A Survey about LLM Applications to Financial Texts

Peng Chen
Georgia Institute of Technology

Maria Konte
Georgia Institute of Technology

Abstract

Financial texts, encompassing a wide range of materials such as news articles, reports, and business dialogues, are often unstructured and present a complex challenge for financial analysts who dedicate considerable time to managing them. In recent years, large language models (LLMs) have emerged as powerful tools in natural language processing (NLP), demonstrating exceptional capabilities across various tasks. This survey examines the current research on the application of LLMs in the manipulation of financial texts.

We outline the essential steps for implementing LLMs in financial text processing, which include data collection, data preprocessing, and fine-tuning. Additionally, we explore the state-of-the-art models employed for tasks such as information extraction, sentiment analysis, and question answering. Furthermore, we describe potential security and bias issues that these models may encounter, while also highlighting future research opportunities in the realm of financial text manipulation using LLMs.

1 Introduction

LLM-based applications are proliferating globally, with finance and commerce being significant areas of application [35]. In the finance industry, many employees are reading news, reports, or other sources of information to analyze economic trends, predict stock prices, and write reports. After the emergence of LLMs, many people think that the manual process of financial analysis would be radically simplified and even replaced. LLMs would change every industry a lot, including but not limited to finance, healthcare, education, and so on. Therefore, it is important to survey what tasks LLMs would help with, how much LLMs would change the industries, and at what stage the change is.

This paper focuses on how LLMs would change financial text manipulation, which is a kind of task that lacks survey. Financial analysts read financial reports and news, then they extract related organizations or identities, then they predict what effects these events would bring to them, and predict future with the form of report or answer. Financial texts are usually unstructured, so extracting information and summarization require a lot of manual work.

There are many surveys summarizing what LLMs could bring for the financial industry. YinCheng Li et al [21] summarize how to apply LLMs to financial tasks and tested some models, and create a selection guide for stakeholders. Huaqin Zhao et al [35] also give an overview of how LLMs could be applied to financial tasks, and they also conduct tests on financial tasks through natural language instructions. Similarly, Jean Lee et al [20] summarize FinLLMs, test them with datasets and analyze future challenges, and they creatively summarize general LLMs as well as FinLLMs in chronological order.

However, many of them investigate general financial applications, only a few specific financial task applications are investigated. For example, Han Ding illustrates the LLMs applied to financial trading. [9].No survey specifically investigates how LLMs could help with financial text manipulation, which is an important part of the work of financial employees. Therefore, we choose this topic as the research field.

This survey tries to illustrate these research questions.

Research Question 1: What LLMs could be applied to financial text manipulation? How do they perform?

Research Question 2: What are the steps to apply LLMs to financial text?

Research Question 3: What are the potential issues and future research directions for applying LLMs to financial text?

Our survey has the following contributions:

1. We are the first survey to examine the LLMs application in the field of financial text manipulation.

2. We summarize the steps to apply LLMs to financial text, including data collection, data pre-processing and fine-tuning.

3. We raise a taxonomy of applications of LLMs in the field of financial text manipulation.

4. We analyze the potential security or bias issues LLMs

may face when manipulating financial texts.

Firstly, we illustrate the general application pipeline and dataset requirements for this task, then we analyze how LLMs could help with financial text manipulation from the perspective of both reading and writing. At last, we raise potential security and bias issues this kind of application could face.

The overall structure of this survey can be seen in Figure 1.

2 Background

2.1 Financial Analysis Workflow

In the financial industry, employees deal with many texts every day, such as analyzing news and reports to predict financial trends, responding to clients and other stakeholders by answering their questions, writing reports on new technology or macroeconomics, and providing a credit score to a loan applicant. The general financial analysis pipeline can be seen in Figure 2. The financial analysts can read financial news or reports and then extract company names, predict whether the company would have a better or worse future, and write reports to summarize the content of what they read or answer questions about their point of view for the future.

In the process of analyzing news or reports, financial analysts need to pick out which people, which companies, which organizations are involved. This is usually called Named Entity Recognition. Named Entity Recognition (NER) is a specialized area within Natural Language Processing (NLP) focused on identifying and categorizing key entities in text into specific classes. These entities can include names of people, organizations, locations, dates, quantities, monetary values, and more. The primary objective of NER is to transform unstructured text into structured information, facilitating easier analysis and comprehension. In addition to named entity recognition, they would also extract other information from the texts, such as sentiment classification. Financial analysts always predict that mentioned companies would go down or up after the publication of some specific news or reports, and this is called sentiment analysis. Sentiment analysis is a technique in natural language processing (NLP) that aims to identify and extract the emotional tone from text. It involves evaluating text data to classify sentiments as positive, negative, or neutral and can be applied to various types of content, such as reviews and social media posts.

2.2 LLM Background

LLMs are large language models based on transformers and trained with a huge amount of data from the web or other sources. The models usually contain billions of parameters. A notable application of LLMs is ChatGPT, which builds on the GPT series to enable dialogue, demonstrating an impressive capacity for human-like conversation. The architecture can be seen in Figure 3.

When applied to industry, LLMs often need to retrieve data from the web or other data sources to enhance LLMs' ability. RAG stands for Retrieval-Augmented Generation, and this is a method in natural language processing (NLP) that enhances the effectiveness of language models by integrating retrieval-based approaches with generative models.

Text is not the only source of LLMs. Since humans naturally perceive the world and communicate through various modalities, it is crucial to develop any multimodal LLMs that can process and generate content in any modality, moving us closer to human-level AI. Many LLMs now have integrated multimodal functions, such as GPT-4 (Multimodal) by OpenAI, Google's PaLM-E, and Flamingo by DeepMind.

3 Review Process

We use keywords such as "LLM Finance NER/ LLM NER" to search for the application of LLM used for Information Extraction, "Sentiment Analysis LLM" to search for the application of LLM used for sentiment analysis, "LLM Writing Report/ LLM Financial Report" to search for the application of LLM used for generating reports, "LLM Finance Question Answering / LLM Question Answering" to search for the application of LLM used for financial question answering, "LLM Security Issues" + "LLM Bias Issues/Fairness" to search for the potential security and bias issues of LLM. The overall match could be seen in Table 1. If some keywords limited with finance can't find enough papers, then we use keywords that wipe "finance" to search for papers that have general applications and see whether they can be used in the financial industry.

Keywords	Paper Topics
LLM Finance NER / LLM NER	Information Extraction
Sentiment Analysis LLM	Sentiment Analysis
LLM Writing Report / LLM Financial Report	Generating Reports
LLM Finance Question Answering / LLM Question Answering	Question Answering
LLM Security/ LLM Security Issues	Security Issues of General LLMs
LLM Bias/ LLM Fairness/ LLM Bias Issues	Bias Issues of General LLMs

Table 1: Keywords and Paper Topic

We have such selection and exclusion criteria as:

1) The application should better be closely related to the financial industry, at least could be transferred to the financial

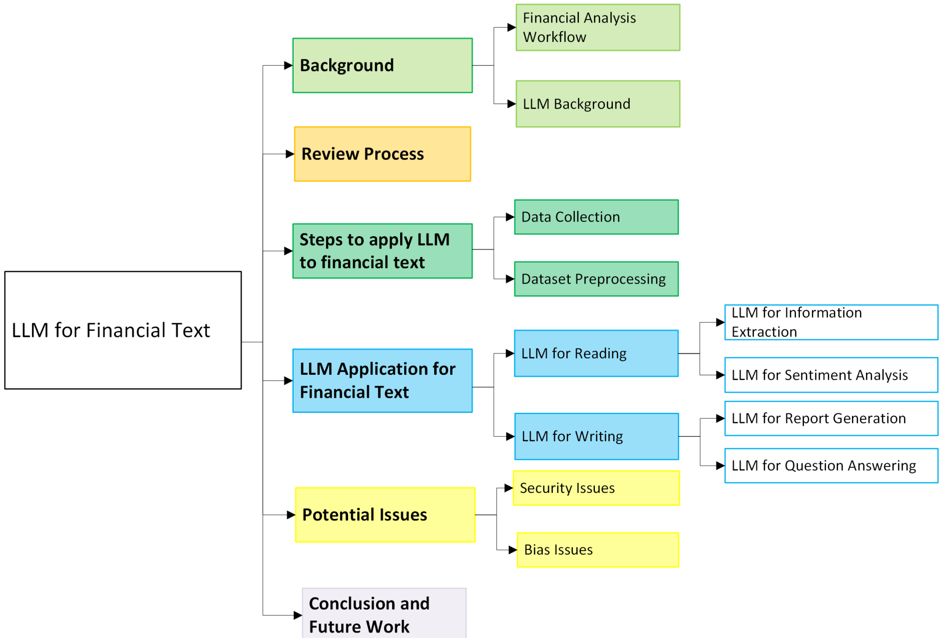

Figure 1: Overall Structure of This Survey

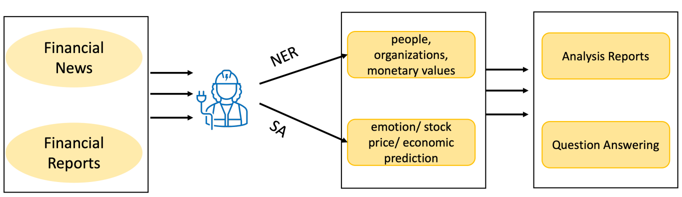

Figure 2: Financial Analysis Pipeline

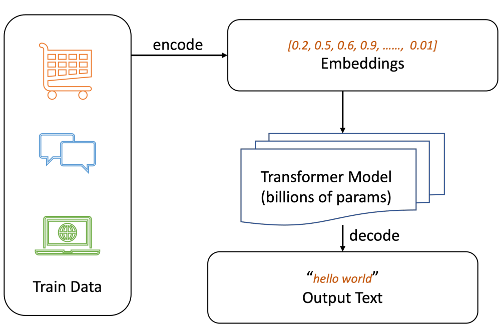

Figure 3: LLM Architecture

industry, like general text manipulation.

2) The majority of the papers should be about applications in text manipulation, but we would simply discuss future research directions, so the multi-modal application in the financial industry would also be included.

The database is IEEE Xplore Digital Library, ACM Digital Library, Science, mediaTUM, SSRN, Arxiv, and so on.

4 Steps to apply LLMs to financial text

The process of applying LLMs to financial text usually consists of 1) data collection, 2) data pre-processing, and 3) fine-tuning or direct test. Data collection can be divided as fetching data from the Internet and direct utilization of other public datasets. Data pre-processing includes data annotating and instruction augmentation. The general architecture can be seen in Figure 3.

4.1 Data Collection

Firstly, doing experiments or fine-tuning requires a huge amount of raw financial texts.

However, official news sites often limit access rates, so it is hard to directly collect a huge amount of news at a time and now news aggregator websites are useful. Dolphin et al [10] collect a live aggregate news feed from Google News for a select list of providers and topics, and then fetch the full paper from the link provided in the response. Similarly, Harish Bhat et al [2] collected news about a list of 32 mega-cap companies

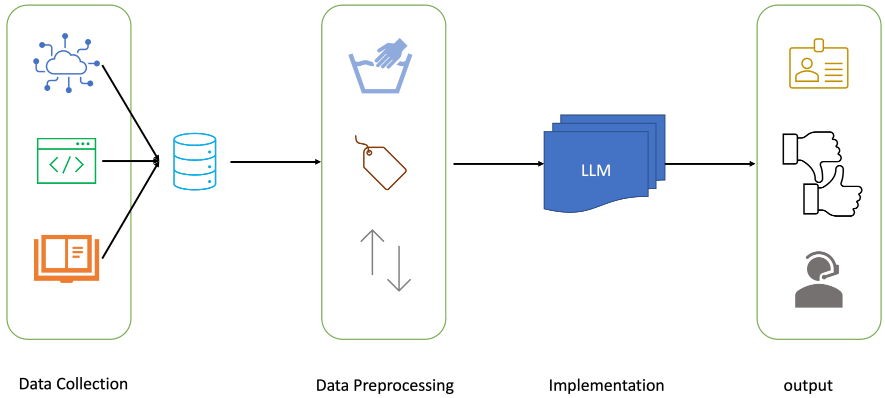

Figure 4: Steps to Apply LLMs on Financial Text

from similar news aggregators. Deng et al [8] select posts according to the popularity of the referenced stock using an internal system and randomly sample 20,000 posts for further analysis. Given the absence of existing datasets for market sentiment analysis on Reddit, they also sample an additional 100 posts for evaluation. These posts are annotated by three in-house experts.

Meanwhile, direct utilization of public data is also popular. Zhang et al directly use a public Chinese financial dataset ChFinAnn [37] and a business punishment dataset from an App called Enterprise Early Warning. The datasets are publicly available or directly collected from enterprises.Inserte [17] also utilizes a public dataset including EDGAR Files, Reuters News, In-house Dataset and The Pile [12]. Zhang et al [33] combine Twitter Financial News dataset [23] and FiQA dataset [24].

We can use annotated datasets to fine-tune LLMs. A simple pipeline is fine-tuning a pre-trained model to solve financial problems. For example, Wei Luo fine-tuned the open-source Llama2-7B model to do financial sentiment analysis. [22]

We can also use chain-of-thought to fine-tune LLMs to learn how to label raw social media texts. Xiang Deng et al [8] found that LLM-assisted labels help a small model to perform better than supervised counterparts.

4.2 Dataset Preprocessing

Zhang and Zhang [34] randomly picked some data points from ChFinAnn and AdminPunish datasets, which is called a resampling scheme. They repeat resampling for 5 times. They also properly merge entities that have the same meaning. Deng et al [8] also split the datasets into training and testing. Inserte et al [17] not only minimize the dataset by picking a portion of the dataset, but also concatenate all the documents.

In most cases, raw data without any label could be used to feed LLMs and let LLMs output texts according to inputs. But there are also some cases where LLMs need labeled datasets like fine-tuning.

Wei Luo et al [25] used the dataset containing 4845 English sentences randomly extracted from financial news found on LexisNexis database and split the datasets into train, validation, and test.

For dataset label annotation, there are some easier methods, like [27] [31] [30] [36]. Zhang et al [33] create instruction datasets by methods from [36] to fine-tune models.

5 LLM Applications to Financial Texts

5.1 LLM for Reading

5.1.1 LLM for Information Extraction

Information extraction is one of the most important tasks financial employees take. People often extract keywords such as company names or key sentences from financial news or reports.

The traditional method of extracting information is by using a domain dictionary, but this kind of work requires a lot of manual work. Then the machine learning-based method rises like Random Forest or SVM, but this kind of method requires a lot of data labeling.

When deep learning emerged, in 2003, Hammerton et al [16] first applied unidirectional LSTM to the field of NER. In 2005, Graves et al [13] first proposed a model called BiLSTM to incorporate context in speech recognition.

Then BiLSTM, CNN, and CRF are always combined and experimented with in different experiments.

However, for the BiLSTM-CRF model, there are more parameter settings and the model training time is longer. To solve the problems of ignoring context information and low model efficiency, BERT-BiGRU-CRF was proposed. [28].

The incapacity to extract structured information from unstructured data is solved by [34].

For financial NER, Zhang and Zhang [34] proposed a new model called FinBERT-MRC and performed experiments on this model. Their results have F1 scores of 92.78% and

96.80% on Chinese Financial datasets, which outperform many state-of-the-art NER models, including BiLSTM-CRF, BERT-Tagger, and BERT-CRF. Rian Dolphin [10] improved financial NER by extracting relevant company tickers in addition to company names from the content of raw news articles.

5.1.2 LLM for Sentiment Analysis

The same as NER tasks, sentiment analysis also went through a history of transfer from rule-based to machine learning, then to deep learning methodology.

After the emergence of deep learning, the context is better incorporated. Min-Yuh Day was the first researcher to apply deep learning to financial sentiment analysis. [7]

With the emergence of pre-trained models, pre-trained models could solve such sentiment analysis problems more easily without labeled datasets for training. Wenxuan Zhang and Yue Deng performed experiments on how LLMs and small language models (SLMs) trained on domain-specific datasets perform on sentiment analysis tasks. They found that LLMs perform quite well on simpler tasks in a zero-shot setting, but they underperform when meeting more complex tasks. Then many LLM-assisted methods or augmented LLM methods are tested on the financial sentiment analysis task. [2, 8, 17, 33] Given certain conditions, many models can achieve an accuracy of over 80%. Pau Rodriguez Inserte et al [17] found that even well-adapted small models related to financial tasks outperform famous general Large Language Models, which paves the way for small companies to make their own useful models.

5.2 LLM for Writing

5.2.1 LLM for Report Generation

Report generation should be classified as two categories. Sometimes financial employees are asked to generate summaries based on given information, while sometimes they need to write new content given a title or topic.

The generation of summarization is usually the same as Automatic Text Summarization(ATS) applied in the finance industry.

The generation of expository content based on a topic or idea is more like automatic writing. Nishant Balepur et al [1]raised a framework to combine RAG, content planning, and paraphrasing. Wenjin Yao et al [29] introduced a methodology to automatically write a survey, and this method significantly outperforms the generation of naive RAG-based LLM and matches human performance in content and citation quality. Although these methods are not designed for general text, they can also be applied to financial text.

5.2.2 LLM for Question Answering

Question answering is one of the most important tasks of financial advisory, such as providing basic knowledge of investment to customers or making predictions for some events.

For example, a customer wants to invest in the stock market, and he needs basic knowledge of what stocks are, how to purchase them, and what stocks are recommended. There are also some more advanced cases. A financial analyst may want to avoid cumbersome calculations, and he or she would need LLMs to do the calculations.

Stephen Choi et al [6] created Conversational Factor Information Retrieval Method(ConFIRM), which is helpful for fine-tuning LLMs for question answering and proved that it can achieve 91% accuracy in classifying financial queries.

Pranab Islam and researchers from Patronus AI and Contextual AI [18] want to know how existing LLMs perform on financial questions. They created a first-of-its-kind test suite called FINANCEBENCH and found that existing LLMs have clear limitations for financial QA, which are not suitable for enterprise usage.

When it comes to reasoning or calculations, Ethan Callanan et al [4] perform a comprehensive assessment of ChatGPT and GPT-4 using the CFA exams, revealing that ChatGPT fails to pass, whereas GPT-4 manages to succeed under certain few-shot and chain-of-thought prompting conditions.

6 Potential Issues

6.1 Security Issues

The membership inference attack is the most possible leakage for finance LLM. Some inference attack methods could even achieve a mean accuracy of 82% and a median certainty of 82%. The MIA can determine whether a patient is a member of the diabetic dataset of all models tested [19].

For data leakage, two of the mitigations are synthetic data and differential privacy.

LLM-based agents are prevalent in the financial industry to provide portfolio strategies or recommend financial products. Backdoor attacks can be exploited to change the outcome of the financial product recommendation or to forecast the outcome. [32]

Between 2020 and 2023, "pig butchering" scams have reportedly caused financial losses exceeding $75 billion to victims. [15] LLMs are reported to replace manual work to initiate text, voice, and video dialogue with victims. [14]

For fraud and scams, the most important defense is to detect them. For fraud and phishing emails, some LLM could achieve a detection precision of more than 90%. [5]

Application	Papers
Financial NER	10, 13, 16, 28, 34
Financial Sentiment Analysis	2, 7, 8, 17, 33
Financial Report Generation	1, 29
Finance Question Answering	4, 6, 18

Table 2: Keywords and Paper Topic

6.2 Bias Issues

Hui Zhong et al [38] did experiments and found that LLMs have bias in economic and financial decision making because the data used for training has bias.

For example, for the task of understanding user data and predicting user behavior, LLM tends to have bias. Duanyu Feng et al [11] found that LLMs tend to be biased when giving credit scores to different genders, races, or ages. Their experiments show that GPT-4 is more serious than ChatGPT.

In order to mitigate bias issues of LLMs when applying LLMs to financial texts, we can incorporate more samples to train the models or adopt fairness-aware algorithms, such as WEAT [3], to audit how the models train [26].

7 Conclusion and Future Work

We mainly analyzed how the LLMs could be applied to financial text manipulation, how they perform, and what problems they might bring.

While most of the LLMs designed for finance claim to be very accurate and stable, there are still not enough empirical experiments of their actual performance and accuracy.

Some applications of LLMs on specific tasks still lack research, such as financial report generation. There are many researches about report generation, but what problems LLMs would face and how to address them in finance are still unresolved.

In some scenarios like answering financial questions, a benchmark is raised for the industry to check whether the models are effective. However, some scenarios such as the generation of financial reports still lack such a benchmark which is verified by financial experts to allow researchers or financial analysts to know whether a LLM is adoptable. One research direction is to test LLM models on these tasks by empirical experiments and to raise benchmark test suites.

LLM has encompassed many other kinds of data, such as voice and video, and how these multi-modal LLMs would affect the financial industry is a good future research direction.

When applied to financial text manipulation, we should notice the security and bias issues mentioned above. The mitigation of the problems still requires research.

Acknowledgments

Thanks for Dr Maria Konte and all the staff from CS 8803 O23: Special Topics: Modern Internet Res. Methods. With the help of this course, I finished my first survey, and know how to start research.

References

[1] Nishant Balepur, Jie Huang, and Kevin Chen-Chuan Chang. Expository text generation: Imitate, retrieve, paraphrase. *arXiv preprint arXiv:2305.03276*, 2023.

[2] Rithesh Bhat and Bhanu Jain. Stock price trend prediction using emotion analysis of financial headlines with distilled llm model. In *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 67–73, 2024.

[3] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, 2017.

[4] Ethan Callanan, Amarachi Mbakwe, Antony Papadimitriou, Yulong Pei, Mathieu Sibue, Xiaodan Zhu, Zhiqiang Ma, Xiaomo Liu, and Sameena Shah. Can gpt models be financial analysts? an evaluation of chatgpt and gpt-4 on mock cfa exams. *arXiv preprint arXiv:2310.08678*, 2023.

[5] Joymallya Chakraborty, Wei Xia, Anirban Majumder, Dan Ma, Walid Chaabene, and Naveed Janvekar. Detoxbench: Benchmarking large language models for multitask fraud & abuse detection. *arXiv preprint arXiv:2409.06072*, 2024.

[6] Stephen Choi, William Gazeley, Siu Ho Wong, and Tingting Li. Conversational financial information retrieval model (confirm). *arXiv preprint arXiv:2310.13001*, 2023.

[7] Min-Yuh Day and Chia-Chou Lee. Deep learning for financial sentiment analysis on finance news providers. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1127–1134. IEEE, 2016.

[8] Xiang Deng, Vasilisa Bashlovkina, Feng Han, Simon Baumgartner, and Michael 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*, pages 107–110, 2023.

[9] Han Ding, Yinheng Li, Junhao Wang, and Hang Chen. Large language model agent in financial trading: A survey. *arXiv preprint arXiv:2408.06361*, 2024.

[10] Rian Dolphin, Joe Dursun, Jonathan Chow, Jarrett Blankenship, Katie Adams, and Quinton Pike. Extracting structured insights from financial news: An augmented llm driven approach. *arXiv preprint arXiv:2407.15788*, 2024.

[11] Duanyu Feng, Yongfu Dai, Jimin Huang, Yifang Zhang, Qianqian Xie, Weiguang Han, Zhengyu Chen, Alejandro Lopez-Lira, and Hao Wang. Empowering many, biasing a few: Generalist credit scoring through large language models. *arXiv preprint arXiv:2310.00566*, 2023.

[12] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

[13] Alex Graves and Jürgen Schmidhuber. Framewise phoneme clasïsification with bidirectional lstm networks. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 4, pages 2047–2052. IEEE, 2005.

[14] Gilad Gressel, Rahul Pankajakshan, and Yisroel Mirsky. Discussion paper: Exploiting llms for scam automation: A looming threat. In *Proceedings of the 3rd ACM Workshop on the Security Implications of Deepfakes and Cheapfakes*, pages 20–24, 2024.

[15] John Griffin and Kevin Mei. How do crypto flows finance slavery? the economics of pig butchering. *SSRN Electronic Journal*, 01 2024.

[16] James Hammerton. Named entity recognition with long short-term memory. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003*, pages 172–175, 2003.

[17] Pau Rodriguez Inserte, Mariam Nakhlé, Raheel Qader, Gaëtan Caillaut, and Jingshu Liu. Large language model adaptation for financial sentiment analysis. *arXiv preprint arXiv:2401.14777*, 2024.

[18] Pranab Islam, Anand Kannappan, Douwe Kiela, Rebecca Qian, Nino Scherrer, and Bertie Vidgen. Financebench: A new benchmark for financial question answering. *arXiv preprint arXiv:2311.11944*, 2023.

[19] Balder Janryd and Tim Johansson. Preventing health data from leaking in a machine learning system: Implementing code analysis with llm and model privacy evaluation testing, 2024.

[20] Jean Lee et al. A survey of large language models in finance (finllms). *arXiv preprint arXiv:2402.02315*, 2024.

[21] Yinheng Li et al. Large language models in finance: A survey. In *Proceedings of the fourth ACM international conference on AI in finance*, 2023.

[22] Wei Luo and Dihong Gong. Pre-trained large language models for financial sentiment analysis. *arXiv preprint arXiv:2401.05215*, 2024.

[23] Neural Magic. Twitter financial news sentiment, 2022.

[24] Macedo Maia, Siegfried Handschuh, Andre Freitas, Brian Davis, Ross McDermott, Manel Zarrouk, and Alexandra Balahur. Www '18: Companion proceedings of the web conference 2018. In *Proceedings of the International World Wide Web Conference*, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[25] Pekka Malo, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, and Pyry Takala. Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, 65(4):782–796, 2014.

[26] Nneka Adaobi Ochuba, Adetumi Adewunmi, and David Olanrewaju Olutimehin. The role of ai in financial market development: enhancing efficiency and accessibility in emerging economies. *Finance & Accounting Research Journal*, 6(3):421–436, 2024.

[27] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744, 2022.

[28] Qiuli Qin, Shuang Zhao, and Chunmei Liu. A bert-bigru-crf model for entity recognition of chinese electronic medical records. *Complexity*, 2021(1):6631837, 2021.

[29] Yidong Wang, Qi Guo, Wenjin Yao, Hongbo Zhang, Xin Zhang, Zhen Wu, Meishan Zhang, Xinyu Dai, Min Zhang, Qingsong Wen, et al. Autosurvey: Large language models can automatically write surveys. *arXiv preprint arXiv:2406.10252*, 2024.

[30] Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. Self-instruct: Aligning language model with self generated instructions. *arXiv preprint arXiv:2212.10560*, 2022.

[31] Jason Wei, Maarten Bosma, Vincent Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*, 2022.

[32] Wenkai Yang, Xiaohan Bi, Yankai Lin, Sishuo Chen, Jie Zhou, and Xu Sun. Watch out for your agents! investigating backdoor threats to llm-based agents. *arXiv preprint arXiv:2402.11208*, 2024.

[33] Boyu Zhang, Hongyang Yang, Tianyu Zhou, Muhammad Ali Babar, and Xiao-Yang Liu. Enhancing financial sentiment analysis via retrieval augmented large language models. In *Proceedings of the fourth ACM international conference on AI in finance*, pages 349–356, 2023.

[34] Yuzhe Zhang and Hong Zhang. Finbert–mrc: financial named entity recognition using bert under the machine reading comprehension paradigm. *Neural Processing Letters*, 55(6):7393–7413, 2023.

[35] Huaqin Zhao et al. Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*, 2024.

[36] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[37] Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. Doc2edag: An end-to-end document-level framework for chinese financial event extraction. *arXiv preprint arXiv:1904.07535*, 2019.

[38] Hui Zhong, Songsheng Chen, and Mian Liang. Gender bias of llm in economics: An existentialism perspective. *arXiv preprint arXiv:2410.19775*, 2024.