

Large Language Models: Business Applications and Business Research

Prof. Jing Wu

jingwu@cuhk.edu.hk

About Myself

About Myself

- ▶ Jing Wu, Associate Professor in the Chinese University of Hong Kong Business School
- ▶ Education:
 - ▶ Ph.D. in Economics and Management Science: University of Chicago Booth School of Business
 - ▶ Bachelor of Electronic Engineering and Computer Science: Tsinghua University
- ▶ Research Interests:
 - ▶ Global Supply Chain
 - ▶ FinTech and Business Intelligence
 - ▶ Operations-Finance Interface
- ▶ Publications: over 10 papers in leading business journals such as Management Science, M&SOM, and POMS.
- ▶ Media Coverage: Over 400 media outlets in over 20 countries worldwide (*MIT Sloan Management Review*, *the Economist*, *Forbes*...)
- ▶ Website: <https://www.jingwulab.org/>

Media Coverage (400+ media, 20+ countries)

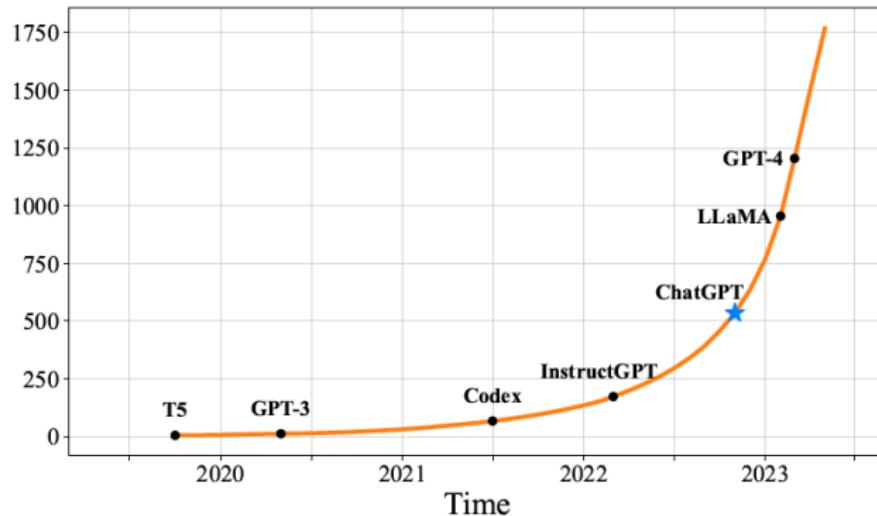
- ▶ *Barron's*, "The dirty secret behind companies' carbon-emissions disclosures", July 17, 2023
- ▶ *CCTV*, "Shanghai Supply Chains Battle COVID-19", April 16, 2022
- ▶ *WSJ*, "Amazon is proof of a major weakness in ESG ratings", Nov 11, 2021
- ▶ *BBC*, "From a rare negative growth in Vietnam to a surge in Chinese orders: the reconstruction of trade in the epidemic era", Oct 22, 2021.
- ▶ *Chicago Booth Review*, "Global supply chains can hurt a company's credit", August 11, 2021
- ▶ *South China Morning Post*, "How to spot and avoid greenwashing in supply chains", June 12, 2021
- ▶ *Sloan Management Review*, "The Right Way to Mix and Match Your Customers", Summer 2021
- ▶ *Economist*, "A wave of supply-chain reshoring around the corner", Dec, 2020
- ▶ *Reuters*, "Decoupling from China does not help companies mitigate risks.", Oct 29, 2020
- ▶ *China Business Knowledge*, "The Chinese Economic Impact of COVID-19 and The Future of Manufacturing Chains", May 7, 2020
- ▶ *Forbes*, "Financial Supply Chain in the COVID-19 Pandemic: Fuel or Wildfire?", Apr 30, 2020
- ▶ *VOA*, "Made in Vietnam' Goods Increasingly Popular in US", March 26, 2020.

About This Talk

- ▶ Let's discuss **Large Language Model (LLM)** and its applications in business and business research.
- ▶ Cross-disciplinary Innovation: it can process and generate human-like text based on the vast amounts of data they have been trained on.
 - ▶ **LLM Applications in FinTech:** financial reasoning, smart contract auditing, and intelligent advisory services.
 - ▶ **LLMs Applications in Business Process:** customer support, collaborative work, product innovation, price forecast.
- ▶ Exploring LLM has the potential to generate high-performing models, redefining business models and decision frameworks.

LLM Development

- ▶ ChatGPT: Trained on the largest textual material to date.

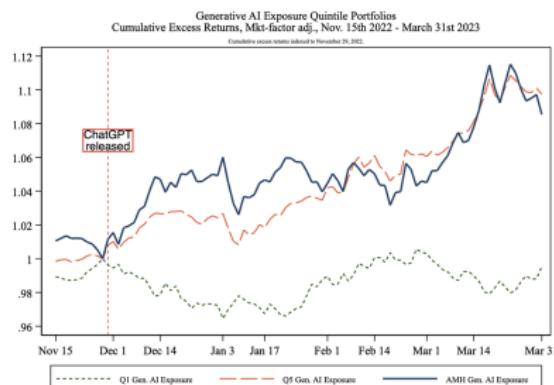


Source: Zhao et al. (2023)

Figure: The trends of the cumulative numbers of arXiv papers that contain the keyphrases “large language model”

LLM and Firm Value

- ▶ NBER working paper: with higher exposure of generative AI earned excess returns that are 0.4% higher on a daily basis than returns of firms with lower exposures following the release of ChatGPT.



Source: Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). Generative ai and firm values. National Bureau of Economic Research working paper.

BloombergGPT¹: First LLM for Finance

- ▶ FinPile (private dataset)
 - ▶ Financial domain web pages
 - ▶ Renowned news sources in finance
 - ▶ Corporate finance reports
 - ▶ Publications from finance-related companies
 - ▶ Bloomberg analysts
- ▶ Costs
 - ▶ 1.3 million GPU hours on an A100 GPU of 40G GPU memory
(actually, 512 A100 of 40G * 53 days)
 - ▶ \$ 2.67 million of training process costs
- ▶ Summary: The high computational cost and scarcity of data can make it challenging for individuals or organizations to develop their domain-specific LLMs like BloombergGPT from ZERO ground.

¹Wu, Shijie, et al. "Bloomberggpt: A large language model for finance." arXiv preprint arXiv:2303.17564 (2023).

Alternative: Two Efficient Approaches to Using LLMs

- ▶ Method 1: Fine-tuning Domain-specific LLMs
 - ▶ A few general-purpose LLMs, maintained by specialized organizations, act as foundations.
 - ▶ Other professionals, with domain-specific knowledge/data, further fine-tune LLMs for domain-specific tasks.
 - ▶ SFT (Supervised Fine-Tuning)²: Re-training all layers of LLM (next token prediction on domain datasets).
 - ▶ LoRA³: Parameter-efficient method by freezing a pre-trained LLM and only updating replacement matrices (1/100 parameters).
- ▶ Method 2: Using LLMs as “Brain” with External Tools
 - ▶ LLMs interacting with external tools (search engine, solver, visualization...) for diverse tasks.
 - ▶ Toolformer⁴: A LLM trained to learn to use tools.
- ▶ Summary: Both methods serve to deploy efficient LLMs in real-world applications, with different focuses (**long-tail vs. standardized**).

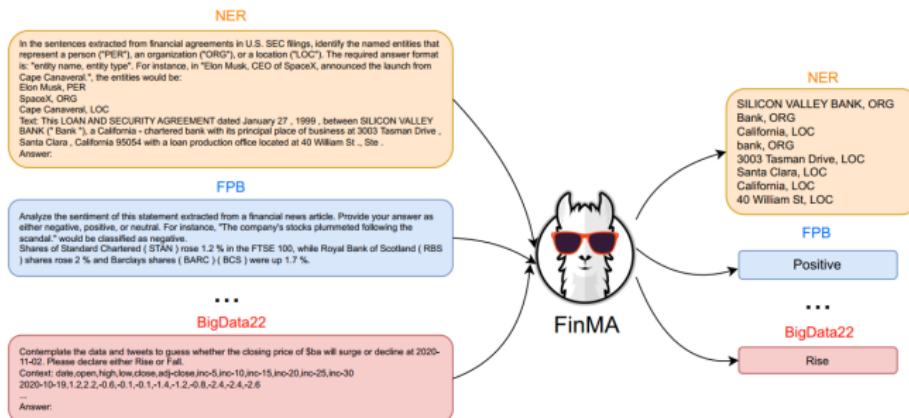
²Ouyang, Long, et al. "Training language models to follow instructions with human feedback." Advances in Neural Information Processing Systems 35 (2022).

³Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." arXiv preprint arXiv:2106.09685 (2021).

⁴Schick, Timo, et al. "Toolformer: Language models can teach themselves to use tools." arXiv preprint arXiv:2302.04761 (2023).

Financial LLM with SFT Method: FinMA⁶

- ▶ Motivation: Lack of publicly available financial LLMs.
- ▶ Methods
 - ▶ Create a large-scale **multi-task instruction dataset** for diverse financial tasks, document types, and data modalities.
 - ▶ Full fine-tuning **LLaMA⁵** with the constructed dataset to FinMA.

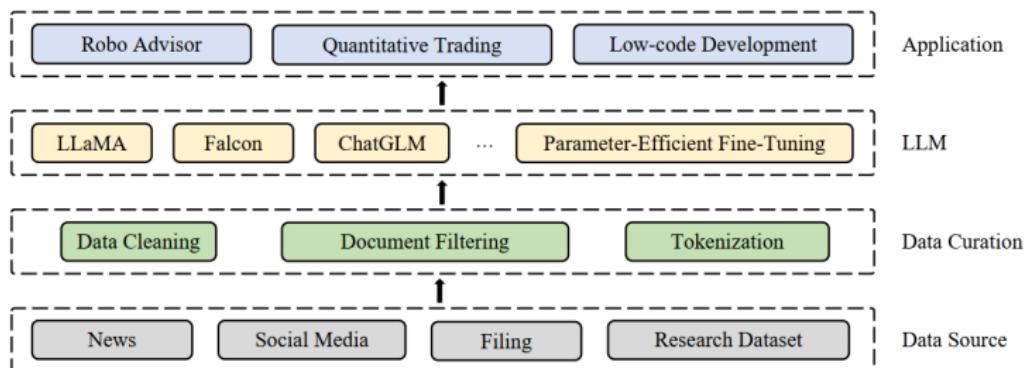


⁵Touvron, Hugo, et al. "Llama: Open and efficient foundation language models." arXiv preprint arXiv:2302.13971 (2023).

⁶Xie, Qianqian, et al. "PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance." arXiv preprint arXiv:2306.05443 (2023).

Financial LLM with LoRA Method: FinGPT⁷

- ▶ Motivation: Financial LLM for frequent data update.
- ▶ Methods
 - ▶ Collection and curation of real-time financial data from ≥ 34 sources.
 - ▶ Low-rank Adaptation (LoRA) beats SFT in speed and cost.
- ▶ Summary: FinGPT democratizes FinLLMs, enables its application in robo-advisory and algorithmic trading.



⁷Liu, Xiao-Yang, Guoxuan Wang, and Daochen Zha. "FinGPT: Democratizing Internet-scale Data for Financial Large Language Models." arXiv preprint arXiv:2307.10485 (2023).

LLM Applications in FinTech

Financial Reasoning⁸

- ▶ Motivation
 - ▶ Financial reasoning is very important for investment decisions.
 - ▶ Despite demonstrated abilities in complex tasks, the use of LLMs in financial reasoning remains uncharted, e.g., to generate logically coherent and persuasive investment opinions.
 - ▶ Apply LLMs to generate financial reasoning and provide a domain dataset for LLM training?
- ▶ Methods
 - ▶ Generate synthetic dataset by GPT-3.5.
 - ▶ Build in-context prompts to test various GPT variants with parameter scales ranging from 2.8B to 13B.
- ▶ Summary: Coherent financial reasoning first emerges at 6B parameters, and in-context prompts can improve the performance.

Base Model	Instruction-Tuned	Param.
Pythia	dolly-v2-3b	2.8B
GPT-J	dolly-v1-6b	6B
Galactica	galpaca-6.7b	6.7B
LLama	vicuna-13b-delta-v1.1	13B

⁸Son, Guijin, et al. "Beyond Classification: Financial Reasoning in State-of-the-Art Language Models." arXiv preprint arXiv:2305.01505 (2023).

Financial Reasoning*

- ▶ Synthetic dataset
 - ▶ 1. Collect 1,087 analyst reports, written by experts from various institutions, covering 752 U.S. stock market companies.
 - ▶ 2. From each report, extract "Investment Thesis" and "Related Risk" sections, creating a set of expert-written investment theses.
 - ▶ 3. Construct Full-Text type inputs by collating the abstracts from each analyst report.
 - ▶ 4. Q&A type input question generation by feeding the Full-Text type inputs into the system and prompting it to generate questions that targeted important information.
 - ▶ 5. The same API was used to produce dummy answers for the questions generated in step 4. We then employed human annotators to remove any answers that significantly deviated from reality.
 - ▶ 6. Lastly, use the GPT3.5-Turbo API to generate investment opinions for both input types.

Financial Reasoning

► In-context Prompt

- Input: prompt template (left top)
- In-context: a list of Q&A pairs (left bottom)
- Output: generated example (right)

In our experiment, we use the following template for generations.

```
prompt = f"Assume you are a professional financial analyst. Read the provided question and answer pair about {company} and write an investment thesis be logical and argumentative.\nQA: {QApair} Please write in English language. \n### Investment Thesis:"
```

Q1: What are the primary business segments of Nvidia Corporation?

A1: Nvidia operates through two segments: GPU and Tegra Processor.

Q2: How does the Tegra brand contribute to Nvidia's portfolio?

A2: The Tegra brand integrates an entire computer onto a single chip, and incorporates GPUs and multi-core CPUs to drive supercomputing for mobile gaming and entertainment devices, as well as autonomous robots, drones, and cars.

Q3: What markets does Nvidia address with its processor platforms?

A3: Nvidia's processor platforms address four markets: Gaming, Professional Visualization, Datacenter, and Automotive.

Q4: What is the predicted earnings improvement for Nvidia in 2021 and 2022?

A4: Nvidia earnings are expected to improve to \$9.50 a share in 2021, and anticipate a +15% increase into 2022.

Nvidia is a leading player in the semiconductor and computer technology industry, with a strong presence in the gaming, professional visualization, datacenter, and automotive markets. The company's primary business segments are GPU and Tegra Processor, which are designed to address the growing demand for data and cloud computing, as well as the increasing popularity of gaming and entertainment devices.

The acquisition of Mellanox is expected to contribute to Nvidia's long-term growth, despite having a negative impact on near-term margin numbers. The company's partnerships in machine learning and self-driving also hold significant growth potential, as these technologies are expected to become consumer worthy in the coming decade.

Nvidia is committed to producing energy-efficient chipsets and has set a goal to source 65% of its global power usage from renewable sources by 2025, making a positive contribution to the fight against climate change. The company's corporate governance is considered disciplined and transparent, with a diverse board of directors and few flags in accounting or reporting practices.

Overall, Nvidia is well-positioned to benefit from the growth in demand for data and cloud computing, as well as the increasing popularity of gaming and entertainment devices. The company's acquisition of Mellanox and partnerships in machine learning and self-driving are expected to drive long-term growth, and the company's commitment to sustainability is a positive factor. Given these factors, we believe that Nvidia is a strong investment opportunity, with a potential for a high ROI.

Smart Contract Audit⁹

- ▶ Motivation
 - ▶ Current smart contracts audit is time-consuming and costly.
 - ▶ Explore the possibility of employing LLM for this task.
- ▶ Methods
 - ▶ **Prompt engineering** for enhanced security analysis.
 - ▶ Collect a dataset of 52 DeFi smart contracts that have previously been compromised.
- ▶ Summary: LLM show promise in automating the process of smart contract security audits. Although false positives remain an issue, there is significant potential for further enhancements.

⁹David, Isaac, et al. "Do you still need a manual smart contract audit?." arXiv preprint arXiv:2306.12338 (2023).

Smart Contract Audit

- ▶ Prompt engineering

You are an AI smart contract auditor that excels at finding vulnerabilities in blockchain smart contracts. Review the following smart contract code in detail and very thoroughly. Think step by step, carefully. Is the following smart contract vulnerable to '{vulnerability_type}' attacks? Reply with YES or NO only. Do not be verbose. Think carefully but only answer with YES or NO! To help you, find here a definition of a '{vulnerability_type}' attack: {
vulnerability_description}\n\n

Source code: {source_code}

Smart Contract Audit

► 52 DeFi smart contract datasets

Address	Attack Type Ground Truth
0xe952..1659	Reentrancy, Token standard incompatibility
0x55dB..3830	On-chain oracle manipulation, Absence of code logic or sanity check
0x4e3F..9D2B	Governance attack
0x833e..743D	On-chain oracle manipulation
0x74BC..1fD6	Absence of code logic or sanity check
0x0624..5286	On-chain oracle manipulation
0x7B3B..c6Ca	Visibility errors, including unrestricted action
0x3212..8923	Token standard incompatibility, Reentrancy
0x5bD6..deE4	Unfair slippage protection, Absence of code logic or sanity check
0x32e5..de43	Flash liquidity borrow, purchase, mint or deposit
0x39b1..Db15	On-chain oracle manipulation, Frontrunning
0x17e8..Ca5f	Deployment mistake, Other unsafe DeFi protocol dependency
0x35C6..7810	Flash liquidity borrow, purchase, mint or deposit, Absence of code logic or sanity check
0x66e7..eE63	Absence of code logic or sanity check
0x55dB..3830	On-chain oracle manipulation, Absence of code logic or sanity check
0xDdD7..1101	On-chain oracle manipulation
0xC9f2..14ef	Reentrancy, Visibility errors, including unrestricted action, Delegatecall injection
0x0eEe..F5ea	Token standard incompatibility, Reentrancy
0x88Cc..8b17	Absence of code logic or sanity check
0xae46..c8cF	On-chain oracle manipulation
0x818E..B755	On-chain oracle manipulation
0xAcbD..c747	Flash liquidity borrow, purchase, mint or deposit
0x6b7a..1522	Unsafe call to phantom function, Absence of code logic or sanity check
0xA231..7eF5	Direct call to untrusted contract, Insider trade or other activities

Financial Advisor: Explainable Time Series Forecasting¹⁰

- ▶ Motivation
 - ▶ Financial time series challenges, e.g., cross-sequence reasoning, incorporating multi-modal signals, and model interpretation.
- ▶ Methods
 - ▶ Dataset: NASDAQ-100 stocks, leveraging publicly accessible data and historical economic / financial news.
 - ▶ Prompt of few-shot inference with GPT-4.
- ▶ Summary
 - ▶ The LLM approach outperforms several baselines, including ARMA-GARCH and gradient-boosting tree models.
 - ▶ LLMs effectively reason over information from both textual news and price time series, extracting insights, leveraging cross-sequence information, and utilizing inherent knowledge.

¹⁰Yu, Xinli, et al. "Temporal Data Meets LLM–Explainable Financial Time Series Forecasting." arXiv preprint arXiv:2306.11025 (2023).

Financial Advisor: Explainable Time Series Forecasting

- ▶ An example of a stock's company profile consisting of the company description, the general positive/negative factors impacting the company's stock prices.

Prompt: Generate a short description for stock AAPL's company. Also list general positive and negative factors that might impact the stock price; be brief and use keywords. Consider diverse general factors, such as macro economic situation (e.g. inflation, CPI growth), business factors (e.g. sales, investment, products), technology factors (e.g. innovation), and others. Use format Description: ..., Positive Factors: ..., Negative factors: ...

GPT-4 Response:

Description:

AAPL (Apple Inc.) is a world-renowned technology company based in Cupertino, California. Known for its sleek designs and innovative technologies, the company produces a wide range of products including the iPhone, iPad, Mac, Apple Watch, and services like iCloud, App Store, and more.

Positive Factors:

Macroeconomic: Economic growth, increased consumer spending

Business: Strong sales, expansion into new markets, positive earnings reports, strategic acquisitions

Technology: Innovation, new product launches, advancements in artificial intelligence and augmented reality

Brand: Strong global brand, loyal customer base

Legal/Regulatory: Favorable policy changes, successful patent applications

Negative Factors:

Macroeconomic: Inflation, economic downturn, trade restrictions

Business: Decreased sales, poor earnings reports, lack of new investments

Technology: Slow pace of innovation, delays in product launches

Brand: Damage to brand reputation, loss of customer loyalty

Legal/Regulatory: Antitrust investigations, unfavorable policy changes, patent disputes

Environmental: Failure to meet sustainability goals, negative environmental impact

Financial Advisor: Explainable Time Series Forecasting

- ▶ An example of news summary and keywords extracted from one news article for a stock's company.

Prompt [template, those inside {} are placeholders]: Please summarize the following noisy but possible news data extracted from web page HTML, and extract keywords of the news. The news text can be very noisy due to it is HTML extraction. Give formatted answer such as Summary: ..., Keywords: ... The news is supposed to be for {symbol} stock. You may put 'N/A' if the noisy text does not have relevant information to extract.

News: {news}

GPT-4 Response:

Summary: Apple Inc is postponing perks for some corporate divisions and expanding a cost-cutting effort, joining its peers in streamlining operations. The company proposes to forego bonuses or promotions for April, with the payments occurring only in October. Apple is also restricting hiring and leaving additional positions open after employee departure. This is triggering fear of downsizing among employees.

Keywords: Apple, bonus, hiring, cost-cutting, promotions, downsizing, employees

- ▶ An example of one week's meta summary and keywords condensed from all the company's summaries and keywords from the week.

Summary: Apple Inc.'s stock is displaying rising relative strength, although shy of a key benchmark, and the company has been highlighted as a top-performing stock due to its sales of 1.3 billion iPhones and a significant \$19.4 billion profit last quarter. As the NASDAQ surges with Apple as a big cap, Apple expands cost-cutting measures, postpones perks, restricts hiring, and leaves positions open, instigating downsizing concerns. The overall stock market performance is boosted by a \$30 billion deposit infusion for struggling firms.

Keywords: Apple Inc., stock, relative strength, cost-cutting, downsizing, NASDAQ, \$30 billion deposit infusion, iPhones, profit.

Financial Advisor: Explainable Time Series Forecasting

- We notice the cross-sequence information and macro-economy information are obviously considered in the LLM's reasoning.

Prompt [template, those inside [] are placeholders]:

Instruction: Forecast next week stock return (price change) for symbol, given the company profile, historical weekly news summary, keywords, and stock returns, and optionally the examples from other stocks of a similar company.

The trend is represented by bins "D5+", "D5", "D4", "D3", "D2", "D1", "U1", "U2", "U3", "U4", "U5", "U5+", where "D5+" means price dropping more than 5%, D5 means price dropping between 4% and 5%, "D4" means price dropping between 3% and 4%, "U5+" means price rising more than 5%, "U5" means price rising between 4% and 5%, "D4" means price rising between 3% and 4%, etc.

Company Profile: [company_profile] [company description, general positive/negative factors shown in Figure 1.]

Recent News: News are ordered from oldest news to latest news.

8 Weeks ago.

[meta_news_summary_company] [meta news summary and keywords of the week for the company]

[meta_news_summary_macro] [meta news summary and keywords for the macro economy and finance status of the week]

—

7 weeks ago.

—

Last week.

Forecasting Examples: [few_shot_learning_examples_from_similar_stocks] [We use forecasting examples from top-3 similar stocks as the few-shot learning examples. The example is structured similar to this prompt, with company profile and recent news, but without the instruction texts.]

Now predict what could be the next week's Summary, Keywords, and forecast the Stock Return. The predicted Summary/Keywords should explain the stock return forecasting. You should predict what could happen next week. Do not just summarize the history. The next week stock return need not be the same as the previous week. Use format Summary: ..., Keywords: ..., Stock Return: ...

GPT-4 Response [Forecasting week 05/01/2023 to 05/07/2023]:

Summary: Apple's stock is expected to see fluctuations in the week to come against a backdrop of mixed news. The company's announced expansion into India, coupled with the introduction of high-yield savings accounts, initially provided a boost to the stock. However, growing concerns surrounding reduced PC shipments and apprehensions about the company's upcoming earnings injected a degree of uncertainty. Despite these challenges, Apple's unwavering focus on cost control and the anticipation surrounding innovative offerings, such as the forthcoming mixed-reality headset, shored up investor confidence. Consequently, the stock maintained a general upward trend throughout the week, albeit punctuated by a few dips. The broader economic climate on Wall Street also had a role to play, with a recovering sentiment in Q1 towards high-tech companies like Microsoft and Alphabet. These companies released pivotal financial reports and ground-breaking products during the period, influencing the tech sector's performance, potentially impacting Apple's stock trajectory.

Keywords: AAPL, mixed-reality headset, India expansion, high-yield savings account, cost control strategies, Bernstein, Wall Street, economic recovery, tech sector performance

Stock Return: U1

Ground Truth [Summary from actual news from 05/01/2023 to 05/07/2023 and the actual return]:

Summary: Apple's stock saw significant gains after outperforming Wall Street's fiscal second quarter targets, driven by robust iPhone sales and strong services, despite posting a decrease in revenue for the second consecutive quarter and facing potential competition in the VR headset market from Reliance Jio's JioDive. The company also noted an improvement in supply and strength in emerging markets, further boosting investor confidence.

Keywords: Apple, stock, Wall Street, fiscal second quarter, iPhone, services sales, quarterly earnings, revenue, declining revenue, emerging markets, improved supply, Virtual Reality, competition, JioDive, Reliance Jio.

Stock Return: U3

Financial Advisor: Explainable Time Series Forecasting

- GPT-4 outputs detailed reasoning steps if we add an instruction "Can you reason step by step before the finalized output?".

Prompt: {The same prompt as in Figure 4.}
Can you reason step by step before the finalized output?

GPT-4 Response [Forecasting week 05/01/2023 to 05/07/2023]:

Given the information provided, we can see that Apple (AAPL) has had a number of positive developments and a few potential concerns in the recent past.

Positive Developments:

- Recent sales and earnings have been strong, and the company appears to be exploring new markets such as India.
- The upcoming mixed-reality headset and the "Apple Pay Later" service show that Apple is still innovating and pushing into new product areas.
- The cost-cutting measures and the plan to use 100% recycled cobalt show that Apple is taking steps to improve its efficiency and sustainability, which could be attractive to certain investors.
- Winning the lawsuit against Britain's antitrust regulator and avoiding potential penalties.
- AAPL stock has been trending upwards recently, with multiple weeks of upward movement (U5, U4, U3, U2), and only a single day of slight downward movement (D1).

Potential Concerns:

- Apple is cutting costs, delaying bonuses, and slowing down hiring in certain areas, which could potentially affect employee morale and productivity.
- The negative sentiment from analyst Toni Sacconaghi about the overestimation of Q3 and Q4 revenues by Wall Street could potentially impact the stock price.

In addition to above factors, the general direction of the stock market appears to be on an upward trajectory, as evidenced by the recent increases in the Dow Jones, S&P 500, and Nasdaq. In recent weeks, the stock performances of high-tech giants such as Microsoft and Alphabet have also shown considerable strength. Furthermore, recent positive developments in the macroeconomic environment, including a surge in consumer spending, seem encouraging.

Now, we can generate a prediction for next week's Summary, Keywords, and Stock Return. However, it's important to note that while the information provided is used as a basis, this is a prediction and the actual outcome may be different.

Summary: Apple's mixed-reality headset previews have received positive feedback from industry experts, increasing investor confidence. However, apprehension about cost-cutting measures and fears of slower hiring might add some uncertainty. Wall Street anticipates a strong earnings report, boosting stock morale. Furthermore, Apple's expansion into India continues to show promise.

Keywords: Apple, mixed-reality headset, investor confidence, cost-cutting, hiring, Wall Street, earnings report, expansion, India.

Stock Return: U2. Given the positive trends and promising developments, the stock is likely to experience upward movement. However, cost-cutting concerns may moderate gains, leading to a prediction of "U2", suggesting a rise of 1% 2% next week.

Financial Advisor: Explainable Time Series Forecasting

► Model Performance Comparison

	Weekly			Monthly (Every 4 Weeks)		
	Binary Precision	Bin Precision	MSE	Binary Precision	Bin Precision	MSE
Most-Frequent Historical Bin	50.7%	16.4%	43.5	51.4%	17.2%	155.1
ARMA-GARCH	52.4%	11.1%	22.1	50.5%	6.2%	90.1
Gradient Boosting Tree Model	60.8%	26.4%	24.3	56.4%	17.7%	85.6
GPT-4 Zero-Shot	64.5%	31.2%	20.5	64.8%	26.0%	60.1
GPT-4 Few-Shot	65.8%	32.7%	20.6	65.3%	26.5%	58.2
GPT-4 Few-Shot w/ COT	66.5%	35.2%	18.7	69.5%	28.6%	50.4
Open LLaMA (13B) Fine-Tuned	62.2%	26.5%	23.3	60.1%	22.6%	63.3
Open LLaMA (13B) Fine-Tuned w/ COT	64.7%	30.7%	21.0	62.2%	24.4%	63.5

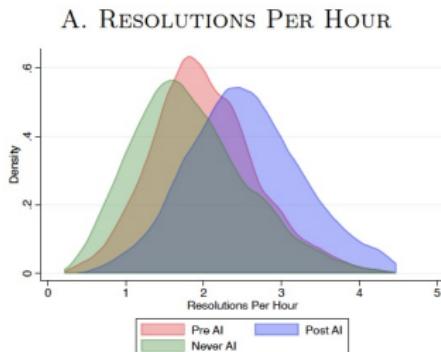
Table 1: Performance comparison between the baseline models and LLMs for stock price weekly/monthly forecasting.

How LLMs Benefit Business Process

Customer Support Efficiency¹¹

- ▶ The research utilizes data from 5,179 customer support agents and studies the staggered introduction of the conversational LLM tool.
- ▶ Summary
 - ▶ Introduction of the AI tool increases productivity by 14% on average.
 - ▶ Greater impact on novice and low-skilled workers, minimal impact on experienced and highly skilled workers.
 - ▶ AI assistance improves customer sentiment, reduces requests for managerial intervention, and improves employee retention.

FIGURE 3: RAW PRODUCTIVITY DISTRIBUTIONS, BY AI TREATMENT



¹¹Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. Generative AI at work. No. w31161. National Bureau of Economic Research, 2023.

Location Decision based on Geographic Knowledge¹²

- ▶ Motivation
 - ▶ This study investigates how GPT-4 has acquired factual geographic knowledge and its ability to use this knowledge for interpretative reasoning, and whether it can help geospatial analysis, supply chain management, and disaster response.
- ▶ Methods
 - ▶ It examines a range from factual tasks (location, distance, and elevation estimation) to complex tasks (generating country outlines and travel networks, route finding under constraints, and supply chain analysis).
- ▶ Summary
 - ▶ The study demonstrates that GPT-4 has impressive capabilities and potential in applications involving geographic data.

¹²Roberts, Jonathan, et al. "GPT4GEO: How a Language Model Sees the World's Geography." arXiv preprint arXiv:2306.00020 (2023).

Idea Creation and Product Innovation¹³

- ▶ Lab experiments at Wharton, Cornell, and INSEAD courses.
- ▶ GPT-4 can generate creative ideas much faster, cheaper, and better than students, the ideas are on average of higher quality (as measured by purchase-intent) and exhibit higher variance in quality.
 - ▶ Compact Printer, Solar-Powered Charger, QuickClean Vacuum...

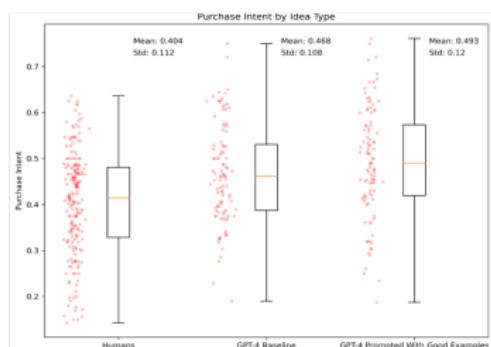


Figure 1 - Distribution of idea quality for three sets of ideas. Purchase intent is the weighted average of the five-box response scale per Jameson and Bass (1989).

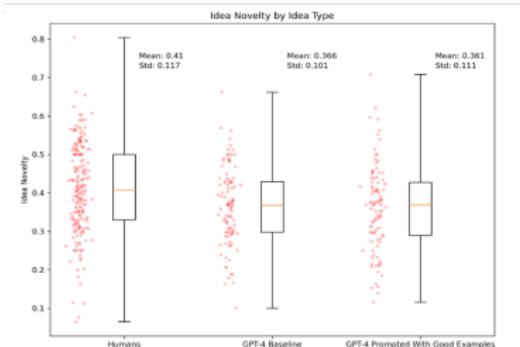


Figure 2 - Distribution of novelty ratings for three samples of ideas. Novelty based on mTurk assessment per Kwon, Kim, and Lee (2009).

¹³Girotra, K., Meincke, L., Terwiesch, C. and Ulrich, K.T. Ideas are dimes a dozen: Large language models for idea generation in innovation. Available at SSRN 4526071.

My Paper using GPT

“The Bullwhip Effect in Servitized Manufacturers”

- ▶ *Management Science*, forthcoming
https://papers.ssrn.com/abstract_id=3988802
- ▶ Empirical question:
 - ▶ To construct a measure on manufacturer servitization.
 - ▶ Apple offers complementary repair service.
 - ▶ Ford offers substitutable rental service.
 - ▶ To examine the impact of manufacturer servitization on demand variability and the intra-firm bullwhip.
- ▶ Manufacturers' service offerings reduce the bullwhip in 2 steps:
 - ▶ Basic complementing services lead to decreased demand variability.
 - ▶ Advanced substituting services reduce the intra-firm bullwhip.

Servitization Measurement by GPT4

- ▶ Call GPT-4 API to process 10-K samples with 742,582 tokens.
- ▶ Refined zero-shot prompts through iterative adjustments:

```
system_prompt = """  
You are a researcher in operations management and you need to identify servitization in company's annual reports.  
Consider 13 types of manufacturers' services, including  
"Trading and retail services", "Logistic and procurement services", "Parts and technical support", "Maintenance and repair services", "Certification and testing services", "R&D services", "Consultancy and training services",  
"Contracting Operations Services", "Leasing and performance-based services", "Renewal and upgrade services",  
"Installation and implementation services", "Integrated solutions", "Value-added digital and streaming service".  
"""  
  
user_context = """  
Given a company's annual report, determine whether it contains 13 types of services manufacturing firms offer.  
You should first retrieve sentences or sub-sentences related to a service provided by the company.  
If there is a sentence or sub-sentence related to any service provided by the company, that service is included and the indicator is 1. Otherwise, the indicator is 0. While performing the task think step-by-step and justify your steps. Answer in the format {"index": index, "service_type": service_name, "sentences": [sentences or sub-sentences], "indicator": 1 or 0}  
without any explanation. The final result is in json format. Note that json is output directly without any other texts,  
as is shown in following:  
[  
{"index": 1, "service_type": "Trading and retail services", "sentences": [sentences or sub-sentences], "indicator": 1},  
 {"index": 2, "service_type": "Logistic and procurement services", "sentences": [], "indicator": 0},  
 {"index": 3, "service_type": "Parts and technical support", "sentences": [sentences or sub-sentences], "indicator": 1},  
 {"index": 4, "service_type": "Maintenance and repair services", "sentences": [], "indicator": 0},  
 {"index": 5, "service_type": "Certification and testing services", "sentences": [], "indicator": 0},  
 {"index": 6, "service_type": "R&D services", "sentences": [sentences or sub-sentences], "indicator": 1},  
 {"index": 7, "service_type": "Consultancy and training services in basic uses", "sentences": [], "indicator": 0},  
 {"index": 8, "service_type": "Contracting Operations Services", "sentences": [], "indicator": 0},  
 {"index": 9, "service_type": "Leasing and performance-based services", "sentences": [], "indicator": 0},  
 {"index": 10, "service_type": "Renewal and upgrade services", "sentences": [], "indicator": 0},  
 {"index": 11, "service_type": "Installation and implementation services", "sentences": [], "indicator": 0},  
 {"index": 12, "service_type": "Integrated solutions", "sentences": [sentences or sub-sentences], "indicator": 1},  
 {"index": 13, "service_type": "Value-added digital and streaming service", "sentences": [sentences or sub-sentences], "indicator": 0}  
]"""  
  


Declaring contexts and tasks



Trigger chain-of-thoughts


```

Servitization Measurement by GPT4

- ▶ Performance Evaluation:
 - ▶ Precision, Recall and F_1 scores all above 0.7.
 - ▶ Surpassed the performance benchmark (0.68) from a recent CS conference paper [David et al. 2023](#).

Table OA10: Evaluation of GPT Classification of Two Service Categories

	Precision	Recall	$F1$ Score
Micro Average Value	0.75	0.86	0.80
Macro Average Value	0.73	0.83	0.78
Weighted Average Value	0.74	0.86	0.79

Note. *Precision* is the number of truly positive over all positive predictions. *Recall* is the number of truly positive over the sum of truly positive and false negative predictions. *$F1$ score* is the harmonic mean of precision and recall. Micro average metrics are precisions/recalls/f1 scores when regarding all samples as one sample. Macro average metrics are the arithmetic mean of precisions/recalls/f1 scores of all samples. Weighted average metrics are the weighted mean of precisions/recalls/f1 scores with weights equal to the proportion of label 1 in a sample relative to the total count of label 1 across all samples. A value approaching 1 signifies heightened measurement precision.

My Ongoing Projects with LLM

Machine Psychology: “Biases in LLM Investment”

- ▶ Generative AI is fundamentally different from decision-making AI, as the training data has personality and emotions.
- ▶ Research questions:
 - ▶ Does LLM investment decision-making have cognitive biases such as disposition effect, herding effect, and overconfidence?
 - ▶ Can we prove that such biases originate from human language rather than random components in experimental design?
 - ▶ Can SFT methods systematically correct such biases?
- ▶ Experimental design:
 - ▶ Based on market data of CSI 500 stocks, let LLM make trading decisions for the next stage, and measure various behavioral biases.
 - ▶ Control group: random buying and selling, momentum reversal, ARMA-GARCH, machine learning strategy...
- ▶ Preliminary findings:
 - ▶ LLM can effectively outperforms simple trading strategies.
 - ▶ However it does show behavioral biases, which can be reduced by improving professionalism, setting stop losses, and reference control.
 - ▶ Opportunity: **LLM quantitative investment strategy, LLM alternative Alpha factor.**

High-quality Text Generation with LLM

- ▶ "Uncovering Unseen Biases: Can LLM Correct Idiosyncratic Relationships Bias in Analyst Recommendations?"
- ▶ EPS forecast for institutions vs. stock advice for retail investors
- ▶ Possible reasons for idiosyncratic bias:
 - ▶ Buy side's fund holdings, served by the brokerage firms
 - ▶ Geographical distance between the company and the brokerage firm?
 - ▶ The impact of key customer relationships on recommendations
- ▶ Research questions:
 - ▶ Can LLM identify idiosyncratic biases in each analyst report?
 - ▶ Can the fine-tuned LLM correct such biases?
- ▶ Preliminary findings:
 - ▶ LLM can proactively identify idiosyncratic biases and provide correction suggestions
 - ▶ LLM can generate research report content and provide more consistent and accurate investment recommendation.
 - ▶ Opportunity: **LLM robo-advisor service for retail investors.**

Supplemental

Traits of LLMs¹⁴

- ▶ Bias in Generative AI: LLMs' gender and race stereotypes present mostly white males.

Prompt: A Portrait of Personal Financial Advisors



Prompt: A Portrait of Chief Executives



Prompt: A Portrait of Dishwashers



Prompt: A Portrait of Anesthesiologist Assistants



Fig. 1. Examples of Images Generated by Midjourney with Different Prompts.

¹⁴Zhou, M., Abhishek, V. and Srinivasan, K., Bias in Generative AI, 2023.

Impact of LLMs on Collective Opinion ¹⁵

- ▶ The output opinion of LLMs has a unique and positive effect on the collective opinion difference.
- ▶ When people partially rely on LLMs, the exchange process of opinion becomes more intense and the diversity of opinion becomes more favorable.

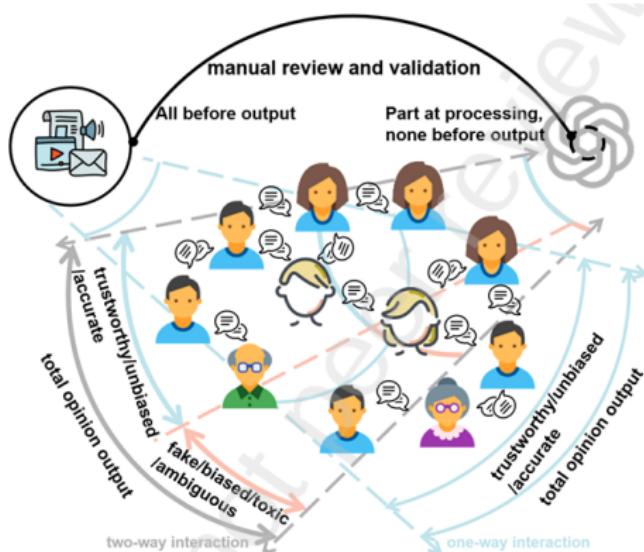


Fig.1. Schematic diagram of the difference between LLMs and traditional media. The left side represents the pattern of opinion dissemination in the interactions between traditional media and people, the right side represents the pattern of opinion dissemination in the interactions between LLMs and people, and the center part represents face-to-face interactions in opinion networks.