Large Language Models for Predicting Investor Sentiment and Market Behavior Based on Self-Narrative Analysis

Viktor R. Kennedy
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
kennedy894@ucf.edu

Nicolas A. Johansson
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
nicolasajohansson@unr.edu

Abstract

The prediction of investor sentiment and its influence on market dynamics has long challenged both financial economists and computational researchers. Traditional sentiment analysis tools, while effective at capturing coarse affective signals, often fail to uncover the layered cognitive and motivational structures embedded in the narratives that investors construct about themselves and their economic environments. This paper argues that large language models (LLMs) open a transformative pathway for extracting such self-narrative signals and translating them into actionable predictions of market sentiment and aggregate behavior. We adopt a system-level perspective that moves beyond algorithmic novelty to examine the architectural trade-offs, data governance frameworks, infrastructure design, and fairness considerations essential for deploying LLM-based sentiment prediction in real-world financial ecosystems. The analysis integrates insights from behavioral finance, narrative psychology, and sociotechnical systems engineering to delineate how self-narrative features—ranging from future-oriented planning statements to justificatory reasoning—can serve as leading indicators of market shifts. We further explore the deployment challenges associated with latency-sensitive trading environments, the sustainability implications of large-scale model inference, and the regulatory pressures shaping the transparency and explainability of such systems. By foregrounding structural robustness, epistemic uncertainty quantification, and fairness across market participants, we provide a comprehensive blueprint for the responsible integration of LLM-driven narrative analytics into institutional investment processes. The paper concludes with a discussion of the broader policy ramifications, including the potential for systemic risk amplification and the necessity for co-designed governance mechanisms that align technological capabilities with market stability objectives.

Keywords

large language models, investor sentiment, self-narrative analysis, market behavior, financial NLP, system architecture, fairness, governance.

1. Introduction

Financial markets operate as complex adaptive systems in which the beliefs, expectations, and emotions of heterogeneous agents coalesce into price movements and volatility patterns that frequently deviate from the predictions of purely rational models. The gap between fundamental valuation and observed market behavior has motivated decades of research into sentiment analysis, demonstrating that textual data from news media, social platforms, and corporate disclosures carry predictive information about future returns and trading volumes. The maturation of natural language processing has brought increasingly powerful tools to this domain, beginning with lexicon-based dictionaries and culminating in the transformer architectures that now underpin state-of-the-art language understanding. The attention mechanism introduced by Vaswani et al. [1] established a foundational shift in sequence modeling, enabling the bidirectional contextual representations of BERT [2] and the autoregressive generative capacities of GPT-3 [3]. Subsequent refinements such as RoBERTa [4] further optimized pretraining dynamics, allowing models to capture subtle linguistic cues that escape simpler approaches. Despite these advances, much financial NLP research has remained confined to shallow sentiment classification, measuring positive or negative affect without probing the deeper narrative structures that shape investor identity and forward-looking intention.

The central premise of this paper is that self-narrative—the way investors story their own economic roles, justify past decisions, and project future states—constitutes a rich, multidimensional signal that large language models are uniquely positioned to decode. Traditional sentiment proxies, such as the media-based measures developed by Tetlock [5] or the financial dictionaries refined by Loughran and McDonald [6], have proven valuable but are limited to coarse affective valence. Even more dynamic approaches, such as the Twitter mood tracking of Bollen et al. [7], still abstract away from the narrative coherence and self-referential reasoning that individual and institutional investors continuously produce in earnings calls, investment forums, and personal financial blogs. By shifting the analytic focus toward self-narrative analysis, we argue that LLMs can recover not merely how positive or negative an investor feels, but what temporal horizons, causal attributions, and identity commitments structure their sentiment. This perspective aligns with the narrative economics framework articulated by Shiller [8], which treats contagion of stories as a macroeconomic force. Yet operationalizing narrative economics at scale has remained elusive precisely because of the linguistic depth and contextual reasoning required. The present work addresses this gap through a system-level examination of how LLMs can be architected, governed, and deployed to extract self-narrative signals in a manner that is robust, fair, and consistent with financial market constraints. In doing so, we extend the discourse from algorithm-centric performance metrics to the socio-technical infrastructure that must surround such predictive systems if they are to function responsibly within live trading environments.

2. Self-Narrative as a Multidimensional Sentiment Signal

Psychological research has long established that individuals construct and revise personal narratives to make sense of experience, maintain self-coherence, and regulate emotion. McAdams [9] conceptualized life stories as integrative frameworks through which people articulate identity, while Bruner [10] emphasized the irreducible narrative structure of self-understanding. Applied to financial domains, the stories that investors tell about their portfolios, the market, and their own decision-making processes encode far richer information than isolated sentiment words. An investor who describes a recent loss by attributing it to temporary external shocks and expresses confidence in a long-term recovery is projecting a qualitatively different future orientation than one who internalizes the loss as a personal failure and withdraws from risk. These self-narrative dimensions include causal attribution, temporal extension, agentic framing, and the presence of counterfactual reasoning, each of which carries distinct implications for subsequent trading behavior. Pennebaker and King [11] demonstrated that linguistic style, including pronoun use and cognitive process words, serves as a stable marker of individual differences, further supporting the idea that self-narrative features are psychometrically meaningful and not reducible to transient mood. The work of Solanki et al. [12] on how individuals think about themselves provides additional empirical grounding for the claim that reflective self-statements contain latent signals about future behavior, as the introspective framing adopted in personal narratives correlates with downstream decision-making patterns. When aggregated across market participants, these signals may anticipate shifts in collective risk appetite, herding dynamics, or capitulation points.

Translating this psychological insight into a computable financial signal demands models that can parse and represent narrative structure rather than simply annotating sentences for polarity. Coarse sentiment scores treat the sentence "I know I should hold, but I keep imagining a crash" as ambiguous, whereas a narrative-aware model can identify the clash between normative self-guidance and emotionally charged anticipation. This clash signals internal conflict and potential volatility in position adjustments. Moreover, the temporal dimension inherent in self-narratives aligns with behavioral finance findings on prospect theory and mental accounting, where reference points and framing effects profoundly influence risk-taking [19]. The reflexive nature of self-narrative—where the act of articulating a story can reinforce or alter the underlying beliefs—introduces a feedback loop that is absent from simpler sentiment indices. When institutional earnings call transcripts are examined, the rhetorical strategies employed by executives to frame disappointing results reveal not only current sentiment but also the projected trajectory of corporate action, which in turn shapes investor expectations. Thus, self-narrative analysis leverages the full generative capability of LLMs to model not only what is said but how it is embedded within identity-preserving and future-constructing discursive practices.

3. Architectural Foundations and Model Adaptation

The successful extraction of self-narrative signals rests on careful architectural choices that balance representation depth with computational tractability. While early transformer models demonstrated the ability to capture contextual semantics, the scale and specialization required for financial narrative analysis demand further adaptation. Open-source foundation models such as LLaMA [13] provide a customizable base that can be fine-tuned on domain-specific corpora, allowing institutions to retain data sovereignty while adapting to the specialized lexicon and narrative conventions of financial discourse. Proprietary large-scale models like PaLM [14] illustrate the opposite end of the trade-off spectrum, where massive parameter counts enable nuanced few-shot reasoning but introduce substantial infrastructure dependencies and opaque update cycles. The architectural decision matrix thus extends beyond accuracy on a static benchmark to encompass considerations of inference latency, memory footprint, model ownership, and the ability to retrain or align models as market conditions evolve.

Fine-tuning strategies for self-narrative analysis differ from standard sentiment classification because the target signal is not a scalar sentiment score but a structured representation of narrative features. Approaches that combine instruction tuning with chain-of-thought prompting allow the model to decompose a text into attributional statements, temporal projections, and efficacy assessments before aggregating these into a sentiment trajectory. This decomposition mirrors clinical narrative coding schemas but is implemented at machine scale. The architecture must support multi-step reasoning while maintaining consistency across long documents, as self-narratives in annual reports or investor letters can span thousands of words. Retrieval-augmented generation pipelines can be layered onto these models to ground the narrative analysis in historical statements by the same entity, thereby enabling the detection of shifts in narrative framing that may precede strategic changes. The substantial energy requirements of training and serving such large models have prompted a parallel focus on efficiency; Strubell et al. [15] highlighted the environmental costs of NLP scaling, reinforcing the need for model distillation, quantization, and sparse activation techniques when deploying narrative analysis at the scale of global equity markets. These sustainability pressures intersect with technical design choices, as the computational budget allocated to narrative depth must be justified by predictive gains that survive transaction cost considerations.

4. Data Infrastructure, Governance, and Fairness

Ingesting the textual sources that carry self-narrative content—social media posts, earnings call transcripts, retail investor forum discussions, and financial news interviews—requires a data infrastructure that addresses volume, velocity, and veracity challenges while respecting privacy regulations. Real-time streaming architectures built on distributed log systems must filter, deduplicate, and annotate millions of documents daily, aligning them with entity identifiers and timestamped market data. However, the very richness of self-narrative data amplifies fairness concerns, because the linguistic patterns on which the models are trained are shaped by socioeconomic, cultural, and demographic factors that correlate with market access and financial literacy. Barocas and Selbst [16] articulated the mechanisms through which data-driven systems can reproduce and amplify historic disparities, a danger acutely relevant when predictive models influence capital allocation. If an LLM-based sentiment predictor systematically misinterprets the narrative conventions of underrepresented investor groups, its outputs may channel liquidity toward segments of the market already privileged by linguistic homogeneity, thereby distorting price discovery and exacerbating inequality.

Addressing these fairness challenges requires governance frameworks that span the entire data lifecycle, from collection policies that ensure representational diversity to model evaluation protocols that disaggregate performance across demographic and linguistic strata. The opacity of large neural models also raises explainability demands that intersect with legal and regulatory expectations. Wachter et al. [17] examined the contested status of a right to explanation under data protection law, concluding that meaningful transparency requires more than feature importance scores; it demands counterfactual accounts that a regulated entity can audit. In the context of self-narrative sentiment prediction, this could translate into the requirement that a trading system justify a position change not merely by citing a sentiment score but by articulating the narrative pattern that drove the signal, such as an increase in counterfactual regret statements. Such explanations become part of the institutional record and must be archived alongside trading decisions for regulatory review. Governance bodies must also determine data retention policies, as the discarding of narrative records may undermine audit trails while indefinite storage amplifies surveillance risks in ways that Zuboff [23] characterizes as a threat to human autonomy. The tensions between predictive value, privacy, and fairness are not merely operational details but constitutive of the system's legitimacy, and their resolution requires participatory design processes that involve regulators, platform operators, and representatives of impacted communities.

5. Deployment and Integration in Financial Systems

Deploying LLM-based self-narrative predictors within live trading environments introduces a distinct set of system engineering constraints that differ markedly from those encountered in academic experimentation. Financial markets exhibit extreme latency sensitivity, with competitive advantages often measured in microseconds, yet narrative analysis requires processing sequences of text that may span hundreds of tokens, placing minimum bounds on inference time. Architecturally, this tension is managed through a separation of concerns in which batch-oriented narrative enrichment pipelines precompute structured narrative signatures for monitored entities, while real-time event-triggered inference refines these signatures only when new narrative material appears. The resulting microservice topology must be fault-tolerant, capable of graceful degradation if a language model endpoint becomes unavailable, and instrumented with circuit breakers that prevent cascading failures from propagating into order execution systems. The adaptive markets hypothesis advanced by Lo [18] provides a useful conceptual lens for understanding why fixed model parameters are insufficient; as market ecologies shift, the relationship between narrative features and price dynamics evolves, necessitating continuous model monitoring and periodic retraining cycles that respect capital allocation risk budgets.

The behavioral foundations of self-narrative prediction also demand integration with theories of investor psychology that have been extensively mapped in the literature. Kahneman and Tversky [19] established the centrality of reference-dependent preferences and loss aversion, while Barberis and Thaler [20] synthesized a broad survey of behavioral finance anomalies that narrative signals can help contextualize. A model that detects an uptick in personal narrative themes of regret and counterfactual processing across a retail trading platform may predict a withdrawal of risk capital that manifests as reduced trading volume before price-based indicators register the shift. However, acting on such signals requires fusing narrative intelligence with market microstructure realities. O'Hara [21] detailed how order flow fragmentation, dark pool activity, and high-frequency quoting strategies complicate the transmission belt from sentiment signal to executable trade, underscoring that narrative analytics must be embedded within a broader market access architecture that adjusts for liquidity conditions and execution certainty. The deployment system thus becomes a socio-technical hybrid in which the hermeneutic outputs of language models are translated into quantitative forecasts through calibration layers that learn, from historical data, the time-varying mapping between narrative states and order flow characteristics.

6. Robustness, Explainability, and Epistemic Risks

The reliance on self-narrative analysis introduces distinctive robustness challenges that go beyond standard model accuracy degradation. Narratives can be strategically manipulated by sophisticated actors who understand the linguistic patterns the model associates with future behavior. A corporate executive aware that a rising frequency of agentic language forecasts positive momentum might artfully inflate such language in public statements, inducing a feedback loop in which the model's predictions reinforce the very narrative it measures. Such adversarial narrative dynamics constitute a form of epistemic risk that is structurally analogous to Goodhart's law, wherein a measure loses its informational content once it becomes a target. The stochastic nature of large language models further complicates the picture; Bender et al. [22] warned that the seemingly coherent outputs of LLMs can mask a lack of true understanding, creating the illusion of narrative insight where only surface-level statistical patterning exists. In financial contexts, this illusion can be perilous, as a model that confidently misidentifies a firm's forward-looking narrative may generate signals that trigger automated trading cascades.

Mitigating these risks demands a multilayer robustness strategy that combines adversarial training, epistemic uncertainty quantification, and human-in-the-loop oversight mechanisms. Architectural techniques that expose model confidence through ensemble disagreement or conformal prediction can flag narrative interpretations that fall below a critical threshold of reliability, redirecting them for human review. Explainability frameworks must be designed not only for regulatory compliance but also for operational debiasing; a portfolio manager who receives a narrative-derived sell signal must be able to interrogate the specific narrative clauses that drove the prediction and compare them against alternative readings. Floridi [24] analyzed the European Union's approach to AI regulation, highlighting the centrality of human oversight and transparency obligations that apply with particular force when narrative analysis influences markets. Furthermore, Hovy and Spruit [25] examined the broader social impact of NLP technologies, noting that the deployment of language models in high-stakes domains can shift power asymmetries by concentrating interpretive authority among institutions that control the models. In the market context, this concentration could lead to a situation where a small number of firms with advanced narrative analytics gain an insurmountable informational advantage, undermining the level playing field that securities regulation seeks to preserve.

7. Sustainability and Policy Implications

The systemic diffusion of LLM-based sentiment prediction systems carries sustainability burdens that extend beyond the energy consumption of model training to encompass the lifecycle management of data, models, and the market structures they influence. The carbon footprint of continuously retraining large models on streaming narrative corpora can be significant, although counterbalanced by the efficiency gains from shared foundational pretraining and the application of sparse mixture-of-experts architectures. A broader sustainability perspective must also consider the cognitive load placed on market participants who increasingly interact with narrative-driven analytics interfaces; if the interpretability of model outputs is poor, traders may experience automation surprise and disengage critical judgment, transferring epistemic responsibility to opaque systems. The governance mechanisms discussed earlier must thus be extended into a full lifecycle framework that specifies decommissioning protocols for models whose narrative templates become stale or whose fairness metrics drift beyond acceptable bounds.

Policy responses to the rise of narrative-based market prediction are still nascent, but several directions are discernible. Securities regulators will need to determine whether the use of LLMs to parse and trade on the self-narratives of public companies constitutes a form of material non-public information, particularly when the models detect subtle linguistic shifts that are not yet reflected in conventional news. The boundary between permissible sentiment analysis and manipulative narrative engineering must be clarified, potentially through safe harbors for transparency reporting analogous to those that govern algorithmic trading. International coordination will be required because narrative signals cross jurisdictional boundaries and because differing privacy regimes, such as the European General Data Protection Regulation, impose constraints on the processing of personal narrative data that may conflict with the global nature of financial markets. The co-evolution of technical systems and regulatory frameworks suggests that adaptive governance, informed by ongoing empirical studies of narrative analytics impacts, will be more effective than static rulemaking.

Ultimately, the responsible integration of self-narrative prediction into market infrastructure depends on a collective willingness to treat the technology not as a neutral tool but as a constitutive element of the financial system's sensemaking apparatus, one that must be continuously aligned with values of fairness, stability, and transparency.

8. Conclusion

This paper has advanced a system-level analysis of how large language models can be employed to predict investor sentiment and market behavior through the lens of self-narrative analysis. By shifting the unit of analysis from shallow affect to the multidimensional narrative structures that shape investor identity and forward-looking intention, LLMs can capture latent psychological dynamics that traditional sentiment tools miss. The architectural choices involved in fine-tuning these models, managing trade-offs between proprietary scale and domain-specific customization, and embedding them within latency-sensitive trading infrastructures demand a holistic engineering perspective that extends from data ingestion to model serving. Governance frameworks must be woven throughout the system to ensure fairness across diverse investor populations, robust explainability that satisfies regulatory demands, and resilience against adversarial narrative manipulation. The deployment of narrative prediction systems inevitably restructures informational asymmetries in financial markets, and therefore must be accompanied by adaptive policy mechanisms that safeguard market integrity and participant autonomy. As narrative analytics mature, interdisciplinary collaboration among computer scientists, financial economists, psychologists, and legal scholars will be indispensable to navigate the epistemic and ethical terrain. Large language models are not simply instruments for more accurate prediction; they are components of an evolving cognitive infrastructure through which markets come to know themselves, and their design must reflect the full weight of that responsibility.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171–4186.

3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

4. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

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

6. Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65.

7. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8.

8. Shiller, R. J. (2017). Narrative economics. American Economic Review, 107(4), 967–1004.

9. McAdams, D. P. (2001). The psychology of life stories. Review of General Psychology, 5(2), 100–122.

10. Bruner, J. (2004). Life as narrative. Social Research, 71(3), 691–710.

11. Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. Journal of Personality and Social Psychology, 77(6), 1296–1312.

12. Solanki, D., Hsu, H. M., Zhao, O., Zhang, R., Bi, W., & Kannan, R. (2020, July). The way we think about ourselves. In International Conference on Human-Computer Interaction (pp. 276-285). Cham: Springer International Publishing.

13. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

14. Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., … Fiedel, N. (2023). PaLM: Scaling language modeling with pathways. Journal of Machine Learning Research, 24(240), 1–113.

15. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650.

16. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104, 671–732.

17. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99.

18. Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. Journal of Portfolio Management, 30(5), 15–29.

19. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291.

20. Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), Handbook of the Economics of Finance (Vol. 1, pp. 1053–1128). Elsevier.

21. O'Hara, M. (2015). High frequency market microstructure. Journal of Financial Economics, 116(2), 257–270.

22. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.

23. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

24. Floridi, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. Philosophy & Technology, 34, 215–222.

25. Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 591–598.