

FinGPT: Open-Source Financial Large Language Models

Hongyang Yang¹, Xiao-Yang Liu², Christina Dan Wang^{3*}

¹AI4Finance Foundation[†];

²Columbia University;

³New York University Shanghai

contact@ai4finance.org

Abstract

Large language models (LLMs) have shown the potential of revolutionizing natural language processing in diverse domains, sparking great interest in finance. However, the finance domain presents unique challenges, including high temporal sensitivity, constant dynamism, and a low signal-to-noise ratio (SNR). 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 customize their financial LLMs (FinLLMs). We highlight the importance of an automatic data curation pipeline and the lightweight low-rank adaptation technique in building FinGPT. Furthermore, we provide fundamental tasks as building blocks for benchmarking and showcase potential applications as stepping stones for users, such as robo-advising and sentiment analysis. 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 LLMs [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 across diverse

domains. This sweeping change has engendered keen interest in the potential applications of financial LLMs (FinLLMs). 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 efficacious and efficient FinLLMs.

Utilizing LLMs in the finance sector reveals intricate hurdles. Firstly, there's the issue of 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). Secondly, the financial landscape is marked by high dynamism. It is in a constant state of flux due to the ceaseless flow of news, social media updates, and other market-related information. Given these constant changes, retraining LLMs frequently is not only expensive but also impractical. Lastly, financial data is often characterized by a low signal-to-noise ratio (SNR) [Yang *et al.*, 2020]. The useful information is often hidden amongst a significant amount of irrelevant or noisy data. Extracting valuable insights from this sea of information necessitates advanced techniques.

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

- **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.

*Corresponding author.

[†]AI4Finance Foundation: ai4finance.org

- **End-to-end framework:** FinGPT embraces a full-stack framework with five layers:
 - *Data source 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 takes care of the highly dynamic nature of financial data, ensuring the model’s relevance and accuracy.
 - *Tasks Layer:* This layer is responsible for executing fundamental tasks. These tasks serve as the benchmarks for performance evaluations and cross-comparisons in the realm of FinLLMs.
 - *Applications layer:* Showcasing practical applications and demos, this layer highlights the potential capability of FinGPT in the finance sector.

- **Democratization:** FinGPT, as an open-source framework, aims to democratize financial data and FinLLMs, uncovering untapped potentials in open finance. We envision FinGPT as a catalyst for stimulating innovation within the finance domain. FinGPT is not limited to providing technical contributions, but 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 refine our understanding and application of FinLLMs.

2 Related Work

2.1 The Raise of FinLLMs

Large Language Models (LLMs) have been recognized as a technological breakthrough in NLP, such as GPT-3 and GPT-4 [Brown *et al.*, 2020; Jiang *et al.*, 2023; OpenAI, 2023; Team *et al.*, 2023; Liu *et al.*, 2024]. They take transformer-based architectures, demonstrating impressive performance across various text-generation tasks. As an offshoot of the GPT family developed by OpenAI, ChatGPT was designed to produce human-like texts based on input prompts. It has shown significant utility in diverse applications, from drafting emails to writing code and even in creating art content.

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 FinLLMs is BloombergGPT [Wu *et al.*, 2023], which was trained on a mixed dataset of financial and general data 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 the aforementioned hurdles, presenting an open-source FinLLM. 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 open finance.

2.2 Why Open-Source FinLLMs?

AI4Finance Foundation¹ is a non-profit, open-source organization that integrates Artificial Intelligence (AI) and financial applications. With a proven track record of nurturing an innovative ecosystem of FinTech tools, such as FinRL [Yang *et al.*, 2020] and FinRobot [Yang *et al.*, 2024], the foundation is poised to accelerate the evolution of FinLLMs. Steadfast commitment and cutting-edge contributions may pave the way for AI’s transformative applications in open 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 with the prospect of exploring the complexities of FinLLMs through direct engagement with fully operational models.
- **Upgrade foundation infrastructure for financial text data by community collaboration:** This collaborative participation bolsters the model’s long-term durability and effectiveness.

3 Overview of FinGPT: An Open-Source Framework for FinLLMs

FinGPT represents an innovative open-source framework designed specifically for FinLLMs. As delineated in Fig. 1, FinGPT consists of four components: Data Source, Data Engineering, LLMs, and Applications. Each plays a crucial role in maintaining the functionality and adaptability of FinGPT.

- **Data source layer:** The starting point 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 the

¹<https://ai4finance.org>. The AI4Finance Foundation is a U.S.-registered 501(c)(3) nonprofit public charity focused on promoting open scientific research in financial AI, building open-source infrastructure, and supporting a global community of researchers through shared datasets, benchmarks, and educational programs.

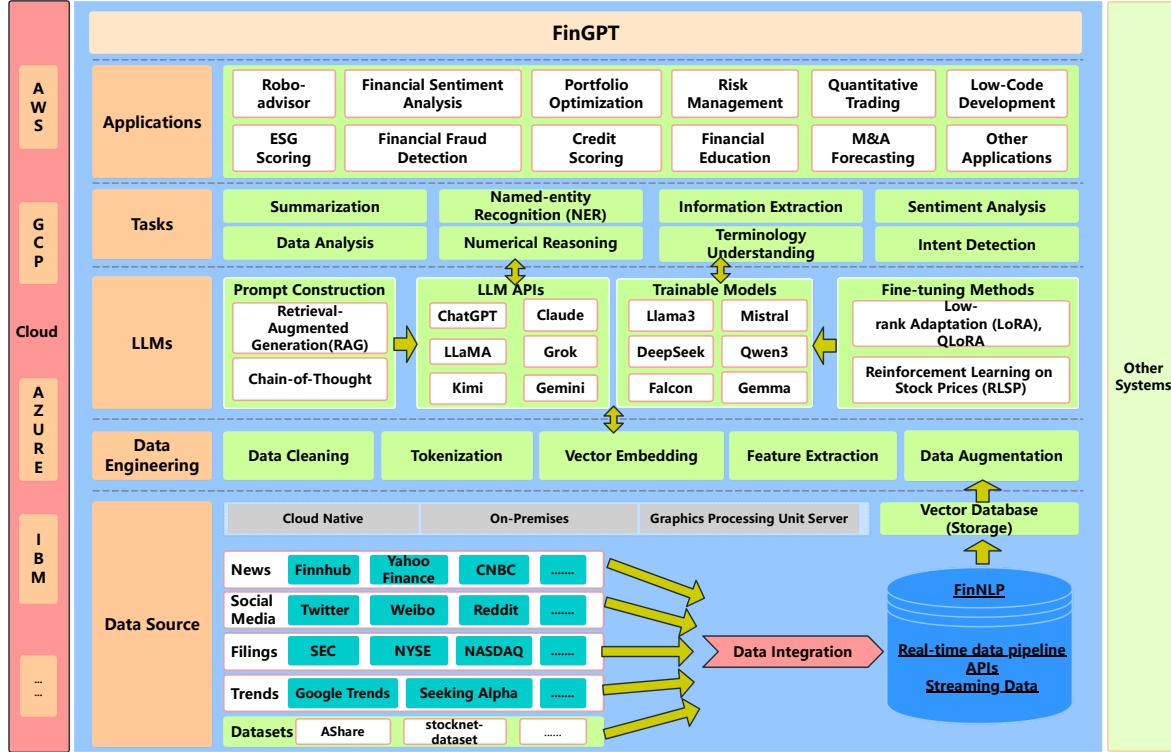


Figure 1: Overall framework of FinGPT.

nuance of the market, thereby addressing the inherent temporal sensitivity.

- **Data engineering layer:** This layer focuses on the real-time processing of text 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, prioritizing lightweight adaptation, to keep the model updated and pertinent. By maintaining an updated model, FinGPT can take care of the highly dynamic nature of financial data, ensuring its responses are in sync with the current financial climate.
- **Tasks layer:** The tasks layer is designed to provide building blocks. This layer serves a dual purpose: first, it executes a variety of fundamental tasks that are crucial in the FinLLMs landscape, such as sentiment analysis, content summarization, and numerical reasoning. Second, it establishes a standardized set of metrics and attributes. These standardized elements act not only as indicators but also as benchmarks, facilitating both performance evaluation and comparative analysis in the domain of FinLLMs.
- **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 and sentiment analysis. These practical

demos not only serve as a guide to potential users but also underscore the transformative potential of FinLLMs.

3.1 Data Sources

The first stage of FinGPT 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 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 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.

3.2 Real-Time Data Curation 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.

Vector embedding: FinGPT encodes curated financial text into dense semantic vectors using domain-adapted embedding models. The embedding process incorporates entity-aware representations (tickers, ratios, events) and temporal metadata, allowing the system to capture fine-grained financial meaning. All embeddings are indexed in a vector database for low-latency retrieval, supporting RAG, event clustering, and market-aligned RLSP training.

Feature extraction: Feature extraction involves transforming raw data into an input that can be understood by ML 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.

Data augmentation: In the dynamic landscape of financial markets, enhancing the variety and volume of training data is crucial for building robust NLP models. Data augmentation strategies will be employed to generate synthetic data that can mimic the characteristics of actual financial data.

3.3 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:** Established LLM APIs offer foundational language capabilities that serve as the base for further model development and customization.
- **Trainable models:** Users can fine-tune FinGPT’s trainable models on private data for personalized financial applications, ensuring relevance and accuracy in specific use cases.
- **Fine-tuning methods:** FinGPT supports various fine-tuning methodologies, facilitating its adaptation into personalized robo-advisors efficiently and effectively.
- **Prompt Engineering:** Prompt Engineering is crucial for optimizing input queries to LLMs, enhancing the extraction of accurate financial information. This iterative process requires careful crafting of prompts for nuanced responses, necessitating a deep understanding of both finance and language model characteristics.

Why lightweight fine-tuning LLMs for finance?

Fine-tuning or Instruction tuning of pre-existing LLMs for finance, as described in [Ouyang *et al.*, 2022], presents a cost-efficient and time-saving alternative to the expensive and lengthy process of retraining models from scratch.

BloombergGPT [Wu *et al.*, 2023], 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 [Wu *et al.*, 2023], 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 around \$300 per fine-tuning.

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 a pre-trained LLM utilizing a 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.

When the application of LLMs is envisioned for the scrutiny of financial texts and facilitation of quantitative trading strategies, it is imperative to contemplate the utilization of the intrinsic labeling mechanisms available within the financial marketplace. In light of this, FinGPT adopts the percentage of relative stock price changes corresponding to individual news articles as output labels. By assigning predetermined thresholds, these continuous labels are categorized into three discrete sentiment classes: positive, negative, and neutral.

Simultaneously, during the prompt engineering phase, the model is meticulously instructed to elect one among the three sentiment classes as its output. This meticulous approach ensures that the information gleaned during pre-training is maximally exploited, fostering the generation of insightful and reliable predictions on financial sentiment. The implementation of Low-Rank Adaptation (LoRA) for LLMs [Hu *et al.*, 2021; Dettmers *et al.*, 2023], along with its quantized variant, QLoRA [Dettmers *et al.*, 2023], significantly streamlines the model by reducing the count of trainable parameters from an overwhelming 6.17 billion to a manageable 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.

Retrieval Augmented Generation (RAG)

Retrieval-augmented generation (RAG) is a pivotal technique incorporated within FinGPT [Zhang *et al.*, 2023], as it seamlessly amalgamates the prowess of both context retrieval mechanisms and Large Language Models (LLMs) to optimize language generation tasks. This meticulous process ensures that the LLMs are not generating content in a vacuum but are rather informed and nuanced in their output, drawing from a rich tapestry of context provided by the retrieved documents. These documents, working in tandem with the input prompt, steer the LLMs effectively towards crafting responses that are not only accurate but deeply ingrained in the relevant context, thereby increasing the utility and reliability of the generated text.

3.4 Fundamental Tasks

FinGPT serves as a versatile tool in the financial sector, providing valuable assistance to both professionals and individuals by effectively filtering and analyzing information. The model excels in the following fundamental tasks:

- **Summarization:** FinGPT can efficiently condense lengthy financial documents into concise summaries, preserving the crucial information and insights. This function is invaluable for quickly understanding the essence of comprehensive reports, news articles, or financial statements without going through the entire content.
- **Named-entity recognition (NER):** The model is adept at identifying and classifying named entities within the text, such as company names, stock tickers, monetary values, and percentages. This ability is crucial for extracting specific data points from unstructured text, facilitating more structured and informed analysis.
- **Information extraction:** FinGPT can meticulously extract relevant information from various sources, providing users with valuable insights. This capability is crucial for decision-making processes, as it sifts through the noise to highlight essential data and trends.
- **Sentiment analysis:** Sentiment Analysis is pivotal as a fundamental task due to its dual application in both identifying market sentiment, namely financial sentiment analysis and being utilized within robo-advisory platforms to discern client emotions during product recommendations.
- **Data analysis:** FinGPT can process and analyze vast datasets, identifying patterns, anomalies, and significant changes in the data. This feature supports data-driven decision-making, offering a clearer understanding of market dynamics and financial performance.
- **Numerical reasoning:** The model can perform calculations and numerical analysis based on the data provided in the text, supporting users in evaluating financial metrics, making projections, and assessing risks effectively.
- **Terminology understanding:** FinGPT is proficient in understanding and interpreting complex financial terminology and jargon, making it a valuable assistant for both seasoned professionals and individuals new to the financial sector.
- **Intent detection:** The model can accurately identify the user's intent behind a query, facilitating more effective and relevant responses. This feature is particularly useful in developing intuitive and user-friendly financial advisory applications and services.

Various open-source datasets serve as benchmarks, effectively engaging in a multitude of fundamental tasks. Examples include BloombergGPT[Wu *et al.*, 2023], which utilizes a curated selection of financial datasets derived from the FLUE benchmark[Shah *et al.*, 2022]. These datasets are employed for a spectrum of essential tasks such as Sentiment Analysis and NER. Other noteworthy datasets include FinRED [Sharma *et al.*, 2022], instrumental for information extraction tasks, FINQA [Chen *et al.*, 2021] for numerical reasoning assessments, and FinRAD [Ghosh *et al.*, 2021], which is crucial for understanding and identifying financial terms.

3.5 Potential Applications

FinGPT may find wide applications in financial services, aiding professionals and individuals as a powerful information filter. The potential applications include:

- **Financial sentiment analysis:** Evaluating sentiments across different financial platforms for insightful investment guidance.
- **Robo-advisor:** The Robo-advisor function within FinLLMs plays a pivotal role in providing personalized financial advice, minimizing the necessity for continual human consultations.
- **Quantitative trading:** Producing trading signals for informed trading decisions.
- **Portfolio optimization:** Utilizing numerous economic indicators and investor profiles for optimal investment portfolio construction.
- **Credit scoring:** Predicting creditworthiness from financial data to aid lending decisions.
- **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.
- **Risk management:** Formulating effective risk strategies by analyzing various risk factors.
- **Fraud detection:** Identifying potential fraudulent transaction patterns for enhanced financial security.
- **Automating KYC Processes:** FinGPT can streamline KYC procedures by analyzing documents for identity validation, cross-checking information against databases, and detecting inconsistencies. It can also interpret complex legal documents using its NLP capabilities.
- **Enhancing Anti-Money Laundering (AML) Measures:** FinGPT can be a valuable tool in ML operations. It can be used to analyze the flow of funds, identify suspicious patterns, and highlight transactions that require further investigation.
- **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.

4 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 financial data.

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 filings and announcements, social media Discussions, and trends.

4.1 Financial News

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.

4.2 Company Filings and Announcements

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 filings contain reliable and verified data vetted by regulatory bodies.
- **Periodicity:** Company filings 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.

4.3 Social Media Discussions

Social media discussions related to finance will 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.

4.4 Trends

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.

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

5 Experiments: Financial Sentiment Analysis

In this section, we evaluate the sentiment analysis capability of FinGPT. This experiment demonstrates the effectiveness of FinGPT’s data-centric design and lightweight adaptation methodology in real-world financial text classification.

5.1 Dataset

We utilize a large-scale financial news sentiment dataset curated through the FinGPT real-time data pipeline. The dataset contains:

- Over **620,000** cleaned financial news headlines;
- Sources including *CNBC*, *Reuters*, *Yahoo Finance*, *MarketWatch*, etc., collected through the FinNLP pipeline;
- Time span from **2016–2024**;
- Market-driven labels generated using short-term price movement:

$$\text{label} = \begin{cases} \text{Positive,} & r > \theta_p \\ \text{Negative,} & r < -\theta_n \\ \text{Neutral,} & |r| \leq \theta \end{cases}$$

where r denotes the stock’s percentage price change following the news. This “self-labeled” approach aligns sentiment with true market reactions and avoids costly manual annotation.

5.2 Model and Training Setup

We adopt a lightweight two-stage adaptation process.

LoRA-based Supervised Fine-Tuning

We fine-tune the pretrained Llama-3.1-8B-Instruct model using Low-Rank Adaptation (LoRA). Under a standard configuration of rank $r = 8$ and scaling factor $\alpha = 16$, the total number of trainable parameters introduced by LoRA is approximately **8.3M**, which is well below 0.1% of the original 8B-parameter model.

The fine-tuning configuration is as follows:

- Trainable parameters: **8.3M**;
- LoRA rank: $r = 8$, scaling factor $\alpha = 16$;
- Batch size: 64;
- Learning rate: 2×10^{-4} ;
- Training epochs: 3.

LoRA enables FinGPT to acquire domain-specific sentiment classification ability efficiently.

Reinforcement Learning on Stock Prices (RLSP)

To align the model with real market behavior, we further apply RLSP, where the environment is the financial market and the reward is the stock price’s post-news reaction.

$$R = f(\Delta p)$$

This aligns the sentiment output with actual financial outcomes and enhances generalization.

5.3 Baselines

We compare FinGPT against standard financial NLP baselines:

- FinBERT [Araci, 2019];
- BloombergGPT [Wu *et al.*, 2023];
- ChatGPT (zero-shot) [OpenAI, 2023];
- Llama3.1-8B (zero-shot) [Grattafiori *et al.*, 2024].

5.4 Evaluation Metrics

We evaluate performance using the following metrics:

- Accuracy;
- Precision, Recall, and F1-score for each class;
- Macro-F1 (to mitigate class imbalance);
- AUC for binary (positive/negative) subsets.

5.5 Results

Overall Performance

FinGPT outperforms all baselines significantly, demonstrating the benefits of data-centric labeling and RLSP reinforcement alignment.

Ablation Study

LoRA performs most of the heavy lifting, while RLSP further improves market alignment.

Case Study

We illustrate the model’s financial reasoning capability using the following headline:

“Tesla cuts prices again in China as EV competition intensifies.”

- **Human Annotation:** Negative (Price reductions are commonly interpreted as a sign of weakening pricing power and intensified competitive pressure, both of which imply potential margin compression and typically induce negative investor sentiment.)
- **Base Llama3:** Neutral (The model captures the surface-level wording but fails to infer the underlying financial implications of price competition.)
- **FinGPT (SFT):** Negative
- **FinGPT (RLSP):** Negative (*with stronger alignment to the subsequent price reaction*)

This case highlights FinGPT’s ability to incorporate domain-specific financial reasoning and to produce sentiment predictions that are more consistent with market-impactful interpretations.

Table 1: Sentiment Classification Performance

Model	Acc.	Macro-F1	Pos-F1	Neg-F1	Neu-F1
ChatGPT (0-shot)	63.4	61.7	64.0	59.1	62.0
Llama3.1-8B (0-shot)	57.9	54.4	56.1	53.2	54.0
FinBERT	71.2	69.9	73.0	69.1	67.5
FinGPT (LoRA-SFT)	78.8	77.3	79.6	76.8	75.4
FinGPT (SFT+RLSP)	82.1	80.9	83.4	81.5	77.8

Table 2: Ablation on LoRA and RLSP

Configuration	Macro-F1
Base Llama3	54.4
+ LoRA SFT	77.3
+ RLSP	80.9

5.6 Discussion

Key observations:

- Market-driven labels (self-labeled data) strongly improve real-world applicability;
- LoRA reduces adaptation cost by $\sim 1000\times$ compared to full fine-tuning;
- RLSP incorporates financial market feedback, distinguishing FinGPT from traditional supervised models.

This experiment confirms that FinGPT provides a scalable and effective foundation for financial sentiment analysis.

6 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.

7 Future Work

Future development of FinLLMs will focus on establishing open, industry-level standards for financial large language models. This includes advancing parameter-efficient fine-tuning methods such as LoRA and QLoRA to support low-cost, domain-specific customization across diverse financial institutions. Furthermore, FinLLMs will continue to expand its unified data curation pipeline, promoting high-quality, standardized financial datasets to streamline training

and evaluation. By integrating open-source tooling, reproducible benchmarks, and transparent workflows, FinLLMs aims to provide a foundation for reliable, scalable, and interoperable financial AI systems.

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.

[Chen *et al.*, 2021] Zhiyu Chen, Wenhui Chen, Charese Smiley, Sameena Shah, Iana Borova, Dylan Langdon, Reema Moussa, Matt Beane, Ting-Hao Huang, Bryan Routledge, et al. Finqa: A dataset of numerical reasoning over financial data. *arXiv preprint arXiv:2109.00122*, 2021.

[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 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.

[Ghosh *et al.*, 2021] Sohom Ghosh, Shovon Sengupta, Sudip Naskar, and Sunny Kumar Singh. FinRead: A transfer learning based tool to assess readability of definitions of financial terms. In *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*, pages 658–659, National Institute of Technology Silchar, Silchar, India, December 2021. NLP Association of India (NLPAI).

[Grattafiori *et al.*, 2024] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

[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.

[Jiang *et al.*, 2023] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023.

[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.*, 2024] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

[OpenAI, 2023] OpenAI. Chatgpt. <https://chat.openai.com/>, 2023. Large language model accessed via ChatGPT interface.

[Ouyang *et al.*, 2022] 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. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

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

[Shah *et al.*, 2022] Raj Sanjay Shah, Kunal Chawla, Dheeraj Eidnani, Agam Shah, Wendi Du, Sudheer Chava, Nattraj Raman, Charese Smile, Jiaao Chen, and Diyi Yang. When flue meets flang: Benchmarks and large pre-trained language model for financial domain. *arXiv preprint arXiv:2211.00083*, 2022.

[Sharma *et al.*, 2022] Soumya Sharma, Tapas Nayak, Arusarka Bose, Ajay Kumar Meena, Koustuv Dasgupta, Niloy Ganguly, and Pawan Goyal. Finred: A dataset for relation extraction in financial domain. In *Companion Proceedings of the Web Conference 2022*, WWW ’22, page 595–597, New York, NY, USA, 2022. Association for Computing Machinery.

[Team *et al.*, 2023] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricu, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

[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, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

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

[Yang *et al.*, 2020] Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid. Deep reinforcement learning for automated stock trading: An ensemble strategy. In *Proceedings of the first ACM international conference on AI in finance*, pages 1–8, 2020.

[Yang *et al.*, 2024] Hongyang Yang, Boyu Zhang, Neng Wang, Cheng Guo, Xiaoli Zhang, Likun Lin, Junlin Wang, Tianyu Zhou, Mao Guan, Runjia Zhang, et al. Fin-robot: An open-source ai agent platform for financial applications using large language models. *arXiv preprint arXiv:2405.14767*, 2024.

[Zhang *et al.*, 2023] Boyu Zhang, Hongyang Yang, Tianyu Zhou, Ali Babar, and Xiao-Yang Liu. Enhancing financial sentiment analysis via retrieval augmented large language models. *ACM International Conference on AI in Finance (ICAIF)*, 2023.