Democratizing Alpha: LLM-Driven Portfolio Construction for Retail Investors Using Public Financial Media

Daesan Oh, Taehwan Kim, Junkyu Jang, Sung-Hyuk Park*
College of Business
KAIST
{daesan__5, taehwan, jbkjsm, sunghyuk.park}@kaist.ac.kr

Abstract

Recent technological advancements and macroeconomic changes have led to a surge in individual investors' participation in capital markets. However, these investors face challenges in making investment decisions due to bounded rationality, characterized by constraints on their time, cognitive capacity, and ability to process vast information. This study empirically examine whether publicly accessible large language models (LLMs) and financial media information to enhance the investment performance of retail investors. We employ daily market commentary video transcripts from publicly available YouTube channels, including Bloomberg Television and Yahoo Finance, to prompt four LLMs (LLaMA 3, Qwen2, Gemma, GPT 4o-mini) to construct investment portfolios. These portfolios are then backtested against the S&P 500 and NASDAQ from June 2024 to July 2025. The analysis demonstrates that LLM-based portfolios exhibited consistently outperform market benchmarks across critical performance metrics, including CAGR, Sharpe ratio, and Calmar ratio. Qualitative analysis further confirms that LLMs successfully extract coherent and economically meaningful investment rationales from unstructured video content. Our findings provide a practical methodology for retail investors to leverage accessible AI, democratizing advanced analytical techniques once exclusive to institutional investors and demonstrating that AI-based tools can effectively support rational decision-making.

1 Introduction

The recent surge in retail investor participation, driven by technological and societal shifts, has highlighted significant risks associated with their market activities [20]. These risks are rooted in the principle of bounded rationality, where investors face constraints in information access, cognitive processing, and available time, leading to suboptimal financial decisions [17, 25]. While digital media has democratized access to financial information, it has also created an environment of information overload. Retail investors, often lacking the time and expertise of institutional players, struggle to convert this vast sea of unstructured data into actionable insights, making them vulnerable to financial losses.

This study investigates whether recent advancements in LLMs can bridge this gap by transforming publicly available information into high-performing investment strategies for retail investors. While prior research has applied advanced technologies to finance, many studies use proprietary data or computationally intensive methods inaccessible to the average person. Our research addresses this

*Corresponding author

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Generative AI in Finance Workshop.

gap by asking a practical question: can a combination of publicly accessible LLMs and readily available public media be used to construct portfolios that outperform standard passive benchmarks on a risk-adjusted basis?

To answer this, we design a transparent and replicable pipeline using only open resources. We feed daily market commentary transcripts from public YouTube channels (Bloomberg Television and Yahoo Finance) into four different LLMs (Llama 3, Qwen2, Gemma, and GPT-4o-mini). The models' stock selections are then used to construct portfolios, which are evaluated in a historical backtest. Our findings demonstrate that LLM-constructed portfolios consistently and significantly outperform the S&P 500 and NASDAQ indices. For instance, a mean-variance optimized portfolio built on selections from LLaMA3 achieved a compound annual growth rate (CAGR) of 42.4% and a Sharpe ratio of 1.5. A qualitative analysis further confirms that the models can extract coherent and economically meaningful investment rationales, underscoring the potential for LLMs to democratize advanced investment strategies for retail investors.

2 Background and Related Works

Numerous studies have established that individual retail investors often deviate from the rational decision-making posited by traditional finance. Empirical evidence consistently shows that retail investors tend to trade more frequently, hold under-diversified portfolios, and exhibit poor stock selection, behaviors that contribute to their documented underperformance [4, 8, 14]. These actions are largely driven by a range of cognitive and emotional factors. Prominent among these are the disposition effect, leading investors to sell profitable assets prematurely while holding on to losing ones [15, 19]; overconfidence, which results in excessive trading and risk-taking [5]; herding behavior, where investors mimic group actions rather than relying on their own analysis [16]; and attention bias, which causes a disproportionate focus on salient information [6].

The challenges faced by retail investors are compounded by the modern information environment. News and media coverage serve as crucial conduits of information that significantly influence investor sentiment, market volatility, and asset prices [3, 12, 18]. While sources like news articles, social media, and even YouTube videos are vital information channels [2, 7], their sheer volume creates a new problem: information overload. Individual investors, often with limited time and expertise, struggle to filter valuable signals from this massive influx of noisy and unstructured content. This environment can exacerbate behavioral biases, making it difficult to make well-informed, rational decisions.

Recent advancements in LLMs present a promising solution to mitigate these challenges [11, 13]. With their powerful capabilities in natural language processing and pattern recognition, LLMs are increasingly being used to process the vast amounts of textual data inherent in financial markets [22, 26]. In this landscape, specialized Financial LLMs (FinLLMs) have emerged, fine-tuned for the nuances of financial language. These range from early models like FinBERT, used for sentiment analysis [1], to more recent instruction-tuned models like FinMA (PIXIU) [23] and large-scale models like BloombergGPT, which shows superior performance in financial NLP tasks [21]. Studies demonstrate that LLMs can interpret news to predict stock returns, summarize complex financial reports, and analyze sentiment from diverse media sources [9, 10, 24]. By systematically digesting information and formulating strategies, LLMs have the potential to compensate for human cognitive limitations, helping investors navigate information overload and mitigate the impact of behavioral biases.

3 Methodology

This study empirically evaluates whether LLMs can enhance retail investors' portfolio performance by analyzing publicly available online market commentary. We conduct a historical backtest from June 3, 2024, to July 7, 2025, to compare the performance of LLM-driven investment strategies against passive benchmarks. The primary data source consists of 6,177 video transcripts from curated market commentary playlists on the YouTube channels of Bloomberg Television and Yahoo Finance, covering the period from June 2024 to June 2025. This dataset represents timely information readily accessible to retail investors.

Our methodology employs four publicly available LLMs. To prevent look-ahead bias and ensure temporal validity, we select model versions with knowledge cutoffs that precede our evaluation period. We use a two-stage prompting framework (Figure 1) to extract investment insights. In the first stage, operating daily, each video transcript is individually summarized by the LLM into structured components: market themes, risk factors, and key points on specific stocks within the S&P 500 index. In the second stage, these daily summaries from the preceding week are aggregated and provided as input to the LLM. The model is then prompted to synthesize this information, select 10 stocks for a long-only portfolio, assign respective weights, and provide a clear rationale for each selection, as depicted in Figure 1b. This hierarchical process allows the LLM to first distill granular daily information before forming a comprehensive weekly investment strategy.

To evaluate the LLM-generated recommendations, we construct two types of portfolios that are rebalanced weekly: an equal-weighted (EW) portfolio of the 10 selected stocks and a mean-variance optimized (MVO) portfolio based on the same stocks. The performance of these strategies is compared against the S&P 500 and NASDAQ indices. We assess all portfolios using a comprehensive suite of metrics, including CAGR, annualized Sharpe ratio, maximum drawdown (MDD), volatility, and Calmar, Sortino, VaR, and CVaR ratios to provide a thorough risk-return analysis. The backtest assumes zero transaction costs, no market impact, and no taxes to isolate the performance of the stock selection and weighting strategies.

```
LLM Prompt

You are analyzing a financial news video transcript. Summarize investment-relevant information for this script.

{{TRANSCRIPT}}

IMPORTANT:
- market_themes, risk_factors, key_points must be pulled word-for-word from the transcript.
- If there is no investment-relevant content, return exactly: "No relevant investment discussion"
- Each `key_points` list MUST contain 2-3 verbatim sentences from the transcript.
```

(a) Daily-Level Video Summarization

```
LLM Prompt

You are an expert financial analyst specializing in identifying investment opportunities from transcript
summaries between {{STARTDATE}} and {{ENDDATE}}. Your job is to select {{K}} stocks for the portfolio
rebalancing on {{REBALANCEDATE}}.

You may choose only among the following S&P 500 components:
  {{WHITELIST}}

Summaries:
  {{SUMMARIES}}

TASKS:
  1. Identify and select exactly {{K}} unique stocks for your weekly rebalancing on {{REBALANCEDATE}}.
  2. For each pick, output an object with:
  - symbol: the ticker
  - weight: a percentage, rounded to two decimals
  - select_rationale: an object containing:
    - Evidence: extract 2-3 quotes that appear under that stock's section summaries
  3. Assign portfolio weights to each selected stock.

REQUIREMENTS:
  - S&P 500 Exclusivity: You may choose any S&P 500 component, whether named in the transcripts or not,
based solely on the analysis of the provided summaries.
  - Unique Picks: No duplicate symbols only select exactly {{K}} stocks.
  - Strictly Data-Driven: Your analysis, selections, rationales, and weights must be derived exclusively
from the content of the provided transcripts. Do not use any outside knowledge or real-world financial
data.
  - Market Context: Recommendations should reflect the market environment as of {{REBALANCEDATE}}.
```

(b) Weekly Portfolio Construction

Figure 1: LLM prompt templates

Table 1: Comparative analysis of portfolio performance

Model		CAGR (↑)	MDD (↓)	Sharpe (↑)	Volatility (↓)	Calmar Ratio (↑)	Sortino Ratio (↑)	VaR 95% (↑)	CVaR 95% (↑)
Benchmark	S&P 500	0.1592	0.1741	0.8361	0.2008	0.9144	0.7986	-0.0410	-0.0648
	NASDAQ	0.1890	0.2266	0.7965	0.2592	0.8341	0.7338	-0.0501	-0.0742
LLaMA3	LLaMA3	0.4192	0.1838	1.4759	0.2604	2.2813	1.5938	-0.0473	-0.0726
	LLaMA3-EW	0.3824	0.1944	1.3797	0.2591	1.9671	1.4863	-0.0495	-0.0744
	LLaMA3-MVO	**0.4236**	0.1509	1.5308	0.2517	2.8063	1.5780	-0.0511	-0.0711
Qwen2	Qwen2	0.3661	0.1565	1.4186	0.2405	2.3388	1.5163	-0.0385	-0.0701
	Qwen2-EW	0.3878	0.1419	**1.5383**	0.2306	2.7322	1.5933	-0.0385	-0.0673
	Qwen2-MVO	0.4049	**0.1229**	1.4998	0.2471	**3.2944**	**1.6992**	-0.0394	-0.0670
Gemma	Gemma	0.1657	0.1678	0.8688	0.1994	0.9879	0.8518	**-0.0351**	-0.0580
	Gemma-EW	0.1144	0.1504	0.6536	**0.1944**	0.7603	0.6770	-0.0367	**-0.0546**
	Gemma-MVO	0.1786	0.2124	0.7699	0.2537	0.8410	0.9335	-0.0393	-0.0663
GPT 4o mini	GPT 4o mini	0.2099	0.2539	0.8182	0.2804	0.8267	0.8112	-0.0619	-0.0805
	GPT 4o-mini-EW	0.2029	0.2555	0.7944	0.2819	0.7942	0.7921	-0.0619	-0.0805
	GPT 4o-mini-MVO	0.2488	0.2633	0.9354	0.2778	0.9449	1.1096	-0.0567	-0.0654

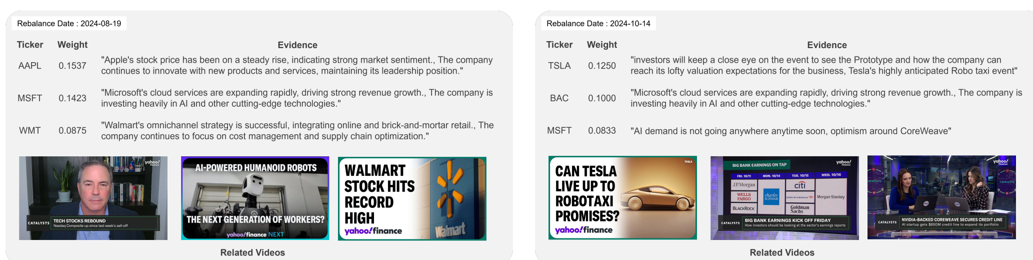
* The best-performing metrics are in bold; the best-performing instances for each model are underlined.

4 Experimental Result

Our empirical analysis reveals that portfolios constructed using LLMs significantly outperformed passive benchmarks. As detailed in Table 1, the strategies driven by LLaMA3 and Qwen2 were particularly effective, achieving Compound Annual Growth Rates (CAGRs) of 41.9% and 36.6%, respectively—substantially higher than the S&P 500 (15.9%) and NASDAQ (18.9%). This strong performance was matched by superior risk-adjusted returns, with LLaMA3 and Qwen2 recording the highest Sharpe ratios (1.48 and 1.42). In contrast, Gemma and GPT 4o-mini delivered more conservative, benchmark-like returns, highlighting performance variability across different models.

The superior returns of the top-performing models were not achieved by taking on excessive risk. The LLaMA3 and Qwen2 portfolios demonstrated robust downside protection, evidenced by lower Maximum Drawdowns (MDD) and higher Calmar and Sortino ratios compared to the benchmarks. A deeper analysis into these results shows that this success stems from two key factors. First, an evaluation of equal-weighted (EW) portfolios confirms that models like LLaMA3 and Qwen2 possess superior underlying stock-selection capabilities. Second, a comparison against these EW portfolios suggests that LLaMA3 also demonstrates an effective weighting mechanism, as its self-assigned weights further amplified its performance.

The synergy between AI-driven insight and quantitative finance was further confirmed by applying a Mean-Variance Optimization (MVO) strategy to the LLM-selected stocks. This hybrid approach often enhanced the risk-return profile, particularly for the best-performing models, demonstrating that systematic optimization can effectively complement the qualitative signals extracted by LLMs. To provide transparency into this signal extraction process, we conducted a qualitative analysis of the models' outputs.



(a) Rebalance on 2024-08-19 (Qwen2)

(b) Rebalance on 2024-10-14 (LLaMA 3)

Figure 2: Examples of LLM-Generated Portfolios and Investment Insights from Youtube Videos

This qualitative review shows that the LLMs are capable of converting unstructured video commentary into structured, actionable investment rationales (Figure 2). For instance, models justified stock selections like Apple, Microsoft, and Tesla by citing specific, coherent themes identified in the video transcripts, such as "sustained AI infrastructure demand," "effective omnichannel retail strategy," or "investor anticipation for new product cycles." These findings illustrate a clear and interpretable pipeline from public media narrative to portfolio decision, reinforcing the conclusion that LLMs can serve as powerful tools for generating high-performing and transparent investment strategies.

5 Conclusion

This study demonstrates that publicly available LLMs can successfully process financial media transcripts to construct investment portfolios that significantly outperform passive benchmarks on both absolute and risk-adjusted terms. Our qualitative analysis confirms these models extract coherent, economically meaningful rationales from unstructured content, establishing a viable pipeline from public information to actionable insight.

The practical implications of these findings are substantial, particularly for retail investors who face constraints on time, cognitive capacity, and informational access. By democratizing access to sophisticated analytical capabilities previously exclusive to institutional players, our methodology offers a scalable framework to mitigate the challenges of bounded rationality and navigate information overload. This work points toward a hybrid future of financial advising—where AI handles large-scale data analysis and human experts provide strategic oversight—ultimately fostering a more robust and inclusive financial ecosystem.

While the results are promising, this research is subject to limitations that provide clear directions for future work. These include a relatively short backtest period due to LLM knowledge cutoffs, the use of a limited set of media sources, and potential systemic risks such as herding behavior from widespread AI adoption. Addressing these challenges will be crucial for advancing the responsible application of AI in finance.

References

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

[2] M. Arias, A. Arratia, and R. Xuriguera. Forecasting with twitter data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):1–24, 2014.

[3] L. Barbaglia, S. Consoli, and S. Manzan. Forecasting with economic news. *Journal of Business & Economic Statistics*, 41(3):708–719, 2023.

[4] B. M. Barber and T. Odean. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, 55(2):773–806, 2000.

[5] B. M. Barber and T. Odean. The internet and the investor. *Journal of Economic Perspectives*, 15(1):41–54, 2001.

[6] B. M. Barber and T. Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818, 2008.

[7] S. Crawford, B. Kim, M. Koo, and T. Le. The effect of corporate twitter, instagram and youtube activity on investor attention and market liquidity. *Accounting & Finance*, 64(3):2605–2633, 2024.

[8] M. Grinblatt and M. Keloharju. What makes investors trade? *The journal of Finance*, 56(2):589–616, 2001.

[9] A. Gueta, A. Feder, Z. Gekhman, A. Goldstein, and R. Reichart. Can llms learn macroeconomic narratives from social media? *arXiv preprint arXiv:2406.12109*, 2024.

[10] S. Kim, S. Kim, Y. Kim, J. Park, S. Kim, M. Kim, C. H. Sung, J. Hong, and Y. Lee. Llms analyzing the analysts: Do bert and gpt extract more value from financial analyst reports? In *Proceedings of the Fourth ACM International Conference on AI in Finance*, pages 383–391, 2023.

[11] Y. Li, S. Wang, H. Ding, and H. Chen. Large language models in finance: A survey. In *Proceedings of the fourth ACM international conference on AI in finance*, pages 374–382, 2023.

[12] F. Lillo, S. Miccichè, M. Tumminello, J. Piilo, and R. N. Mantegna. How news affects the trading behaviour of different categories of investors in a financial market. *Quantitative Finance*, 15(2):213–229, 2015.

[13] Y. Nie, Y. Kong, X. Dong, J. M. Mulvey, H. V. Poor, Q. Wen, and S. Zohren. A survey of large language models for financial applications: Progress, prospects and challenges. *arXiv preprint arXiv:2406.11903*, 2024.

[14] T. Odean. Do investors trade too much? *American economic review*, 89(5):1279–1298, 1999.

[15] H. Shefrin and M. Statman. The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of finance*, 40(3):777–790, 1985.

[16] R. J. Shiller, S. Fischer, and B. M. Friedman. Stock prices and social dynamics. *Brookings papers on economic activity*, 1984(2):457–510, 1984.

[17] H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118, 1955.

[18] P. C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168, 2007.

[19] M. Weber and C. F. Camerer. The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, 33(2):167–184, 1998.

[20] World Economic Forum in collaboration with Boston Consulting Group and Robinhood Markets. 2024 Global Retail Investor Outlook. `https://www.weforum.org/publications/global-retail-investor-outlook-2025/`, March 2025. Accessed: 2025-07-15.

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

[22] Y. Xiao, E. Sun, D. Luo, and W. Wang. Tradingagents: Multi-agents llm financial trading framework. *arXiv preprint arXiv:2412.20138*, 2024.

[23] Q. Xie, W. Han, X. Zhang, Y. Lai, M. Peng, A. Lopez-Lira, and J. Huang. Pixiu: A comprehensive benchmark, instruction dataset and large language model for finance. *Advances in Neural Information Processing Systems*, 36:33469–33484, 2023.

[24] X. Yu, Z. Chen, Y. Ling, S. Dong, Z. Liu, and Y. Lu. Temporal data meets llm–explainable financial time series forecasting. *arXiv preprint arXiv:2306.11025*, 2023.

[25] S. Zahera and R. Bansal. Do investors exhibit behavioral biases in investment decision making? a systematic review. *Qualitative Research in Financial Markets*, 10(2):210–251, 2018.

[26] H. Zhao, Z. Liu, Z. Wu, Y. Li, T. Yang, P. Shu, S. Xu, H. Dai, L. Zhao, G. Mai, et al. Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*, 2024.