, Artif. Intell. Rev. 50 (1) (2018) 49–73.

[51] F. Xing, E. Cambria, R. Welsch, Intelligent asset allocation via market sentiment views, IEEE Comput. Intell. Mag. 13 (4) (2018) 25–34.

[52] A. Mudinas, D. Zhang, M. Levene, Market trend prediction using sentiment analysis: lessons learned and paths forward, 2019, arXiv preprint arXiv:1903.05440.

[53] Y. Ma, R. Mao, Q. Lin, P. Wu, E. Cambria, Multi-source aggregated classification for stock price movement prediction, Inf. Fusion 91 (2023) 515–528.

[54] N. Oliveira, P. Cortez, N. Areal, The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices, Expert Syst. Appl. 73 (2017) 125–144.

[55] A.K. Nassirtoussi, S. Aghabozorgi, T.Y. Wah, D.C.L. Ngo, Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment, Expert Syst. Appl. 42 (1) (2015) 306–324.

[56] G.M. Mian, S. Sankaraguruswamy, Investor sentiment and stock market response to earnings news, Account. Rev. 87 (4) (2012) 1357–1384.

[57] X. Ding, Y. Zhang, T. Liu, J. Duan, Deep learning for event-driven stock prediction, in: Proceedings of the 24th International Conference on Artificial Intelligence, 2015, pp. 2327–2333.

[58] Y. Liu, Q. Zeng, H. Yang, A. Carrio, Stock price movement prediction from financial news with deep learning and knowledge graph embedding, in: Pacific Rim Knowledge Acquisition Workshop, Springer, 2018, pp. 102–113.

[59] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, A. Sakurai, Combining technical analysis with sentiment analysis for stock price prediction, in: IEEE DASC, 2011, pp. 800–807.

[60] R. Frankel, M. Johnson, D.J. Skinner, An empirical examination of conference calls as a voluntary disclosure medium, J. Account. Res. 37 (1) (1999) 133–150.

[61] R.M. Bowen, A.K. Davis, D.A. Matsumoto, Do conference calls affect analysts' forecasts? Account. Rev. 77 (2) (2002) 285–316.

[62] L. Cohen, D. Lou, C.J. Malloy, Casting conference calls, Manage. Sci. 66 (11) (2020) 5015–5039.

[63] K. Keith, A. Stent, Modeling financial analysts' decision making via the pragmatics and semantics of earnings calls, in: Proceedings of ACL, 2019, pp. 493–503.

[64] Y. Chen, Z. Wei, X. Huang, Incorporating corporation relationship via graph convolutional neural networks for stock price prediction, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 1655–1658.

[65] R. Sawhney, S. Agarwal, A. Wadhwa, R. Shah, Deep attentive learning for stock movement prediction from social media text and company correlations, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2020, pp. 8415–8426.

[66] J. Yoo, Y. Soun, Y.-c. Park, U. Kang, Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 2037–2045.

[67] Y. Soun, J. Yoo, M. Cho, J. Jeon, U. Kang, Accurate stock movement prediction with self-supervised learning from sparse noisy tweets, in: IEEE International Conference on Big Data, 2022, pp. 1691–1700.

[68] Z. Zhou, L. Ma, H. Liu, Trade the event: Corporate events detection for news-based event-driven trading, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online, 2021, pp. 2114–2124.

[69] M. Baker, J. Wurgler, Investor sentiment and the cross-section of stock returns, J. Finance 61 (4) (2006) 1645–1680.

[70] K. Du, R. Mao, F. Xing, E. Cambria, A dynamic dual-graph neural network for stock price movement prediction, in: 2024 International Joint Conference on Neural Networks, IJCNN, Yokohama, Japan, 2024.

[71] T.H. Nguyen, K. Shirai, Topic modeling based sentiment analysis on social media for stock market prediction, in: Proceedings of ACL-IJCNLP, 2015, pp. 1354–1364.

[72] Y. Xu, S.B. Cohen, Stock movement prediction from tweets and historical prices, in: Proceedings of ACL, 2018, pp. 1970–1979.

[73] H. Wu, W. Zhang, W. Shen, J. Wang, Hybrid deep sequential modeling for social text-driven stock prediction, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM, 2018, pp. 1627–1630.

[74] X. Guo, J. Li, A novel Twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency, in: 2019 Sixth International Conference on Social Networks Analysis, Management and Security, SNAMS, IEEE, 2019, pp. 472–477.

[75] Z. Jin, Y. Yang, Y. Liu, Stock closing price prediction based on sentiment analysis and LSTM, Neural Comput. Appl. 32 (2020).

[76] S. Wu, Y. Liu, Z. Zou, T.-H. Weng, S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis, Connect. Sci. 34 (1) (2022) 44–62.

[77] F. Jin, N. Self, P. Saraf, P. Butler, W. Wang, N. Ramakrishnan, Forex-foreteller: Currency trend modeling using news articles, in: Proceedings of ACM SIGKDD, 2013, pp. 1470–1473.

[78] S. Seifollahi, M. Shajari, Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to FOREX market prediction, J. Intell. Inf. Syst. 52 (1) (2019) 57–83.

[79] H.N. Semiromi, S. Lessmann, W. Peters, News will tell: Forecasting foreign exchange rates based on news story events in the economy calendar, North Am. J. Econom. Finance 52 (2020) 101181.

[80] F. Xing, D.H. Hoang, D.-V. Vo, High-frequency news sentiment and its application to forex market prediction, in: Proceedings of HICSS, 2020, pp. 1583–1592.

[81] H. Laakkonen, The Impact of Macroeconomic News on Exchange Rate Volatility, Bank of Finland Discussion Paper 24, 2004.

[82] M.D. Evans, R.K. Lyons, Do currency markets absorb news quickly? J. Int. Money Finance 24 (2) (2005) 197–217.

[83] E. Fedorova, S. Druchok, P. Drogovoz, Impact of news sentiment and topics on IPO underpricing: US evidence, Int. J. Account. Inf. Manage. 30 (1) (2022) 73–94.

[84] S.J. Benjamin, P.K. Biswas, M.S. M Marathamuthu, M. Arunachalam, Social media sentiments and firm value, Appl. Econ. 54 (26) (2022) 2983–2997.

[85] F. Xing, E. Cambria, R. Welsch, Growing semantic vines for robust asset allocation, Knowl.-Based Syst. 165 (2019) 297–305, http://dx.doi.org/10.1016/j.knosys.2018.11.035.

[86] L. Malandri, F.Z. Xing, C. Orsenigo, C. Vercellis, E. Cambria, Public mood–driven asset allocation: The importance of financial sentiment in portfolio management, Cogn. Comput. 10 (6) (2018) 1167–1176.

[87] S. Koyano, K. Ikeda, Online portfolio selection based on the posts of winners and losers in stock microblogs, in: 2017 IEEE Symposium Series on Computational Intelligence, SSCI, IEEE, 2017, pp. 1–4.

[88] W. Tu, D.W. Cheung, N. Mamoulis, M. Yang, Z. Lu, Investment recommendation using investor opinions in social media, in: Proceedings of ACM SIGIR, 2016, pp. 881–884.

[89] S.Y. Yang, Y. Yu, S. Almahdi, An investor sentiment reward-based trading system using Gaussian inverse reinforcement learning algorithm, Expert Syst. Appl. 114 (2018) 388–401.

[90] R. Sawhney, A. Wadhwa, S. Agarwal, R. Shah, Quantitative day trading from natural language using reinforcement learning, in: Proceedings of NAACL-HLT, 2021, pp. 4018–4030.

[91] Y. Ye, Y. Pei, B. Wang, P.-Y. Chen, Y. Zhu, J. Xiao, B. Li, Reinforcement-learning based portfolio management with augmented asset movement prediction states, in: Proceedings of AAAI, Vol. 34, 2020, pp. 1112–1119.

[92] X. Du, K. Tanaka-Ishii, Stock embeddings acquired from news articles and price history, and an application to portfolio optimization, in: Proceedings of ACL, 2020, pp. 3353–3363.

[93] P. Koratamaddi, K. Wadhwani, M. Gupta, S.G. Sanjeevi, Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation, Eng. Sci. Technol. Int. J. (2021).

[94] M. Chen, Z. Zhang, J. Shen, Z. Deng, J. He, S. Huang, A quantitative investment model based on random forest and sentiment analysis, J. Phys. Conf. Ser. 1575 (1) (2020) 012083, http://dx.doi.org/10.1088/1742-6596/1575/1/012083.

[95] R. Sawhney, A. Wadhwa, S. Agarwal, R.R. Shah, FAST: Financial news and tweet based time aware network for stock trading, in: Proceedings of EACL, 2021, pp. 2164–2175.

[96] M.-C. Hung, P.-H. Hsia, X.-J. Kuang, S.-K. Lin, Intelligent portfolio construction via news sentiment analysis, Int. Rev. Econ. Finance 89 (2024) 605–617.

[97] Y. Ma, R. Mao, Q. Lin, P. Wu, E. Cambria, Quantitative stock portfolio optimization by multi-task learning risk and return, Inf. Fusion 104 (2024) 102165.

[98] Y. Sun, M. Fang, X. Wang, A novel stock recommendation system using guba sentiment analysis, Pers. Ubiquitous Comput. 22 (2018) 575–587.

[99] P. Ghosh, A. Neufeld, J.K. Sahoo, Forecasting directional movements of stock prices for intraday trading using LSTM and random forests, Finance Res. Lett. (ISSN: 1544-6123) 46 (2022) 102280.

[100] M. El-Haj, P. Rayson, A. Moore, The first financial narrative processing workshop (FNP 2018), in: Proceedings of the LREC 2018 Workshop, 2018, pp. 1–65.

[101] A.T. Oyewole, O.B. Adeoye, W.A. Addy, C.C. Okoye, O.C. Ofodile, C.E. Ugochukwu, Automating financial reporting with natural language processing: A review and case analysis, World J. Adv. Res. Rev. 21 (3) (2024) 575–589.

[102] A. Kim, M. Muhn, V.V. Nikolaev, Financial Statement Analysis with Large Language Models, Chicago Booth Research Paper Forthcoming, Fama-Miller Working Paper, 2024.

[103] V. Gupta, G.S. Lehal, A survey of text summarization extractive techniques, J. Emerg. Technol. Web Intell. 2 (3) (2010) 258–268.

[104] N. Moratanch, S. Chitrakala, A survey on abstractive text summarization, in: 2016 International Conference on Circuit, Power and Computing Technologies, ICCPCT, IEEE, 2016, pp. 1–7.

[105] M. El-Haj, M. Litvak, N. Pittaras, G. Giannakopoulos, et al., The financial narrative summarisation shared task (FNS 2020), in: Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, 2020, pp. 1–12.

[106] N. Zmandar, M. El-Haj, P. Rayson, M. Litvak, G. Giannakopoulos, N. Pittaras, et al., The financial narrative summarisation shared task fns 2021, in: Proceedings of the 3rd Financial Narrative Processing Workshop, 2021, pp. 120–125.

[107] E. Zavitsanos, A. Kosmopoulos, G. Giannakopoulos, M. Litvak, B. Carbajo-Coronado, A. Moreno-Sandoval, M. El-Haj, The financial narrative summarisation shared task (FNS 2023), in: 2023 IEEE International Conference on Big Data, BigData, IEEE, 2023, pp. 2890–2896.

[108] M. El-Haj, A. Ogden, Financial narrative summarisation using a hybrid tf-idf and clustering summariser: Ao-lancs system at fns 2022, in: Proceedings of the 4th Financial Narrative Processing Workshop@ LREC2022, 2022, pp. 79–82.

[109] T. Gokhan, P. Smith, M. Lee, Extractive financial narrative summarisation using sentencebert based clustering, in: Proceedings of the 3rd Financial Narrative Processing Workshop, 2021, pp. 94–98.

[110] M. La Quatra, L. Cagliero, End-to-end training for financial report summarization, in: Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, 2020, pp. 118–123.

[111] S. Abdaljalil, H. Bouamor, An exploration of automatic text summarization of financial reports, in: Proceedings of the Third Workshop on Financial Technology and Natural Language Processing, 2021, pp. 1–7.

[112] M. Orzhenovskii, T5-LONG-EXTRACT at FNS-2021 shared task, in: Proceedings of the 3rd Financial Narrative Processing Workshop, 2021, pp. 67–69.

[113] M. Litvak, N. Vanetik, Z. Puchinsky, Hierarchical summarization of financial reports with RUNNER, in: Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, 2020, pp. 213–225.

[114] N. Shukla, A. Vaid, R. Katikeri, S. Keeriyadath, M. Raja, Dimsum: Distributed and multilingual summarization of financial narratives, in: Proceedings of the 4th Financial Narrative Processing Workshop@ LREC2022, 2022, pp. 65–72.

[115] N.K. Shukla, R. Katikeri, M. Raja, G. Sivam, S. Yadav, A. Vaid, S. Prabhakararao, Generative AI approach to distributed summarization of financial narratives, in: 2023 IEEE International Conference on Big Data, BigData, IEEE, 2023, pp. 2872–2876.

[116] H. Li, Q. Peng, X. Mou, Y. Wang, Z. Zeng, M.F. Bashir, Abstractive financial news summarization via transformer-bilstm encoder and graph attention-based decoder, IEEE/ACM Trans. Audio Speech Lang. Process. (2023).

[117] S. Liu, J. Cao, R. Yang, Z. Wen, Long text and multi-table summarization: Dataset and method, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2022, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 1995–2010.

[118] N. Zmandar, A. Singh, M. El-Haj, P. Rayson, Joint abstractive and extractive method for long financial document summarization, in: Proceedings of the 3rd Financial Narrative Processing Workshop, 2021, pp. 99–105.

[119] A. Singh, Point-5: Pointer network and t-5 based financial narrativesummarisation, 2020, arXiv preprint arXiv:2010.04191.

[120] R. Mukherjee, A. Bohra, A. Banerjee, S. Sharma, M. Hegde, A. Shaikh, S. Shrivastava, K. Dasgupta, N. Ganguly, S. Ghosh, P. Goyal, ECTSum: A new benchmark dataset for bullet point summarization of long earnings call transcripts, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 10893–10906.

[121] A. See, P.J. Liu, C.D. Manning, Get to the point: Summarization with pointer-generator networks, 2017, arXiv preprint arXiv:1704.04368.

[122] R. Paulus, C. Xiong, R. Socher, A deep reinforced model for abstractive summarization, 2017, arXiv preprint arXiv:1705.04304.

[123] A.P. Widyassari, S. Rustad, G.F. Shidik, E. Noersasongko, A. Syukur, A. Affandy, et al., Review of automatic text summarization techniques & methods, J. King Saud Univ.-Comput. Inf. Sci. 34 (4) (2022) 1029–1046.

[124] V. Tretyak, D. Stepanov, Combination of abstractive and extractive approaches for summarization of long scientific texts, 2020, arXiv preprint arXiv:2006.05354.

[125] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P.J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, J. Mach. Learn. Res. 21 (140) (2020) 1–67.

[126] Z. Xiao, Z. Mai, Z. Xu, Y. Cui, J. Li, Corporate event predictions using large language models, in: 2023 10th International Conference on Soft Computing & Machine Intelligence, ISCMI, IEEE, 2023, pp. 193–197.

[127] G. Jacobs, V. Hoste, SENTiVENT: enabling supervised information extraction of company-specific events in economic and financial news, Lang. Resour. Eval. 56 (1) (2022) 225–257.

[128] G. Jacobs, E. Lefever, V. Hoste, Economic event detection in company-specific news text, in: 1st Workshop on Economics and Natural Language Processing (ECONLP) At Meeting of the Association-for-Computational-Linguistics, ACL, Association for Computational Linguistics (ACL), 2018, pp. 1–10.

[129] L. Ein-Dor, A. Gera, O. Toledo-Ronen, A. Halfon, B. Sznajder, L. Dankin, Y. Bilu, Y. Katz, N. Slonim, Financial event extraction using wikipedia-based weak supervision, in: U. Hahn, V. Hoste, Z. Zhang (Eds.), Proceedings of the Second Workshop on Economics and Natural Language Processing, Association for Computational Linguistics, Hong Kong, 2019, pp. 10–15.

[130] S. Carta, S. Consoli, L. Piras, A.S. Podda, D.R. Recupero, Event detection in finance using hierarchical clustering algorithms on news and tweets, PeerJ Comput. Sci. 7 (2021) e438.

[131] K. Guo, T. Jiang, H. Zhang, Knowledge graph enhanced event extraction in financial documents, in: 2020 IEEE International Conference on Big Data, Big Data, IEEE, 2020, pp. 1322–1329.

[132] D. Cheng, F. Yang, X. Wang, Y. Zhang, L. Zhang, Knowledge graph-based event embedding framework for financial quantitative investments, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 2221–2230.

[133] A. Balashankar, S. Chakraborty, S. Fraiberger, L. Subramanian, Identifying predictive causal factors from news streams, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, 2019, pp. 2338–2348.

[134] K. Izumi, H. Sakaji, Economic causal-chain search using text mining technology, in: Artificial Intelligence. IJCAI 2019 International Workshops: Macao, China, August 10–12, 2019, Revised Selected Best Papers 28, Springer, 2020, pp. 23–35.

[135] T. Takayanagi, M. Suzuki, R. Kobayashi, H. Sakaji, K. Izumi, Is ChatGPT the future of causal text mining? A comprehensive evaluation and analysis, 2024, arXiv preprint arXiv:2402.14484.

[136] P. Rajpoot, A. Parikh, GPT-finRE: In-context learning for financial relation extraction using large language models, in: C.-C. Chen, H.-H. Huang, H. Takamura, H.-H. Chen, H. Sakaji, K. Izumi (Eds.), Proceedings of the Sixth Workshop on Financial Technology and Natural Language Processing, Association for Computational Linguistics, Bali, Indonesia, 2023, pp. 42–45.

[137] L. Hillebrand, T. Deußer, T. Dilmaghani, B. Kliem, R. Loitz, C. Bauckhage, R. Sifa, Kpi-bert: A joint named entity recognition and relation extraction model for financial reports, in: 2022 26th International Conference on Pattern Recognition, ICPR, IEEE, 2022, pp. 606–612.

[138] Y. Zhang, H. Zhang, FinBERT–MRC: financial named entity recognition using BERT under the machine reading comprehension paradigm, Neural Process. Lett. 55 (6) (2023) 7393–7413.

[139] B. Oral, E. Emekligil, S. Arslan, G. Eryiğit, Information extraction from text intensive and visually rich banking documents, Inf. Process. Manage. 57 (6) (2020) 102361.

[140] Y. Yang, J. Wu, Y. Yang, S. Lian, F. Guo, Z. Wang, A survey of information extraction based on deep learning, Appl. Sci. 12 (19) (2022) 9691.

[141] Q. Li, Q. Zhang, A unified model for financial event classification, detection and summarization, in: Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, 2021, pp. 4668–4674.

[142] D. Mariko, H. Abi Akl, K. Trottier, M. El-Haj, The financial causality extraction shared task (FinCausal 2022), in: Proceedings of the 4th Financial Narrative Processing Workshop@ LREC2022, 2022, pp. 105–107.

[143] A. Moreno-Sandoval, J. Porta-Zamorano, B. Carbajo-Coronado, D. Samy, D. Mariko, M. El-Haj, The financial document causality detection shared task (FinCausal 2023), in: 2023 IEEE International Conference on Big Data, BigData, IEEE, 2023, pp. 2855–2860.

[144] S. Kaur, C. Smiley, A. Gupta, J. Sain, D. Wang, S. Siddagangappa, T. Aguda, S. Shah, REFinD: Relation extraction financial dataset, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023, pp. 3054–3063.

[145] S. Sharma, T. Nayak, A. Bose, A.K. Meena, K. Dasgupta, N. Ganguly, P. Goyal, FinRED: A dataset for relation extraction in financial domain, in: Companion Proceedings of the Web Conference 2022, 2022, pp. 595–597.

[146] G. Becquin, GBe at FinCausal 2020, task 2: Span-based causality extraction for financial documents, in: Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation, 2020, pp. 40–44.

[147] P. Yang, X. Cong, Z. Sun, X. Liu, Enhanced language representation with label knowledge for span extraction, in: M.-F. Moens, X. Huang, L. Specia, S.W.-t. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 4623–4635.

[148] R. Shah, K. Chawla, D. Eidnani, A. Shah, W. Du, S. Chava, N. Raman, C. Smiley, J. Chen, D. Yang, When FLUE meets FLANG: Benchmarks and large pretrained language model for financial domain, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 2322–2335.

[149] J.C.S. Alvarado, K. Verspoor, T. Baldwin, Domain adaption of named entity recognition to support credit risk assessment, in: Proceedings of the Australasian Language Technology Association Workshop 2015, 2015, pp. 84–90.

[150] J. Kang, A. Ait Azzi, S. Bellato, B. Carbajo Coronado, M. El-Haj, I. El Maarouf, M. Gan, A. Gisbert, A. Moreno Sandoval, The financial document structure extraction shared task (finTOC 2022), in: M. El-Haj, P. Rayson, N. Zmandar (Eds.), Proceedings of the 4th Financial Narrative Processing Workshop @LREC2022, Marseille, France, 2022, pp. 83–88.

[151] W. Au, A. Ait-Azzi, J. Kang, FinSBD-2021: The 3rd shared task on structure boundary detection in unstructured text in the financial domain, in: Companion Proceedings of the Web Conference 2021, 2021.

[152] K.S.D. Ishwari, A.K.R.R. Aneeze, S. Sudheesan, H.J.D.A. Karunaratne, A. Nugaliyadde, Y. Mallawarrachchi, Advances in natural language question answering: A review, 2019, arXiv:1904.05276.

[153] E. Cambria, Understanding Natural Language Understanding, Springer, ISBN: 978-3-031-73973-6, 2024.

[154] Z. Chen, W. Chen, C. Smiley, S. Shah, I. Borova, D. Langdon, R. Moussa, M. Beane, T.-H. Huang, B. Routledge, W.Y. Wang, FinQA: A dataset of numerical reasoning over financial data, in: M.-F. Moens, X. Huang, L. Specia, S.W.-t. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 3697–3711.

[155] Z. Wang, Modern question answering datasets and benchmarks: A survey, 2022, arXiv:2206.15030.

[156] P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang, SQuAD: 100,000+ questions for machine comprehension of text, in: J. Su, K. Duh, X. Carreras (Eds.), Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 2383–2392.

[157] M. Joshi, E. Choi, D. Weld, L. Zettlemoyer, TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension, in: R. Barzilay, M.-Y. Kan (Eds.), Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 1601–1611.

[158] Z. Chen, S. Li, C. Smiley, Z. Ma, S. Shah, W.Y. Wang, ConvFinQA: Exploring the chain of numerical reasoning in conversational finance question answering, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 6279–6292.

[159] C. Li, W. Ye, Y. Zhao, Finmath: Injecting a tree-structured solver for question answering over financial reports, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022, pp. 6147–6152.

[160] Q. Zhang, S. Chen, D. Xu, Q. Cao, X. Chen, T. Cohn, M. Fang, A survey for efficient open domain question answering, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 14447–14465.

[161] F. Zhu, W. Lei, Y. Huang, C. Wang, S. Zhang, J. Lv, F. Feng, T.-S. Chua, TAT-QA: A question answering benchmark on a hybrid of tabular and textual content in finance, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 3277–3287.

[162] Y. Deng, W. Lei, W. Zhang, W. Lam, T.-S. Chua, PACIFIC: Towards proactive conversational question answering over tabular and textual data in finance, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 6970–6984.

[163] M. Bulla, L. Hillebrand, M. Lübbering, R. Sifa, Knowledge graph based question answering system for financial securities, in: KI 2021: Advances in Artificial Intelligence: 44th German Conference on AI, Virtual Event, September 27–October 1, 2021, Proceedings 44, Springer, 2021, pp. 44–50.

[164] R. Nararatwong, N. Kertkeidkachorn, R. Ichise, KIQA: Knowledge-infused question answering model for financial table-text data, in: E. Agirre, M. Apidianaki, I. Vulić (Eds.), Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, Association for Computational Linguistics, Dublin, Ireland and Online, 2022, pp. 53–61.

[165] S. Panwar, A. Bansal, F. Zareen, Comparative analysis of large language models for question answering from financial documents, in: International Conference on Communication and Intelligent Systems, Springer, 2023, pp. 297–308.

[166] P. Srivastava, M. Malik, T. Ganu, Assessing LLMs' mathematical reasoning in financial document question answering, 2024, arXiv preprint arXiv:2402.11194.

[167] T. Okuda, S. Shoda, AI-based chatbot service for financial industry, Fujitsu Sci. Tech. J. 54 (2) (2018) 4–8.

[168] V. Kobets, K. Kozlovskyi, Application of chat bots for personalized financial advice, Herald Adv. Inf. Technol. 3 (5) (2022) 229–242.

[169] Q. Xie, D. Tan, T. Zhu, Q. Zhang, S. Xiao, J. Wang, B. Li, L. Sun, P. Yi, Chatbot application on cryptocurrency, in: 2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics, CIFEr, IEEE, 2019, pp. 1–8.

[170] J.-W. Chang, N. Yen, J.C. Hung, Design of a NLP-empowered finance fraud awareness model: the anti-fraud chatbot for fraud detection and fraud classification as an instance, J. Ambient Intell. Humaniz. Comput. 13 (10) (2022) 4663–4679.

[171] S.E. Finch, J.D. Choi, Towards unified dialogue system evaluation: A comprehensive analysis of current evaluation protocols, in: O. Pietquin, S. Muresan, V. Chen, C. Kennington, D. Vandyke, N. Dethlefs, K. Inoue, E. Ekstedt, S. Ultes (Eds.), Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics, 1st virtual meeting, 2020, pp. 236–245.

[172] F. Zhu, W. Lei, C. Wang, J. Zheng, S. Poria, T.-S. Chua, Retrieving and reading: A comprehensive survey on open-domain question answering, 2021, arXiv:2101.00774.

[173] A. Balaguer, V. Benara, R.L.d. Cunha, R. de M. Estevão Filho, T. Hendry, D. Holstein, J. Marsman, N. Mecklenburg, S. Malvar, L.O. Nunes, R. Padilha, M. Sharp, B. Silva, S. Sharma, V. Aski, R. Chandra, RAG vs fine-tuning: Pipelines, tradeoffs, and a case study on agriculture, 2024, arXiv:2401.08406.

[174] C. Packer, S. Wooders, K. Lin, V. Fang, S.G. Patil, I. Stoica, J.E. Gonzalez, MemGPT: Towards LLMs as operating systems, 2024, arXiv:2310.08560.

[175] T. Munkhdalai, M. Faruqui, S. Gopal, Leave no context behind: Efficient infinite context transformers with infini-attention, 2024, arXiv:2404.07143.

[176] L. Xu, H. Xie, S.-Z.J. Qin, X. Tao, F.L. Wang, Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment, 2023, arXiv:2312.12148.

[177] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, LoRA: Low-rank adaptation of large language models, 2021, arXiv:2106.09685.

[178] T. Kaufmann, P. Weng, V. Bengs, E. Hüllermeier, A survey of reinforcement learning from human feedback, 2024, arXiv:2312.14925.

[179] N. Mecklenburg, Y. Lin, X. Li, D. Holstein, L. Nunes, S. Malvar, B. Silva, R. Chandra, V. Aski, P.K.R. Yannam, T. Aktas, T. Hendry, Injecting new knowledge into large language models via supervised fine-tuning, 2024, arXiv:2404.00213.

[180] Z.Z. Yu, L.J. Jaw, Z. Hui, B.K.H. Low, Fine-tuning language models with generative adversarial reward modelling, 2024, arXiv:2305.06176.

[181] S. Khan, M.R. Rabbani, Chatbot as islamic finance expert (CaIFE) when finance meets artificial intelligence, in: Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control, 2020, pp. 1–5.

[182] M.S.I. Bhuiyan, A. Razzak, M.S. Ferdous, M.J.M. Chowdhury, M.A. Hoque, S. Tarkoma, BONIK: A blockchain empowered chatbot for financial transactions, in: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom, IEEE, 2020, pp. 1079–1088.

[183] S.F. Suhel, V.K. Shukla, S. Vyas, V.P. Mishra, Conversation to automation in banking through chatbot using artificial machine intelligence language, in: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), IEEE, 2020, pp. 611–618.

[184] H.A. Pandya, B.S. Bhatt, Question answering survey: Directions, challenges, datasets, evaluation matrices, 2021, arXiv:2112.03572.

[185] S. Kogan, D. Levin, B.R. Routledge, J.S. Sagi, N.A. Smith, Predicting risk from financial reports with regression, in: Proceedings of NAACL-HLT, 2009, pp. 272–280.

[186] C.-J. Wang, M.-F. Tsai, T. Liu, C.-T. Chang, Financial sentiment analysis for risk prediction, in: Proceedings of IJCNLP, 2013, pp. 802–808.

[187] M.-F. Tsai, C.-J. Wang, Financial keyword expansion via continuous word vector representations, in: Proceedings of EMNLP, 2014, pp. 1453–1458.

[188] C. Nopp, A. Hanbury, Detecting risks in the banking system by sentiment analysis, in: Proceedings of EMNLP, 2015, pp. 591–600.

[189] W.Y. Wang, Z. Hua, A semiparametric Gaussian copula regression model for predicting financial risks from earnings calls, in: Proceedings of ACL, 2014, pp. 1155–1165.

[190] N. Rekabsaz, M. Lupu, A. Baklanov, A. Dür, L. Anderson, Volatility prediction using financial disclosures sentiments with word embedding-based IR models, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017.

[191] F.Z. Xing, E. Cambria, Y. Zhang, Sentiment-aware volatility forecasting, Knowl.-Based Syst. 176 (2019) 68–76.

[192] J. Deveikyte, H. Geman, C. Piccari, A. Provetti, A sentiment analysis approach to the prediction of market volatility, 2020, arXiv:2012.05906.

[193] K.G. Al-Hashedi, P. Magalingam, Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019, Comp. Sci. Rev. 40 (2021) 100402.

[194] J.F. Rodríguez, M. Papale, M. Carminati, S. Zanero, et al., A natural language processing approach for financial fraud detection, in: CEUR workshop proceedings, Vol. 3260, CEUR-WS. org, 2022, pp. 135–149.

[195] P. Boulieris, J. Pavlopoulos, A. Xenos, V. Vassalos, Fraud detection with natural language processing, Mach. Learn. (2023) 1–22.

[196] X. Yang, C. Zhang, Y. Sun, K. Pang, L. Jing, S. Wa, C. Lv, FinChain-BERT: A high-accuracy automatic fraud detection model based on NLP methods for financial scenarios, Information 14 (9) (2023) 499.

[197] Y.-J. Chen, C.-H. Wu, Y.-M. Chen, H.-Y. Li, H.-K. Chen, Enhancement of fraud detection for narratives in annual reports, Int. J. Account. Inf. Syst. 26 (2017) 32–45.

[198] W. Dong, S. Liao, Z. Zhang, Leveraging financial social media data for corporate fraud detection, J. Manage. Inf. Syst. 35 (2) (2018) 461–487.

[199] P. Craja, A. Kim, S. Lessmann, Deep learning for detecting financial statement fraud, Decis. Support Syst. 139 (2020) 113421.

[200] M.A.K. Achakzai, J. Peng, Detecting financial statement fraud using dynamic ensemble machine learning, Int. Rev. Financ. Anal. 89 (2023) 102827.

[201] X. Mao, H. Sun, X. Zhu, J. Li, Financial fraud detection using the related-party transaction knowledge graph, Procedia Comput. Sci. 199 (2022) 733–740.

[202] P. Seemakurthi, S. Zhang, Y. Qi, Detection of fraudulent financial reports with machine learning techniques, in: 2015 Systems and Information Engineering Design Symposium, IEEE, 2015, pp. 358–361.

[203] Z. Ahmadi, P. Martens, C. Koch, T. Gottron, S. Kramer, Towards bankruptcy prediction: Deep sentiment mining to detect financial distress from business management reports, in: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA, IEEE, 2018, pp. 293–302.

[204] A. Markov, Z. Seleznyova, V. Lapshin, Credit scoring methods: Latest trends and points to consider, J. Finance Data Sci. 8 (2022) 180–201.

[205] C. Wang, D. Han, Q. Liu, S. Luo, A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM, IEEE Access 7 (2018) 2161–2168.

[206] J. Kriebel, L. Stitz, Credit default prediction from user-generated text in peer-to-peer lending using deep learning, European J. Oper. Res. 302 (1) (2022) 309–323.

[207] O. Netzer, A. Lemaire, M. Herzenstein, When words sweat: Identifying signals for loan default in the text of loan applications, J. Mar. Res. 56 (6) (2019) 960–980.

[208] M. Stevenson, C. Mues, C. Bravo, The value of text for small business default prediction: A deep learning approach, European J. Oper. Res. 295 (2) (2021) 758–771.

[209] R. Matin, C. Hansen, C. Hansen, P. Mø lgaard, Predicting distresses using deep learning of text segments in annual reports, Expert Syst. Appl. 132 (2019) 199–208.

[210] F. Mai, S. Tian, C. Lee, L. Ma, Deep learning models for bankruptcy prediction using textual disclosures, European J. Oper. Res. 274 (2) (2019) 743–758.

[211] T. Nieß ner, D.H. Gross, M. Schumann, Evidential strategies in financial statement analysis: a corpus linguistic text mining approach to bankruptcy prediction, J. Risk Financ. Manage. 15 (10) (2022) 459.

[212] K. Yang, H. Yuan, R.Y. Lau, PsyCredit: an interpretable deep learning-based credit assessment approach facilitated by psychometric natural language processing, Expert Syst. Appl. 198 (2022) 116847.

[213] Q. Gao, M. Lin, R. Sias, Words matter: The role of readability, tone, and deception cues in online credit markets, J. Financ. Quant. Anal. 58 (1) (2023) 1–28.

[214] A. El-Qadi, M. Trocan, P. Conde-Cespedes, T. Frossard, N. Díaz-Rodríguez, Credit risk scoring using a data fusion approach, in: International Conference on Computational Collective Intelligence, Springer, 2023, pp. 769–781.

[215] B.-H. Nguyen, V.-N. Huynh, Textual analysis and corporate bankruptcy: A financial dictionary-based sentiment approach, J. Oper. Res. Soc. 73 (1) (2022) 102–121.

[216] M. Hassan, L.A.-R. Aziz, Y. Andriansyah, The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance, Rev. Contemp. Bus. Anal. 6 (1) (2023) 110–132.

[217] T. Butler, L. O'Brien, Artificial intelligence for regulatory compliance: Are we there yet? J. Financ. Compliance 3 (1) (2019) 44–59.

[218] V. Jain, A. Balakrishnan, D. Beeram, M. Najana, P. Chintale, Leveraging artificial intelligence for enhancing regulatory compliance in the financial sector, Int. J. Comput. Trends Technol. 72 (5) (2024) 124–140.

[219] R. Al-Shabandar, G. Lightbody, F. Browne, J. Liu, H. Wang, H. Zheng, The application of artificial intelligence in financial compliance management, in: Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, 2019, pp. 1–6.

[220] S. Abualhaija, M. Ceci, N. Sannier, D. Bianculli, D.A. ZETZSCHE, M. Bodellini, Toward automated change impact analysis of financial regulations, in: 2024 Workshop on Software Engineering Challenges in Financial Firms, FinanSE'24, ACM, 2024, pp. 31–32.

[221] E. Cambria, R. Mao, M. Chen, Z. Wang, S.-B. Ho, Seven pillars for the future of artificial intelligence, IEEE Intell. Syst. 38 (6) (2023) 62–69.

[222] P. Weber, K.V. Carl, O. Hinz, Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature, Manage. Rev. Q. (2023) 1–41.

[223] W.J. Yeo, W. van der Heever, R. Mao, E. Cambria, R. Satapathy, G. Mengaldo, A comprehensive review on financial explainable AI, 2023, arXiv preprint arXiv:2309.11960.

[224] D. Kumar, G.W. Taylor, A. Wong, Opening the black box of financial ai with clear-trade: A class-enhanced attentive response approach for explaining and visualizing deep learning-driven stock market prediction, 2017, arXiv preprint arXiv:1709.01574.

[225] S. Deng, N. Zhang, W. Zhang, J. Chen, J.Z. Pan, H. Chen, Knowledge-driven stock trend prediction and explanation via temporal convolutional network, in: Companion Proceedings of the 2019 World Wide Web Conference, 2019, pp. 678–685.

[226] S. Carta, A.S. Podda, D. Reforgiato Recupero, M.M. Stanciu, Explainable AI for financial forecasting, in: International Conference on Machine Learning, Optimization, and Data Science, Springer, 2021, pp. 51–69.

[227] K. Ong, W. van der Heever, R. Satapathy, G. Mengaldo, E. Cambria, FinX-ABSA: Explainable finance through aspect-based sentiment analysis, in: 2023 IEEE International Conference on Data Mining Workshops, ICDMW, 2023, pp. 773–782.

[228] B. Liang, H. Su, L. Gui, E. Cambria, R. Xu, Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks, Knowl.-Based Syst. 235 (2022) 107643.

[229] H. Bandi, S. Joshi, S. Bhagat, D. Ambawade, Integrated technical and sentiment analysis tool for market index movement prediction, comprehensible using xai, in: 2021 ICCICT, IEEE, 2021, pp. 1–8.

[230] S. Gite, H. Khatavkar, K. Kotecha, S. Srivastava, P. Maheshwari, N. Pandey, Explainable stock prices prediction from financial news articles using sentiment analysis, PeerJ Comput. Sci. 7 (2021) e340.

[231] J. Yuan, Z. Zhang, Connecting the dots: forecasting and explaining short-term market volatility, in: Proceedings of ACM ICAIF, 2020, pp. 1–8.

[232] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, Language models are unsupervised multitask learners, OpenAI 1 (8) (2019) 9.

[233] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, P. Sen, A survey of the state of explainable AI for natural language processing, in: K.-F. Wong, K. Knight, H. Wu (Eds.), Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Suzhou, China, 2020, pp. 447–459.

[234] T. Lim, Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways, Artif. Intell. Rev. 57 (4) (2024) 1–64.

[235] P. Capelli, F. Ielasi, A. Russo, Forecasting volatility by integrating financial risk with environmental, social, and governance risk, Corp. Soc. Responsib. Environ. Manage. 28 (5) (2021) 1483–1495.

[236] C.-C. Chen, Y.-M. Tseng, J. Kang, A. Lhuissier, M.-Y. Day, T.-T. Tu, H.-H. Chen, Multi-lingual ESG issue identification, in: Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI for Financial Forecasting, 2023, pp. 111–115.

[237] P. Glenn, A. Gon, N. Kohli, S. Zha, P.P. Dakle, P. Raghavan, Jetsons at the finnlp-2023: Using synthetic data and transfer learning for multilingual esg issue classification, in: Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI for Financial Forecasting, 2023, pp. 133–139.

[238] W. Wang, W. Wei, Q. Song, Y. Wang, Leveraging contrastive learning with bert for esg issue identification, in: Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI for Financial Forecasting, 2023, pp. 116–120.

[239] S. Pasch, D. Ehnes, Nlp for responsible finance: Fine-tuning transformer-based models for esg, in: 2022 IEEE International Conference on Big Data, Big Data, IEEE, 2022, pp. 3532–3536.

[240] N. Raman, G. Bang, A. Nourbakhsh, Mapping ESG trends by distant supervision of neural language models, Mach. Learn. Knowl. Extract. 2 (4) (2020) 453–468.

[241] B. Koloski, S. Montariol, M. Purver, S. Pollak, Knowledge informed sustainability detection from short financial texts, in: Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing, FinNLP, 2022, pp. 228–234.

[242] N. Smeuninx, B. De Clerck, W. Aerts, Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and NLP, Int. J. Bus. Commun. 57 (1) (2020) 52–85.

[243] D. Kouloukoui, N. de Marcellis-Warin, S.M. da Silva Gomes, T. Warin, Mapping global conversations on twitter about environmental, social, and governance topics through natural language processing, J. Clean. Prod. 414 (2023) 137369.

[244] T. Nugent, N. Stelea, J.L. Leidner, Detecting environmental, social and governance (ESG) topics using domain-specific language models and data augmentation, in: Flexible Query Answering Systems: 14th International Conference, FQAS 2021, Bratislava, Slovakia, September 19–24, 2021, Proceedings 14, Springer, 2021, pp. 157–169.

[245] S. Ghosh, S.K. Naskar, Ranking environment, social and governance related concepts and assessing sustainability aspect of financial texts, in: Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing (FinNLP), 2022, pp. 243–249.

[246] B. Sandwidi, S.P. Mukkolakal, Transformers-based approach for a sustainability term-based sentiment analysis (STBSA), in: Proceedings of the Second Workshop on NLP for Positive Impact, NLP4PI, 2022, pp. 157–170.

[247] A. Sokolov, K. Caverly, J. Mostovoy, T. Fahoum, L. Seco, Weak supervision and black-litterman for automated ESG portfolio construction, J. Financ. Data Sci. (2021).

[248] T. Guo, N. Jamet, V. Betrix, L.-A. Piquet, E. Hauptmann, ESG2risk: A deep learning framework from ESG news to stock volatility prediction, ERN: Stock Market Risk (Topic) (2020).

[249] M. Apel, A. Betzer, B. Scherer, Real-time transition risk, Finance Res. Lett. 53 (2023) 103600.

[250] C.-L. Jan, Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry, Sustainability 13 (17) (2021) 9879.

[251] S. Mehra, R. Louka, Y. Zhang, Esgbert: Language model to help with classification tasks related to companies environmental, social, and governance practices, 2022, arXiv preprint arXiv:2203.16788.

[252] E. Haase, R. Sassen, Uncovering lobbying strategies in sustainable finance disclosure regulations using machine learning, J. Environ. Manag. 356 (2024) 120562.

[253] H. Lee, S.H. Lee, H. Park, J.H. Kim, H.S. Jung, ESG2PreEM: Automated ESG grade assessment framework using pre-trained ensemble models, Heliyon (2024).

[254] T. Schimanski, A. Reding, N. Reding, J. Bingler, M. Kraus, M. Leippold, Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication, Finance Res. Lett. 61 (2024) 104979.

[255] A. Sokolov, J. Mostovoy, J. Ding, L. Seco, Building machine learning systems for automated ESG scoring, J. Impact ESG Invest. 1 (3) (2021) 39–50.

[256] A.H. Huang, H. Wang, Y. Yang, FinBERT: A large language model for extracting information from financial text, Contemp. Account. Res. 40 (2) (2023) 806–841.

[257] G. Fan, X. Wu, Going Green: The Governance Role of Environmental Regulations on Firm Innovation and Value, Singapore Management University School of Accountancy Research Paper 2023–163, 2022.

[258] K. Król, D. Zdonek, Digital assets in the eyes of generation z: Perceptions, outlooks, concerns, J. Risk Financ. Manage. 16 (1) (2022) 22.

[259] K.H. Al-Yahyaee, M.U. Rehman, W. Mensi, I.M.W. Al-Jarrah, Can uncertainty indices predict bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches, North Am. J. Econom. Finance 49 (2019) 47–56.

[260] D. Shen, A. Urquhart, P. Wang, Does twitter predict bitcoin? Econom. Lett. 174 (2019) 118–122.

[261] O. Kraaijeveld, J. De Smedt, The predictive power of public Twitter sentiment for forecasting cryptocurrency prices, J. Int. Financ. Mark. Inst. Money 65 (2020) 101188.

[262] S. Oikonomopoulos, K. Tzafilkou, D. Karapiperis, V. Verykios, Cryptocurrency price prediction using social media sentiment analysis, in: 2022 13th International Conference on Information, Intelligence, Systems & Applications, IISA, IEEE, 2022, pp. 1–8.

[263] M.A. Naeem, I. Mbarki, S.J.H. Shahzad, Predictive role of online investor sentiment for cryptocurrency market: Evidence from happiness and fears, Int. Rev. Econ. Finance 73 (2021) 496–514.

[264] K. Kim, S.-Y. Lee, S. Assar, The dynamics of cryptocurrency market behavior: sentiment analysis using Markov chains, Ind. Manage. Data Syst. ahead-of-print (2021).

[265] X. Huang, W. Zhang, X. Tang, M. Zhang, J. Surbiryala, V. Iosifidis, Z. Liu, J. Zhang, LSTM based sentiment analysis for cryptocurrency prediction, in: Database Systems for Advanced Applications: 26th International Conference, Springer, 2021, pp. 617–621.

[266] Z. Leitter, E. Cambria, Non-fungible tokens: What makes them valuable? in: 2023 IEEE International Conference on Data Mining Workshops, ICDMW, 2023, pp. 750–756.

[267] S.C. Meyns, F. Dalipi, What users tweet on NFTs: mining Twitter to understand NFT-related concerns using a topic modeling approach, IEEE Access 10 (2022) 117658–117680.

[268] X. Gao, W. Huang, H. Wang, Financial Twitter sentiment on bitcoin return and high-frequency volatility, Virtual Econom. 4 (1) (2021) 7–18.

[269] J. Saha, S. Patel, F. Xing, E. Cambria, Does social media sentiment predict bitcoin trading volume? in: Proceedings of the 43rd International Conference on Information Systems, ICIS, 2022, pp. 1–9.

[270] R. Mao, G. Chen, X. Zhang, F. Guerin, E. Cambria, GPTEval: A survey on assessments of ChatGPT and GPT-4, in: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC-COLING 2024, ELRA and ICCL, Torino, Italia, 2024, pp. 7844–7866.

[271] B. Chen, Z. Wu, R. Zhao, From fiction to fact: the growing role of generative AI in business and finance, J. Chin. Econ. Bus. Stud. 21 (4) (2023) 471–496.

[272] N. Rane, Role and challenges of ChatGPT and similar generative artificial intelligence in finance and accounting, 2023, Available at SSRN 4603206.

[273] G. Shabsigh, E.B. Boukherouaa, Generative artificial intelligence in finance, FinTech Not. 2023 (006) (2023).