https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 71, No. 11, November 2025, pp. 9831–9855
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The News in Earnings Announcement Disclosures: Capturing Word Context Using LLM Methods

Federico Siano[a]

[a] Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080
Contact: federico.siano@utdallas.edu, https://orcid.org/0000-0001-9493-8363 (FS)

Received: March 25, 2024
Revised: December 10, 2024
Accepted: February 4, 2025
Published Online in Articles in Advance: April 7, 2025

https://doi.org/10.1287/mnsc.2024.05417

Abstract. This study examines the information content of textual disclosures in firms' earnings announcements. Using a large language model (LLM) to capture information in both words and word context, I show that the news in earnings press releases (i) explains three times more variation in short-window stock returns than a host of textual measures based on dictionary and non-LLM machine learning methods; (ii) doubles the R^2 of an array of financial statement surprises, modeled with conventional regression or machine learning approaches; and (iii) accounts for a large fraction of immediate price revisions within just five minutes of release. LLM-modeled conference calls further enhance R^2 by one fourth compared with press releases and financial surprises. Textual disclosures are more informative when earnings are less persistent and during periods of aggregate uncertainty. Most news arises from text describing numbers, at the beginning of the disclosure, and including novel contents. These findings highlight the role of firms' textual disclosures in moving stock prices and advance our understanding of how investors utilize corporate disclosures.

History: Accepted by Suraj Srinivasan, accounting.
Funding: The author gratefully acknowledges financial support from the Naveen Jindal School of Management.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2024.05417.

Keywords: disclosure • earnings announcements • textual analysis • large language models (LLMs) • machine learning

1. Introduction

Starting from Ball and Brown (1968), a prolific stream of empirical literature has explored the valuation role of accounting information, with a focus on reported earnings (Kothari 2001). Despite the perception that accounting information is widely utilized by equity investors, researchers have documented a relatively low association of earnings with stock returns (Lev 1989, Lev and Zarowin 1999). The overall conclusion from this literature is that investors do not respond strongly to earnings numbers because their information content is already embedded in prices. Consequently, more recent work investigates the valuation role of textual information (Li 2010, Loughran and McDonald 2016, Bochkay et al. 2022). However, this literature faces a major empirical challenge because language underlies a complex structure that is hard to model. In particular, traditional and recent textual analysis approaches (Frankel et al. 2022) largely ignore *word context*—that is, the sequential semantic relations across words. Such limitation has likely contributed to the seemingly low explanatory power of earnings press releases for stock market outcomes.

To overcome the limitations of prior textual analysis methods, this study uses a large language model (LLM) to capture information in both words and word context within earnings announcement disclosures and thus better characterize the information content of such disclosures. I use this new approach to address two questions. First, how much news is there in earnings announcement text beyond accounting numbers? Second, under what economic circumstances and which textual information in earnings announcements do investors value? Revisiting these questions is important because, using standard approaches, researchers have had limited success in explaining investor response to earnings announcements based on accounting numbers or simple measures of textual disclosures. My findings suggest that textual disclosures are substantially more informative than previously documented and that LLMs can shed light on *how* investors utilize accounting text for valuation.

The limitations of traditional textual analysis approaches stem from characterizing disclosures based on individual words, largely ignoring the complex linguistic structure underlying corporate disclosures or focusing on single-dimension attributes of text

(e.g., "disclosure tone/sentiment"). These approaches lead to (i) measurement error—because words within disclosures are strongly interconnected—and (ii) an incomplete characterization of information within text. For example, the word "loss" is classified as a "negative" term by traditional approaches; however, if used in the context of a one-time (i.e., transitory) charge, the same term might not convey "bad news." Similarly, a "loss" resulting from aggressive investments in new technology can signal a strategic move to secure future growth and thus be considered "good news." Therefore, ignoring word context in corporate disclosures leads to underestimating the information content of accounting text.

The new approach I implement involves fine-tuning a current-generation LLM to predict short-window market reactions to firms' earnings announcement disclosures. Specifically, I infer the "news" in the text of earnings announcement press releases by fine-tuning an LLM to predict the magnitude and direction of abnormal stock returns in the two-day window surrounding the earnings announcement.[1] Thus, I derive a single summary information variable synthesizing the news in the text, based on words and word context, within such disclosures.

Fine-tuning an LLM to directly predict the magnitude and direction of short-window stock returns provides several key benefits, beyond capturing information in both words and word context. First, this approach allows the LLM to learn the linkages between information contained in textual disclosures and how market participants incorporate this information into stock prices *without* the need to specify an investor expectation model. Second, this empirical choice is relevant because while a theory exists linking income numbers to the market valuation of firms' equity (Ohlson 1995), the mapping of qualitative disclosure to market valuation remains underdeveloped. For example, it is theoretically unclear how "positive" or "negative" words in a sentiment proxy map into firm value. Third, training an LLM to explain returns, rather than using the returns themselves, improves our ability to identify the information content of textual disclosures, because returns capture a mixture of private and public information, including reported numbers (French and Roll 1986, Boudoukh et al. 2018).

I begin the empirical analysis by examining the information content of earnings announcement text. Using an LLM approach, I find that the text in earnings press releases explains at least *three* times more variation in two-day abnormal stock returns around earnings announcements than a host of conventional textual proxies derived from simple dictionary-based approaches or other non-LLM machine learning techniques. As a comparison, using non-LLM machine learning methods, Frankel et al. (2022, p. 5516) report that earnings press releases explain 4.5% of the variation in two-day abnormal stock returns. In contrast, the LLM approach shows that these textual disclosures explain 15% of the variation in short-window abnormal returns, which corresponds to a 200% gain in (or three-fold) explanatory power. Notably, the extracted textual news offers a 100% average improvement in information content relative to a wide array of financial statement surprises (Hand et al. 2022) alternatively modeled using traditional regression or machine learning methods. Because the relation between announcement-window returns and earnings news is highly nonlinear (Kothari 2001), these latter machine learning methods are especially important to isolate the incremental information content of the extracted textual news relative to reported numbers. Overall, this analysis indicates that prior studies and measurement methods *significantly underestimate* the information content of earnings announcement disclosures because prior approaches do not capture complex word context within the textual disclosures.[2]

I extend the analysis of the news content and timeliness of textual disclosures to other settings. To begin with, I use LLM-modeled earnings press releases to explain *immediate* price revisions within minutes of the disclosure release. Using granular stock price data, I find that textual disclosures provide an R^2 of 8% for five-minute market reactions, enhancing the explanatory power of financial statement surprises by 15%. The incorporation of textual information into prices occurs rapidly and continues as the event window extends. In fact, the contribution of textual disclosures to explaining returns grows steadily, increasing from 15% in the 5-minute window to 18% in the 30-minute window, and reaching 25% in the 2-hour window—almost doubling the value added by textual information. This finding provides valuable insights into whether, and to what extent, investors process textual information quickly—an important issue given recent literature highlighting underreaction and inattentiveness to such information (Cohen et al. 2020).

I further extend the analysis to a complementary disclosure channel shaping investor reactions. Earnings conference calls represent another key medium for corporate disclosures during earnings announcements (Matsumoto et al. 2011, Mayew et al. 2020, Frankel et al. 2022). Thus, I apply the LLM approach to analyze conference call transcripts, evaluating their incremental information relative to press release text and financial surprises. My findings indicate that, although press releases and financial surprises contain most announcement period news, conference call text provides substantial added value. Specifically, it enhances R^2 by up to 30% and helps explain both two-day and one-hour conference call returns.

The information content of textual disclosures likely depends on economic factors. To better understand

when—or under what economic conditions—investors find textual disclosures particularly informative, I examine both cross-sectional and time series variation in the information content of press releases. First, I conjecture that textual disclosures will convey greater incremental news to investors, relative to financial statement numbers, in economic scenarios in which accounting earnings exhibit low persistence and therefore are expected to be less useful for valuation. I find that text has up to 70% higher information content for firms reporting (i) losses (Hayn 1995), (ii) large intangible assets (Lev and Zarowin 1999), and (iii) earnings with low autocorrelation (Collins and Kothari 1989). Second, I expect that textual disclosures become particularly useful at the onset of macroeconomic disruptions and other events that increase aggregate uncertainty, as financial metrics may be less effective in predicting firms' future performance in such conditions (Loh and Stulz 2018). My analyses reveal that the incremental news in textual disclosures is pronounced during economic and corporate crises, such as the onset of the COVID-19 pandemic, the U.S.-China trade war, the 2008 financial crisis, and the Enron scandal.

Moreover, the information content of textual disclosures likely depends on the specific contents being disclosed. To better understand *which* textual contents investors find most useful for price revisions, I analyze variation in information content within press releases. Results document that most news arises from text (i) directly describing numbers, (ii) at the beginning of the disclosures, and (iii) exhibiting significant change in contents relative to the prior quarter disclosure.

My findings suggest that the information content and the role of earnings announcement text is conditional on the nature of reported earnings. On the one hand, the *within*–press release analyses suggest that disclosures help investors assess the implications of accounting numbers for firm value (i.e., complement). On the other hand, the *across*–press release analyses reveal that text is especially useful for valuation when earnings are less persistent and correlate less strongly with future cash flows (i.e., substitute). These findings imply that the role of earnings announcement text can vary with the characteristics and decision usefulness of reported earnings.

Results are robust across various sensitivity tests. First, consistent evidence is obtained with alternative out-of-sample approaches to mitigate look-ahead bias, including a 2021–2023 sample *outside* the LLM pretraining period. Second, removing numbers from the text does not significantly alter the results; thus, most news originates from the discussion of numbers rather than the numbers themselves. Third, similar evidence is observed with point-in-time actuals, suggesting that standardization practices do not drive incremental news in text. Fourth, significant news content is measured during announcement periods without conference calls, or analyst/management forecasts, suggesting the findings are not driven by concurrent events. Finally, the LLM extracts text-based information that predicts the level and uncertainty of future earnings and explains other market outcomes such as abnormal volumes and volatility.

This study contributes in several ways to the capital markets literature and the textual analysis literature in accounting and finance. First, I provide new evidence on the information content of corporate disclosures for equity investors (Healy and Palepu 2001, Holthausen and Watts 2001, Kothari 2001). Lev (1989, p. 173) describes accounting earnings as "… an information variable that explains only about 5% of stock return variability and whose relation with returns is unstable." My study is the first to document that accounting text in earnings press releases offers large absolute (i.e., 15% of R^2) and incremental (i.e., an added 10% of adjusted R^2 relative to income numbers) explanatory power for short-window market outcomes. Recent disclosure work documents a substantially lower decision relevance of earnings press releases for short-window pricing (i.e., about 4.5% of absolute R^2 in Frankel et al. (2022) and a 4% gain in classification accuracy over random guessing in Siano and Wysocki (2021)). Although my study does not alter the conclusion that earnings themselves are not timely (Collins et al. 1994), I find that earnings press releases contain significantly more *new* information than previously observed. Unlike prior work, I utilize machine learning techniques to model news in *both* earnings announcement text and reported numbers. Thus, to the best of my knowledge, my study is the first to better quantify not only the absolute but also the incremental information content of earnings disclosures beyond reported numbers, as opposed to capturing large inconsistencies in modeling approaches. My findings extend prior research suggesting that the information content of earnings announcements can derive from information accompanying reported earnings rather than the earnings themselves (Ball and Shivakumar 2008). Moreover, I connect these findings to valuation theory (Ohlson 1995) and show that disclosures map into stock returns through a channel of revised expectations about firms' future cash flows.

Second, my study contributes to the extensive and influential literature in accounting and financial economics examining the relation between asset prices and information. This body of research has long sought to understand what drives stock prices and the role of public information in this context. Seminal studies such as Ball and Brown (1968), French and Roll (1986), Roll (1984, 1988), and Ball and Shivakumar (2008) laid the groundwork for this research area, which continues to evolve with recent contributions from Boudoukh et al. (2018), Shao et al. (2021), Hand

et al. (2022), and Brogaard et al. (2022). This body of research collectively highlights that price movements on public news days reflect a complex interplay of (i) public firm-specific news (both company- and non–company-sourced), (ii) private information revealed through rational trading, and (iii) mispricing. Hirshleifer (2001, p. 1560) states that "Little of stock price […] variability has been explained empirically by relevant public news." My study highlights that a large fraction of stock price movements around earnings announcements is driven by public news originating from accounting-related *textual disclosures*.

Third, this study enhances our understanding of investors' ability to process textual information quickly and uncover the implications of earnings news in corporate disclosures. While my findings do not alter the conclusion that textual information is gradually incorporated into prices over time and not fully reflected instantaneously, the analysis shows that equity investors respond promptly to corporate text—within minutes of the disclosure release. These insights provide a new perspective on investor inattention to textual disclosures during earnings announcements (Cohen et al. 2020, Meursault et al. 2022).

Fourth, I build on the textual analysis literature in accounting and finance (Li 2010, Loughran and McDonald 2016, Bochkay et al. 2022) to examine *when* (i.e., in which economic circumstances) earnings announcement text contains most news, and *which* textual content is most useful to equity investors. Unlike recent textual analysis work that primarily focuses on predicting returns (Frankel et al. 2022) or accounting performance (Siano and Wysocki 2021), my study conducts a more nuanced analysis of disclosures following the recent call in Bochkay et al. (2022). I show that accounting text is most relevant for valuation when reported earnings are less useful due to lower persistence or aggregate uncertainty and that most news derives from text that explains reported numbers. These findings advance our understanding of disclosure quality, highlighting when and which textual disclosures matter to investors.

Finally, I offer a methodological contribution by explaining and applying a regression-based (rather than the usual binary classification) LLM approach to extract information relevant for valuation from narrative text. Specifically, I use an LLM to extract a single information variable that captures the combined news in both words and word context. My study is the first to show that LLMs can predict not only daily but also *minute-level* signed stock returns using corporate text. This finding is relevant as language models learn from data variability (Bishop 2006), whereas immediate price revisions exhibit only half the variance observed in daily returns. I also demonstrate the broad applicability of LLMs in disclosure analysis, highlighting their ability to process different texts (e.g., press releases, earnings call transcripts) of arbitrary length. Overall, I extend measurement insights into recent "big data" applications for textual accounting disclosures.

2. Background and Prior Literature

2.1. Disclosure News in Earnings Announcements

As noted in Kothari (2001, p. 116), a long strand of capital markets literature has analyzed price changes around accounting events to infer new information about the timing, amount, or uncertainty of firm's future cash flows. This literature has focused extensively on earnings announcement events and analyzed in detail the information content of reported earnings. The general conclusion from these studies is that earnings contain some new information, but the earnings number is also largely anticipated by market participants (Collins et al. 1994). Notably, accounting numbers are usually accompanied by disclosures that explain those numbers (e.g., press releases). Thus, several studies have attempted to analyze the information content of textual disclosures (e.g., see the survey by Loughran and McDonald (2016)). However, characterizing news in such disclosures has been empirically challenging due to the limitations in modeling complex word context. These limitations (i) likely contribute to measurement error, (ii) can lead to underestimating the information content of corporate disclosures, (iii) and can explain why prior studies have found only modest news in textual disclosures released during earnings announcement events.

Notably, the economic magnitude of earnings disclosure news remains a subject of empirical inquiry for several reasons, beyond measurement error. First, news may be limited due to managers' disclosure behavior. Frankel et al. (2022) find modest information content in earnings press releases, suggesting that these disclosures include cautious and legally protective language making them less spontaneous and dynamic compared with other corporate text. Second, investors may need significant time to acquire and process textual information. Cohen et al. (2020) and Meursault et al. (2022) show that prices tend to react with delay to textual disclosures. Finally, finance research shows that public news explains only a limited portion of stock price movements—alternatively driven by private information revealed through rational trading (Roll 1988) or irrational trading (Hirshleifer 2001)—and a relevant fraction of this public news includes information sources that are external to the firm, such as media (Boudoukh et al. 2018).[3]

2.2. Conventional and Recent Textual Analysis Approaches

The limitations of traditional textual analysis approaches have long been recognized. Loughran and McDonald (2016, p. 1190) point out that a significant limitation of the textual analysis literature is the issue of "moving beyond assuming words occur as independent units." In fact, conventional techniques employ "bag-of-words" methods, including dictionary approaches, which typically count word frequencies and treat words without considering their position in a document, thereby losing contextual connections. Also, traditional approaches often focus only on subsets of keywords to characterize single-dimension attributes of accounting text, such as "disclosure tone/sentiment." This results in an incomplete characterization of the information content of corporate disclosures—because accounting text includes more information than simply "positive" or "negative" words. Finally, the design of textual attributes (e.g., "disclosure tone" defined as [positive–negative]/total words) largely reflects researchers' pragmatism rather than theory.[4]

To overcome these limitations, recent studies have applied machine learning methods to extract information from disclosures (e.g., see the survey by Bochkay et al. (2022)). Recent approaches generally (i) extract one or two-word phrases (i.e., "N-Grams") from disclosure text, (ii) model their occurrence and co-occurrence through nonlinear functions, and (iii) fit the text to a classification or regression outcome (Jegadeesh and Wu 2013, Donovan et al. 2021, Frankel et al. 2022). Nonlinear functions can capture some interactions across words; text can be characterized based on a notion of "local word context" (e.g., two consecutive words). However, these machine learning methods are still unable to capture the semantic role of and similarities among words in a document (e.g., whether "reported" is a verb and whether it is similar to "recognized"), and they largely ignore word order (e.g., "did not report a loss related to" and "did report a loss, not related to" might be treated equally). Therefore, these approaches are unable to overcome the fundamental barrier of characterizing complex word context, which likely results in *underestimating* the information content of corporate disclosures.

2.3. Benefits of LLMs for Disclosure Analysis

Recent advances in LLM show promise in mitigating measurement error and offering a more complete characterization of the information content of corporate disclosures. An LLM represents a type of deep neural network that has been pretrained on a vast amount of textual data (e.g., the universe of Wikipedia posts) and can perform a wide range of natural language processing tasks (e.g., text classification). LLMs combine *two* key advancements over traditional and recent textual analysis methods.

First, LLMs can capture the stand-alone "meaning" of words. They achieve this goal by leveraging "word embeddings" (Mikolov et al. 2013). Word embeddings represent each word in a language as a high-dimensional vector (e.g., with more than a thousand dimensions) that captures the stand-alone meaning of words by analyzing their semantic relationships with other words in a language. For example, the vector representation of the term "stock" will be relatively closer to that of "equity" and "shareholders" compared with "bond," which is closer to "maturity" and "credit." This nuanced representation allows LLMs to interpret the meaning of words more accurately, enhancing their ability to process and understand corporate disclosures. In contrast, most textual analysis approaches to measure the information content of text do not utilize embeddings and, as a result, characterize words based on a unidimensional representation (Loughran and McDonald 2011, Frankel et al. 2022). In fact, traditional methods often rely on word counts, lacking the nuanced understanding of semantic relationships across words that embeddings provide.

Second, large language models can capture the "conditional meaning" of words in a disclosure, conditional on the surrounding words and their order. Saying this differently, LLMs can capture complex *word context*. Specifically, recent LLMs employ "attention-based transformers" (Vaswani et al. 2017), which represent each word in a document, not only based on its semantic relationship to other words in a language (through embeddings), but also in terms of its ordered relationship to any other words in the same disclosure. This architecture allows for a rich and contextualized characterization of linguistic features within textual documents. For instance, although the term "volatility" can signal both an uptick in revenues and an increase in downside risk, an LLM can "understand" that, when used in the word context of supply chain logistics, the same term identifies a downside risk factor rather than a market opportunity. Notably, capturing complex ordered relationships across words in a disclosure has represented a key barrier for traditional and recent textual analysis methods in accounting and finance. Nonetheless, sensitivity to word order is an especially desirable feature for a language model because it allows to more closely mimic investors' processing of text and therefore better understand the nuances of their decision process.

LLMs have additional benefits and costs. They are usually readily and freely available to users through open-source versions: This represents a major advantage given the high computational intensiveness of large-scale pretraining. However, LLMs have disadvantages as well. Even pretrained models, given their

size, tend to require more resources, in terms of computational time, hardware, and energy consumption, compared with simpler textual analysis techniques. Thus, their utilization is justified for complex applications.

2.4. Which LLM? BERT-Based Models

Two well-known classes of LLMs are *BERT* (bidirectional encoder representations from transformers; Devlin et al. 2019) and *GPT* (generative pretrained transformer; Brown et al. 2020). Although they share common features, such as the large-scale pretraining, these LLMs are also characterized by structural differences. Unlike *GPT* models, *BERT*-based LLMs are pretrained using a "bidirectional" approach, wherein for each linguistic token, both the preceding and following portions of text (i.e., word context) characterize its semantic role. On the other hand, *GPT* models are pretrained based on a "left-to-right" approach. For this reason, *BERT*-based models are particularly well suited to capture complex word context based on surrounding words, whereas *GPT* models are designed to generate coherent text based on preceding words only. Thus, in this study I employ a *BERT*-based LLM to better characterize disclosure news by capturing complex word context.

BERT-based LLMs operate through "transfer learning" (see the survey by Pan and Yang 2010). The large language model (i.e., deep neural network) is first pretrained on large-scale texts to learn generalizable features of language (such as whether a word is a noun or a verb, or subject-object relations; Jawahar et al. 2019, Tenney et al. 2019). Subsequently, the model's "knowledge" (i.e., the set of learned features and parameters) is transferred or "fine-tuned" (i.e., adapted) to a particular domain of interest (e.g., financial accounting disclosures). As previously noted, the pretrained model is freely available to users—a major advantage given the high computational intensiveness of large-scale pretraining. Thus, researchers only need to fine-tune the model. Fine-tuning corresponds to a supervised learning task; the neural network is fit to narrative disclosures labeled with a chosen outcome (e.g., two-day returns), and it learns a set of parameters mapping linguistic features to that outcome. This adapts the final layers of the pretrained neural network's parameters to the linguistic nuances of the reference domain, a step particularly important for domains characterized by technical terminology such as financial accounting.

2.5. BERT-Based Applications in Accounting Research

Some recent studies in accounting have applied *BERT*-based LLMs. The primary goal of these studies has been mostly limited to binary classification involving the tone of earnings disclosures (e.g., positive versus negative) or accounting performance (e.g., above median versus below median).[5] Moreover, these applications have usually focused on a limited portion of financial disclosures (e.g., the initial sentences) or on single-dimension attributes of text (e.g., disclosure tone/sentiment). As a result, although valuable, these applications still led to an incomplete characterization of the information content of textual disclosures. For example, Siano and Wysocki (2021) use a *BERT*-based LLM to classify the direction (i.e., above median or below median) of contemporaneous financial performance (i.e., sales and net income) using the initial sentences of a small sample of earnings announcement disclosures. Interestingly, this study also documents that the LLM's ability to classify the *sign* of stock returns is significant only for long-window (as opposed to short-window) returns—a finding that is primarily attributed to prices leading disclosures. Huang et al. (2022) utilize a similar LLM to classify the tone of earnings conference calls. Again, they find that an LLM-based tone measure performs better than dictionary approaches but offers very modest explanatory power for short-window market reactions. Kim and Nikolaev (2024) apply a *BERT*-based LLM to a subset of sentences within 10-K reports to classify the direction of changes (i.e., increase or decrease) in future performance and long-window stock returns.

My work advances existing literature in four significant ways. First, I apply a *BERT*-based LLM for predicting both the magnitude and direction of market reactions, thereby adapting the LLM to perform regression analysis instead of its usual task of classification. This approach is crucial because binary classification (typical in prior applications) fails to take advantage of the information in the magnitude of the market reactions, leading to underestimating the information content of textual disclosures. Notably, training an LLM to learn how financial text maps into market reactions is a valuable approach to quantify the information content of earnings disclosures, especially in the absence of (i) a clear theory mapping disclosures into a firm's value and (ii) an observable investor expectation model (see Section 1). Second, I focus on short-window stock returns around earnings announcements rather than accounting performance or longer-window returns. This choice aims to better identify information that triggers changes in investors' beliefs. Third, I examine the full text of earnings disclosures—instead of a select subset of sentences—by employing new techniques to overcome some of the structural limitations of current-generation LLMs. Finally, I employ interpretability tools, specifically the "SHAP (SHapley Additive exPlanations)" framework (Lundberg and Lee 2017), to bridge the gap between "machine understanding" and "human understanding" by highlighting the words and words-in-context the LLM identifies as most influential.

3. Data and Methodology

3.1. Setting and Sample

I examine the extent to which text in earnings press releases, characterized using an LLM, explains contemporaneous investors' reactions. I primarily focus on a short-window setting that includes the two days surrounding earnings announcements (Ball and Shivakumar 2008). However, I also extend the analysis to include the immediate price response within 5 minutes, 30 minutes, and 2 hours of the disclosure release because prior research documents a substantial market reaction to earnings announcements in the minutes following their release (Patell and Wolfson 1984, Jiang et al. 2012). This latter setting (i) enhances identification and (ii) offers insights into the timeliness of corporate disclosures and investors' processing of firms' textual information (Collins et al. 1994, Cohen et al. 2020). I examine signed stock returns to assess the extent to which accounting text helps investors distinguish good from bad news (Ball and Brown 1968) and focus on the earnings announcement setting, because prior research highlights the relatively stronger news content of earnings announcements compared with other disclosures, such as 10-K reports (Li and Ramesh 2009).

In the context of earnings announcements, my focus is on press releases as they are the primary disclosures, yet prior research indicates they provide limited textual information content (Frankel et al. 2022). In addition, compared with conference calls, press releases are available for a broader cross-section of firms and are less subject to self-selection issues. Nevertheless, given the growing importance of earnings conference calls, I also examine the incremental news provided by their textual transcripts.

I source earnings press releases from 8-K filings obtained through the SEC EDGAR (U.S. Securities and Exchange Commission - Electronic Data Gathering, Analysis, and Retrieval system) database and earnings call transcripts from Capital IQ Transcripts. I gather daily stock returns from CRSP (Center for Research in Security Prices), whereas after-hour and before-hour returns are collected from TAQ (Trade and Quote). Quarterly fundamentals are sourced from Compustat, analyst quarterly estimates from I/B/E/S (Institutional Brokers' Estimate System), and management guidance from I/B/E/S Guidance. Earnings announcement (conference calls) timestamps are sourced from I/B/E/S (Capital IQ).[6]

I retain unique firm-quarter earnings announcements, excluding texts with fewer than 10 sentences or 250 total words. I also exclude generic cautionary statements unlikely to affect investors' decisions, as well as tables due to noncomparable variation in their layout across firms and quarter-years. I use the remaining text of the selected disclosures.[7] Finally, I exclude observations that suggest a weaker expected connection between disclosures and market reactions, including (i) "penny stocks" (i.e., priced below $1 per share), (ii) low-volume stocks (i.e., with daily trading volumes in the bottom 1% of the distribution), and (iii) small companies (i.e., with total assets in the bottom 1% of the distribution).

The final sample includes 98,171 quarterly earnings press releases with available stock returns and financial statement variables over 2014–2023. Of this sample, 56,670 observations are characterized by concurrent (i.e., same-day) earnings conference calls with available textual transcripts. Table 1 presents the sample and the selection criteria.

Table 1. Sample Selection

	Number
EDGAR 8-K filings including quarterly earnings announcement press releases (2006-2023)	253,727
Less:	
Observations with fewer than 10 sentences or 250 total words	(3,451)
Observations with missing CRSP or TAQ return data	(4,540)
"Penny stocks," low volume stocks (i.e., bottom 1%), and small stocks (i.e., bottom 1%)	(15,831)
Total financial disclosures available for fine-tuning and out-of-sample modeling	**229,905**
Fine-tuning observations (2006–2013)	109,687
Out-of-sample observations (2014–2023)	120,218
Less:	
Observations with missing I/B/E/S match or Compustat items	(22,031)
Total number of firm-quarter-press releases available for empirical tests	**98,171**
Available firm-quarters with concurrent press releases and conference calls	**56,670**

Notes. This table presents selection criteria for the primary sample of 98,171 firm-quarter observations with available (i) SEC EDGAR 8-K filings including earnings announcements press releases, (ii) daily CRSP or TAQ stock return data, and (iii) financial variables. This sample comprises textual observations characterized by a minimum of 10 sentences and 250 words. "Penny stocks" are defined as stocks with a unit price below $1. Low volume stocks are those in the bottom 1% of the pooled distribution of two-day cumulative trading volume in the announcement window. Small stocks are those in the bottom 1% of the pooled distribution of average announcement-window market value of equity. Fine-tuning observations refer to the years 2006–2013 and are antecedent in time to the out-of-sample observations which refer to the years 2014–2023. In total, 229,905 firm-quarter observations are available for fine-tuning and out-of-sample modeling. Of the 98,171 primary observations, 56,670 feature earnings announcement press releases and conference calls occurring on the same day.

3.2. LLM Implementation and Out-of-Sample Approach

I apply a *BERT*-based LLM to each of the quarterly earnings press releases to extract a single information variable—solely from disclosure text, based on words *and* word context—predicting the magnitude and direction of the disclosure news. I then use the predicted textual news to explain out-of-sample signed abnormal returns during the two-day announcement window.

Specifically, I fit the LLM to the text of quarterly earnings press releases, labeled with two-day cumulative abnormal stock returns around the earnings announcements (i.e., *CAR [0,1]*). This fine-tuned model has learned how text in earnings announcement press releases likely maps into investors short-window reactions (see also Section 2). Once fine-tuned, I use the *BERT* LLM to predict contemporaneous (i.e., same-quarter) short-window stock returns based solely on the out-of-sample earnings announcement texts (i.e., 98,171 disclosures never processed before by the model).[8] I define abnormal returns predicted out-of-sample using press releases and a *BERT* LLM as *CAR LLM EA*.

I similarly implement the previous process for analyses involving (i) immediate price responses within 5 minutes, 30 minutes, and 2 hours following the press release (i.e., 63,249 disclosure observations) and (ii) earnings conference call text (i.e., 56,670 earnings call transcripts).

To mitigate look-ahead bias, I use observations *before* 2014 to fine-tune the model and disclosures *from* 2014 for out-of-sample prediction. I choose 2014 as a cutoff year to ensure that approximately 50% of the observations are used for fine-tuning and the remaining 50% for prediction. A similar protocol is used by Donovan et al. (2021). Notably, this latter approach is especially valuable to mitigate look-ahead bias. In fact, the model is fine-tuned using disclosures and stock returns that are *antecedent* to the out-of-sample disclosures and market reactions.[9]

Compared with prior applications of *BERT* LLMs (Siano and Wysocki 2021, Huang et al. 2022), I introduce three key innovations. First, I address *BERT*'s 512-token sequence limit by splitting earnings announcements into 512-token segments, processing each, and averaging the predicted outputs (e.g., contemporaneous short-window market reactions). Second, I modify the neural network by adding a linear layer to generate continuous outputs, enabling regression tasks instead of the typical classification. Details on hyperparameters, software, and computational time are provided in the Online Appendix. Third, I use the SHAP interpretability algorithm (Lundberg and Lee 2017) to identify which words and words-in-context most influence the LLM's short-window returns predictions. SHAP assigns each word in a disclosure a score reflecting its impact on the predicted stock return. The Online Appendix presents this analysis on select disclosures, illustrating how the *BERT*-based LLM can meaningfully map earnings announcement text to market reactions. Although computationally intensive, this approach offers valuable insights and improves the model's interpretability.

3.3. Other Measures of Information Content in Earnings Announcements

3.3.1. Alternative Textual Measures. I benchmark LLM-based disclosure news to a host of alternative textual proxies, namely (i) simple dictionary-based textual attributes and (ii) recent non-LLM machine learning measures of narrative information content employed by prior research.

The dictionary-based textual proxies include five narrative attributes that prior literature finds associated with investors' reactions (Loughran and McDonald 2016), namely *Tone* (i.e., (positive – negative)/total words using the Henry (2008) dictionary), *Length* (i.e., the natural logarithm of total words in the earnings announcement), *Fog* (i.e., the readability Gunning Fog 1952 index), *Numbers* (i.e., the total number of dollar and percentage amounts disclosed within text), and *Future* (i.e., future tense verbs scaled by total words to capture forward-looking statements).

The machine learning–based textual proxies are developed applying a primary and three alternative algorithms to model the text of earnings announcement press releases. I choose "gradient boosting" (GB; Krupa and Minutti-Meza 2022) as the primary algorithm because it offers the best performance among "random forest" (RF; Frankel et al. 2022), "support vector regressions" (SVR; Frankel et al. 2016), and "supervised latent Dirichlet allocation" (sLDA; Frankel et al. 2022).[10] The goal is predicting out-of-sample market reactions around the release of the earnings disclosures (similar to the LLM prediction task). As primary inputs to these recent non-LLM machine learning models, I employ "N-grams" (in particular, the 1,500 most relevant "unigrams" and "bigrams" or one- and two-word phrases, respectively), which have been used extensively in prior accounting research (Frankel et al. 2022).[11] The resulting prediction is labeled as *CAR GB EA*. The implementation of these recent machine learning techniques follows prior work (Frankel et al. 2022) and its details are described within the Online Appendix. Notably, all the above machine learning techniques applied to textual inputs still face the barrier of capturing complex word context within corporate disclosures.

In additional analyses described within Section 6.2, I also employ simple "word embeddings" (Mikolov et al. 2013) as inputs to gradient boosting models. Embeddings have the advantage of capturing some semantic relationships across words (see also Section 2),

thus representing a more challenging benchmark compared with "N-grams."[12]

3.3.2. Financial Statement Surprises. My empirical analyses also include measures of earnings announcement information content based on reported numbers within financial statements. Again, I model reported numbers using *both* (a) traditional regression, and (b) recent machine learning methods to better capture absolute and incremental information content and mitigate the effect of modeling inconsistencies. This empirical choice is important because recent work documents substantial gains in predictive power when stock returns are modeled through nonlinear machine learning techniques using reported numbers as inputs (Barth et al. 2022).

Specifically, following Hand et al. (2022), I evaluate 12 influential financial statement variables—comprising (i) analyst forecast surprises and (ii) management guidance surprises—that have been identified as the most significant in explaining market reactions around earnings announcements. Analyst forecast surprises are computed as actual minus median forecast, scaled by lagged market capitalization; they include *Earn Surp* (based on street earnings); *Sales Surp* (based on sales); *Ebitda Surp* (based on Ebitda); *Ebit Surp* (based on Ebit); *Pre-Tax Surp* (based on pretax income); *Gaap Surp* (based on Gaap earnings); and *Cfo Surp* (based on cash flows from operations). Management guidance surprises are computed as the new management guidance for year/quarter $t+1/q+1$, disclosed on the earnings announcement date, minus the median analyst consensus forecast for year/quarter $t+1/q+1$ as of just prior to the earnings announcement, scaled by lagged market capitalization. These include the following annual (GA) and quarterly (GQ) guidance surprises: *Earn Surp (GA)* (based on annual earnings); *Sales Surp (GA)* (based on annual sales); *Ebitda Surp (GA)* (based on annual Ebitda); *Earn Surp (GQ)* (based on quarterly earnings); and *Sales Surp (GQ)* (based on quarterly sales).

Financial surprises are used both as in-sample variables, similar to the alternative textual measures, and alternatively to train a GB machine learning algorithm to predict out-of-sample short-window returns. The resulting prediction is labeled as *CAR GB FSA*.[13]

3.4. Sample Statistics

Table 2 presents the descriptive statistics for the primary set of 98,171 out-of-sample firm-quarter observations. The mean (median) market capitalization is $9.9 billion ($1.9 billion); the mean (median) quarterly unscaled earnings amount to $107 million ($13 million); 30% of firms report losses; and the median firm is followed by eight analysts. Finally, the median earnings press release includes 1,160 words, 55 disclosed numbers, and exhibits a positive disclosure tone.

Table 3 presents further univariate evidence. Column (1) shows a strong correlation between *CAR LLM EA* and *CAR [0,1]* (i.e., 0.38). Textual attributes (e.g., *Tone*) and alternative machine learning predictions (i.e., *CAR GB EA*) are also significantly associated with the outcome of interest, although these associations are comparatively weaker.

4. How Much News Is There in the Text of Earnings Announcement Disclosures?

4.1. Primary Analyses

I estimate the following OLS (Ordinary Least Squares) model to test the explanatory power (i.e., adjusted and within-R^2) of LLM-based disclosure news for short-window market reactions:

$$CAR\ [0,1]_{iq} = \alpha_1 + \beta_1 \boldsymbol{CAR\ LLM\ EA_{iq}} + Fixed\ Effects_i + \varepsilon_{iq}. \quad (1)$$

The outcome variable is *CAR [0,1]* or the two-day cumulative abnormal stock return around earnings announcements. The experimental variable is *CAR LLM EA*, which represents *CAR [0,1]* predicted out-of-sample using a *BERT* LLM that captures the capital market news contained solely in the text—characterized using words and word context—of a firm's earnings announcement press release. The variables are measured each calendar quarter q, from 2014 to 2023, for each firm i in the sample. I estimate the model both across and within firm (i.e., without and with firm fixed effects, respectively) focusing primarily on adjusted and within R^2 as evaluation metrics (Breuer and deHaan 2023).

I start analyzing the *absolute* R^2 of LLM-based disclosure news. Table 4, Panel A, shows that *CAR LLM EA* explains about 15% of both the across-firm (i.e., adjusted R^2 in column 1) and within-firm variation (i.e., within R^2 in column 6) in short-window price reactions. This result is economically large; as a comparison, Frankel et al. (2022) documents that earnings press releases explain 4.5% of short-window stock returns when the press releases are characterized using non-LLM machine learning methods.

Next, I focus on the *incremental* explanatory power of *CAR LLM EA* relative to (i) alternative textual proxies and (ii) financial statement surprises. As noted in Section 3, I model both of the latter measures using traditional regression and machine learning approaches. Thus, I estimate and compare the following two OLS models first including text measures and financial statement items characterized using traditional approaches and then alternatively modeled using recent machine learning techniques.

$$CAR\ [0,1]_{iq} = \alpha_1 + \gamma_1\, Alternative\ Text\ Measure_{iq} + \gamma_2\, Financial\ Statement\ Surprises_{iq} + \gamma_3\, Controls_{iq} + Fixed\ Effects_i + \eta_{iq} \quad (2)$$

Table 2. Descriptive Statistics

Primary sample	*N*	Mean	Standard deviation	p25	Median	p75
Market outcome						
CAR [0,1]	98,171	0.021	6.380	−3.467	0.064	3.638
Experimental variable						
CAR LLM EA	98,171	0.259	2.427	−0.795	0.307	1.476
Textual attributes						
Tone	98,171	1.898	1.405	0.907	1.745	2.784
Fog	98,171	19.347	3.127	17.249	18.950	20.908
Length	98,171	7.044	0.597	6.653	7.056	7.435
Numbers	98,171	67.623	49.842	33.000	55.000	87.000
Future	98,171	0.593	0.416	0.293	0.516	0.810
Financial statement surprises and controls						
Earn Surp	98,171	−0.001	0.043	−0.001	0.001	0.003
Sales Surp	98,171	0.018	0.233	−0.012	0.003	0.038
Ebitda Surp	98,171	0.005	0.097	−0.002	0.000	0.016
Ebit Surp	98,171	0.008	0.013	−0.001	0.000	0.027
Pre-Tax Surp	98,171	0.005	0.147	−0.006	0.000	0.025
Gaap Surp	98,171	−0.010	0.101	−0.002	0.000	0.002
Cfo Surp	98,171	−0.001	0.023	0.000	0.000	0.000
Earn Surp (GA)	98,171	1.201	19.859	0.000	0.000	0.000
Earn Surp (GQ)	98,171	−0.364	15.377	0.000	0.000	0.000
Sales Surp (GA)	98,171	0.019	0.184	0.000	0.000	0.000
Sales Surp (GQ)	98,171	−0.001	0.074	0.000	0.000	0.000
Ebitda Surp (GA)	98,171	0.003	0.039	0.000	0.000	0.000
Spec Items	98,171	0.353	1.137	0.000	0.017	0.214
Size	98,171	7.590	1.837	6.319	7.550	8.760
Alternative machine learning predictions						
CAR GB EA	98,171	0.137	1.855	−0.899	0.208	1.235
CAR GB FSA	98,171	0.238	2.713	−1.565	0.263	1.923

Notes. This table reports the summary statistics for the primary sample of 98,171 firm-quarter observations with available disclosure and financial data spanning 2014–2023. All variables are defined in the appendix. Continuous variables (excluding LLM and other machine learning–based predictions) are winsorized at the 1st and 99th percentiles of their quarterly distribution. Return-based variables, special items, tone, and future are expressed as percentages, whereas surprise variables are multiplied by 10,000 to enhance readability.

$$\begin{aligned} CAR\ [0,1]_{iq} = {} & \alpha_2 + \delta_1 \boldsymbol{CAR\ LLM\ EA_{iq}} \\ & + \delta_2\, Alternative\ Text\ Measure_{iq} \\ & + \delta_3\, Financial\ Statement\ Surprises_{iq} \\ & + \delta_4\, Controls_{iq} + Fixed\ Effects_i + \chi_{iq} \end{aligned} \quad (3)$$

Alternative textual measures represent disclosure text characterized using methodologies that are unable to fully capture word context. These include the dictionary-based narrative attributes and non-LLM machine learning proxies described in Section 3. The models also include controls for the firm's (i) general risk and information environment, namely *Size* (market capitalization), and (ii) special business circumstances with valuation implications, namely *Spec Items* (the absolute value of special items).[14] Standard errors are clustered by firm and date. I formally test differences in explanatory power for Models (2) and (3) by bootstrapping R^2 using 10,000 repetitions (with replacement).

Table 4, Panel A, columns 2–5, and Figure 1 show that *CAR LLM EA* provides almost a 10-fold increase in explanatory power relative to conventional narrative attributes (i.e., 1.6% R^2). It also offers a large incremental explanatory power (12%) relative to financial statement surprises in explaining investors' reactions (5%). Column 5 reports that the LLM-based news proxy provides dimensions of informativeness that are distinct and incremental to both traditional text attributes and reported income numbers, adding 11% of adjusted R^2 (versus 6% provided by narrative attributes and reported numbers taken together). Similar results are obtained for within-firm estimations (columns 7 and 8).

Table 4, Panel B, and Figure 1 compare LLM-based disclosure news with news derived from alternative text measurements and relevant financial surprises modeled using recent machine learning (i.e., nonlinear) techniques. Text and surprises are used as input to machine learning models to predict short-window market reactions out-of-sample. Consistent with Frankel et al. (2022), column 2 shows that largely a-contextual "N-grams" offer twice as much explanatory power than traditional narrative attributes (i.e., 3.1% versus 1.6%). However, *CAR LLM EA*

Table 3. Correlations

Primary sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
CAR [0,1]	**1.00**																					
CAR LLM EA	**0.38**	**1.00**																				
Tone	**0.12**	**0.32**	**1.00**																			
Fog	**−0.02**	**−0.05**	**−0.12**	**1.00**																		
Length	0.00	0.01	**−0.18**	**0.04**	**1.00**																	
Numbers	**0.02**	**0.10**	**−0.08**	**−0.05**	**0.74**	**1.00**																
Future	**−0.03**	**−0.06**	**−0.13**	**0.14**	**0.05**	**−0.09**	**1.00**															
Earn Surp	**0.09**	**0.10**	**0.03**	**−0.01**	0.00	**0.03**	**−0.02**	**1.00**														
Sales Surp	**0.13**	**0.17**	**0.11**	0.01	0.00	0.01	**0.00**	**0.12**	**1.00**													
Ebitda Surp	**0.17**	**0.21**	**0.07**	**−0.01**	**−0.02**	0.01	**−0.02**	**0.25**	**0.27**	**1.00**												
Ebit Surp	**0.15**	**0.18**	**0.06**	**−0.01**	**−0.02**	0.01	**−0.02**	**0.36**	**0.26**	**0.57**	**1.00**											
Pre-Tax Surp	**0.15**	**0.18**	**0.06**	**−0.01**	**−0.01**	**0.02**	**−0.02**	**0.41**	**0.22**	**0.45**	**0.63**	**1.00**										
Gaap Surp	**0.07**	**0.10**	**0.08**	**−0.02**	**−0.02**	**0.03**	**−0.04**	**0.45**	**0.10**	**0.12**	**0.18**	**0.22**	**1.00**									
Cfo Surp	**0.04**	**0.03**	0.00	0.01	0.00	**0.01**	**−0.01**	**0.03**	**0.02**	**0.06**	**0.05**	**0.03**	**0.02**	**1.00**								
Earn Surp (GA)	**0.03**	**0.05**	**0.05**	−0.01	**0.01**	**0.02**	0.00	0.00	0.00	**0.01**	**0.01**	**0.01**	**0.01**	0.00	**1.00**							
Earn Surp (GQ)	**0.05**	**0.03**	0.01	0.00	−0.00	**−0.01**	**−0.01**	0.00	**0.01**	**0.02**	**0.01**	**0.01**	**0.01**	0.00	**−0.01**	**1.00**						
Sales Surp (GA)	**0.02**	**0.04**	**0.07**	**−0.01**	**0.01**	**0.03**	0.00	0.00	**0.02**	**0.01**	**0.01**	**0.01**	0.01	0.00	**0.24**	**−0.04**	**1.00**					
Sales Surp (GQ)	**0.04**	**0.05**	**0.03**	0.00	**−0.01**	**−0.02**	0.00	0.00	**0.03**	**0.01**	**0.01**	**0.01**	0.01	0.00	**−0.01**	**0.39**	**−0.02**	**1.00**				
Ebitda Surp (GA)	**0.02**	**0.03**	**0.03**	**0.00**	**0.02**	**0.02**	0.00	0.00	**0.01**	0.00	**−0.01**	**−0.01**	**0.00**	0.00	**0.19**	**−0.03**	**0.32**	**−0.03**	**1.00**			
Spec Items	**−0.04**	**−0.09**	**−0.05**	**0.03**	0.00	**−0.03**	**0.02**	**−0.05**	**−0.02**	**−0.03**	**−0.06**	**−0.08**	**−0.19**	**−0.01**	0.00	**−0.02**	**0.01**	**−0.03**	**0.02**	**1.00**		
Size	**0.03**	**0.17**	**0.18**	**−0.03**	**0.07**	**0.14**	**−0.01**	**0.05**	**0.03**	**0.04**	**0.02**	**0.03**	**0.10**	0.00	**0.04**	**−0.01**	**0.02**	**0.01**	**−0.01**	**−0.06**	**1.00**	
CAR GB EA	**0.18**	**0.43**	**0.32**	**−0.05**	**−0.04**	**0.08**	**−0.09**	**0.07**	**0.09**	**0.11**	**0.09**	**0.10**	**0.08**	**0.01**	**0.03**	**0.01**	**0.03**	**0.03**	**0.01**	**−0.07**	**0.13**	**1.00**
CAR GB FSA	**0.34**	**0.40**	**0.16**	**−0.02**	**−0.03**	**0.03**	**−0.03**	**0.34**	**0.34**	**0.51**	**0.51**	**0.51**	**0.18**	**0.12**	**0.08**	**0.12**	**0.05**	**0.11**	**0.05**	**−0.04**	**0.01**	**0.20**

Notes. This table reports Pearson correlations for the primary sample of 98,171 firm-quarter observations with available disclosure and financial data spanning 2014–2023. All variables are defined in the appendix. Continuous variables (excluding LLM and other machine learning–based predictions) are winsorized at the 1st and 99th percentiles of their quarterly distribution. Bold coefficients indicate significance at <0.01 (two-tailed tests).

Table 4. Disclosure News and Price Revisions Around Earnings Announcements

	Dependent variable: *CAR [0,1]*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Comparison with dictionary-based measures and financial surprise variables								
CAR LLM EA	1.01***				0.95***	1.11***		1.00***
Tone		0.56***		0.46***	0.01		0.66***	0.05
Fog		−0.01		−0.01 *	−0.01		−0.03***	−0.03**
Length		−0.10		−0.02	0.14***		−0.14	0.00
Numbers		0.01***		0.01***	−0.01***		0.01***	0.00
Future		−0.07		−0.04	−0.05		−0.14	0.04
Earn Surp			2.39**	2.85***	2.63***		2.46***	2.57***
Sales Surp			2.07***	1.84***	1.10***		1.79***	1.12***
Ebitda Surp			6.55***	6.31***	3.51***		6.48***	3.60***
Ebit Surp			1.56***	1.54***	1.06***		1.73***	1.20***
Pre-Tax Surp			2.25***	2.18***	1.20**		2.20***	1.18***
Gaap Surp			1.07***	0.69 *	0.32		0.57	0.18
Cfo Surp			8.46***	8.70***	7.12***		8.21***	6.70***
Earn Surp (GA)			0.01***	0.01***	0.01**		0.01***	0.01**
Earn Surp (GQ)			0.01***	0.01***	0.01***		0.01***	0.01***
Sales Surp (GA)			0.29***	0.08	−0.01		0.14	0.11
Sales Surp (GQ)			1.98***	1.84***	0.75		1.62***	0.69
Ebitda Surp (GA)			2.04**	1.84**	1.10		1.67**	0.92
Spec Items			−0.11***	−0.09***	−0.00		−0.03	0.01
Size			0.05**	−0.02	−0.13***		−0.99***	−1.17***
N	98,171	98,171	98,171	98,171	98,171	97,942	97,942	97,942
Fixed effects	No	No	No	No	No	Firm	Firm	Firm
Adjusted R^2	14.9%	1.6%	4.9%	5.9%	16.6%	16.7%	8.6%	18.3%
Within R^2	—	—	—	—	—	14.4%	6.2%	16.3%
Incremental adjusted and within R^2	—	13.3%***	11.6%***	10.7%***	—	—	10.1%***	—
Panel B: Comparison with non-LLM machine learning measures derived from text and financial surprise variables								
CAR LLM EA	1.01***				0.78***	1.07***		0.81***
CAR GB EA		0.61***		0.39***	0.01		0.39***	0.04 *
CAR GB FSA			0.80***	0.74***	0.52***		0.78***	0.55***
N	98,171	98,171	98,171	98,171	98,171	97,942	97,942	97,942
Fixed effects	No	No	No	No	No	Firm	Firm	Firm
Adjusted R^2	14.9%	3.1%	11.5%	12.7%	19.0%	16.1%	14.6%	20.3%
Within R^2	—	—	—	—	—	13.9%	12.4%	18.3%
Incremental adjusted and within R^2	—	11.8%***	7.5%***	6.3%***	—	—	5.9%***	—

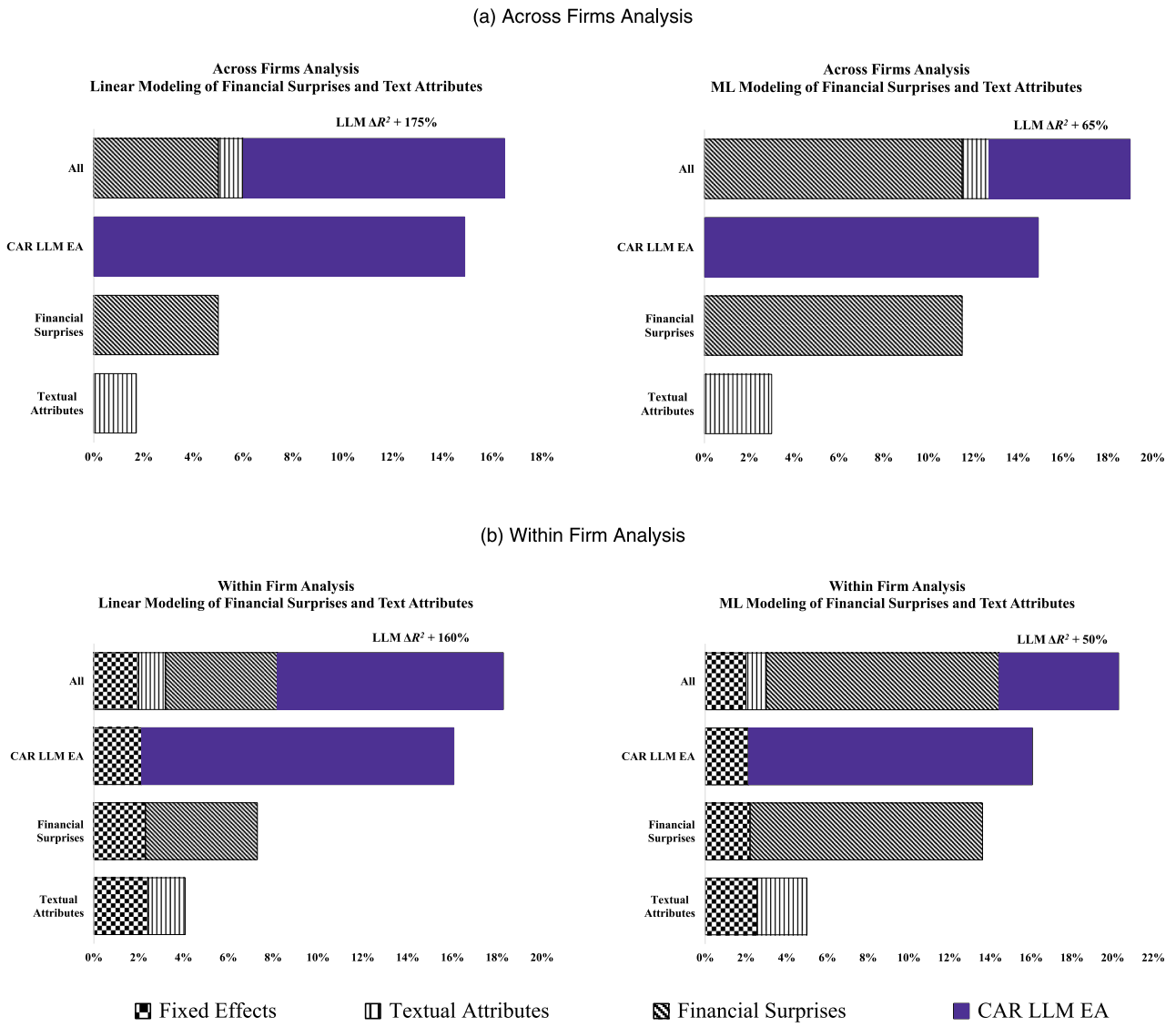
Notes. This table reports results of regressions in which the dependent variable is *CAR [0,1]*, the cumulative abnormal return for the two-day period surrounding an earnings announcement. The experimental variable is *CAR LLM EA*, the "news" contained within the text of a firm's quarterly earnings press release obtained from a *BERT*-based LLM that is fine-tuned to predict *CAR [0,1]* based solely on the text of a firm's earnings press releases. The sample includes 98,171 firm-quarter observations with available disclosure and financial data spanning 2014–2023 (i.e., out-of-sample observations). Panel A reports absolute and incremental adjusted R^2 and within R^2 of *CAR LLM EA* compared with those obtained from dictionary-based approaches and financial surprise and controls variables. Panel B reports this comparison relative to Gradient Boosting–based machine learning measures alternatively capturing the information content of text within press releases (i.e., *CAR GB EA*) and of financial surprise and control variables (i.e., *CAR GB FSA*). Continuous variables (excluding LLM and other machine learning–based predictions) are winsorized at the 1st and 99th percentiles of their quarterly distribution. Standard errors are clustered by firm and quarter-year. The statistical significance of the incremental adjusted R^2 (in columns 1–5) and within R^2 (in columns 6–8) is assessed using a bootstrapping protocol based on 10,000 iterations (with replacement). All models are estimated with an intercept (unreported). All variables are described in the appendix.

***, **, and *Significance at <0.01, <0.05, and <0.10, respectively.

significantly outperforms *CAR GB EA* (i.e., N-grams–based predicted *CAR*); in fact, it generates a 12% incremental adjusted R^2, representing a 300% improvement. Of note, column 3 shows that CAR *LLM EA* also offers an incremental adjusted R^2 of 7.5%, representing a 65% improvement, compared with *CAR GB FSA* (i.e., the predicted *CAR* using financial statement surprises). All the above differences are formally confirmed using bootstrapping.

Overall, the text in earnings announcement press release, characterized using an LLM that can capture complex word context, explains an economically relevant fraction of short-window price revisions—offering a *three*-fold average increase in R^2 compared with methods employed by prior research. These results suggest that prior research has substantially underestimated the information content of earnings announcement text that investors appear to utilize

Figure 1. (Color online) Disclosure News Around Earnings Announcements



Notes. This figure presents the across-firm (a) and within-firm (b) adjusted R^2 provided by *CAR LLM EA*, financial surprise variables, and both, when regressing *CAR [0,1]* on such variables. Financial statement surprises are modeled both linearly and nonlinearly, using a gradient boosting model to predict *CAR [0,1]* based on these surprises. The sample includes 98,171 observations that the LLM has not used for fine-tuning (i.e., they are out of sample). All variables are defined in the appendix.

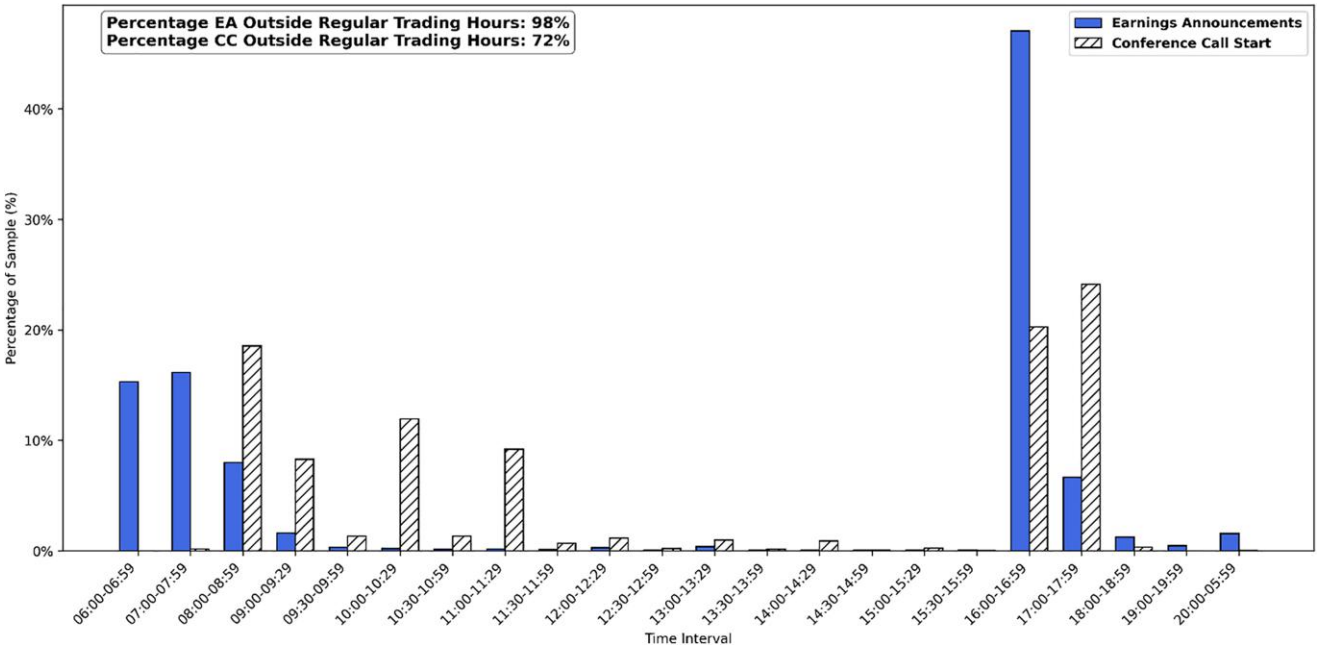
for valuation significantly more than previously observed.

4.2. Immediate Price Revisions

Next, I test if LLM-modeled text in earnings press releases can explain immediate price revisions occurring within minutes of corporate disclosures. Prior research documents a significant market reaction to earnings announcements, including *both* reported numbers and textual disclosures, within minutes of their release (Patell and Wolfson 1984, Jiang et al. 2012). Thus, I fine-tune the LLM to predict buy-and-hold stock returns over 5-minute, 30-minute, and 2-hour windows immediately following the announcement, using only the text of the earnings press release. As shown in Figure 2, 98% of earnings announcements occur outside regular trading hours, consistent with prior literature (Jiang et al. 2012). Therefore, this analysis utilizes after-hours and before-hours granular price data from TAQ.[15]

Of note, I consider normalizing returns (i.e., deriving abnormal returns) to account for market-wide shocks. However, the sample of firms with available 5-minute, 30-mintue, or 2-hour returns significantly varies at each point in time for a given announcing company, likely introducing noise into the normalization factor. Furthermore, the narrow within-minutes

Figure 2. (Color online) Distribution of Earnings Announcements and Conference Call Times



Note. This figure shows the frequency of earnings announcements and conference calls throughout a day for the primary sample of 2014–2023 earnings announcement press releases (98,171) and conference calls transcripts (56,670).

return window reduces the likelihood of concurrent aggregate shocks influencing signed returns, making the calculation of abnormal returns less critical.

For these tests, I model both textual disclosures and financial surprises nonlinearly and retain a constant sample observation with nonmissing 5-minute, 30-minute, and 2-hour stock returns, immediately following the earnings announcement.[16] Table 5 reveals two key insights. First, textual disclosures account for a substantial portion of the market response to earnings announcements, captured within minutes of the release. In fact, earnings press releases provide an

Table 5. Immediate Price Revisions

	Sample with available 5-minute, 30-minute, and 2-hour returns immediately after the earnings announcements ($N = 63{,}249$)								
	Dependent variable: *RET 5M EA*			Dependent variable: *RET 30M EA*			Dependent variable: *RET 2H EA*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RET LLM EA	2.04***		1.16***	2.18***		1.28***	1.55***		0.99***
RET GB FSA		1.08***	0.91***		1.12***	0.92***		1.09***	0.86***
N	63,249	63,249	63,249	63,249	63,249	63,249	63,249	63,249	63,249
Fixed effects	No	No	No	No	No	No	No	No	No
Adjusted R^2	8.0%	14.6%	16.8%	9.4%	15.5%	18.3%	10.2%	14.2%	17.7%
Incremental adjusted R^2	—	—	2.2%***	—	—	2.8%***	—	—	3.5%***
% ΔR^2	—	—	15.0%	—	—	18.0%	—	—	25.0%

Notes. This table presents regression results using returns outside of regular trading hours. The sample consists of 63,249 firm-quarter observations with (a) earnings announcement press releases issued outside regular trading hours (i.e., after-hours or before-hours) and (b) with available stock returns for the following three event windows: (i) 5-minute, (ii) 30-minute, and (iii) 2-hour immediately following the earnings announcement. Notably, 98% of the earnings announcements in the primary sample occur outside of regular trading hours. Returns are calculated as the ratio of the last to the first available price closest to or immediately following the event window endpoints, minus one, and are restricted to the same day as the announcement. The experimental variable, *RET LLM EA*, measures the predicted stock returns (i.e., 5-minute, 30-minute, and 2-hour) based solely on the press release text and using a *BERT*-based LLM. *RET GB FSA* represents the gradient boosting–based machine learning predictions of stock returns based on financial surprise variables. The statistical significance of the incremental adjusted R^2 is evaluated using a bootstrapping procedure with 10,000 iterations (with replacement). All models include an intercept (unreported). All variables are described in the appendix.

***, **, and *Significance at <0.01, <0.05, and <0.10, respectively.

absolute R^2 of 8% for five-minute market reactions, increasing the explanatory power of financial surprises by 15%. Second, the incorporation of textual information into prices happens quickly and continuously as the event window is extended. The contribution of textual disclosures to the adjusted R^2 increases *monotonically*, rising from 15% in the 5-minute window to 18% in the 30-minute window, and reaching 25% in the 2-hour window (almost doubling the value of the text in explaining immediate price responses).[17]

These findings enhance our understanding of (i) the role of public, company-sourced accounting disclosures in moving stock prices amidst competing forces and (ii) whether investors process textual information quickly—an important insight given recent literature that highlights underreaction and inattentiveness to such information (Cohen et al. 2020).

4.3. Conference Calls Analyses

I also apply the LLM to conference call transcripts to analyze the value added by these textual disclosures using both two-day and one-hour pricing data.[18] Specifically, I analyze 56,670 transcripts from conference calls held on the same day as the earnings announcement press release. Figure 2 shows that nearly two thirds of conference calls take place outside regular trading hours. Although the analysis is not limited to calls outside regular trading hours, the tenor of the results does not change when focusing on this majority of observations.[19]

Table 6 reveals two key insights. First, Panel A shows that the conference call text explains a large fraction of *CAR [0,1]*, providing an absolute R^2 of approximately 19%. Notably, the conference call text increases the adjusted R^2 by 16% relative to both the financial surprise variables and the textual information content of the earnings press release (both modeled nonlinearly). Second, Panel B reveals that the text of the conference calls also explains a significant portion of the one-hour return around the conference call. I select one-hour returns because most calls conclude within one hour, as observed in Matsumoto et al. (2011) and confirmed in my sample. I find that the improvement in adjusted R^2 (again, relative to both the financial surprises and the press release) is 30%, highlighting the importance of the conference call, particularly in the immediate time window when it takes place, and emphasizing its timeliness.[20]

This analysis offers new insights into the complementary roles of different textual disclosure channels in shaping immediate investor responses. Prior studies generally use returns as a comprehensive measure of earnings announcement information without attempting to disentangle the contributions of text vis-à-vis numerical data (Matsumoto et al. 2011, Brochet et al. 2018); moreover, they do not directly compare the textual content of earnings press releases and conference calls (Frankel et al. 2022). By employing nonlinear modeling and granular hourly price data, my approach isolates the news content of earnings calls text beyond contemporaneous textual news and financial statement surprises. My findings contribute to the literature seeking to better understand the information content of conference calls in real time (Mayew et al. 2020).

5. Under What Economic Circumstances and Which Disclosures Do Investors Value?

5.1. Cross-Sectional Analyses: Differences in Earnings Persistence

I apply the LLM-based disclosure news variable to better understand investors' utilization of corporate disclosures for valuation. I conjecture that textual disclosures can be particularly useful in those economic contexts in which accounting earnings exhibit low persistence and therefore are expected to be less useful for estimating firms' future cash flows (Collins and Kothari 1989). Following prior literature, I identify firms characterized by lower earnings persistence and test for differences in the incremental information content of text within earnings press releases for such companies, relative to other companies. Specifically, I examine firms reporting losses (which investors perceive as transitory; Hayn 1995), large intangible assets (that may highlight a timing mismatch between recognized costs and revenues; Lev and Zarowin 1999), and earnings with low time series autocorrelation (a key cross-sectional determinant of the decision usefulness of earnings; Collins and Kothari 1989).

First, I divide the primary sample into three alternative partitions to capture firms with relatively lower vis-à-vis higher earnings persistence. Specifically, I separate companies reporting (a) losses versus profits, (b) high versus low research and development expenses, and (c) earnings with low versus high AR (1) autocorrelation. Second, for each partition I examine the incremental adjusted R^2 provided by *CAR LLM EA* relative to the information content of financial statement surprises—*CAR GB FSA*.[21] I finally test the statistical difference using a bootstrapping protocol with 10,000 iterations (with replacement).[22] Table 7 presents the results. In line with expectations, Panel A shows that *CAR LLM EA* (i.e., the news in the text extracted using a LLM that characterizes both words and word context) offers up to 70% higher incremental information content for firms characterized by low earnings persistence.

Overall, this evidence shows that investors leverage text more, likely to better understand earnings

Table 6. Incremental News Content of Conference Calls

Sample with concurrent press releases and conference calls ($N = 56{,}670$)			
Panel A: Two-day price revisions			
	Dependent variable: *CAR [0,1]*		
	(1)	(2)	(3)
CAR LLM CC	1.54***		0.89***
CAR LLM EA		0.83***	0.43***
CAR GB FSA		0.55***	0.45***
N	56,670	56,670	56,670
Fixed effects	No	No	No
Adjusted R^2	18.9%	20.8%	24.1%
Incremental adjusted R^2	—	—	3.3%***
% ΔR^2	—	—	16.0%
Panel B: One-hour price revisions			
	Dependent variable: *RET 1H CC*		
	(1)	(2)	(3)
RET LLM CC	2.91***		2.08***
RET LLM EA		0.54***	0.11***
RET GB FSA		0.80***	0.66***
N	33,618	33,618	33,618
Fixed effects	No	No	No
Adjusted R^2	12.1%	11.9%	15.5%
Incremental adjusted R^2	—	—	3.6%***
% ΔR^2	—	—	30.0%

Notes. This table examines the incremental information content of conference call disclosures relative to earnings announcement press releases and financial surprise variables. Panel A presents the analysis for a two-day event window. The sample consists of 56,670 firm-quarter observations where both the earnings announcement press release and the conference call occur on the same day. Notably, in this sample, the earnings press release always precedes the conference call. Panel B provides a return analysis for 36,072 firm-quarters, focusing on a one-hour window starting at the beginning of the conference call and ending one hour later. Observations are reduced to 36,072 due to TAQ data availability. Returns are calculated as the ratio of the last to the first available price closest to or immediately following the event window endpoints, minus one, and are restricted to the same day as the announcement. The experimental variable, *RET LLM CC*, measures the predicted stock returns (i.e., one-hour) based solely on the conference call text and using a *BERT*-based LLM. *RET LLM EA* captures predicted stock returns (i.e., one-hour) based solely on the text of earnings announcement press releases using a *BERT*-based LLM, whereas *RET GB FSA* represents gradient boosting–based machine learning predictions of stock returns based on financial surprise variables. The statistical significance of the incremental adjusted R^2 is evaluated using a bootstrapping procedure with 10,000 iterations (with replacement). All models include an intercept (unreported). All variables are described in the appendix.

***, **, and *Significance at <0.01, <0.05, and <0.10, respectively.

persistence when earnings are expected to map less strongly into firms' future cash flows.

5.2. Time Series Analyses: Changes in Aggregate Uncertainty

To better understand in which aggregate economic conditions textual information in earnings announcements is most relevant to equity investors, I conduct specific time series analyses. Loh and Stulz (2018) show that during periods of heightened uncertainty, investors find it more difficult to assess how aggregate changes affect firm prospects and therefore rely more heavily on public information, such as analyst reports, to complement reported earnings. Thus, I conjecture that disclosure news can be higher at the onset of macroeconomic disruptions and other events that increase uncertainty, as historical financial numbers may be relatively less useful to estimate future cash flows.

To expand this analysis beyond the primary sample period of 2014–2023, I implement a backward-rolling-window approach to fine-tune the LLM. For each year from 2000 to 2023, I use an LLM fine-tuned with disclosures from the prior five years to predict *CAR [0,1]*. Given the significant computational demand of this task—which involves fine-tuning an LLM for each year within the 24-year sample—I conduct the analysis on a subsample. I randomly select 25% of the disclosures from each year, ensuring the subsample is balanced by firm size. Next, for each month-year in my sample, I examine average incremental R^2 separately for textual disclosures and financial surprises, assessing their evolution over time. This approach

Table 7. Cross-Sectional and Time Series Analysis of Disclosure News

Panel A: Cross-sectional tests: Incremental disclosure news and earnings persistence

	Loss (1)	Profit (2)	High R&D firms (3)	Other firms (4)	Low earn AR (1) firms (5)	Other firms (6)
Incremental Adjusted R^2 of CAR LLM EA	9.4%	5.6%	9.0%	6.7%	8.8%	6.5%
p-value (1) > (2)	0.00***					
p-value (3) > (4)			0.00***			
p-value (5) > (6)					0.00***	
N (firm-quarters)	29,803	68,368	18,980	63,074	17,771	71,053

Panel B: Time series tests: Incremental disclosure news and changes in aggregate uncertainty

	Dependent variable: Δ *Incremental Adjusted R^2* Changes in aggregate uncertainty			
	COVID-19 (1)	Trade war (2)	Financial crisis (3)	Enron scandal (4)
I(Event)	0.07**	0.06**	0.06**	0.06 *
p-value	(0.04)	(0.01)	(0.05)	(0.08)
N (month-years)	288	288	288	288

Notes. This table presents cross-sectional and time series analyses of incremental disclosure news derived from quarterly earnings press releases, as captured by a *BERT*-based LLM trained to predict *CAR [0,1]* based solely on press release text. Incremental disclosure news is measured by the increase in adjusted R^2 when adding *CAR LLM EA* to *CAR GB FSA* in OLS models with *CAR [0,1]* as the dependent variable. Panel A compares incremental adjusted R^2 for (i) loss versus profit firms, (ii) firms with high R&D expenses (top quintile of R&D-to-sales ratio) versus others, and (iii) firms with low earnings autocorrelation (first quintile of AR(1) coefficients in time series earnings regressions) versus others, using a sample of 98,171 firm-quarters. Missing R&D values are replaced with zero, and at least 10 observations are needed to calculate AR(1) coefficients. Panel B examines time series differences in incremental adjusted R^2 between textual disclosures and financial surprises across periods, focusing on four key events: (i) COVID-19 (2020), (ii) U.S.-China trade war (2019), (iii) financial crisis (2008), and (iv) Enron scandal (2001). For each event, the LLM is fine-tuned annually (between 2000 and 2023) using a five-year backward-rolling window, with 25% of disclosures randomly sampled each year. *I(Event)* represents a dummy variable equal to one for the relevant month-years capturing each crisis event. All regressions are estimated without fixed effects. The statistical significance of differences in incremental adjusted R^2 is assessed using a bootstrapping protocol based on 10,000 iterations (with replacement) upon which a one-sided *p*-value is reported. Standard errors are clustered by firm and quarter-year in Panel A and by year in Panel B.

***, **, and *Significance at <0.01, <0.05, and <0.10, respectively.

balances computational efficiency and model accuracy while minimizing the risk of look-ahead bias (Bishop 2006).[23]

Using this extended sample, I examine key events likely to influence the incremental news content of earnings disclosures. Specifically, I focus on crisis events associated with shifts in aggregate uncertainty: (a) the COVID-19 pandemic (2020), (b) the U.S.-China trade war (2019), (c) the financial crisis (2008), and (d) the Enron scandal (2001). For each event, I analyze the months characterized by the peak of these crises to enhance identification.[24]

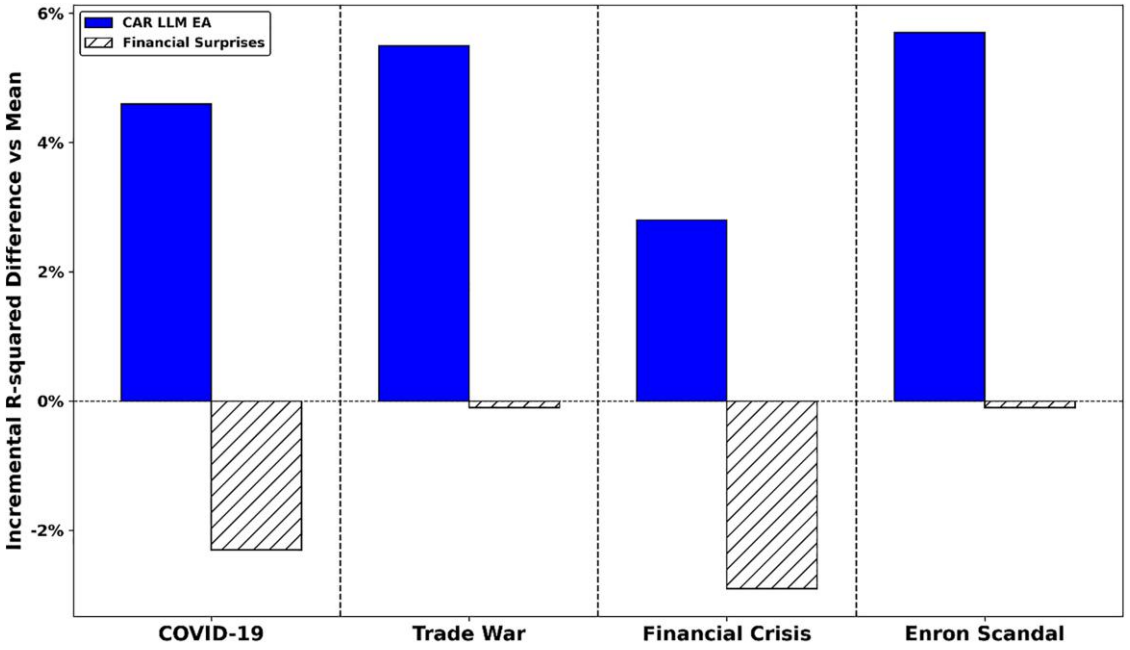
Table 7, Panel B, shows that the incremental information content of earnings announcement disclosures is higher than that of financial surprises during periods of heightened aggregate uncertainty. For example, at the onset of the COVID-19 crisis, the incremental R^2 difference between textual disclosures and financial surprises is 3.6 percentage points above the average—a 60% increase. Figure 3 shows the incremental adjusted R^2 of textual disclosures (*CAR LLM EA*) and financial surprises (*CAR GB FSA*) relative to their average over the sample period for each event analyzed. During these crisis events, the incremental R^2 of disclosures is above average, whereas that of financial surprises is usually below average

Overall, this evidence suggests the heightened role of the firms' textual disclosures during periods of elevated macroeconomic uncertainty, likely because historical financial statement numbers are relatively less useful to estimate the distribution of firms' future cash flows.

5.3. Disclosure Characteristics and Contents

I implement *within*-document analyses to better understand which content, within press releases, is most useful to investors for short-window equity pricing. I focus on three foundational characteristics of disclosures that, based on general economic arguments, can be associated with higher news content. Specifically, I partition each earnings disclosure three times: first, I separate the sentences that (more) directly discuss reported numbers from other sentences; second, I distinguish text at the beginning from text at the end of the disclosure; third, I separate the sentences that

Figure 3. (Color online) Disclosure News and Changes in Aggregate Uncertainty



Notes. This figure shows the incremental adjusted R^2 values from regressing *CAR [0,1]* on *CAR LLM EA* and financial surprise variables, with the latter modeled nonlinearly using a gradient boosting model. The incremental adjusted R^2 is presented relative to the mean incremental adjusted R^2 over 288 months from 2000 to 2023. For each event, the LLM is fine-tuned annually (2000–2023) using a five-year backward-rolling window, with 25% of disclosures randomly sampled each year, balanced by firm size. All variables are defined in the appendix.

exhibit substantially different contents, relative to the prior quarter disclosure (for a given firm), from other sentences. Next, for each partition, I predict contemporaneous short-window returns twice: using text *with* and *without* the key disclosure characteristic of interest. I finally test differences in R^2 using a bootstrapping protocol.[25]

The primary advantage of this *within*-document analysis is that it allows to hold constant any company-and-time-varying features of an earnings announcement press release. Therefore, such analysis can be especially useful to test and understand which information, within earnings announcement text, is the most useful to investors (i.e., contains most news).

Table 8 presents the results. In columns 1–3, I find that the text which directly discusses quantitative accounting information contains *twice* the amount of news content compared with other text (11.7% versus 6.1% in terms of R^2).[26] In line with word context driving information content, the news arising from the entire disclosure (15%) is higher than the news arising from quantitative and nonquantitative texts taken individually. Column 2 reports results after deleting numbers within text (i.e., the median press release contains 55 numbers). Although the *BERT* LLM is not explicitly designed or fine-tuned to understand numbers (e.g., perform arithmetic operations), it is relevant to assess if most news in earnings disclosures derives from discussions about reported numbers vis-à-vis the presence of the numbers themselves. Results show that numerical tokens account for about 7% of the information content of quantitative sentences; however, their deletion does not alter substantially the incremental news.

Columns 4 and 5 reveal that most decision useful contents are concentrated in the first 15 sentences as opposed to the last 15 sentences of the disclosure. This evidence aligns with Cheng et al. (2021) and demonstrates that, even when the entire (as opposed to selected) textual contents are considered, investors react more to prioritized information—possibly because managers present more relevant information earlier in the disclosure and/or investors find it easier to process salient information. Finally, columns 6 and 7 partition each disclosure into new and stale sentences. I measure new sentences as sentences that are substantially different (i.e., 90% or more) compared with any other sentences in the prior quarter disclosure for the same company. Conversely, I classify sentences that are largely similar to prior-quarter disclosures as stale.[27] This analysis reveals that novel contents in earnings press releases contain two times more information content compared with stale contents (12.7% versus 4.0% in terms of R^2). Repetitive contents appear to be nonetheless relevant (e.g., similar to the entire disclosure characterized using non-LLM machine learning algorithms

Table 8. Disclosure Characteristics and Contents

	Dependent variable: *CAR [0,1]*						
	First partition: Discussion of numbers ($N = 98{,}171$)			Second partition: Location of text ($N = 87{,}406$)		Third partition: Novel contents ($N = 85{,}342$)	
	Sentences discussing numbers (#s included) (1)	Sentences discussing numbers (#s excluded) (2)	Other sentences (3)	First 15 sentences (4)	Last 15 sentences (5)	New sentences (6)	Stale sentences (7)
R^2 *of CAR LLM EA*	11.7%	10.9%	6.1%	15.1%	2.4%	12.7%	3.9%
p-value (1) > (3)		0.00***					
p-value (2) > (3)		0.00***					
p-value (4) > (5)				0.00***			
p-value (6) > (7)						0.00***	

Notes. This table presents absolute R^2 of *CAR LLM* for three different sets of textual inputs. In columns 1–3 (columns 4 and 5) [columns 6 and 7], the sentences that directly discuss numbers and other sentences (the first 15 sentences and the last 15 sentences) [the sentences that include novel contents and those that are similar to the prior quarter press release] are alternatively used as inputs to the *BERT* LLM model to predict *CAR [0,1]*. For the first partition, sentences are classified as *directly* describing numbers whenever they include a dollar amount/percentage or precede/follow a sentence that directly describes numbers. The sample includes $n = 98{,}171$ firm-quarter observations. In column 1, the analysis involves sentences that directly discuss numbers including the numbers themselves. In column 2, the numbers are stripped out from the text that surrounds them. For the second partition, each earnings announcement is divided into three parts of 512 tokens (i.e., approximately 15 sentences). Next, the initial 512 tokens (from sentence 1 to 15) [final 512 tokens (from sentence 30 to 45)] are used as inputs to the *BERT* LLM. Only earnings disclosures that are long enough to include at least three sequences of 512 tokens are analyzed (this leads to 87,406 observations). For the third partition, sentences are classified as new (stale) if they are substantially different from (similar to) the sentences used in the earnings press release issued by the same firm in the immediately preceding quarter. Sentence similarity is assessed using a string similarity algorithm based on the "Gestalt pattern matching" approach proposed by Metzener and Ratcliff (1988). The algorithm identifies the longest contiguous substring, after removing elements such as white spaces, between two texts. The output is a score capturing their similarity. I apply this approach to each pair of sentences across consecutive quarterly press releases (for a firm). If the score is above 90%, I classify the sentence as stale or repeated. The sample includes only earnings disclosures for which a lagged press release, in quarter $q - 1$, is available within the out-of-sample set of observations (this leads to 85,342 observations). All regressions are estimated without fixed effects. The statistical significance of differences in explanatory power is assessed using a bootstrapping protocol based on 10,000 iterations (with replacement). All variables are described in the appendix. "#s" stands for "numbers."

***, **, and *Significance at <0.01, <0.05, and <0.10, respectively.

based on N-grams). This evidence confirms the usefulness of fine-tuning LLMs to directly explain investor reactions to accounting information because investors' expectation models are largely unobservable.

6. Sensitivity and Additional Analyses

6.1. Sensitivity Tests

I examine the robustness of the primary findings to four types of sensitivity analyses. First, I replicate the primary tests using alternative out-of-sample approaches. This analysis mitigates potential look-ahead biases, because there may be a valid concern about overlap between the primary sample period and the LLM's pretraining period. The LLM used in this study, known as *RoBERTa* (Liu et al. 2019), completed its pretraining in 2019. Thus, I restrict the sample I examine to earnings announcements from calendar years 2021, 2022, and 2023. Starting from 2021 establishes a clear temporal separation from the LLM's pretraining period, minimizing the risk of information leakage and ensuring the evaluation reflects entirely out-of-sample performance.[28] Table OA-1, Panel A, in the Online Appendix, shows that the LLM captures substantial (sometimes even higher) disclosure news when evaluated on observations *outside* the pretraining period. Thus, the results do not appear to suffer from evident look-ahead biases.[29]

Second, I use select point-in-time fundamentals to mitigate the concern that the financials provided by data vendors may not align with those used by investors at the time of the announcement (Du et al. 2023, Lyle et al. 2025). Untabulated results indicate that standardization practices are unlikely to drive the incremental informational content of textual earnings disclosures. Third, I replicate the primary analyses for two alternative samples: one that *only* includes earnings announcements within the two-day period surrounding the announcement and another that *also* comprises concurrent events, such as conference calls, and analyst or management forecasts. This test helps better identify the specific role of disclosure news in earnings press releases. Unreported analyses reveal comparable incremental news within the two subsamples. Finally, I confirm that the *BERT* LLM can be applied to alternative market outcomes. I fine-tune the LLM to predict, out-of-sample, abnormal trading volumes and abnormal volatility. Unreported tests show a strong absolute and incremental explanatory power of *BERT*-based predictions.

6.2. Additional Analyses

I conduct seven additional analyses, several of which are reported in the Online Appendix. First, leveraging valuation theory (Ohlson 1995), I show that earnings disclosures predict both the *level* and *uncertainty* of future earnings. Using only press releases, predicted earnings for quarter $q+1$ improve the model's R^2 by 20 percentage points, whereas the predicted standard deviation of $q+1$ to $q+4$ earnings (i.e., uncertainty) adds 15 percentage points, indicating that textual disclosures help investors revise expectations about firms' future cash flows (Table OA-2 in the Online Appendix). This test extends the evidence provided by Kim and Nikolaev (2024) on using MD&A (Management's Discussion and Analysis) disclosures to predict the sign of changes (i.e., increase or decrease) in future earnings. I add to this work by demonstrating that LLMs can (i) predict *both* the direction and the amount of future earnings and (ii) effectively capture earnings *uncertainty*.

Second, Fama-MacBeth regressions across 288 monthly panels confirm that earnings announcement news content is robust and not driven by influential observations, with 82% of months exhibiting substantial textual news content (Table OA-1, Panel B, Online Appendix). Third, incorporating additional financial statement analysis variables, such as dividends and leverage changes (Ou and Penman 1989), does not alter the tenor of the results (Table OA-1, Panel C, Online Appendix). Fourth, an analysis of how textual attributes and fundamentals relate to *BERT* LLM's accuracy reveals that the model performs better with larger companies and firms with stronger accounting performance, whereas its accuracy is not negatively impacted by higher linguistic complexity (Table OA-3 in the Online Appendix). Fifth, using simple "word embeddings" (Mikolov et al. 2013) rather than N-grams marginally improves accuracy but does not alter *BERT* LLM results significantly. Sixth, randomizing disclosure word order reduces explanatory power by 80%, highlighting the importance of word context, which N-grams cannot fully capture.[30] Finally, fine-tuning *BERT* LLM with *only* the 100 most frequent performance or tone words diminishes its predictive ability, emphasizing the importance of sequential and contextual information within earnings announcement disclosures.[31]

7. Conclusion

This study offers novel empirical evidence about the information content of earnings announcement disclosures. I use an LLM approach, capturing information in both words and word context, to address measurement limitations in prior textual analysis literature and to better quantify and understand news in corporate disclosures.

Applying this approach, I extract a single information variable from the text of earnings disclosures and then use this information variable to explain, out-of-sample, price revisions around earnings announcements. The text of earnings press releases explains 15% of the variation in signed two-day earnings announcement returns, representing a 300% improvement in R^2 compared with conventional and recent textual measures. In addition, such disclosures provide large incremental explanatory power relative to a wide array of financial surprises, generating a 100% average improvement in R^2. Although press releases and financial surprises contain the most news, conference call disclosures increase the R^2 by up to 30%.

These findings suggest that prior textual analysis methods significantly underestimate the economic relevance of earnings disclosures for equity pricing decisions. This underestimation stems primarily from the limitations of both traditional and recent textual analysis techniques in capturing complex word context.

Notably, earnings disclosures explain a relevant part of the immediate price response, even within the first five minutes of their release. Although investors may take time to fully incorporate textual information into prices, the process begins immediately, indicating that investors are attentive to such information.

In addition, textual disclosures are particularly relevant when stand-alone accounting numbers are less useful for valuation—specifically, when they exhibit low persistence and during periods of heightened aggregate uncertainty. Furthermore, investors respond most strongly to text that directly discusses numbers, appears at the beginning of the disclosure, and presents novel contents.

Through a richer and context-based characterization of earnings disclosures, this study advances our understanding of investors' use of accounting text, and shows that LLMs can offer insights into attributes, properties, and outcomes of corporate disclosures (Leuz and Wysocki 2016). As such, my findings have implications for researchers, regulators, and practitioners.

Acknowledgments

This paper is based on the author's dissertation titled "Contextualized News in Corporate Disclosures: A Neural Language Approach." The author thanks the dissertation committee for insightful guidance: Peter Wysocki (chair), Francois Brochet, Dokyun Lee, and Eddie Riedl. The author thanks Suraj Srinivasan (department editor), an anonymous associate editor, and two anonymous reviewers for thoughtful comments; workshop participants at the 2021 AAA (American Accounting Association)/Deloitte Foundation/J. Michael Cook Doctoral Consortium, 2023 FARS (Financial Accounting and Reporting Section) Meeting, 2023 SARAC (Swiss Accounting Research Alpine Camp), 2023 HARC (Hawaii Accounting Research Conference), Bocconi University, Boston University, INSEAD, Northeastern University, Northwestern University, The Ohio

State University, The University of Texas at Dallas, Universidad Carlos III de Madrid, University of Kentucky, University of Washington, and University of Wisconsin-Madison for valuable comments and discussions; and Daniel Bens, Oliver Binz, Elizabeth Blankespoor, Ed deHaan, Ronald Dye, Nargess Golshan, Charles Lee, Maria Loumioti, Russell Lundholm, Matthew Lyle, Stanimir Markov, Miguel Minutti-Meza (Discussant), Darren Roulstone, Sugata Roychowdhury, Gil Sadka, Jason Schloetzer, Shiwon Song, Dan Stone, Stephen Stubben, Andrew Van Buskirk, and Daniel Wangerin.

Appendix. Variable Definitions

Variable	Description	Source
Market outcome		
CAR [0,1]	Cumulative abnormal stock returns computed as actual security returns minus the return on the overall market (including dividends), in the two-day window surrounding an earnings announcement event (from day 0 to 1).	CRSP
Experimental variable		
CAR LLM EA	The "news" contained within the text of a firm's quarterly earnings press release. The "news" is obtained from a *BERT*-based large language model (LLM) that is fine-tuned to predict, out-of-sample, *CAR [0,1]* based solely on the text of a firm's quarterly earnings announcement press release. The *BERT* LLM uses both words *and* word context—i.e., the ordered sequence of all elements in the disclosure—to predict *CAR [0,1]*.	EDGAR CRSP
Alternative textual measures		
of which: *Dictionary-based textual attributes*		
Tone	Difference between Henry (2008) positive and negative words scaled by the total number of words in the document.	EDGAR
Fog	Gunning (1952) Fog Index computed as ([average words per sentence + percentage of complex words] * 0.4).	EDGAR
Length	Natural logarithm of the total number of words found in an earnings announcement text.	EDGAR
Numbers	Total number of numbers disclosed within the text of an earnings announcement.	EDGAR
Future	Total number of future-tense verbs scaled by the total number of words in the earnings announcement.	EDGAR
of which: *Machine learning–based predictions*		
CAR GB EA	*CAR [0,1]* predicted, out-of-sample, using top 1,500 "unigrams" and "bigrams" (i.e., "N-Grams"), by TF-IDF, extracted from quarterly earnings announcement press releases. "N-Grams" are characterized nonlinearly using a gradient boosting machine learning model.	EDGAR CRSP
Financial statement surprises and controls		
of which: *Stand-alone variables*		
Analyst Forecast Surprises	Analyst forecast error (actual – median forecast) scaled by the market value of equity immediately before the earnings announcement window. The forecast is taken from the latest quarter *q* consensus period prior to the earnings announcement. The following analyst forecast surprises are computed: *Earn Surp* (based on street earnings); *Sales Surp* (based on sales); *Ebitda Surp* (based on Ebitda); *Ebit Surp* (based on Ebit); *Pre-Tax Surp* (based on pretax income); *Gaap Surp* (based on Gaap earnings); *Cfo Surp* (based on cash flows from operations).	I/B/E/S
Management Guidance Surprises (GA)	New management guidance for year *t*+1, reported at the earnings announcement, minus the median analyst consensus forecast for year *t*+1, as of just prior to the earnings announcement, scaled by the market value of equity immediately before the announcement window. The following management guidance surprises are computed: *Earn Surp (GA)* (based on annual earnings); *Sales Surp (GA)* (based on annual sales); *Ebitda Surp (GA)* (based on annual Ebitda).	I/B/E/S

Appendix. (Continued)

Variable	Description	Source
Management Guidance Surprises (GQ)	New management guidance for quarter q+1, reported at the earnings announcement, minus the median analyst consensus forecast for quarter q+1, as of just prior to the earnings announcement, scaled by the market value of equity immediately before the announcement window. The following management guidance surprises are computed: *Earn Surp (GQ)* (based on quarterly earnings); *Sales Surp (GQ)* (based on quarterly sales).	I/B/E/S
Spec Items	The absolute value of special items scaled by quarterly assets.	Compustat
Size	Natural logarithm of a firm's market capitalization.	CRSP
of which: *Machine learning–based predictions*		
CAR GB FSA	*CAR [0,1]* predicted, out-of-sample, using all the financial surprise and control variables (see above). Financial surprise and control variables are characterized nonlinearly using a gradient boosting machine learning model.	I/B/E/S Compustat CRSP
Main variables for subsample analyses		
Immediate Price Revision EA	The stock return measured for three event window intervals immediately following the earnings announcement press release when released outside of regular trading hours (i.e., after-/before-hour). The return is calculated as the ratio of the last available price to the first available price closest to or immediately following the event window endpoints, minus one. Three returns are considered: *RET 5M EA* (5-minute window), *RET 30M EA* (30-minute window), and *RET 2H EA* (2-hour window).	TAQ
Immediate Price Revision LLM EA	Immediate price revision (following earnings announcements) predicted, out-of-sample, based solely on the text of a firm's quarterly earnings press release using a *BERT*-based large language model (LLM).	TAQ EDGAR
CAR LLM CC	*CAR [0,1]* predicted, out-of-sample, based solely on the text of a firm's earnings conference call transcript using a *BERT*-based LLM.	CRSP Capital IQ Transcripts
Immediate Price Revision CC	The stock return measured during earnings conference calls that occur on the same day as the earnings press release. The return is calculated as the ratio of the last available price to the first available price closest to or immediately following the event window endpoints, minus one. One return measure is considered: *RET 1H CC* (1-hour window).	TAQ
Immediate Price Revision LLM CC	Immediate price revision (following earnings conference calls) predicted, out-of-sample, based solely on the text of a firm's earnings conference call transcript using a *BERT*-based large language model (LLM).	TAQ Capital IQ Transcripts

Endnotes

[1] As described in Section 6.1, sensitivity analyses show that inferences do not change when concurrent news events (e.g., conference calls, analyst, or management forecasts) are alternatively included or excluded from the analysis.

[2] The economic magnitude of earnings disclosure news remains a subject of empirical inquiry. Prior studies suggest that the low information content of accounting (textual) disclosures may jointly derive from (i) managers' disclosure behavior (Frankel et al. 2022); (ii) investors' processing constraints (Meursault et al. 2022); (iii) the prevalence of private or other public information during earnings announcement reactions (Boudoukh et al. 2018); and (iv) measurement error (Loughran and McDonald 2011, Bochkay et al. 2022).

[3] This stream of the finance literature also suggests that an upper-bound limit for the information content of earnings announcement disclosures, measured using R^2, is significantly lower than 100%. Therefore, while performance improvements over current-generation LLMs are likely, news extraction from corporate disclosures is constrained.

[4] For example, Frankel et al. (2022) note that "The meaning of "sentiment" is unclear" (p. 5531), further highlighting the theoretical and empirical limitations of conventional tone or sentiment proxies.

[5] Other studies apply (i) *BERT*-based LLMs for topic modeling/classification (Lee and Zhong 2022) or (ii) shallow(er) neural networks, which are relatively less powerful at capturing complex word context compared to *BERT*-based LLMs, primarily for classification tasks (Li 2022).

[6] Other studies, such as Jiang et al. (2012) and Brochet et al. (2018), have utilized I/B/E/S earnings announcement timestamps. To validate the accuracy of the I/B/E/S and Capital IQ timestamps, I manually cross-reference 50 timestamps with Dow Jones Newswire timestamps obtained from Factiva, finding a 97% accuracy rate.

[7] Tables are characterized by differences in formatting, separation from the main earnings announcement text, types, and formatting

of numbers. These differences vary across and within firm over time. The *BERT* LLM is not explicitly pre-trained on tabular information and thus lacks training to maximize information extraction from tables. In untabulated tests, I use the entire text of the earnings announcements (inclusive of tables and cautionary statements) and find a 1.5-percentage-point lower explanatory power, on average; however, the tenor of my results does not change. The Online Appendix offers more details about this parsing strategy.

[8] My study does not extract an ex-ante trading signal. In fact, I investigate how text maps into contemporaneous, rather than future, investors' reactions.

[9] Although *BERT* LLMs are pretrained to predict the next word in a sequence rather than forecast stock market movements, the overlap between the LLM pretraining period (ending in 2019) and my out-of-sample period motivated additional analyses using data entirely *outside* the pretraining period. These analyses yield consistent evidence, as detailed in Section 6.1.

[10] "gradient boosting" provides, on average, a 0.7% higher adjusted R^2 compared to other machine learning techniques (i.e., with an average improvement of 30% in terms of model fit); moreover, it requires a generally lower computational time (i.e., 15 minutes for the main prediction task as opposed to an average of 25 minutes).

[11] Results are similar when the number of unigrams and bigrams is extended to 3,000.

[12] This study employs a range of traditional and more recent machine learning techniques that can map textual information into returns, through linear and non-linear functions. The goal is *not* to include a comprehensive set of benchmarks, but rather a range of approaches, alternative to *BERT* LLM, that can reflect the general evolution of textual analysis methods and offer insights into the context-based modeling of corporate disclosures.

[13] I replace missing surprises with zero as several of these variables are sparsely populated. This is also noted in Hand et al. (2022) that apply a similar protocol (p. 1403). However, I find consistent results without such a replacement. Notably, the machine learning–based modeling of surprises yields a higher accuracy when missing items *are* replaced.

[14] In a series of robustness analyses (not reported), I also include analyst following, and the release of analysts' and management's forecasts as additional controls to account for relevant concurrent news events. The results are unaffected. See also the sensitivity analyses described in Section 6.

[15] Consistent with prior literature (Francis et al. 1992), I anticipate selection issues among the 2% of announcements made during regular trading hours. In my sample, these announcements are associated with smaller firms and less pronounced market reactions. Therefore, I exclude them from the analysis.

[16] Consistent with prior research (Brochet et al. 2018), I impose the following data requirements of PRICE > 0, SIZE > 0, CORR < 2, and COND not equal to A/C/D/N/O/R/Z. Additionally, in most cases, the first available preannouncement price corresponds to the prior market close for earnings announcements released before regular trading hours and the current market close for those released after regular trading hours. Therefore, in such cases, I use these prices. Sensitivity analyses yield similar results.

[17] Similar results are obtained *excluding* earnings announcements that are released outside of regular trading hours, but which immediate price reaction occurs during regular trading hours (e.g., a press release issued at 9:28 a.m.).

[18] In some cases, the first available price after the end of the call occurs two or more hours after the end of the call or on the next day. I exclude these observations from the calculation of hourly returns.

[19] As noted in Section 4.2, valid selection concerns may arise when examining calls held during trading hours. In my sample, these calls are associated with smaller firms and less pronounced market reactions.

[20] Sensitivity analyses indicate that the results remain consistent when calculating one-hour returns starting 10 minutes after the call begins.

[21] *CAR GB FSA* represents a restrictive benchmark, relative to traditional surprises; hence, it allows stronger isolation of incremental information in disclosures. In unreported tests, I also use traditional surprises finding similar results.

[22] I test for statistical differences in R^2 using a nonparametric bootstrap protocol as outlined in Davison and Hinkley (1997). Specifically, I repeat each regression analysis 10,000 times, each time with a different sample (of size equal to the total number of observations available for each regression). The different samples are obtained from independently resampling data points for each sample partition (with replacement).

[23] I generate 100 bootstrap samples with replacement for each month-year, ensuring at least 15 observations, following a method similar to Section 5.1 to address the limited sample size in monthly regressions.

[24] February–April 2020 for COVID-19 (when the WHO declared a global pandemic), June–December 2019 for the U.S.–China trade war (a time of major tariff escalations), September–December 2008 for the financial crisis (when Lehman Brothers declared bankruptcy), and November 2001–January 2002 for the Enron scandal (coinciding with the firm's bankruptcy). Restricting these windows to a fixed number of month-years generally yields similar results. Furthermore, the tenor of the findings remains consistent with the inclusion of yearly fixed effects or clustering.

[25] I also consider analyzing *specific* contents (e.g., discussions about M&A). However, disclosure contents vary substantially across firms and within a firm over time, thus leading to inferences that are less generalizable.

[26] I classify sentences that include dollar amounts or percentages, and that are immediately adjacent to them, as sentences that discuss numbers directly.

[27] I apply a string similarity algorithm based on "gestalt pattern matching" (Metzener and Ratcliff 1988). The algorithm identifies the longest contiguous substring, after removing elements such as white spaces, between two texts. The output is a score capturing their similarity. I apply this approach to each pair of sentences across consecutive quarterly press releases (for a firm). If the score is above 90%, I classify the sentence as "stale." The choice of the threshold is empirical and balances type I and type II errors. I manually check 10 earnings announcements finding an accuracy of 80%. Because disclosures in Q4 are longer and richer in contents than those in Q1–Q3, I also test the sensitivity of results comparing year-on-year disclosures; results (unreported) are similar.

[28] This approach is conservative and parsimonious. Alternatively, one could pretrain the LLM excluding certain years.

[29] Textual news is high in recent years, including 2023. This evidence may underlie the role of LLMs in enhancing investors' use of earnings disclosures. Exploring this further presents a promising avenue for future research.

[30] The randomization only involves out-of-sample observations. The LLM is *not* fine-tuned on modified texts.

[31] Performance words include terms such as "income" or "ebitda," whereas "tone" words derive from Henry (2008).

References

Ball R, Brown P (1968) An empirical evaluation of accounting income numbers. *J. Accounting Res.* 6(2):159–178.

Ball R, Shivakumar L (2008) How much new information is there in earnings? *J. Accounting Res.* 46(5):975–1016.

Barth ME, Li K, McClure C (2022) Evolution in value relevance of accounting information. *Accounting Rev.* 98(1):1–28.

Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer, New York).

Bochkay K, Brown SV, Leone AJ, Tucker JW (2022) Textual analysis in accounting: What's next? *Contemporary Accounting Res.* 40(2): 765–805.

Boudoukh J, Feldman R, Kogan S, Richardson M (2018) Information, trading, and volatility: Evidence from firm-specific news. *Rev. Financial Stud.* 32(3):992–1033.

Breuer M, deHaan E (2023) Using and interpreting fixed effects models. *J. Accounting Res.* 62(4):1183–1226.

Brochet F, Kolev K, Lerman A (2018) Information transfers and conference calls. *Rev. Accounting Stud.* 23(3):907–957.

Brogaard J, Nguyen TH, Putnins TJ, Wu E (2022) What moves stock prices? The roles of news, noise, and information. *Rev. Financial Stud.* 35(9):4341–4386.

Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, et al. (2020) Language models are few-shot learners. Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds. *Advances in Neural Information Processing Systems*, vol. 33 (Curran Associates, Inc., Red Hook, NY), 1877–1901.

Cheng L, Roulstone D, Van Buskirk A (2021) Are investors influenced by the order of information in earnings press releases? *Accounting Rev.* 96(2):413–433.

Cohen L, Malloy C, Nguyen Q (2020) Lazy prices. *J. Finance* 75(3): 1371–1415.

Collins DW, Kothari SP (1989) An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. *J. Accounting Econom.* 11(2–3):143–181.

Collins DW, Kothari SP, Shanken J, Sloan RG (1994) Lack of timeliness and noise as explanations for the low contemporaneous return-earnings association. *J. Accounting Econom.* 18(3): 289–324.

Davison AC, Hinkley DV (1997) *Bootstrap Methods and Their Application* (Cambridge University Press, Cambridge, UK).

Devlin J, Chang M, Lee K, Toutanova K (2019) Pre-training of deep bi-directional transformers for language understanding. Preprint, submitted May 24, https://arxiv.org/abs/1810.04805.

Donovan J, Jennings J, Koharki K, Lee J (2021) Measuring credit risk using qualitative disclosure. *Rev. Accounting Stud.* 26 (2):815–863.

Du K, Huddart S, Jiang XD (2023) Lost in standardization: Effects of financial statement database discrepancies on inference. *J. Accounting Econom.* 76(1):101573.

Francis J, Pagach D, Stephan J (1992) The stock market response to earnings announcements released during trading versus nontrading periods. *J. Accounting Res.* 30(2):165–184.

Frankel R, Jennings J, Lee J (2016) Using unstructured and qualitative disclosures to explain accruals. *J. Accounting Econom.* 62(2–3): 209–227.

Frankel R, Jennings J, Lee J (2022) Disclosure sentiment: Machine learning vs. dictionary approaches. *Management Sci.* 68(7):5514–5532.

French KR, Roll R (1986) Stock return variances: The arrival of information and the reaction of traders. *J. Financial Econom.* 17(1): 5–26.

Gunning R (1952) *The Technique of Clear Writing* (McGraw-Hill International Book Co., New York).

Hand JRM, Laurion H, Lawrence A, Martin N (2022) Explaining firms' earnings announcement stock returns using FactSet and I/B/E/S data feeds. *Rev. Accounting Stud.* 27:1389–1420.

Hayn C (1995) The information content of losses. *J. Accounting Econom.* 20(2):125–153.

Healy PM, Palepu KG (2001) Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *J. Accounting Econom.* 31(1):405–440.

Henry E (2008) Are investors influenced by how earnings press releases are written? *J. Bus. Comm.* 45(4):363–407.

Hirshleifer D (2001) Investor psychology and asset pricing. *J. Finance* 56(4):1533–1597.

Holthausen RW, Watts RL (2001) The relevance of the value-relevance literature for financial accounting standard setting. *J. Accounting Econom.* 31(1–3):3–75.

Huang AH, Wang H, Yang Y (2022) FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Res.* 40(2):806–841.

Jawahar G, Sagot B, Seddah D (2019) What does BERT learn about the structure of language? *Proc. 57th Ann. Meeting Assoc. Comput. Linguistics* (Association for Computational Linguistics, Florence, Italy), 3651–3657.

Jegadeesh N, Wu D (2013) Word power: A new approach for content analysis. *J. Financial Econom.* 110(3):712–729.

Jiang CX, Likitapiwat T, McInish TH (2012) Information content of earnings announcements: Evidence from after-hours trading. *J. Financial Quant. Anal.* 47(6):1303–1330.

Kim AG, Nikolaev VN (2024) Context-based interpretation of financial information. *J. Bus. Finance Accounting*, ePub ahead of print December 18, https://doi.org/10.1111/1475-679X.12593.

Kothari SP (2001) Capital markets research in accounting. *J. Accounting Econom.* 31(1–3):105–231.

Krupa J, Minutti-Meza M (2022) Regression and machine learning methods to predict discrete outcomes in accounting research. *J. Financial Rep.* 7(2):131–178.

Lee MC, Zhong Q (2022) Shall we talk? The role of interactive investor platforms in corporate communication. *J. Accounting Econom.* 74(2–3):101524.

Leuz C, Wysocki PD (2016) The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *J. Accounting Res.* 54(2):525–622.

Lev B (1989) On the usefulness of earnings and earnings research: Lessons and directions from two decades of empirical research. *J. Accounting Res.* 27(2):153–192.

Lev B, Zarowin P (1999) The boundaries of financial reporting and how to extend them. *J. Accounting Res.* 37(2):353–385.

Li F (2010) Textual analysis of corporate disclosures: A survey of the literature. *J. Accounting Literature* 29:143–165.

Li K (2022) Textual fundamentals in earnings press releases. *Adv. Accounting* 6(57):100591.

Li EX, Ramesh K (2009) Market reaction surrounding the filing of periodic SEC reports. *Accounting Rev.* 84(4):1171–1208.

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, et al. (2019) RoBERTa: A robustly optimized BERT pretraining approach. Preprint, submitted July 26, https://arxiv.org/abs/1907.11692.

Loh RK, Stulz RM (2018) Is sell-side research more valuable in bad times? *J. Finance* 73(3):909–1458.

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

Loughran T, McDonald B (2016) Textual analysis in accounting and finance: A survey. *J. Accounting Res.* 54(4):1187–1230.

Lundberg S, Lee S (2017) A unified approach to interpreting model predictions. Preprint, submitted November 25, https://arxiv.org/abs/1705.07874v2.

Lyle MR, Siano F, Yohn T (2025) Re-adjusted financial statement data: Challenges in replicating research. Preprint, submitted January 22, http://dx.doi.org/10.2139/ssrn.5107985.

Matsumoto D, Pronk M, Roelofsen E (2011) What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *Accounting Rev.* 86(4): 1383–1414.

Mayew WJ, Sethuraman M, Venkatachalam M (2020) Individual analysts' stock recommendations, earnings forecasts, and the informativeness of conference call question and answer sessions. *Accounting Rev.* 95(6):311–337.

Metzener DE, Ratcliff JW (1988) Pattern matching: The gestalt approach. *Dobb's J.* 13:355–358.

Meursault V, Liang PJ, Routledge BR, Scanlon MM (2022) PEAD.txt: Post-earnings-announcement drift using text. *J. Financial Quant. Anal.* 58(6):2299–2326.

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. Preprint, submitted September 7, https://arxiv.org/abs/1301.3781v3.

Ohlson J (1995) Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Res.* 11(2):661–687.

Ou J, Penman S (1989) Financial statement analysis and the prediction of stock returns. *J. Accounting Econom.* 11(4):295–329.

Pan SJ, Yang Q (2010) A survey of transfer learning. *IEEE Trans. Knowledge Data Engrg.* 22(10):1345–1359.

Patell JM, Wolfson MA (1984) The intraday speed of adjustment of stock prices to earnings and dividends announcements. *J. Financial Econom.* 13:223–252.

Roll R (1984) Orange juice and weather. *Amer. Econom. Rev.* 74(5):861–880.

Roll R (1988) R-squared. *J. Finance* 43(3):541–566.

Shao S, Stoumbos R, Zhang XF (2021) The power of firm fundamental information in explaining stock returns. *Rev. Accounting Stud.* 26(4):1249–1289.

Siano F, Wysocki PD (2021) Transfer learning and textual analysis of accounting disclosures: Applying big data methods to small(er) datasets. *Accounting Horizons* 35(3):217–244.

Tenney I, Xia P, Chen B, Wang A, Poliak A, McCoy RT, Kim N, et al. (2019) What do you learn from context? Probing for sentence structure in contextualized word representations. Preprint, submitted May 15, https://arxiv.org/abs/1905.06316v1.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. Preprint, submitted December 6, https://arxiv.org/abs/1706.03762v5.