Quantifying Strategic Ambiguity in Corporate Language for AI-Driven Trading Strategies

Vaisnavi Nemala
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
vnemala@andrew.cmu.edu

Abstract

This paper introduces a novel approach for extracting actionable insights from corporate communications by quantifying strategic ambiguity in language. While prior work in natural language analysis has largely focused on sentiment or factual content, we explore how organizations deliberately hedge, obscure, or soften information, using linguistic ambiguity as a rich signal of intent and hidden meaning. We propose the Strategic Ambiguity Score (SAS) which captures deliberate vagueness by integrating hedge frequency, negation patterns, and model-based attention to critical phrases. Unlike traditional sentiment models, SAS measures how and where uncertainty is strategically embedded within the text. We demonstrate that SAS can effectively highlight subtle signals that correlate with subsequent outcomes, and we illustrate its utility through predictive analyses in corporate disclosures. By shifting the focus from simple sentiment interpretation to ambiguity detection, this work provides a generalizable framework for AI applications in decision-making, risk assessment, and strategic communication analysis across diverse domains.

Introduction

Financial disclosures, as defined by (Gibbins, Richardson, and Waterhouse 1990), encompass all forms of financial data release, including mandatory filings like the 10-K/10-Q to voluntary communications such as earnings calls. Voluntary disclosures, including earnings conference calls, are becoming a dominant medium for firms to engage with stakeholders and manage transparency (Beattie, Dhanani, and Jones 2008; Williams 2008), becoming an essential part of the financial genre chain (Camiciottoli 2010). Conference calls typically open with prepared statements by the company's management (which usually restate the press release), and are then open to questions from analysts (Frankel, Johnson, and Skinner 1999; Frankel, Mayew, and Sun 2010; Kimbrough 2005). In particular, the Q&A segments present an opportunity for unscripted, high-stakes dialogue to occur between management and analysts. Shifts in tone, language nuances, and sentiments during these interactions and calls have the potential to convey latent signals not fully captured in quantitative disclosures alone.

One such example of critical qualitative undertones is the use of strategic ambiguity and hedging language during earnings calls. Hedging language are words and/or expressions that are used in communication for introducing uncertainty, 'fuzziness', ambiguity, or vagueness into a statement (Lakoff 1973; Bachenko, Fitzpatrick, and Schonwetter 2008). Example phrases include "it appears that," "likely", "probably", "could", "maybe", "I guess", etc (Duran et al. 2010). In typical communication, it is used as a tool to avoid appearing overconfident, acknowledge limitations, and to provide nuance in a statement. Executives may often utilize hedging language in earnings calls to qualify their statements and avoid making overly confident or specific predictions that may not materialize. It can also be used to temper forward-looking statements, as is dictated by the safe harbor provision (15 U.S.C. § 78u-5 2022), which emphasizes that forward-looking statements made during the call are inherently uncertain and cannot be seen as guarantees of future performance (to protect company from litigation and legal obligations). Furthermore, executives may also strategically employ ambiguous language in order to downplay negative news or uncertain financial circumstances.

Prior research in financial text applications has predominantly focused on sentiment extraction, such as quantifying the polarity (positive or negative) or emotion (such as anxiety) embedded within earnings disclosures (Fatouros et al. 2023; Todd, Bowden, and Moshfeghi 2024). However, while sentiment captures first-order expressions of tone, it fails to fully capture more subtle linguistic choices and maneuvers, such as strategic ambiguity, which are critical components of evasive or ambiguous corporate communications. Existing NLP approaches, such as transformer-based models like FinBERT (Araci 2019) or multimodal architectures (Yang, Xu, and Gao 2020) with audio/visual cues, are able to provide token-level sentiment scoring, yet have not fully captured or explored strategic evasiveness as a quantitative and actionable score that translates into a tradeable financial signal (Todd, Bowden, and Moshfeghi 2024). Our primary contributions in this paper are threefold. First, we introduce the Strategic Ambiguity Score (SAS), which enables systematic detection of obfuscated or hedged language that often precedes market-relevant events. Second, we move beyond traditional sentiment polarity in AI and NLP models by quantifying linguistic opacity. This demonstrates that subtle hedging and ambiguity in corporate communications can serve as a predictive signal for future stock returns. Third, we provide empirical evidence that the SAS signal has significant predictive power in long-short trading portfolios across multiple holding periods, consistently outperforming both naive sentiment-based strategies and randomized benchmarks. Overall, these contributions highlight how AI can extract actionable insights from language beyond traditional sentiment analysis, offering a novel framework to turn linguistic nuance into measurable trading advantages.

Related Works

AI in Earnings Calls and Financial Disclosures

In 2011, Loughran and McDonald introduced their pioneering financial-domain-specific sentiment dictionaries, derived from a comprehensive sample of 10-K filings between 1994 and 2008 (Loughran and McDonald 2011; Theil, Štajner, and Stuckenschmidt 2020). These dictionaries focused on categorizing words into positive, negative, litigious, strong modal, weak modal, and crucially, uncertain terms, with a special emphasis on general imprecision rather than risk-specific vocabulary. Their findings established a significant relationship between the cumulative tf-idf scores of uncertain words and post-filing stock return volatility. Expanding upon their research, they later demonstrated that a file-size based readability measure could outperform traditional formulas like the Gunning Fog Index for explaining volatility, analyst forecast errors, and forecast dispersion (Loughran and McDonald 2014). Our approach builds off the foundations of their approach of applying event studies to assess 10-K impacts on financial uncertainty. However, we specifically focus on uncertainty as the primary independent variable rather than readability, and we further hypothesize that by enriching uncertainty dictionaries with industry-specific terms, it will yield stronger regression results.

Tsai and Wang took the Loughran-McDonald dictionaries further by employing word embeddings trained on 10-K filings from 1994 to 2006 in order to automatically expand the vocabulary set (Tsai and Wang 2014; Theil, Štajner, and Stuckenschmidt 2020). They found that by appending the top 20 cosine-similar terms per dictionary entry, it enhanced SVMrank and SVR model performances, especially for predicting stock return volatility. In following works, Tsai et al. demonstrated that these expanded dictionaries could also be used to predict post-event volatility using the Fama-French 3-factor model (Tsai, Wang, and Chien 2016; Theil, Štajner, and Stuckenschmidt 2020). However, they cautioned that the regression's sensitivity to the parameter k (the number of added terms) leads to needing to keep it at 20 due to diminishing returns. Our results challenge this static k approach; while larger k values benefit short-term volatility regressions, we observe that analyst-based uncertainty measures behave differently, as detailed in the methodology.

Rekabsaz et al. refined dictionary expansion by integrating additional financial features, term weighting strategies, and feature fusion methods (Rekabsaz et al. 2017; Theil, Štajner, and Stuckenschmidt 2020). Using 10-Ks from 2006 to 2015, they combined bag-of-words with market volatility measures, GARCH models, and sector variables, leading to improved volatility prediction. Theil et al. compared domain-specific embedding expansions with general-domain (news) expansions, and confirmed that domain-specific models are superior for financial volatility regressions, though manual filtering of candidate terms added negligible benefit (Theil, Štajner, and Stuckenschmidt 2020). They also demonstrated the model's utility in classifying sentences as sentences as certain or uncertain, which expand upon it by training industry-specific embedding models and linking them to downstream financial uncertainty measures like analyst forecast error and dispersion. Our work builds upon these efforts by embedding specialized industry jargon into uncertainty models and providing a more holistic view of how linguistic uncertainty propagates into financial uncertainty.

Signal Generation in AI-Driven Financial Forecasting

AI techniques have had diverse applications and use in financial trading;deep learning, reinforcement learning, transformer-based, and hybrid model architectures have all been extensively utilized for predictive modeling efforts (Kearney and Liu 2014). Araci released FinBERT in 2019, a language model based on BERT and fine-tuned to tackle NLP and sentiment analysis tasks in the financial domain (Araci 2019). Jiang and Zeng integrated FinBERT into an LSTM-based architecture in order to predict stock movements, outperforming standard BERT, standalone LSTM, and ARIMA baselines (Jiang and Zeng 2023). Their results highlight how sentiment-aware models are a notable source of potential features for forecasting accuracy. Another study applied FinBERT to energy sector news, emphasizing that content sentiment (from the full articles) had significant potential for stock prediction accuracy, supporting the need for context-rich models in sector-specific applications.

FinBERT's development addressed an important gap in pretrained financial-domain language models. Later iterations, such as FinBERT-FOMC, further fine-tuned the model to handle complex central bank communications, utilizing methods like Sentiment Focus (SF) to simplify sentence structures for better sentiment accuracy (Gössi et al. 2023). Comparative studies between ChatGPT and FinBERT revealed that GPT-4o, optimized with prompt engineering, outperformed FinBERT in sentiment classification by up to 10% based on the sector (Fatouros et al. 2023; Kang and Choi 2025). These findings underscore prompt engineering's critical role in enhancing LLM potential in financial applications.

Sentiment scores from BERT were further applied in portfolio optimization within the Black-Litterman framework (Colasanto et al. 2022). By incorporating sentiment as a dynamic "view" into Monte Carlo simulated paths, forecasts were made more robust and aligned with market-moving events extracted from financial news.

In addition to the FinBERT model, FinLlama introduced an LLM framework built on Llama 2, also fine-tuned for financial sentiment analysis in algorithm trading (Iacovides et al. 2024). Drawing upon a generator-discriminator approach, FinLlama provided nuanced sentiment valence and strength estimations, optimizing for low-resource deployment and accuracy. Humor's impact in earnings calls has also been investigated, revealing that executive's strategic use of humor during earnings conference calls can influence market reactions, an often underutilized dimension of earnings call analysis (Call et al. 2024).

There is a critical literature gap in the application of strategic ambiguity by executives in voluntary disclosures in earnings call analysis, and especially their predictive ability in signal generation. We seek to close this gap by exploring strategic ambiguity, hedging language, and other subtle linguistic choices representing uncertainty in managerial statements, beyond simple sentiment analysis, which can also serve as a quantitative factor in improved signal construction and providing new information in investing decisions.

Dataset and Preprocessing

We utilize a large corpus of 18,755 earnings call transcripts sourced from The Motley Fool dataset (Potterer 2023), covering a wide range of publicly traded companies across multiple sectors, time periods, and market cap. A total of 2876 unique companies are represented in the dataset, and it includes earnings call transcripts from 2017 - 2023. Each transcript represents a real-time disclosure of financial and operational performance, described by the earnings narrative from corporate executives and CFOs. Each transcript is structured with the following sections: prepared remarks (management overview), Q&A, list of participants, and other miscellaneous/introductory commentary. Additionally, historical OHLCV data has been sourced from Yahoo! Finance's API (yfinance) in order to align earnings events with stock market reactions.

Earnings calls can contain distinct linguistic patterns across sections. As compared to prepared remarks, the Q&A portion is typically more spontaneous and more likely to contain evasive language, hedging, and strategic ambiguity. A custom parser segments each transcript into four buckets based on typical section headers: prepared remarks, q&a, participants, and other commentary. Standard text normalization was then completed, including lowercasing, collapsing of whitespace, cleaning of punctuation and special characters, and the removal of boilerplate legal disclaimers (e.g. safe harbor statement (15 U.S.C. § 78u-5 2022)). The cleaned transcript is then tokenized into individual sentences, which serve as our primary unit for later analysis.

Methodology

Our goal is to represent and record strategic ambiguity in corporate disclosures as a quantitative and actionable score, and to assess its predictive utility for near-term stock returns. We design this methodology to isolate ambiguity from specific linguistic choices (negation and hedging), while accounting for the direction of sentiment polarity.

Sentiment Classification with FinBERT

Strategic ambiguity involves the expression of a tone with reduced clarity or commitment; therefore, we first extract sentiment from the text in order to understand the direction (is the narrative positive or negative?), and we augment it with hedging and negation to understand how clearly or confidently it is being expressed. Sentiment serves as the semantic base signal, and our contributions lie in quantifying how this signal is being diluted through evasive linguistic devices.

We utilize ProsusAI/FinBERT, a pre-trained and fine-tuned transformer-based BERT language model for sentiment classification on financial communications, such as earnings reports, analyst statements, etc. In comparison to general sentiment models, FinBERT has been trained on the specialized language found in financial contexts, leading to better understanding and accuracy on financial phrasing and sentiment detection. For each sentence in an earnings call transcript, the classifier outputs a probability distribution over three classes (positive, negative, or neutral). The sentiment polarity score for each sentence is summarized as:

$$\begin{aligned}\text{Polarity} &= 1 \cdot P_{\text{positive}} + 0 \cdot P_{\text{neutral}} + (-1) \cdot P_{\text{negative}} \\ &= P_{\text{positive}} - P_{\text{negative}}\end{aligned}$$

The polarity score is used to capture the net emotional tone conveyed, which serves as our foundation for distinguishing ambiguity from sentiment.

Linguistic Feature Annotation

While sentiment is essential for understanding what is being expressed, negation and hedging capture how language is expressed. Each sentence is then annotated for negation (use of negators, such as "not", "never", and "no") and hedging (use of modal words and uncertain language, e.g. "we believe", "may", "it is possible"). Negation, such as in statements like "We do not expect significant margin compression", is included as a feature of interest since it can be used to invert sentiment or hedge responsibility. Something being "not bad" does not necessarily equate "good." Therefore, we explicitly handle negation separately in addition to sentiment scoring. Including negation as an additional feature allows the model to separately learn the contextual inversion and better capture the linguistic variety in uncertain language.

In addition, we also include hedging detection in statements such as "We believe this trend is likely to continue". Hedging language refers to expressions that are used to make statements indefinite, or reduce the strength of the assertion that a speaker is making. This is often used to sound more cautious and avoid sounding overly confident. Hedges are used to reduce the assertiveness of a statement, which is something that is not captured by sentiment analysis alone. Including hedging as an additional feature allows the model to include insight into how the speaker is avoiding direct firm predictions and declarations. We curate a lexicon of hedging language and phrases tailored for financial language, extending upon Bill McDonald's and Tim Loughran's dictionary of uncertain language (Loughran and McDonald 2011), and

Strategy	Mean Return	Sharpe Ratio	Max Drawdown	t-stat	p-value
SAS_1d	**0.004425**	**1.170594**	**-0.477190**	**1.005686**	**0.315880**
Sentiment_1d	0.000007	0.001810	-0.668420	0.001555	0.998761
Random_1d	-0.003454	-0.990535	-0.693735	-0.751369	0.453657
SP500_1d	-0.000276	-0.802122	-0.024141	-0.225972	0.823636
SAS_3d	**0.005571**	**1.220087**	**-0.613668**	**–**	**–**
Sentiment_3d	-0.002741	-0.571031	-0.836351	–	–
Random_3d	0.000490	0.102887	-0.706361	0.078044	0.937901
SP500_3d	0.001973	3.824599	-0.032806	1.022167	0.321028
SAS_5d	**0.006655**	**1.277084**	**-0.546530**	**–**	**–**
Sentiment_5d	0.001116	0.210375	-0.724745	–	–
Random_5d	0.003941	0.872134	-0.497104	–	–
SP500_5d	0.004448	7.277730	-0.027740	1.833815	0.086597

Table 1: Performance metrics of various strategies across different time horizons.

employ rule-based flagging for annotating each sentence. As a result of these annotations, each sentence is transformed into a structured dictionary with the processed text, sentiment prediction (neutral, positive, or negative), negation flag (true or false), and hedge flag (true or false).

Attention Weights

As a transformer-based model, attention mechanisms map probability distributions over the input, indicating the relative importance of different parts of the input text, allowing the model to focus on the most relevant information. While not a direct substitute for explainability or model-agnostic, they provide a soft mapping for which words are driving the model's prediction, with low additional resource usage (Bahdanau, Cho, and Bengio 2014; Bibal et al. 2022; Hao et al. 2021; Vaswani et al. 2017; Liu et al. 2022). We extract token-level attention weights from FinBERT's intermediate layers, where each word/token is assigned an importance score. Tokens with a moderate attention mass (thresholded at ¿ 0.05) are retained in order to focus on linguistically significant phrases. For each sentence, the average attention weight of meaningful tokens is computed. These are used as soft attribution scores, serving as a measure of how much attention is served the model has given to each token during sentiment classification.

Strategic Ambiguity Score

Our goal is to now quantify the degree of strategic ambiguity in corporate discourse,by integrating hedging, negation, and localized attention salience, thus capturing intentional ambiguity or deliberate vagueness in financial communication. This measure enables systematic alpha extraction from *how* firms communicate, beyond *what* is being said.

Given a transcript T, composed of N attributed sentences:

$$T = \{S_1, S_2, \ldots, S_N\}$$

Each sentence S_i is associated with:

- Sentiment probabilities: $P^{(i)}_{\text{pos}}, P^{(i)}_{\text{neu}}, P^{(i)}_{\text{neg}}$
- Binary indicators: $H^{(i)} \in \{0,1\}$ (Hedge presence), $N^{(i)} \in \{0,1\}$ (Negation presence)
- Attention weights over tokens: $\{(w^{(i)}_j, a^{(i)}_j)\}_{j=1}^{M_i}$, where $a^{(i)}_j \in [0,1]$

For each sentence S_i, compute its *Sentiment Polarity* as:

$$\text{Polarity}^{(i)} = P^{(i)}_{\text{pos}} - P^{(i)}_{\text{neg}}$$

We filter token attentions to extract *Meaningful Attention Tokens*:

$$A^{(i)} = \left\{ a^{(i)}_j \;\middle|\; a^{(i)}_j > \theta \right\}, \quad \theta = 0.05$$

If $A^{(i)}$ is empty, or $H^{(i)} = 0$ and $N^{(i)} = 0$, define:

$$SAS^{(i)} = 0$$

Otherwise, compute the average attention mass:

$$\bar{A}^{(i)} = \frac{1}{|A^{(i)}|} \sum_{a \in A^{(i)}} a$$

The sentence-level Strategic Ambiguity Score becomes:

$$\text{SAS}^{(i)} = \text{Polarity}^{(i)} \times \bar{A}^{(i)}$$

For a full transcript T, the aggregate Strategic Ambiguity Score is computed as:

$$\text{SAS}(T) = \sum_{i=1}^{N} \text{SAS}^{(i)}$$

The polarity term $(P_{\text{pos}} - P_{\text{neg}})$ captures the directional sentiment leaning, crucial for differentiating between positively and negatively ambiguous statements. The hedge/negation indicators $H^{(i)}$ and $N^{(i)}$ act as binary gates, ensuring only sentences exhibiting linguistic ambiguity are considered. The attention weights $a^{(i)}_j$ are used as token-level importance scores. The multiplication between sentiment polarity and token-level attention allows the model to capture instances where ambiguity arises from emphasis on hedged or negated sentiments. Finally, aggregating over sentences with a summation aligns with the assumption that strategic ambiguity is an additive effect across the transcript.

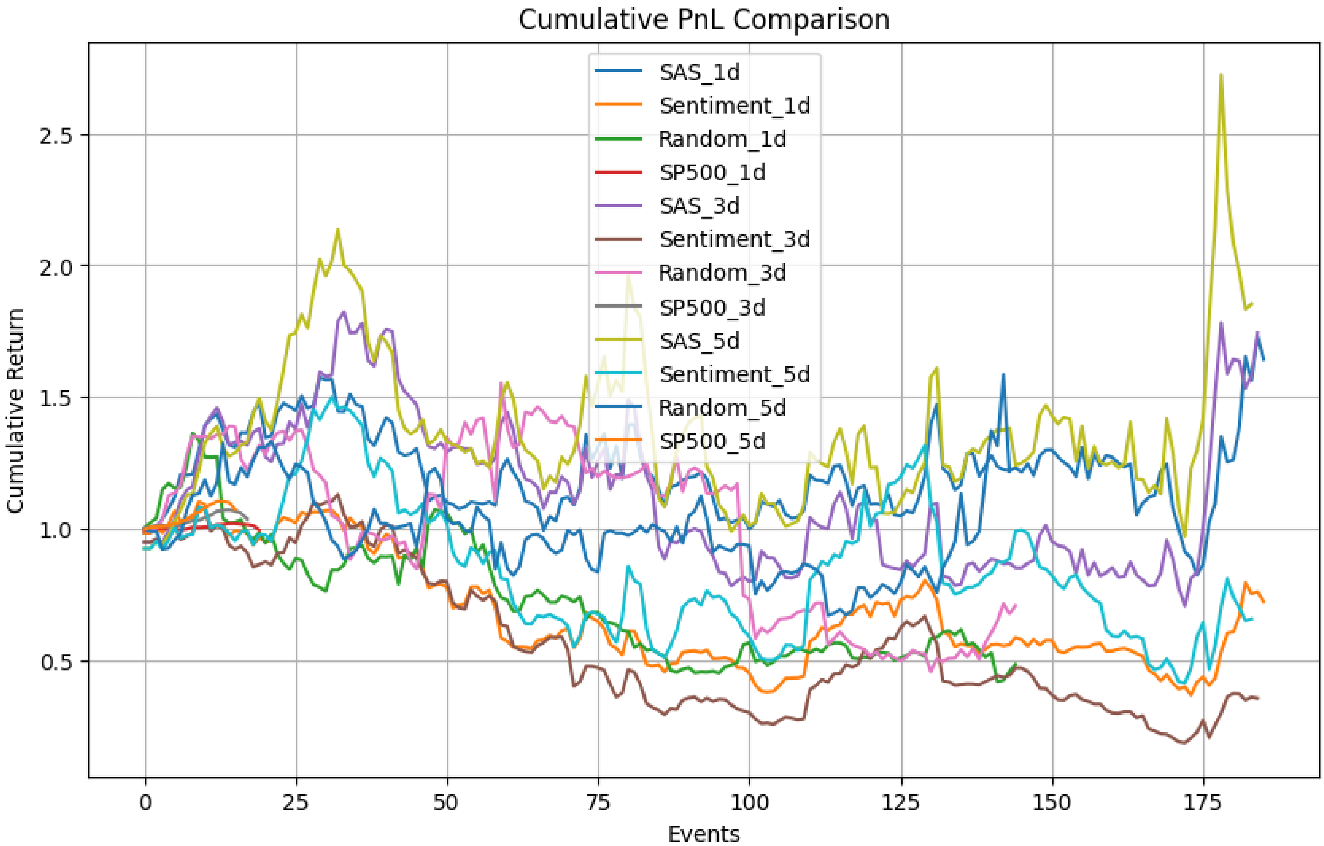

Figure 1: Cumulative PnL comparison for the random benchmark, SP500, sentiment-based, and SAS-based approaches. Cumulative PnL comparison

Vectorized Formulation Defining vectors:

$$\mathbf{p}^{(i)} = P_{pos}^{(i)} P_{neg}^{(i)}, \quad \mathbf{h}^{(i)} = H^{(i)} \vee N^{(i)}, \quad \mathbf{a}^{(i)} = \{a_j^{(i)}\}_{j=1}^{M_i}$$

The sentence-level score can be rewritten as:

$$SAS^{(i)} = \mathbf{h}^{(i)} \cdot \left((P_{pos}^{(i)} - P_{neg}^{(i)}) \times \frac{1}{|A^{(i)}|} \sum_{a_j \in A^{(i)}} a_j \right)$$

The transcript-level SAS remains:

$$SAS(T) = \sum_{i=1}^{N} SAS^{(i)}$$

Alpha Generation Strategy

The central hypothesis of this study is that strategic ambiguity in corporate disclosures contains exploitable alpha, a concept in finance to describe returns that exceed what would be expected based on general market movements or risk exposure. Alpha essentially represents the part of a stock's performance that is due to a trading strategy or informational advantage, beyond overall market trends. In the context of our study, alpha is the ability to generate excess returns by identifying subtle linguistic cues, specifically, strategic hedging and ambiguity in managerial statements that the broader market has not fully priced in. This delayed market reaction is also called post-earnings announcement drift (PEAD), which is a financial phenomenon where the stock prices continue to adjust gradually following an earnings release, rather than immediately reflecting all available information (Fink 2021). By quantifying strategic ambiguity in corporate disclosures, our Strategic Ambiguity Score (SAS) seaks to capture the subtle signals that contribute to PEAD, allowing trading portfolios to exploit these delayed responses and generate measurable alpha.

To test this hypothesis, we construct long-short portfolios based on the Strategic Ambiguity Score (SAS). Long-short portfolios are designed to exploit differences in expected returns: we take "long" positions in firms with high SAS values (betting their stock will outperform) and "short" positions in firms with low SAS values (betting their stock will underperform). These portfolios are then benchmarked against strategies using traditional sentiment measures, random signals, and the SP500 index to determine whether SAS captures unique, actionable signals that translate into excess returns beyond standard market performance.

For each earnings call event, we extract the textual content and compute both the Strategic Ambiguity Score (SAS) and Sentiment Polarity scores. Events are then sorted into quintiles based on their respective scores, and a long-short portfolio is constructed by going long on the top quintile (highest ambiguity or sentiment) and short on the bottom quintile.The cumulative returns are tracked across multiple holding periods (1-day, 3-day, and 5-day post-event) to capture both immediate and delayed market reactions. The performance is benchmarked against:

- Random signals (control group)
- SP500 returns (market baseline)
- Sentiment-based strategies

We evaluate portfolio performance using:

- Mean Daily Return - average return per holding window
- Sharpe Ratio - risk-adjusted return measure
- Maximum Drawdown - largest peak-to-trough loss
- t-statistic and p-values - statistical significance of returns

Results

The empirical results strongly indicate that the Strategic Ambiguity Score (SAS) is a meaningful predictor of short-term stock returns following earnings call events. Across multiple holding periods, SAS-driven portfolios consistently outperform sentiment-based strategies, random portfolios, and the SP500 benchmark. In particular, the 5-day SAS strategy (SAS-5d) achieves the highest cumulative return, exceeding 2.5, which is significantly greater than both sentiment-driven and random strategies. This suggests that the market reacts to strategic ambiguity with some delay, allowing alpha extraction over multiple trading days. Shorter-term strategies, such as SAS-1d and SAS-3d, also demonstrate positive performance relative to controls, indicating that immediate market reactions exist but tend to amplify over subsequent days, likely as analysts and investors interpret the hedging and ambiguity in managerial language.

Risk-adjusted performance metrics further support the value of SAS-based strategies. SAS portfolios achieve Sharpe ratios consistently above 1.0, with SAS-1d at 1.17 and SAS-5d reaching1.28, while sentiment-driven strategies show negligible or even negative Sharpe ratios. This indicates that SAS not only generates positive returns but does so in a risk-efficient manner. Maximum drawdown analysis demonstrates that SAS strategies are relatively resilient to short-term reversals, with drawdowns ranging from -0.477 to -0.613, comparable to or better than sentiment-based portfolios. The observed stability suggests that ambiguity-based signals are less sensitive to noise than naive sentiment measures, reflecting the systematic nature of strategic linguistic choices.

Statistical evaluation provides additional nuance. Although the 1-day SAS strategy has a t-statistic of 1.0057 ($p = 0.3159$) and is not conventionally significant, the overall trend across longer horizons points toward cumulative significance, consistent with the delayed assimilation of nuanced linguistic signals in financial markets. In contrast, sentiment-based strategies fail to deliver consistent alpha. Sentiment-1d, for example, exhibits near-zero mean return and essentially zero Sharpe ratio, while 3-day and 5-day sentiment portfolios show minimal or negative returns, confirming that simple sentiment polarity fails to capture the subtle, forward-looking signals embedded in managerial hedging. The SP500 benchmark, while largely stable, is not alpha-generative in a long-short framework, reinforcing that SAS captures firm-specific, actionable information that broad market indices overlook.

The outperformance of SAS strategies can be interpreted through several mechanisms. Managers' use of hedging and ambiguity may require additional analyst interpretation, which explains why returns tend to increase over multi-day holding periods. Higher SAS values likely reflect deliberate signaling, where executives subtly mask optimism or downplay uncertainty, producing market-relevant effects that are detectable via our structured metric. By isolating linguistic opacity rather than raw sentiment, SAS reduces exposure to generic statements or overt emotional content, improving the signal-to-noise ratio in portfolio construction.

Furthermore, random portfolios underperform consistently across all holding periods, demonstrating that the positive returns of SAS strategies are systematically driven by the underlying linguistic signals. These results support the overall conclusion that strategic ambiguity represents a novel, monetizable alpha factor. By quantifying the deliberate placement of hedges, negation, and attention-salient language, SAS captures dimensions of executive communication beyond just sentiment or factual tone, leading to an actionable value for textual alpha generation.

Conclusion and Future Work

This study introduces strategic ambiguity as a measurable factor using linguistic analysis and machine learning attribution methods. By developing the Strategic Ambiguity Score (SAS), we show that deliberate use of hedges, negations, and nuanced language, especially when combined with model attention, can reveal hidden signals in corporate communications. Our analysis demonstrates that SAS captures subtle patterns in text that are systematically detectable and predictive of subsequent outcomes, providing a new way for AI to extract actionable insights from unstructured language beyond simple sentiment analysis.

The potential applications of this framework extend well beyond finance. SAS offers a generalizable approach for understanding intentional vagueness or hedging in organizational communications, which could support decision-making, risk assessment, or monitoring of corporate messaging. Future work could integrate vocal tone and visual cues through multimodal AI models, expand to non-English communications to study cross-cultural differences in ambiguity, and incorporate adaptive, reinforcement learning-based SAS weighting to respond dynamically to changing contexts. By treating ambiguity as a quantifiable signal, this work lays the foundation for AI systems that can transform subtle linguistic patterns into interpretable, actionable intelligence across a variety of domains.

References

15 U.S.C. § 78u-5. 2022. 15 U.S.C. § 78u-5. Application

of safe harbor for forward-looking statements. https://www.law.cornell.edu/uscode/text/15/78u-5. United States Code, 2018 Edition, Supplement 4, Title 15; Laws in effect as of Jan. 5, 2023.

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

Bachenko, J.; Fitzpatrick, E.; and Schonwetter, M. 2008. Verification and implementation of language-based deception indicators in civil and criminal narratives. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, 41–48.

Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv preprint arXiv:1409.0473*.

Beattie, V.; Dhanani, A.; and Jones, M. J. 2008. Investigating presentational change in UK annual reports: A longitudinal perspective. *The Journal of Business Communication (1973)*, 45(2): 181–222.

Bibal, A.; Cardon, R.; Alfter, D.; Wilkens, R.; Wang, X.; François, T.; and Watrin, P. 2022. Is Attention Explanation? An Introduction to the Debate. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 3889–3900. Dublin, Ireland: Association for Computational Linguistics.

Call, A. C.; Flam, R. W.; Lee, J. A.; and Sharp, N. Y. 2024. Managers' use of humor on public earnings conference calls. *Review of Accounting Studies*, 29(3): 2650–2687.

Camiciottoli, B. C. 2010. Earnings calls: Exploring an emerging financial reporting genre. *Discourse & Communication*, 4(4): 343–359.

Colasanto, F.; Grilli, L.; Santoro, D.; and Villani, G. 2022. BERT's sentiment score for portfolio optimization: a fine-tuned view in Black and Litterman model. *Neural Computing and Applications*, 34(20): 17507–17521.

Duran, N. D.; Hall, C.; McCarthy, P. M.; and McNamara, D. S. 2010. The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied psycholinguistics*, 31(3): 439–462.

Fatouros, G.; Soldatos, J.; Kouroumali, K.; Makridis, G.; and Kyriazis, D. 2023. Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, 14: 100508.

Fink, J. 2021. A review of the Post-Earnings-Announcement Drift. *Journal of Behavioral and Experimental Finance*, 29: 100446.

Frankel, R.; Johnson, M.; and Skinner, D. J. 1999. An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research*, 37(1): 133–150.

Frankel, R.; Mayew, W. J.; and Sun, Y. 2010. Do pennies matter? Investor relations consequences of small negative earnings surprises. *Review of Accounting Studies*, 15(1): 220–242.

Gibbins, M.; Richardson, A.; and Waterhouse, J. 1990. The management of corporate financial disclosure: Opportunism, ritualism, policies, and processes. *Journal of Accounting Research*, 28(1): 121–143.

Gössi, S.; Chen, Z.; Kim, W.; Bermeitinger, B.; and Handschuh, S. 2023. Finbert-fomc: Fine-tuned finbert model with sentiment focus method for enhancing sentiment analysis of fomc minutes. In *Proceedings of the Fourth ACM International Conference on AI in Finance*, 357–364.

Hao, Y.; Dong, L.; Wei, F.; and Xu, K. 2021. Self-attention attribution: Interpreting information interactions inside transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 12963–12971.

Iacovides, G.; Konstantinidis, T.; Xu, M.; and Mandic, D. 2024. Finllama: Llm-based financial sentiment analysis for algorithmic trading. In *Proceedings of the 5th ACM International Conference on AI in Finance*, 134–141.

Jiang, T.; and Zeng, A. 2023. Financial sentiment analysis using FinBERT with application in predicting stock movement. *arXiv preprint arXiv:2306.02136*.

Kang, J.-W.; and Choi, S.-Y. 2025. Comparative investigation of gpt and finbert's sentiment analysis performance in news across different sectors. *Electronics*, 14(6): 1090.

Kearney, C.; and Liu, S. 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33: 171–185.

Kimbrough, M. D. 2005. The effect of conference calls on analyst and market underreaction to earnings announcements. *The Accounting Review*, 80(1): 189–219.

Lakoff, G. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of philosophical logic*, 2(4): 458–508.

Liu, Y.; Li, H.; Guo, Y.; Kong, C.; Li, J.; and Wang, S. 2022. Rethinking Attention-Model Explainability through Faithfulness Violation Test.

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

Loughran, T.; and McDonald, B. 2014. Measuring readability in financial disclosures. *the Journal of Finance*, 69(4): 1643–1671.

Potterer, T. 2023. Scraped Motley Fool Earnings Call Transcripts (Kaggle Dataset). https://www.kaggle.com/datasets/tpotterer/motley-fool-scraped-earnings-call-transcripts. Accessed: 2025-08-04; 18,755 earnings-call transcripts scraped from Motley Fool.

Rekabsaz, N.; Lupu, M.; Baklanov, A.; Hanbury, A.; Dür, A.; and Anderson, L. 2017. Volatility prediction using financial disclosures sentiments with word embedding-based IR models. *arXiv preprint arXiv:1702.01978*.

Theil, C. K.; Štajner, S.; and Stuckenschmidt, H. 2020. Explaining financial uncertainty through specialized word embeddings. *ACM Transactions on Data Science*, 1(1): 1–19.

Todd, A.; Bowden, J.; and Moshfeghi, Y. 2024. Text-based sentiment analysis in finance: Synthesising the existing literature and exploring future directions. *Intelligent Systems in Accounting, Finance and Management*, 31(1): e1549.

Tsai, M.-F.; and Wang, C.-J. 2014. Financial keyword expansion via continuous word vector representations. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1453–1458.

Tsai, M.-F.; Wang, C.-J.; and Chien, P.-C. 2016. Discovering finance keywords via continuous-space language models. *ACM Transactions on Management Information Systems (TMIS)*, 7(3): 1–17.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, ; and Polosukhin, I. 2017. Attention Is All You Need. *arXiv preprint arXiv:1706.03762*.

Williams, C. 2008. Toward a Taxonomy of Corporate Reporting Strategies. *The Journal of Business Communication (1973)*, 45(3): 232–264.

Yang, K.; Xu, H.; and Gao, K. 2020. Cm-bert: Cross-modal bert for text-audio sentiment analysis. In *Proceedings of the 28th ACM international conference on multimedia*, 521–528.