arXiv:2306.06031v1 [q-fin.ST] 9 Jun 2023

# FinGPT: Open-Source Financial Large Language Models

**Hongyang (Bruce) Yang**[1], **Xiao-Yang Liu**[1], **Christina Dan Wang**[2]
[1]Columbia University; [2]New York University (Shanghai)
{HY2500, XL2427}@columbia.edu; christina.wang@nyu.edu

## Abstract

Large language models (LLMs) have shown the potential of revolutionizing natural language processing tasks in diverse domains, sparking great interest in finance. Accessing high-quality financial data is the first challenge for financial LLMs (FinLLMs). While proprietary models like BloombergGPT have taken advantage of their unique data accumulation, such privileged access calls for an open-source alternative to democratize Internet-scale financial data.

In this paper, we present an open-source large language model, FinGPT, for the finance sector. Unlike proprietary models, FinGPT takes a data-centric approach, providing researchers and practitioners with accessible and transparent resources to develop their FinLLMs. We highlight the importance of an automatic data curation pipeline and the lightweight low-rank adaptation technique in building FinGPT. Furthermore, we showcase several potential applications as stepping stones for users, such as robo-advising, algorithmic trading, and low-code development. Through collaborative efforts within the open-source AI4Finance community, FinGPT aims to stimulate innovation, democratize FinLLMs, and unlock new opportunities in open finance. Two associated code repos are https://github.com/AI4Finance-Foundation/FinGPT and https://github.com/AI4Finance-Foundation/FinNLP

## 1 Introduction

The continual expansion and evolution of artificial intelligence have provided a fertile ground for the proliferation of large language models [Vaswani *et al.*, 2017; Radford *et al.*, 2018; Devlin *et al.*, 2018; Ethayarajh, 2019; Lewis *et al.*, 2019; Lewis *et al.*, 2020; Brown *et al.*, 2020; Thoppilan *et al.*, 2022], thereby effecting a transformative shift in the landscape of natural language processing across diverse domains. This sweeping change has engendered keen interest in the potential application of these models in the financial realm. It is, however, evident that the acquisition of high-quality, relevant, and up-to-date data stands as a critical factor in the development of an efficacious and efficient open-source financial language model.

Utilizing language models in the financial arena reveals intricate hurdles. These range from difficulties in obtaining data, dealing with diverse data formats and types, and managing data quality inconsistencies, to the essential requirement of up-to-date information. Especially, historical or specialized financial data extraction proves to be complex due to varying data mediums such as web platforms, APIs, PDF documents, and images.

In the proprietary sphere, models like BloombergGPT [Wu *et al.*, 2023] have capitalized on their exclusive access to specialized data to train finance-specific language models. However, the restricted accessibility and transparency of their data collections and training protocols have accentuated the demand for a more open and inclusive alternative. In response to this demand, we are witnessing a shifting trend towards democratizing Internet-scale financial data in the open-source domain.

In this paper, we address these aforementioned challenges associated with financial data and introduce FinGPT, an end-to-end open-source framework for financial large language models (FinLLMs). Adopting a data-centric approach, FinGPT underscores the crucial role of data acquisition, cleaning, and preprocessing in developing open-source FinLLMs. By championing data accessibility, FinGPT aspires to enhance research, collaboration, and innovation in finance, paving the way for open finance practices.

Our contributions are summarized as follows:

- **Democratization**: FinGPT, as an open-source framework, aims to democratize financial data and FinLLMs, uncovering untapped potentials in open finance.
- **Data-centric approach**: Recognizing the significance of data curation, FinGPT adopts a data-centric approach and implements rigorous cleaning and preprocessing methods for handling varied data formats and types, thereby ensuring high-quality data.
- **End-to-end framework**: FinGPT embraces a full-stack framework for FinLLMs with four layers:
  - *Data source layer*: This layer assures comprehensive market coverage, addressing the temporal sensitivity

of financial data through real-time information capture.

– *Data engineering layer*: Primed for real-time NLP data processing, this layer tackles the inherent challenges of high temporal sensitivity and low signal-to-noise ratio in financial data.
– *LLMs layer*: Focusing on a range of fine-tuning methodologies, this layer mitigates the highly dynamic nature of financial data, ensuring the model's relevance and accuracy.
– *Application layer*: Showcasing practical applications and demos, this layer highlights the potential capability of FinGPT in the financial sector.

Our vision for FinGPT is to serve as a catalyst for stimulating innovation within the finance domain. FinGPT is not limited to providing technical contributions, but it also cultivates an open-source ecosystem for FinLLMs, promoting real-time processing and customized adaptation for users. By nurturing a robust collaboration ecosystem within the open-source AI4Finance community, FinGPT is positioned to reshape our understanding and application of FinLLMs.

## 2 Related Work

### 2.1 LLMs and ChatGPT

Large Language Models (LLMs) have been recognized as a technological breakthrough in natural language processing, such as GPT-3 and GPT-4 [Brown *et al.*, 2020]. They take transformer-based architectures, demonstrating impressive performance across various generative tasks.

As an offshoot of the GPT family developed by OpenAI, ChatGPT was designed to produce human-like text based on input prompts. It has shown significant utility in diverse applications, from drafting emails to writing code and even in creating written content.

### 2.2 LLMs in Finance

LLMs have been applied to various tasks within the financial sector [Dredze *et al.*, 2016; Araci, 2019; Bao *et al.*, 2021; DeLucia *et al.*, 2022], from predictive modeling to generating insightful narratives from raw financial data. Recent literature has focused on using these models for financial text analysis, given the abundance of text data in this field, such as news articles, earnings call transcripts, and social media posts.

The first example of financial LLMs is BloombergGPT [Wu *et al.*, 2023], which was trained on a mixed dataset of financial and general sources. Despite its impressive capabilities, access limitations exist, and the prohibitive training cost has motivated the need for low-cost domain adaptation.

Our FinGPT responds to these challenges, presenting an open-source financial LLM. It employs Reinforcement Learning from Human Feedback (RLHF) to understand and adapt to individual preferences, paving the way for personalized financial assistants. We aim to combine the strengths of general LLMs like ChatGPT with financial adaptation, exploiting LLM's capability in finance.

### 2.3 Why Open-Source FinLLMs?

AI4Finance Foundation is a non-profit, open-source organization that integrates Artificial Intelligence (AI) and financial applications, including financial Large Language Models (FinLLMs). With a proven track record of nurturing an innovative ecosystem of FinTech tools, such as FinRL [Liu *et al.*, 2021] and FinRL-Meta [Liu *et al.*, 2022], the foundation is poised to accelerate the evolution of FinLLMs further. It is steadfast commitment and cutting-edge contributions pave the way for AI's transformative application in finance.

- Advancing equal opportunities via democratizing FinLLMs: Adopting an open-source methodology promotes universal access to state-of-the-art technology, adhering to the ethos of democratizing FinLLMs.
- Cultivating transparency and trust: Open-source FinLLMs offer a comprehensive overview of their foundational codebase, bolstering transparency and trust.
- Accelerating research and innovation: The open-source model fuels progress in research and development within the AI domain. It allows researchers to leverage existing models, thus nurturing a faster progression of innovation and scientific discovery.
- Enhancing education: Open-source FinLLMs serve as robust educational tools, presenting students and researchers with the prospect of exploring the complexities of FinLLMs through direct engagement with fully operational models.
- Promoting community development and collaborative engagement: Open-source promotes a global community of contributors. This collaborative participation bolsters the model's long-term durability and effectiveness.

## 3 Data-Centric Approach for FinLLMs

For financial large language models (FinLLMs), a successful strategy is not solely based on the capability of the model architecture but is equally reliant on the training data. Our data-centric approach prioritizes collecting, preparing, and processing high-quality data.

### 3.1 Financial Data and Unique Characteristics

Financial data comes from a variety of sources, with unique characteristics. We delve into the specifics of different financial data sources, such as Financial News, Company Fillings, Social Media Discussions, and Company Announcements.

**Financial news** carries vital information about the world economy, specific industries, and individual companies. This data source typically features:

- **Timeliness:** Financial news reports are timely and up-to-date, often capturing the most recent developments in the financial world.
- **Dynamism:** The information contained in financial news is dynamic, changing rapidly in response to evolving economic conditions and market sentiment.
- **Influence:** Financial news has a significant impact on financial markets, influencing traders' decisions and potentially leading to dramatic market movements.

for finance we need to develop an ecosystem.

— news is imp but then we need to develop a platform that links old news to new news and then make potential best for short and long term investments.

**Company filings and announcements** are official documents that corporations submit to regulatory bodies, providing insight into a company's financial health and strategic direction. They feature:

- **Granularity**: These documents offer granular information about a company's financial status, including assets, liabilities, revenue, and profitability.
- **Reliability**: Company fillings contain reliable and verified data vetted by regulatory bodies.
- **Periodicity**: Company fillings are periodic, usually submitted on a quarterly or annual basis, offering regular snapshots of a company's financial situation.
- **Impactfulness**: Company announcements often have substantial impacts on the market, influencing stock prices and investor sentiment.

**Social media discussions** related to finance can reflect public sentiment towards specific stocks, sectors, or the overall market. These discussions tend to exhibit:

- **Variability:** Social media discussions vary widely in tone, content, and quality, making them rich, albeit complex, sources of information.
- **Real-time sentiment:** These platforms often capture real-time market sentiment, enabling the detection of trends and shifts in public opinion.
- **Volatility:** Sentiments expressed on social media can be highly volatile, changing rapidly in response to news events or market movements.

**Trends**, often observable through websites like Seeking Alpha, Google Trends, and other finance-oriented blogs and forums, offer critical insights into market movements and investment strategies. They feature:

- **Analyst perspectives:** These platforms provide access to market predictions and investment advice from seasoned financial analysts and experts.
- **Market sentiment:** The discourse on these platforms can reflect the collective sentiment about specific securities, sectors, or the overall market, providing valuable insights into the prevailing market mood.
- **Broad coverage:** Trends data spans diverse securities and market segments, offering comprehensive market coverage.

make an analyst —> that studies the document(earnings calls) and makes the final call.

Each of these data sources provides unique insights into the financial world. By integrating these diverse data types, financial language models like FinGPT can facilitate a comprehensive understanding of financial markets and enable effective financial decision-making.

### 3.2 Challenges in Handling Financial Data

We summarize three major challenges for handling financial data as follows:

- **High temporal sensitivity**: Financial data are characterized by their time-sensitive nature. Market-moving news or updates, once released, provide a narrow window of opportunity for investors to maximize their alpha (the measure of an investment's relative return).
- **High dynamism**: The financial landscape is perpetually evolving, with a daily influx of news, social media posts, and other market-related information. It's impractical and cost-prohibitive to retrain models frequently to cope with these changes.
- **Low signal-to-noise ratio (SNR)**: Financial data often exhibit a low signal-to-noise ratio [Liu *et al.*, 2022], meaning that the useful information is usually dwarfed by a substantial amount of irrelevant or noisy data. Extracting valuable insights from this sea of information necessitates sophisticated techniques.

Addressing these challenges is critical for the effective utilization of financial data and maximizing the potential of FinLLMs. As we navigate these challenges, we propose an open-source framework FinGPT.

## 4 Overview of FinGPT: An Open-Source Framework for FinLLMs

FinGPT represents an innovative open-source framework designed specifically for applying large language models (LLMs) within the financial domain. As delineated in Fig. 1, FinGPT consists of four fundamental components: Data Source, Data Engineering, LLMs, and Applications. Each of these components plays a crucial role in maintaining the functionality and adaptability of FinGPT in addressing dynamic financial data and market conditions.

- **Data source layer**: The starting point of the FinGPT pipeline is the Data Source Layer, which orchestrates the acquisition of extensive financial data from a wide array of online sources. This layer ensures comprehensive market coverage by integrating data from news websites, social media platforms, financial statements, market trends, and more. The goal is to capture every nuance of the market, thereby addressing the inherent temporal sensitivity of financial data.
- **Data engineering layer**: This layer focuses on the real-time processing of NLP data to tackle the challenges of *high temporal sensitivity* and *low signal-to-noise ratio* inherent in financial data. It incorporates state-of-the-art NLP techniques to filter noise and highlight the most salient pieces of information.
- **LLMs layer**: Lying at the heart, it encompasses various fine-tuning methodologies, with a priority on lightweight adaptation, to keep the model updated and pertinent. By maintaining an updated model, FinGPT can deal with the highly dynamic nature of financial data, ensuring its responses are in sync with the current financial climate.
- **Application layer**: The final component of FinGPT is the Applications Layer, designed to demonstrate the practical applicability of FinGPT. It offers hands-on tutorials and demo applications for financial tasks, including robo-advisory services, quantitative trading, and low-code development. These practical demonstrations not only serve as a guide to potential users but also underscore the transformative potential of LLMs in finance.

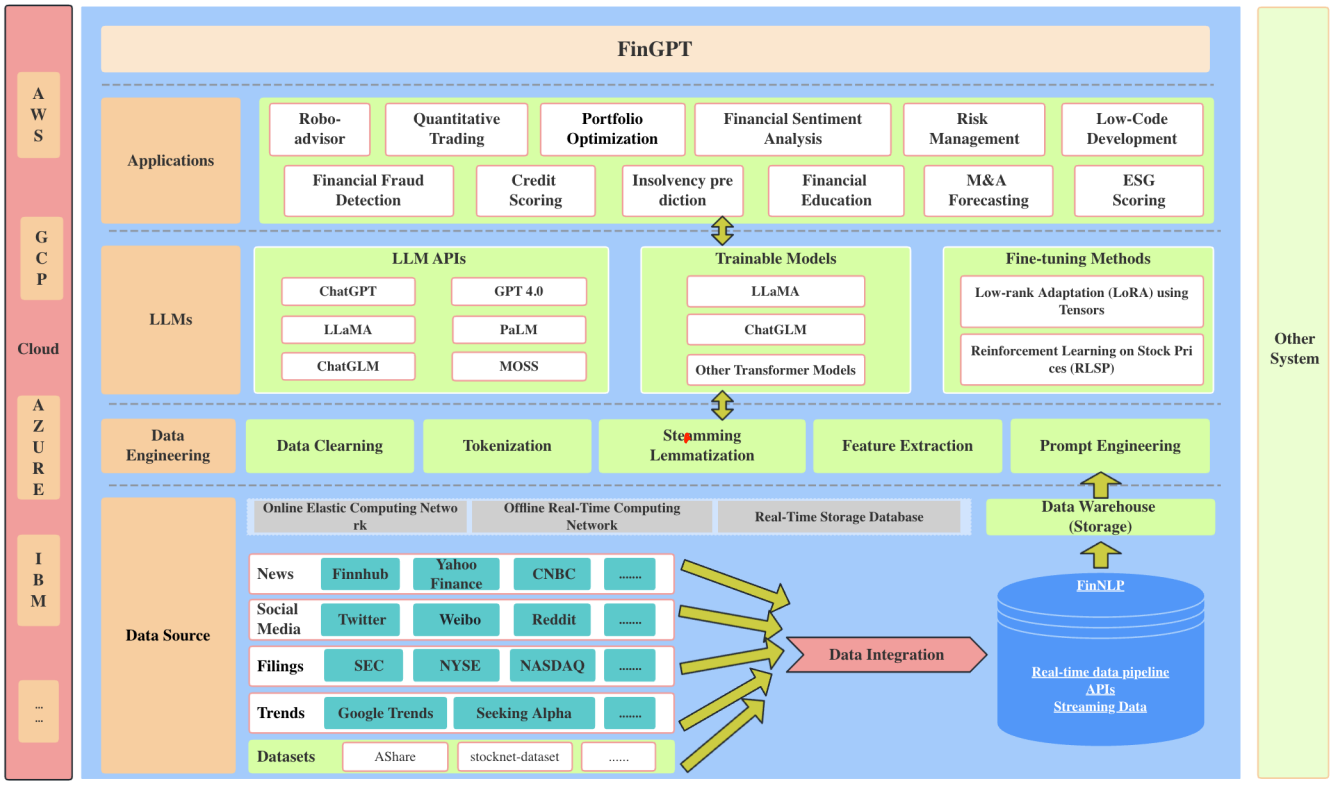


Figure 1: FinGPT Framework.

### 4.1 Data Sources

The first stage of the FinGPT pipeline involves the collection of extensive financial data from a wide array of online sources. These include, but are not limited to:

- **Financial news:** Websites such as Reuters, CNBC, Yahoo Finance, among others, are rich sources of financial news and market updates. These sites provide valuable information on market trends, company earnings, macroeconomic indicators, and other financial events.
- **Social media**: Platforms such as Twitter, Facebook, Reddit, Weibo, and others, offer a wealth of information in terms of public sentiment, trending topics, and immediate reactions to financial news and events.
- **Filings**: Websites of financial regulatory authorities, such as the SEC in the United States, offer access to company filings. These filings include annual reports, quarterly earnings, insider trading reports, and other important company-specific information. Official websites of stock exchanges (NYSE, NASDAQ, Shanghai Stock Exchange, etc.) provide crucial data on stock prices, trading volumes, company listings, historical data, and other related information.
- **Trends**: Websites like Seeking Alpha, Google Trends, and other finance-focused blogs and forums provide access to analysts' opinions, market predictions, the movement of specific securities or market segments and investment advice.
- **Academic datasets**: Research-based datasets that offer curated and verified information for sophisticated financial analysis.

To harness the wealth of information from these diverse sources, FinGPT incorporates data acquisition tools capable of scraping structured and unstructured data, including APIs, web scraping tools, and direct database access where available. Moreover, the system is designed to respect the terms of service of these platforms, ensuring data collection is ethical and legal.

**Data APIs**: In the FinGPT framework, APIs are used not only for initial data collection but also for real-time data updates, ensuring the model is trained on the most current data. Additionally, error handling and rate-limiting strategies are implemented to respect API usage limits and avoid disruptions in the data flow.

### 4.2 Real-Time Data Engineering Pipeline for Financial NLP

Financial markets operate in real-time and are highly sensitive to news and sentiment. Prices of securities can change rapidly in response to new information, and delays in processing that information can result in missed opportunities or increased risk. As a result, real-time processing is essential in financial NLP.

The primary challenge with a real-time NLP pipeline is managing and processing the continuous inflow of data efficiently. The first step in the pipeline is to set up a system to

ingest data in real-time. This data could be streaming from our data source APIs. Below are the steps to design a real-time NLP pipeline for data ingestion.

**Data cleaning**: Real-time data can be noisy and inconsistent. Therefore, real-time data cleaning involves removing irrelevant data, handling missing values, text normalization (like lowercasing), and error corrections.

**Tokenization**: In real-time applications, tokenization has to be performed on the fly. This involves breaking down the stream of text into smaller units or tokens.

**Stop word removal and stemming/lemmatization**: For real-time processing, a predefined list of stop words can be used to filter out these common words from the stream of tokens. Likewise, stemming and lemmatization techniques can be applied to reduce words to their root form.

**Feature extraction and sentiment analysis**: Feature extraction involves transforming raw data into an input that can be understood by machine learning models. In real-time systems, this often needs to be a fast and efficient process. Techniques such as TF-IDF, Bag of Words, or embedding vectors like Word2Vec can be used. Sentiment analysis can also be performed on the cleaned data. This is where we categorize a span of text as positive, negative, or neutral.

**Prompt engineering**: The creation of effective prompts that can guide the language model's generation process toward desirable outputs.

**Alerts/Decision making**: Once the prompt is entered, the results need to be communicated or acted upon. This might involve triggering alerts based on certain conditions, informing real-time decision-making processes, or feeding the output into another system.

**Continuous learning**: In real-time systems, the models should adapt to changes in the data. Continuous learning systems can be implemented, where models are periodically retrained on new data or online learning algorithms are used that can update the model with each new data point.

**Monitoring**: Real-time systems require continuous monitoring to ensure they are functioning correctly. Any delays or issues in the pipeline can have immediate impacts, so it's important to have robust monitoring and alerting in place.

### 4.3 Large Language Models (LLMs)

Once the data has been properly prepared, it is used with LLMs to generate insightful financial analyses. The LLM layer includes:

- **LLM APIs**: APIs from established LLMs provide baseline language capability.
- **Trainable models**: FinGPT provides trainable models that users can fine-tune on their private data, customizing for financial applications.
- **Fine-tuning methods**: Various fine-tuning methods allow FinGPT to be adapted to personalized robo-advisor.

**Why fine-tune LLMs instead of retraining from scratch?**

Leveraging pre-existing Large Language Models (LLMs) and fine-tuning them for finance provides an efficient, cost-effective alternative to expensive and lengthy model retraining from scratch.

BloombergGPT, though remarkable in its finance-specific capabilities, comes with an intensive computational requirement. It used approximately 1.3 million GPU hours for training, which, when calculated using AWS cloud's $2.3 rate, translates to a staggering cost of around $3 million per training. In contrast to the high computational cost of models like BloombergGPT, FinGPT presents a more accessible solution by focusing on the lightweight adaptation of top open-source LLMs. The cost of adaptation falls significantly, estimated at less than $300 per training.

This approach ensures timely updates and adaptability, essential in the dynamic financial domain. Being open-source, FinGPT not only promotes transparency but also allows user customization, catering to the rising trend of personalized financial advisory services. Ultimately, FinGPT's cost-effective, flexible framework holds the potential to democratize financial language modeling and foster user-focused financial services.

**Fine-tuning via Low-rank Adaptation (LoRA)**

In FinGPT, we fine-tune to a pre-trained LLM utilizing a novel financial dataset. It's well recognized that high-quality labeled data is a pivotal determinant for many successful LLMs, including ChatGPT. However, acquiring such top-notch labeled data often proves costly in terms of time and resources and generally requires the expertise of finance professionals.

If our objective is to employ LLMs for analyzing financial-related text data and assisting in quantitative trading, it seems sensible to leverage the market's inherent labeling capacity. Consequently, we use the relative stock price change percentage for each news item as the output label. We establish thresholds to divide these labels into three categories—positive, negative, and neutral—based on the sentiment of the news item.

In a corresponding step, during the prompt engineering process, we also prompt the model to select one from the positive, negative, and neutral outputs. This strategy ensures optimal utilization of the pre-trained information. By deploying the Low-Rank Adaptation (LoRA) of LLMs [Hu *et al.*, 2021; Dettmers *et al.*, 2023], we manage to reduce the number of trainable parameters from 6.17 billion to a mere 3.67 million.

**Fine-tuning via Reinforcement Learning on Stock Prices (RLSP)**

Similarly, we can substitute Reinforcement Learning on Stock Prices (RLSP) for Reinforcement Learning on Human feedback, as utilized by ChatGPT. The reasoning behind this substitution is that stock prices offer a quantifiable, objective metric that reflects market sentiment in response to news and events. This makes it a robust, real-time feedback mechanism for training our model.

Reinforcement Learning (RL) allows the model to learn through interaction with the environment and receiving feedback. In the case of RLSP, the environment is the stock market, and the feedback comes in the form of stock price changes. This approach permits FinGPT to refine its understanding and interpretation of financial texts, improving its ability to predict market responses to various financial events.

By associating news sentiment with the subsequent performance of the related stocks, RLSP provides an effective way to fine-tune FinGPT. In essence, RLSP allows the model to infer the market's response to different news events and adjust its understanding and predictions accordingly.

Therefore, the integration of RLSP into the fine-tuning process of FinGPT provides a powerful tool for improving the model's financial market understanding and predictive accuracy. By using actual stock price movements as feedback, we are directly harnessing the wisdom of the market to make our model more effective.

### 4.4 Applications

FinGPT may find wide applications in financial services, aiding professionals and individuals in making informed financial decisions. The potential applications include:

- **Robo-advisor**: Offering personalized financial advice, reducing the need for regular in-person consultations.
- **Quantitative trading**: Producing trading signals for informed trading decisions.
- **Portfolio optimization**: Utilizing numerous economic indicators and investor profiles for optimal investment portfolio construction.
- **Financial sentiment analysis**: Evaluating sentiments across different financial platforms for insightful investment guidance.
- **Risk management:** Formulating effective risk strategies by analyzing various risk factors.
- **Financial Fraud detection**: Identifying potential fraudulent transaction patterns for enhanced financial security.
- **Credit scoring**: Predicting creditworthiness from financial data to aid lending decisions.
- **Insolvency prediction**: Predicting potential insolvency or bankruptcy of companies based on financial and market data.
- **Mergers and acquisitions (M&A) forecasting**: Predicting potential M&A activities by analyzing financial data and company profiles, helping investors anticipate market movements.
- **ESG (Environmental, Social, Governance) scoring**: Evaluating companies' ESG scores by analyzing public reports and news articles.
- **Low-code development**: Facilitating software creation through user-friendly interfaces, reducing reliance on traditional programming.
- **Financial education**: Serving as an AI tutor simplifying complex financial concepts for better financial literacy.

By linking these distinct yet interconnected components, FinGPT provides a holistic and accessible solution for leveraging AI in finance, facilitating research, innovation, and practical applications in the financial industry.

## 5 Conclusion

In conclusion, the transformative integration of large language models (LLMs) into the financial sector brings unique complexities and vast opportunities. Navigating challenges such as high temporal sensitivity, dynamic financial landscape, and a low signal-to-noise ratio in financial data calls for efficient solutions. FinGPT responds innovatively by leveraging pre-existing LLMs and fine-tuning them to specific financial applications. This approach significantly reduces adaptation costs and computational requirements compared to models like BloombergGPT, offering a more accessible, flexible, and cost-effective solution for financial language modeling. Thus, it enables consistent updates to ensure model accuracy and relevance, a critical aspect in the dynamic and time-sensitive world of finance.

## 6 Future Work

FinLLMs, or Financial Large Language Models, present a vision of the future where personalized robo-advisors or assistants are within everyone's reach. It aims to democratize access to high-quality financial advice, leveraging advanced language modeling techniques to make sense of vast amounts of financial data and transform it into actionable insights. The following blueprint outlines the future direction of FinLLM.

- **Individualization**: At the heart of FinLLM's strategy is the concept of individualized fine-tuning. Using techniques such as LoRA and QLoRA, FinLLM enables users to tailor models to their specific needs, thereby creating a personal robo-advisor or assistant. This aligns with a broader trend towards customization in financial services, as consumers increasingly demand personalized advice that aligns with their unique risk profiles and financial goals.
- **Open-source and low-cost adaptation**: FinLLM champions open-source values, providing users with the tools they need to adapt Large Language Models (LLMs) to their own requirements at a low cost, typically between $100 to $300. This not only democratizes access to advanced financial modeling techniques but also fosters a vibrant community of developers and researchers, collectively pushing the boundaries of what's possible in the field of financial AI.
- **Access to high-quality financial data**: FinLLM goes beyond just providing modeling techniques, also offering access to high-quality financial data. This ensures that users have the data they need to train their models effectively, while also simplifying the data curation process. This access is further enhanced by the provision of a data curation pipeline with demos, empowering users to harness the full potential of their financial data.

**Disclaimer: We are sharing codes for academic purposes under the MIT education license. Nothing herein is financial advice, and NOT a recommendation to trade real money. Please use common sense and always first consult a professional before trading or investing.**

## References


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

[Bao *et al.*, 2021] Siqi Bao, Huang He, Fan Wang, Hua Wu, Haifeng Wang, Wenquan Wu, Zhihua Wu, Zhen Guo, Hua Lu, Xinxian Huang, et al. Plato-xl: Exploring the large-scale pre-training of dialogue generation. *arXiv preprint arXiv:2109.09519*, 2021.

[Brown *et al.*, 2020] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[DeLucia *et al.*, 2022] Alexandra DeLucia, Shijie Wu, Aaron Mueller, Carlos Aguirre, Philip Resnik, and Mark Dredze. Bernice: a multilingual pre-trained encoder for twitter. In *Proceedings of the 2022 conference on empirical methods in natural language processing*, pages 6191–6205, 2022.

[Dettmers *et al.*, 2023] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. QLoRA: Efficient finetuning of quantized llms. *arXiv preprint arXiv:2305.14314*, 2023.

[Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[Dredze *et al.*, 2016] Mark Dredze, Prabhanjan Kambadur, Gary Kazantsev, Gideon Mann, and Miles Osborne. How twitter is changing the nature of financial news discovery. In *proceedings of the second international workshop on data science for macro-modeling*, pages 1–5, 2016.

[Ethayarajh, 2019] Kawin Ethayarajh. How contextual are contextualized word representations? comparing the geometry of bert, elmo, and gpt-2 embeddings. *arXiv preprint arXiv:1909.00512*, 2019.

[Hu *et al.*, 2021] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. *International Conference on Learning Representations*, 2021.

[Lewis *et al.*, 2019] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*, 2019.

[Lewis *et al.*, 2020] Patrick Lewis, Myle Ott, Jingfei Du, and Veselin Stoyanov. Pretrained language models for biomedical and clinical tasks: understanding and extending the state-of-the-art. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*, pages 146–157, 2020.

[Liu *et al.*, 2021] Xiao-Yang Liu, Hongyang Yang, Jiechao Gao, and Christina Dan Wang. FinRL: Deep reinforcement learning framework to automate trading in quantitative finance. *ACM International Conference on AI in Finance (ICAIF)*, 2021.

[Liu *et al.*, 2022] Xiao-Yang Liu, Ziyi Xia, Jingyang Rui, Jiechao Gao, Hongyang Yang, Ming Zhu, Christina Dan Wang, Zhaoran Wang, and Jian Guo. FinRL-Meta: Market environments and benchmarks for data-driven financial reinforcement learning. *NeurIPS*, 2022.

[Radford *et al.*, 2018] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. *OpenAI*, 2018.

[Thoppilan *et al.*, 2022] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, et al. Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239*, 2022.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[Wu *et al.*, 2023] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*, 2023.